



(51) International Patent Classification:

A61K 35/74 (2015.01) C12N 15/75 (2006.01)

(21) International Application Number:

PCT/US2022/051979

(22) International Filing Date:

06 December 2022 (06.12.2022)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/286,736 07 December 2021 (07.12.2021) US

(71) **Applicant:** CORNELL UNIVERSITY [US/US]; Center for Technology Licensing ("CTL"), 395 Pine Tree Road, Suite 310, Ithaca, New York 14850 (US).

(72) **Inventors:** GUO, Chun-Jun; 413 E 69th Street, New York, New York 10021 (US). JIN, Wenbing; 413 E 69th Street, New York, New York 10021 (US). LI, Tingting; 404 East 66th St. Apt. 6M, New York, New York 10044 (US).

(74) **Agent:** EWING, James F. et al.; FOLEY & LARDNER LLP, 3000 K Street, N.W., Suite 600, Washington, District of Columbia 20007 (US).

(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM,

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) **Title:** METHODS AND COMPOSITIONS FOR GENETICALLY MODIFYING HUMAN GUT MICROBES

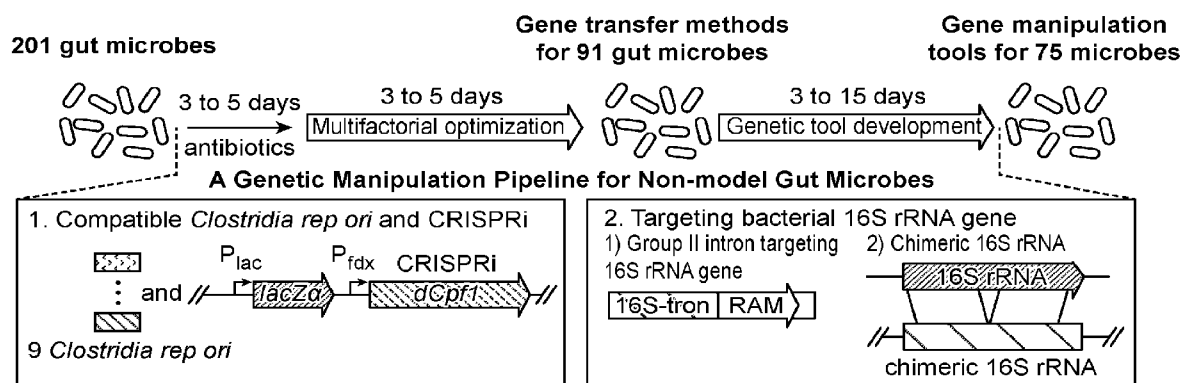


FIG. 1A

(57) **Abstract:** The present technology relates generally to compositions and the methods of preparations thereof for genetically engineering gut-microbiota *in vitro*. The present technology further relates to uses of compositions *in vivo*.

METHODS AND COMPOSITIONS FOR GENETICALLY MODIFYING HUMAN GUT MICROBES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/286,736, filed December 7, 2021, the entire contents of which are incorporated herein by reference.

GOVERNMENT SUPPORT

[0002] This invention was made with government support under DK126871, AI151599, AI095466, AI095608, AI142213 and 1DP2HD101401-01 awarded by the National Institutes of Health. The government has certain rights in the invention.

TECHNICAL FIELD

[0003] The present technology relates generally to compositions and the methods of preparations thereof for genetically engineering gut-microbiota *in vitro*. The present technology further relates to uses of compositions *in vivo*.

BACKGROUND

[0004] The following description of the background of the present technology is provided simply as an aid in understanding the present technology and is not admitted to describe or constitute prior art to the present technology.

[0005] Dysbiosis, or perturbation of the microbiome, has been linked to diseases such as inflammatory bowel disease and obesity. Multi-omics studies uncover many microbiota genes that are associated with host biology. However, it remains challenging to unravel the causal mechanisms underlying microbiota gene-host biology interactions, mainly because many are encoded by non-model gut microbes like *Firmicutes/Clostridia*. While genetic toolsets are readily available for model bacteria like *E. coli* or *B. thetaiotaomicron*, the limitation lies in that the optimal condition identified in one study is not readily applicable to the other. Most of the gut commensals, especially those that are dominant in the gut, are non-model gut *bacteria* (e.g.,

Bacteroides, *Prevotella*, and *Clostridium*) are still resistant to genetic modifications. In addition, engineering therapeutic functions into the microbiome requires targeted genomic edits, which presents a further challenge because many non-model gut *bacteria* (e.g., *Lachnospiraceae*, *Prevotella*) are not genome sequenced, and it is unknown how to introduce exogenous DNA or which gene manipulation tool to select (Waller et al., 2017a).

[0006] There is an urgent need for efficient, standardized, and *in vitro* pipeline to identify their gene transfer methods and build their genetic manipulation systems without prior knowledge of their genome information. Such pipelines are important for three reasons: 1) Multi-omics studies have uncovered significant associations between microbiota genes and diseases. Many of these genes are exclusively expressed in non-model microbes such as *Firmicutes/Clostridia* (Lloyd-Price et al., 2019; Thomas et al., 2019; Wang et al., 2012; Wirbel et al., 2019; Yachida et al., 2019; Zhou et al., 2019). A pipeline addressing this need would be a first step to manipulating these genes *in vivo* and causally connecting them with host diseases. 2) The gut microbiota plays an essential role in regulating host biology, but little is known about which bacteria and genes are responsible. A desirable pipeline would enable gene toggling in previously non-targetable microbes and boost in-depth mechanistic studies of microbiota-host physiology interactions. 3) The microbiota impacts multiple therapies such as fecal microbiota transplantation and cancer immunotherapy (Helmink et al., 2019; Roy and Trinchieri, 2017), but the molecular mechanisms behind them largely remain elusive.

SUMMARY OF THE PRESENT TECHNOLOGY

[0007] In one aspect, the present disclosure provides a bacterial expression vector comprising (a) a nucleic acid encoding a target gene that is conserved in a plurality of human gut commensal gram-negative bacterial species and (b) a heterologous nucleic acid encoding a selectable marker, wherein the selectable marker is an antibiotic resistance gene or an auxotrophic marker, and optionally wherein the target gene is selected from the group consisting of 16s rRNA, 23s rRNA, *mmdA*, *RokA* (Clucokinase gene), and an ABC transporter gene. Additionally or alternatively, in some embodiments, the bacterial expression vector further comprises at least one open reading frame encoding a bioluminescent protein, a chemiluminescent protein, a fluorescent protein, a CRISPR enzyme, a Group II intron-encoded protein, at least one sgRNA,

at least one Group II intron, or any combination thereof. In some embodiments, the 16s rRNA comprises the nucleic acid sequence of SEQ ID NO: 11. Additionally or alternatively, in some embodiments, the bacterial expression vector comprises the nucleic acid sequence of SEQ ID NO: 310.

[0008] In one aspect, the present disclosure provides a bacterial expression vector comprising (a) a gram-positive bacteria replication origin comprising a sequence selected from the group consisting of SEQ ID NOs: 1-9 or 311-319, (b) a heterologous nucleic acid encoding a selectable marker that is an antibiotic resistance gene or an auxotrophic marker, and (c) at least one open reading frame, wherein the at least one open reading frame encodes a bioluminescent protein, a chemiluminescent protein, a fluorescent protein, a CRISPR enzyme, a Group II intron-encoded protein, at least one sgRNA, at least one Group II intron, or any combination thereof. The bacterial expression vector of the present technology may further comprise one or more bacterial conjugation transfer genes and/or an *E. coli* replication origin. Examples of bacterial conjugation transfer genes include *traJ* and *oriT*, and examples of *E. coli* replication origin include *colEI*, *pBR*, and *R6K*. Additionally or alternatively, in some embodiments, the one or more bacterial conjugation transfer genes, the gram-positive bacteria replication origin, and the heterologous nucleic acid encoding the selectable marker are codon optimized. Additionally or alternatively, in some embodiments, the at least one sgRNA or the at least one Group II intron targets one or more genes selected from among 16S rRNA, *porA*, *bcat*, *croA*, *baiA2*, *baiCD*, *baiF*, *baiH*, *baiB*, *baiE*, *baiG* and *bail*.

[0009] In any and all embodiments of the bacterial expression vectors disclosed herein, the antibiotic resistance gene is selected from the group consisting of *catP*, *ermB*, *aad9*, *tetA*, and *ampR*, or the auxotrophic marker is *pyrG*, or *pyrF*.

[0010] In any of the preceding embodiments of the bacterial expression vectors disclosed herein, the CRISPR enzyme is selected from the group consisting of Cas9, dCas9, Cpf1, dCpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, and Csf4.

[0011] Examples of fluorescent proteins include, but are not limited to, GFP, YFP, CFP, RFP, TagBFP, Azurite, EBFP2, mKalama1, Sirius, Sapphire, T-Sapphire, ECFP, Cerulean, SCFP3A, mTurquoise, monomeric Midoriishi-Cyan, TagCFP, mTFP1, EGFP, Emerald, Superfolder GFP, Monomeric Azami Green, TagGFP2, mUKG, mWasabi, EYFP, Citrine, Venus, SYFP2, TagYFP, Monomeric Kusabira-Orange, mKOκ, mKO2, mOrange, mOrange2, mRaspberry, mCherry, dsRed, mStrawberry, mTangerine, tdTomato, TagRFP, TagRFP-T, mApple, mRuby, mPlum, HcRed-Tandem, mKate2, mNeptune, NirFP, TagRFP657, IFP1.4, iRFP, mKeima Red, LSS-mKate1, LSS-mKate2, PA-GFP, PAmCherry1, PAtagRFP, Kaede (green), Kaede (red), KikGR1 (green), KikGR1 (red), PS-CFP2, PS-CFP2, mEos2 (green), mEos2 (red), PSmOrange, or Dronpa. Examples of chemiluminescent proteins include, but are not limited to, β-galactosidase, horseradish peroxidase (HRP), or alkaline phosphatase. Examples of bioluminescent protein include, but are not limited to, Aequorin, firefly luciferase, *Renilla* luciferase, red luciferase, luxAB, or nanoluciferase.

[0012] In any and all embodiments of the bacterial expression vectors disclosed herein, the at least one sgRNA specifically hybridizes with a heterologous or endogenous target gene expressed in a gut bacterial host cell, and/or wherein the at least one sgRNA and/or the CRISPR enzyme is operably linked to a constitutive promoter or a conditional promoter. Additionally or alternatively, in some embodiments of the bacterial expression vectors disclosed herein, the at least one Group II intron specifically targets a heterologous or endogenous gene expressed in a gut bacterial host cell, and/or wherein the at least one Group II intron and/or the Group II intron-encoded protein is operably linked to a constitutive promoter or a conditional promoter.

[0013] In another aspect, the present disclosure provides an engineered gram-negative human gut bacterial cell comprising any and all embodiments of the gram-negative specific bacterial expression vector described herein, wherein the engineered gram-negative human gut bacterial cell is derived from a family selected from the group consisting of *Enterobacteriaceae*, *Bacteroidaceae*, *Tannerellaceae*, and *Prevotellaceae*. In some embodiments, the engineered gram-negative human gut bacterial cell is derived from *Bacteroides cellulosilyticus*, *Bacteroides cellulosilyticus*, *Bacteroides dorei*, *Bacteroides eggerthii*, *Bacteroides finegoldii*, *Bacteroides fragilis*, *Bacteroides intestinalis*, *Bacteroides nordii*, *Bacteroides oleiciplenus*, *Bacteroides ovatus*, *Bacteroides salyersiae*, *Bacteroides sp.*, *Bacteroides thetaiotaomicron*, *Bacteroides*

uniformis, *Bacteroides vulgatus*, *Bacteroides xylanisolvens*, *Parabacteroides faecis*, *Parabacteroides merdae*, or *Prevotella bivia*.

[0014] In one aspect, the present disclosure provides an engineered gram-positive human gut bacterial cell comprising any and all embodiments of the gram-positive specific bacterial expression vector disclosed herein, wherein the engineered gram-positive human gut bacterial cell is derived from a family selected from the group consisting of *Clostridiaceae*, *Lachnospiraceae*, *Eubacteriaceae*, *Erysipelotrichaceae*, *Enterococcaceae*, and *Bifidobacteriaceae*. In some embodiments, the engineered gram-positive human gut bacterial cell is derived from *Blautia hydrogenotrophica*, *Blautia luti*, *Blautia sp.*, *Blautia wexlerae*, *Clostridium bolteae*, *Clostridium innocuum*, *Clostridium paraputrificum*, *Clostridium saccharolyticum*, *Clostridium senegalense*, *Clostridium sp.*, *Clostridium sporogenes*, *Clostridium symbiosum*, *Eubacterium limosum*, *Eubacterium maltosivorans*, *Eubacterium ramulus*, *Eubacterium sp.*, *Roseburia inulinivorans*, *Bifidobacterium catenulatum*, *Enterococcus faecium*, *Escherichia fergusonii*, *Roseburia inulinivorans*, or *Bifidobacterium catenulatum*.

[0015] In one aspect, the present disclosure provides a method for modifying a gram-negative human gut bacteria cell genome comprising transferring at least one gram-negative specific bacterial expression vector described herein into a gram-negative human gut bacteria cell via conjugation. In some embodiments, the at least one bacterial expression vector is integrated into the genome of the gram-negative human gut bacteria cell.

[0016] In another aspect, the present disclosure provides a method for genetically modifying a gram-positive human gut bacteria cell comprising transferring two or more distinct bacterial expression vectors into a gram-positive human gut bacteria cell simultaneously via conjugation, wherein each of the two or more distinct bacterial expression vectors comprise: (a) a gram-positive bacteria replication origin comprising a sequence selected from the group consisting of SEQ ID NOs: 1-9 or 311-319, (b) a heterologous nucleic acid encoding a selectable marker that is an antibiotic resistance gene or an auxotrophic marker, and (c) at least one open reading frame, wherein the at least one open reading frame encodes a bioluminescent protein, a chemiluminescent protein, a fluorescent protein, a CRISPR enzyme, a Group II intron-encoded protein, at least one sgRNA, at least one Group II intron, or any combination thereof. The

antibiotic resistance gene or the auxotrophic marker of each of the two or more distinct bacterial expression vectors may be independently selected from the group consisting of *catP*, *ermB*, *aad9*, *tetA*, *ampR*, *pyrG*, and *pyrF*.

[0017] In some embodiments, each of the two or more distinct bacterial expression vectors further comprise one or more bacterial conjugation transfer genes and/or an *E. coli* replication origin, optionally wherein the one or more bacterial conjugation transfer genes are selected from the group consisting of *traJ*, and *oriT* and/or the *E. coli* replication origin is selected from the group consisting of *colEI*, *pBR*, and *R6K*.

[0018] Additionally or alternatively, in some embodiments, the CRISPR enzyme of each of the two or more distinct bacterial expression vectors is independently selected from the group consisting of Cas9, dCas9, Cpf1, dCpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, and Csf4.

[0019] Additionally or alternatively, in some embodiments of the methods disclosed herein, the fluorescent protein of each of the two or more distinct bacterial expression vectors is independently selected from the group consisting of GFP, YFP, CFP, RFP, TagBFP, Azurite, EBFP2, mKalama1, Sirius, Sapphire, T-Sapphire, ECFP, Cerulean, SCFP3A, mTurquoise, monomeric Midoriishi-Cyan, TagCFP, mTFP1, EGFP, Emerald, Superfolder GFP, Monomeric Azami Green, TagGFP2, mUKG, mWasabi, EYFP, Citrine, Venus, SYFP2, TagYFP, Monomeric Kusabira-Orange, mKOκ, mKO2, mOrange, mOrange2, mRaspberry, mCherry, dsRed, mStrawberry, mTangerine, tdTomato, TagRFP, TagRFP-T, mApple, mRuby, mPlum, HcRed-Tandem, mKate2, mNeptune, NirFP, TagRFP657, IFP1.4, iRFP, mKeima Red, LSS-mKate1, LSS-mKate2, PA-GFP, PAmCherry1, PATagRFP, Kaede (green), Kaede (red), KikGR1 (green), KikGR1 (red), PS-CFP2, PS-CFP2, mEos2 (green), mEos2 (red), PSmOrange, and Dronpa. Additionally or alternatively, in certain embodiments of the methods disclosed herein, the chemiluminescent protein of each of the two or more distinct bacterial expression vectors is independently β-galactosidase, horseradish peroxidase (HRP), or alkaline phosphatase. Additionally or alternatively, in some embodiments of the methods of the present technology, the

bioluminescent protein of each of the two or more distinct bacterial expression vectors is independently Aequorin, firefly luciferase, *Renilla* luciferase, red luciferase, luxAB, or nanoluciferase.

[0020] In any and all embodiments of the methods disclosed herein, the at least one sgRNA sequence of the two or more distinct bacterial expression vectors specifically hybridizes with a heterologous or endogenous target gene expressed in a gut bacterial host cell, and/or wherein the at least one sgRNA and/or the CRISPR enzyme is operably linked to a constitutive promoter or a conditional promoter. Additionally or alternatively, in some embodiments of the methods disclosed herein, the at least one Group II intron of the two or more distinct bacterial expression vectors specifically targets a heterologous or endogenous gene expressed in a gut bacterial host cell, and/or wherein the at least one Group II intron and/or the Group II intron-encoded protein is operably linked to a constitutive promoter or a conditional promoter. In some embodiments, three or four distinct bacterial expression vectors are simultaneously transferred into a gram-positive human gut bacteria cell simultaneously via conjugation.

[0021] In any and all embodiments of the methods disclosed herein, the gram-negative or gram-positive human gut bacteria cell is isolated from a colonic mucosa-enriched lavage sample, a fecal sample, a rectal swab, or an intestinal sample obtained from a human subject.

[0022] Also disclosed herein are engineered human gut bacterial cells generated by any and all embodiments of the methods of the present technology.

[0023] Also provided herein are kits comprising any and all embodiments of the bacterial expression vectors of the present technology and instructions for using the bacterial expression vectors to genetically modify human gut bacteria. The kits may further comprise one or more primers and/or gRNAs comprising the sequence of any one of SEQ ID NOs: 23-287.

BRIEF DESCRIPTION OF THE DRAWINGS

[0024] **Figs. 1A-1C. Overview of the genetic manipulation (GM) pipeline for non-model gut commensals.** **Fig. 1A:** A total of 201 human gut isolates from >140 species and 5 phyla were subject to the GM pipeline. The pipeline identifies gene transfer methods for 91 non-model gut microbes (of 72 species) and build gene manipulation tools for 72 of them. For Gram-

negative gut microbes, identifying their gene transfer methods and building their gene insertion tools are achieved in one step via the chimeric-16s rRNA strategy. **Fig. 1B:** Phylogenetic tree (colored by Family) of the 16s rRNA sequences from the 91 genetically targetable microbes identified via the GM pipeline. **Fig. 1C:** Detailed phylogenetic information of the 91 genetically targetable microbes identified in this study. These microbes are from 72 bacterial species in 16 families.

[0025] Figs. 2A-2D. Developing a genetic manipulation pipeline for non-model gut

commensals. Fig. 2A: Schematic view of a multifactorial optimization of the conjugation/transformation parameters to identify gene transfer conditions for 38 non-model gut *Firmicutes/Clostridia* that are mostly untransformed. **Fig. 2B:** Establishment of a dCpf1-*lacZα* platform for non-model gut *Firmicutes/Clostridia*. (left) Schematic view of a dCpf1-*lacZα* system. The promoter and CDS region of *lacZα* are targeted by a duplex gRNA G1 and G2. (right) The dCpf1-*lacZα* system efficiently suppresses *lacZα* expression in 25 *Clostridia* microbes. The panel shows the mean gene expression of three biological replicates as determined by qPCR. The dCpf1-only and gRNA-only vectors are used as negative controls. Three out of 25 qPCR results are shown. The numbering of the strains corresponds to the strain information shown in **Fig. 22**. Error bar: standard deviation. DR: direct repeat, G1: guide RNA-coding sequence 1, G2: guide RNA-coding sequence 2, Ter: terminator. **Fig. 2C:** Schematic view of the 16s-tron strategy for non-model *Clostridia*. The *Clostridia* 16s rRNA sequences were aligned to identify a conserved target site of Group II intron. The 16s targeting Group II intron (16s-tron) was introduced into 19 *Clostridia* commensals due to RAM (retrotransposition-activated marker) availability. We identified 16 *Clostridia* whose chromosomes have been integrated by the 16s-tron. **Fig. 2D:** Schematic view of a *Bacteroidia/Prevotella* GM pipeline. The *Prevotella* 16s rRNA sequences were aligned to generate a ~1kb chimeric 16s (chi-16s) fragment. The chi-16 was assembled to get a suicide vector pGM-NAC₂P. The pGM-NAC₂P (NAC₂B for *Bacteroides*) was conjugated to 21 *Prevotella*, 39 *Bacteroides*, and 6 *Parabacteroides* commensals targeting their chromosomal 16s rRNA genes. We identified 31 targetable *Bacteroidia* whose 16s rRNA genes have been integrated by pGM-NAC₂P (or NAC₂B).

[0026] Figs. 3A-3D. Modulating *Clostridia* gene expression and microbiome-derived metabolites using gene manipulation tools developed via the GM pipeline. Fig. 3A (top): Schematic view of a duplex gRNA targeting the branched-chain amino acid aminotransferase *bcat* in the *Clostridia* commensals. **Fig. 3A (bottom):** The *bcat* gene of 12 *Clostridia* microbes was efficiently repressed using dCpf1. The panel shows the mean gene expression of three biological replicates as determined by qPCR. Only three representative results (**S54**, **S74**, and **S110**, **Fig. 22**) are shown. **Fig. 3B (top):** Schematic view of knocking out the *Bacteroides mmdA* genes using pGM vectors. The pMG vector was assembled with ~1kb fragment of the *mmdA* genes, and the *mmdA* genes of three *Bacteroides* microbes were knocked out via single crossover integration. **Fig. 3B (bottom):** Three *Bacteroides* Δ *mmdA* mutants (**S25**, **S27**, and **S31**, **Fig. 22**) deplete propionate *in vitro*. The bacterial culture supernatant was derivatized and propionate production was examined using LC-MS (EIC: 216.1137). The *Bacteroides* Δ *mmdA* mutant depletes propionate *in vivo*. Germ-free Swiss Webster mice (n = 3 or 4 per group) were mono-colonized with the *B. sp.* 1_1_6 control strain (**Con**, 16s integrated by pGM-NAC₂B) and Δ *mmdA* mutant (**Mut**). Propionate was depleted in the host by *mmdA* deletion. Student's T-test was performed, and the asterisk indicates *p*-value < 0.05 (*) or < 0.01 (**). Error bar: standard deviation. **Fig. 3C (top):** Schematic view of modulating butyrate production in the *Clostridia* commensals using dCpf1 or Group II intron. **Fig. 3C (bottom):** The butyrate production (quantified by LC-MS) was significantly reduced in three *Clostridia* microbes **S110** *C. symbiosum* (by dCpf1), **S115** *E. limosum* (by Group II intron), and **S117** *E. maltosivorans* (by dCpf1) (**Fig. 22**). The cecal butyrate (quantified by LC-MS) in the germ-free Swiss Webster mice mono-colonized with **S117** *E. maltosivorans* mutant (**Mut**, dCpf1+gRNAs) is significantly lower compared to the control (**Con**, dCpf1 only). Student's T-test was performed, and the asterisk indicates *p*-value < 0.05 (*) or < 0.01 (**). Error bar: standard deviation. **Fig. 3D: In vitro and in vivo depletion of branched short-chain fatty acids (BSCFAs) by S107 C. sporogenes using CRISPR-dCpf1. Fig. 3D (top):** Schematic view of targeting the BSCFAs gene *porA* using CRISPR-dCpf1. The dCpf1 gRNA (G1) targets the *porA* promoter region. **Fig. 3D (bottom):** The *porA* expression (by qPCR) is significantly reduced in the mutant (**Mut**, dCpf1 with gRNA) compared to the control (**Con**, dCpf1 only) *in vitro*. Germ-free Swiss Webster mice (n = 4 per group) mono-associated with *porA* repression mutant have much less isovalerate (quantified by LC-MS) in their feces than the control (dCpf1 only). Student's T-test was performed, and the

asterisk indicates p -value < 0.05 (*) or < 0.01 (**). Error bar: standard deviation. For (A), (C), and (D), DR: direct repeat, G1: guide RNA-coding sequence 1, G2: guide RNA-coding sequence 2, Ter: terminator. The numbering of the strains corresponds to the strain information shown in Fig. 22.

[0027] Figs. 4A-4D. Knocking out *baiH* in gnotobiotic mice. Fig. 4A: The orientation of the S122 *bai* operon for bile acid 7 α -dehydroxylation. The mutated gene *baiH* (by Group II intron) is highlighted in red. The S122 *bai* operon is actively transcribed under host colonization, and three representative results of metatranscriptomic analyses of the S122 *bai* operon are shown.

Fig. 4B: The biosynthetic scheme of bile acid 7 α -dehydroxylation. The *baiH* encodes an oxidoreductase that reduces the 6,7-olefinic bond of the intermediate 3-oxo-4,5-6,7-didehydro-DCA (**2**, EIC: 385.2384). The S122 Ω *baiH* mutant accumulates the predicted intermediate (**2**, EIC: 385.2384) and no longer converts CA (**1**, EIC: 407.2803) to DCA (**3**, EIC: 391.2854) *in vitro*. The structure of the intermediate (**2**) was determined by comparing its retention time and exact mass to the published literature. The asterisk indicates a residual amount of DCA that is a contaminant from the CA chemical standard. EIC: extracted ion chromatogram. **Fig. 4C:** Germ-free C57BL/6J mice ($n = 3$ or 4 per group) were co-colonized with S25 plus the S122 control (**Con**) or Ω *baiH* mutant (**Mut**) (by Group II intron) strain. The relative abundances of S122 in the control and mutant group were assessed by 16s rRNA sequencing and were comparable. **Fig. 4D:** Depleting *baiH* using Group II intron abolishes gut 7 α -dehydroxylating activity and modifies gut bile acid pool in gnotobiotic mice. CA, DCA, and 7-oxo CA (see Fig. 20 for their structures) were quantified using LCMS. Data in Fig. 4C and Fig. 4D were analyzed using unpaired two-tailed Student's T-test. The asterisk indicates p -value < 0.05 (*) or < 0.01 (**). The numbering of the strains corresponds to the strain information shown in Fig. 22.

[0028] Figs. 5A-5H. Knocking out *baiH* in the context of a complex microbiota impacts the host bile acid pool and the gut microbiome. Fig. 5A: SPF C57BL/6J mice ($n = 5$ per group) given low dose antibiotic water (15 μ g/ml thiamphenicol and 10 μ g/ml erythromycin) were colonized with genetically tagged S122 control (**Con**) or Ω *baiH* mutant (**Mut**) (by Group II intron) strain. **Fig. 5B:** The relative abundances of S122 in the control and mutant group were assessed by 16s rRNA sequencing and were comparable. The SPF mice are stably colonized with S122 control (**Con**) and the Ω *baiH* mutant (**Mut**) at about the same level with a comparable

total bacterial load. **Fig. 5C:** Principal coordinates analysis (PCoA) of the fecal microbiome of the control and *ΩbaiH* mutant mice. **Fig. 5D:** Targeted metabolomics analyses (quantified by LC-MS) of the stool bile acid (BA) compositions of the control (**Con**) and *ΩbaiH* mutant (**Mut**) colonized SPF mice. **Fig. 5E:** The relative abundance of taxonomic phyla in the gut microbiota of the control and *ΩbaiH* mutant mice. **Fig. 5F:** Relative abundances of inflammation-associated gut microbial taxa in the stool microbiome of the control and *ΩbaiH* mutant mice. **Fig. 5G:** Volcano plot of differential bacterial OTU abundances calculated from 16S rRNA gene sequencing. Significantly different OTUs ($n = 56$, $FDR < 0.05$) are colored and plotted. The *Bacteroidia* OTU and *Erysipelotrichaceae* OTU with high relative abundances ($>10\%$) are marked with an upward pointing arrow. **Fig. 5H:** Gut 7 α -dehydroxylating activity is positively associated with fecal calprotectin level in nonIBD people. In **Fig. 5B**, **Fig. 5D**, and **Fig. 5F** data were analyzed using unpaired two-tailed Student's T-test, and the asterisk indicates p -value < 0.01 (**). The data in **Fig. 5C**, **Fig. 5E**, **Fig. 5F**, and **Fig. 5G** are representative of two independent experiments with $n = 4$ or 5 per group, and only the changes in taxonomic groups that are consistent between the two experiments are shown. Data are shown as mean $\pm$ SEM. The numbering of the strains corresponds to the strain information shown in **Fig. 22**.

[0029] Figs. 6A-6J. *baiH* modulates intestinal inflammation in the context of complex gut microbiota. **Figs. 6A, 6F:** DSS-induced murine colitis model was applied to the SPF or gnotobiotic mice colonized with the genetically tagged S122 control (**Con**) and *ΩbaiH* mutant (**Mut**). Mice were colonized with the control or mutant strain for at least two weeks before giving DSS, SPF mice were given 2.5% DSS (in water supplemented with 15 μ g/ml thiamphenicol and 10 μ g/ml erythromycin) for 8 days, and gnotobiotic mice were given 2.0% DSS (in water supplemented with 15 μ g/ml thiamphenicol) for 7 days. The disease state was monitored by weight loss (**Figs. 6B, 6G**), hematoxylin and eosin (H&E) staining of the distal colon (**Figs. 6C, 6H**), colon shortening, and histopathologic score (**Figs. 6D, 6I**), and fecal lipocalin-2 and daily hematochezia score (**Figs. 6E, 6J**). Data shown in **Figs. 6B-6E, 6G-6J** are representations of $n = 4$ to 5 mice per group replicated in two or more independent experiments. In **Figs. 6B, 6G**, % of starting weight was calculated by normalizing weights at sacrifice to starting weight. In **Figs. 6D, 6I** and **Figs. 6E, 6J**, colon length and LCN2 data were analyzed using unpaired two-tailed Student's T-test. In **Figs. 6B, 6G** and **6E, 6J**, % of starting weight and hematochezia score data were analyzed using Two-way ANOVA followed by the Bonferroni

post hoc test (n=4). In **Figs. 6D, 6I**, histopathologic score data were analyzed using the Mann-Whitney test. Data are shown as mean $\pm$ SEM. The asterisk indicates p -value < 0.05 (*), < 0.01 (**) or < 0.001 (***). The numbering of the strains corresponds to the strain information shown in **Fig. 22**.

[0030] Figs. 7A-7J. The *baiH*-mediated microbiota composition shift exacerbates DSS-induced colitis in gnotobiotic mice.

Fig. 7A: The growth curve of two *Bacteroides* (*bac*) microbes and seven *Erysipelotrichaceae* (*Ery*) microbes in the presence of 500 μ M DCA, 500 μ M 3-oxo DCA, or DMSO control. The *Erysipelotrichaceae* microbes are more resistant to DCA and 3-oxo DCA than the *Bacteroides* microbes. **Fig. 7B:** The *baiH* gene drives expansion of *Erysipelotrichaceae* microbes in an *in vitro* consortium consisting of 2 *Bacteroides* (*Bac*) and 7 *Erysipelotrichaceae* microbes (*Ery*) with either the S122 control or Ω *baiH* strain. 500 μ M CA was supplemented as the substrate for the *bai* pathway. The relative fold change of *Erysipelotrichaceae* was assessed by qPCR. **Fig. 7C:** DCA drives expansion of *Erysipelotrichaceae* microbes in an *in vitro* consortium consisting of 2 *Bacteroides* (*Bac*) microbes and 7 *Erysipelotrichaceae* (*Ery*) microbes. DCA was supplemented at 0, 250, and 500 μ M, respectively. The relative fold change of *Erysipelotrichaceae* was assessed by qPCR. **Fig. 7D:** DSS-induced murine colitis model was applied to the gnotobiotic mice colonized with a synthetic consortium consisting of the genetically tagged S122 control (**Con**) or Ω *baiH* mutant (**Mut**) (by Group II intron) along with 2 *Bacteroides* (*Bac*) microbes and 7 *Erysipelotrichaceae* (*Ery*) microbes tested in **(A)**, **(B)**, and **(C)**. Mice were colonized with the control or mutant strain for at least two weeks followed by 2.5% DSS for 8 days. **Fig. 7E:** The *baiH* gene drives expansion of *Erysipelotrichaceae* microbes in the context of host colonization before and during DSS treatment assessed by qPCR. The disease state was monitored by weight loss (**Fig. 7F**), hematoxylin and eosin (H&E) staining of the distal colon (**Fig. 7G**), colon shortening (**Fig. 7H**), fecal lipocalin-2 (**Fig. 7I**), and daily hematochezia score (**Fig. 7J**). The data in **Figs. 7A to 7C** are from a representative experiment with three technical replicates (**Fig. 7A**), or with six or four biological replicates (**Figs. 7B, 7C**). Data shown in **Figs. 7F, 7H, 7I, 7J** are representations of $n = 4$ mice per group replicated in two or more independent experiments. In **Fig. 7F**, % of starting weight was calculated by normalizing weights at sacrifice to starting weight. In **Fig. 7H** and **Fig. 7I**, colon length and LCN2 data were analyzed using unpaired two-tailed Student's T-test. In **Figs. 7F** and **7J**, % of starting weight and hematochezia score data were analyzed using Two-

way ANOVA followed by the Bonferroni post hoc test (n=4). Data are shown as mean $\pm$ SEM. The asterisk indicates p -value < 0.05 (*), < 0.01 (**) or < 0.001 (***). The numbering of the strains corresponds to the strain information shown in **Fig. 22**.

[0031] Fig. 8. A detailed workflow and general timeline of the genetic manipulation (GM) pipeline. The dominant human gut commensals can be screened via the GM pipeline, and their targetable genetic system can be built within weeks.

[0032] Fig. 9. All the pGM vectors used in this study. Schematics of all the pGM vectors designed and used in this study are listed. “x” in pGM-xBCM (or pGM-xBCD, pGM-xBCL, pGM-xBCF/G, pGM-xBCD-xxx) represents different gram-positive replication origins, and “xxx” in pGM-xBCD-xxx corresponds to plasmid pGM-xBCD harboring different gRNA designs targeting genome of different strains (see **Figs. 28** and **35**).

[0033] Figs. 10A-10B. A mixed-conjugation strategy to identify the compatible *rep ori* for *Clostridia* microbe. **Fig. 10A:** A preliminary test of the mixed-conjugation strategy in a model gut commensal *C. sporogenes* ATCC 15579. The *E. coli* conjugation donors each harboring a single *Clostridium rep ori* and antibiotic marker gene (pMTL82254, *rep ori*: pBP1, antibiotic marker: *ermB*, erythromycin; pMTL83353, *rep ori*: pCB102, antibiotic marker: *aad9*, spectinomycin; pMTL84151, *rep ori*: pCD6, antibiotic marker: *catP*, thiamphenicol) were mixed and conjugated to a single recipient *C. sporogenes* ATCC 15579. After conjugation, the transconjugants were selected on agar plates supplemented with D-cycloserine and the one corresponding antibiotic (erythromycin for *ermB*, spectinomycin for *aad9*, and thiamphenicol for *catP*). **Fig. 10B:** Schematic view of a mixed-conjugation strategy to identify the *Clostridia* that stably maintain exogenous DNA. Ten pGM vectors, each harboring a single *rep ori* (9 *Clostridia* specific and 1 *rep ori*-less), were separated into three sets and mixed-conjugated to a *Clostridia* recipient.

[0034] Figs. 11A-11B. Multiplex PCR strategy to identify the *rep ori* uptaken by the *Clostridia* microbes. **Fig. 11A:** Multiplex PCR strategy was used to identify which *rep ori*-contained plasmid was introduced into which Gram+ *Clostridia* strain in mixed-conjugation. For the mixed-conjugation with set I, primers pMTL_*laz*_diag_F (universal forward primer) + pGM-ABCM_*rep*_R_1500bp + pGM-BBCM_*rep*_R_1000bp + pGM-CBCM_*rep*_R_2000bp were

used for diagnostic PCR. We would see a 1.5 kb (or 1.0 kb, or 2.0 kb) PCR band if pGM-ABCM (or BBCM, or CBCM) is uptaken by the *Clostridia* microbe. The primers for the set II and set III mixed-conjugation are shown in **Fig. 30**. **Fig. 11B:** Distribution of *Clostridial rep ori* tested in this study based on phylogeny. The phylogenetic tree was constructed using the 16s rRNA sequences of the 42 gut microbes (38 *Firmicutes/Clostridia*, 2 *Enterococcus*, and 2 *Actinobacteria*) with a compatible *rep ori* identified in this study. The sequences were aligned using Clustal Omega, and a neighbor-joining tree was constructed with a bootstrap test of 5000.

[0035] Figs. 12A-12B. CRISPRi-dCpf1 precisely and efficiently suppressed *lacZα* expression in Gram-positive *Clostridia* and *Bifidobacterium* microbes. Fig. 12A: qPCR results showing that CRISPRi-dCpf1 precisely and efficiently suppressed *lacZα* expression in Gram-positive *Clostridia* and *Bifidobacterium* microbes, using both dCpf1-only and gRNA-only as controls. **Fig. 12B:** qPCR results showing that CRISPRi-dCpf1 precisely and efficiently suppressed *lacZα* expression in other Gram-positive *Clostridia* microbes, using dCpf1-only as control. For each strain shown in **Fig. 12A**, conjugation was conducted with *E. coli* harboring plasmids with dCpf1-*lacZα* (dCpf1), gRNA-*lacZα* (gRNA), or gRNA-dCpf1-*lacZα* (dCpf1+gRNA) (see **Figs. 2B** and **9** for detailed information). For each strain shown in **Fig. 12B**, conjugation was conducted with *E. coli* harboring plasmids with dCpf1-*lacZα* (dCpf1, Con), or gRNA-dCpf1-*lacZα* (dCpf1+gRNA, Mut) (see **Figs. 2B** and **9** for detailed information). Then transconjugants were cultured, and RNA was extracted and reverse transcribed to cDNA. Quantitative PCR (qPCR) was used to assess the expression of *lacZα* after normalizing to the expression of the 16s rRNA gene of each strain. Data are shown as mean ± SD. The numbering of the strains corresponds to the strain information shown in **Fig. 22**.

[0036] Figs. 13A-13D. Diagnostic PCR strategy of 16s-targeting Group II intron (16s-tron) integration for *Clostridia* and chi-16s-targeting single crossover for *Bacteroidia* and microbes from other phyla. Fig. 13A: Diagnostic PCR strategy to verify the 16s-targeting Group II intron (16s-tron) retrotransposition-activated marker (RAM) integration designed in targeted *Clostridia* commensals. The forward diagnostic primer is the sequence on the retrotransposition-activated marker, which will not bind to the genome. The reverse diagnostic primer binds to the genome and will not bind to the Group II intron plasmid. There will be a PCR product of 2.0-2.5 kb as designed for colonies that have integrated the retrotransposition-

activated marker, whereas no PCR product will be found for control colonies. **Fig. 13B:** Three representative gel images of 16s-targeting Group II intron integration in *Clostridia*. There were bands of ~2.0-2.5 kb in colonies after RAM integration using primers described in **Fig. 13A**, while no band was found in colonies of control. **Fig. 13C:** Diagnostic PCR strategy to verify the single crossover designed in targeted *Bacteroidia* commensals and microbes from other phyla. The diagnostic PCR strategy is the same for identifying the genetically targetable *Bacteroidia* and other phyla microbes using chi-16s strategy and for targeted deletion of *mmdA* in three *Bacteroidia* microbes to deplete propionate production. The forward diagnostic primer is the sequence on the target gene of the genome (16s rRNA gene or *mmdA*), which will not bind to the introduced suicide plasmid, the reverse diagnostic primer binds to the suicide plasmid-specific sequence and will not bind to the genome of targeted strains. There will be a PCR product of ~2.0-2.5 kb as designed for colonies that have integrated the suicide plasmid, whereas no PCR product will be found for wild-type colonies. For screening using chi-16s, the sequencing primer is on the plasmid just downstream the chimeric 16s. As the chimeric 16s is integrated into the genome, the nucleotide sequence (from Sanger sequencing) consists of part of the original 16s rRNA sequence and part of the chimeric 16s rRNA sequence. **Fig. 13D:** Alignment of the nucleotide sequence of the PCR product amplified using DiagF and DiagR (as shown in **Fig. 13C**) with the chimeric 16s rRNA sequence and the microbial 16s rRNA sequence for the genetically targetable *Bacteroidia* and other phyla microbes identified using chi-16s.

[0037] Fig. 14. CRISPRi-dCpf1 precisely and efficiently suppressed *bcat* expression in 12 *Clostridia* microbes with sequenced genomes. For each strain, conjugation was conducted with *E. coli* harboring plasmids with dCpf1 (control, **Con**) or gRNA-dCpf1 (mutant, **Mut**) (different gRNA sequences were designed for each *bcat* gene in each strain, see **Fig. 28** for detailed information). Colonies were cultured, and RNA was extracted and reverse transcribed to cDNA. Quantitative PCR (qPCR) was used to assess the expression of *bcat* after normalizing to the expression of 16s rRNA gene of each strain.

[0038] Figs. 15A-15C. Mono-colonization of germ-free mice with the control and mutant strains of propionate, butyrate, and isovalerate. Fig. 15A: The *Bacteroides* sp. 1_1_6 (S25) $\Delta mmdA$ mutant depletes propionate in the mono-associated germ-free mice compared to that of the control mono-colonized mice. Density of intestinal colonization of gnotobiotic mice by the

control (**Con**) and the $\Delta mmdA$ mutant (**Mut**) of *Bacteroides sp.* 1_1_6. Germ-free mice (n=3 or 4 per group) were mono-colonized with the control *Bacteroides sp.* 1_1_6 (the 16s rRNA has been integrated by pGM-NAC₂B) or the mutant (integrated with pGM-NAC₂B-003). The mice were fed a standard diet and supplied with water containing 15 μ g/mL thiamphenicol and 2 mg/mL sugar. We calculated colony-forming units in fecal pellets to estimate the density of intestinal colonization. n.s.: not statistically significant. Propionate levels in cecal samples of germ-free mice mono-colonized with the control (**Con**) and the $\Delta mmdA$ mutant (**Mut**) of *Bacteroides sp.* 1_1_6. Propionate concentration was calculated using the standard curve according to AUC and normalized to each sample's weight. Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**). **Fig. 15B:** The *Eubacterium maltosivorans* DSM 105863 (S117) *croA* knockdown mutant depletes butyrate in the mono-associated germ-free mice compared to that of the control mono-colonized mice. Density of intestinal colonization of gnotobiotic mice by control (**Con**) and mutant (**Mut**) of *Eubacterium maltosivorans* DSM 105863 (S117). Germ-free mice were mono-colonized with the control *Eubacterium maltosivorans* DSM 105863 (containing pGM-FBCD) or the mutant (containing pGM-FBCD-020) and fed a standard diet and supplied with water containing 15 μ g/mL thiamphenicol and 2 mg/mL sugar. We calculated colony-forming units in fecal pellets to estimate the density of intestinal colonization. n.s.: not statistically significant. Butyrate levels in fecal samples of mice colonized with the control (**Con**) and the *croA* knockdown mutant (**Mut**) of *Eubacterium maltosivorans* DSM 105863 (S117). Concentrations of butyrate were calculated using the standard curve according to AUC and normalized to the weight of each sample. Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**). **Fig. 15C:** The *C. sporogenes* ATCC 15579 *porA* knockdown mutant reduces branched-chain short-chain fatty acids production *in vitro* and in the mono-associated germ-free mice compared to that of the control colonized mice. *C. sporogenes* ATCC 15579 *porA* knockdown mutant (**Mut**, carrying dCpf1 and gRNA, pGM-ABCD-006) depleted isovalerate production *in vitro* compared to that of the control (**Con**, carrying only the dCpf1, pGM-ABCD). Density of intestinal colonization of gnotobiotic mice by control (**Con**) and mutant (**Mut**) of *C. sporogenes* ATCC 15579. Germ-free mice were mono-colonized with control *C. sporogenes* ATCC 15579 (containing pGM-ABCD) or the mutant (containing pGM-ABCD-006) and fed a standard diet and supplied with water containing 15 μ g/mL thiamphenicol

and 2 mg/mL sugar. We calculated colony-forming units in fecal pellets to estimate the density of intestinal colonization. n.s.: not statistically significant. Isovalerate levels in cecal samples of mice colonized with the control (**Con**) and mutant (**Mut**) of *C. sporogenes* ATCC 15579. Concentrations of isovalerate were calculated using the standard curve according to AUC and normalized to the weight of each sample. Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**).

[0039] Figs. 16A-16C. Gut *Clostridia* commensals harboring the *bai* operon and their prevalence and relative metagenomic abundances. **Fig. 16A:** Currently known gut commensals that harbor a *bai* operon based on their genomic sequence information. All these microbes are *Clostridia* commensals and have no gene transfer methodology and tractable genetic tools. Their prevalence (**Fig. 16B**) and percent relative abundance (**Fig. 16C**) were examined using a large publicly available 16s rRNA dataset (Yatsunenko et al., 2012), including stool samples from 528 individuals. The *Faecalicatena contorta* S122 (S122) and its closely related relatives are more prevalent than *C. hylemonae*, *C. scindens*, and *D. sp.* D27_3, but less than that of *C. hiranonis* (**Fig. 16B**). S122 and its closely related strains are the most abundant 7 α -dehydroxylating commensal in this cohort (**Fig. 16C**). For **Fig. 16B**, the prevalence data were analyzed using Fisher's exact test, and mean $\pm$ SEM was plotted. For **Fig. 16C**, the relative abundance data were first analyzed using the D'Agostino & Pearson test for normality. The relative abundance of S122 was compared to other commensals using the Mann-Whitney test. A Median with a 95% confidence interval (CI) was plotted. The asterisk in **Fig. 16B** and **Fig. 16C** indicates p -value < 0.0001 (****).

[0040] Figs. 17A-17E. Co-colonize gnotobiotic mice colonized with S122 and S25 or S122 and a consortium of 55 targetable gut commensals. **Fig. 17A:** Proposed pathway for the 7 α -dehydroxylation of cholic acid (CA) to deoxycholic acid (DCA) (Funabashi et al., 2020). **Fig. 17B:** Successful insertion of *baiH* was determined by amplifying DNA using primers flanking the target gene *baiH*. The expected PCR product for the control is ~2kb, and the PCR product for the S122 Ω *baiH* mutant is ~4kb. **Fig. 17C:** Metabolomics analyses of bile acids in *Bacteroides sp.* 1_1_6 (S25) + *Faecalicatena contorta* S122 (S122 Ω *baiH* mutant) (**Mut**) and S25 + S122 control (**Con**) co-colonized germ-free mice. Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**), n.s.:

not statistically significant, n.d.: not detected. **Fig. 17D:** Co-colonization of *Faecalicatena contorta* S122 with 55 other genetically targetable microbes identified in this study in germ-free mice. Successful colonization of S122 was determined by CFU and the detection of DCA in LCMS, and colonization of other strains was confirmed by 16s rRNA sequencing. **Fig. 17E:** The density of total intestinal bacteria of SPF control mice compared with SPF mice colonized with *Faecalicatena contorta* S122 (S122) control and $\Omega baiH$ mutant. SPF control mice (**SPF**) were maintained with a standard diet and water, SPF mice colonized with *Faecalicatena contorta* S122 (S122) control (**Con**) and $\Omega baiH$ mutant (**Mut**) were maintained with standard diet and water with 15 $\mu\text{g/mL}$ thiamphenicol and 10 $\mu\text{g/mL}$ erythromycin. We calculated colony-forming units in fecal pellets on TSAB agar plates to estimate the density of intestinal bacteria. Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**), n.s.: not statistically significant.

[0041] Fig. 18. Structures of taurine-conjugated and nonconjugated bile acids in this study.

[0042] Figs. 19A-19I. The *Faecalicatena contorta* S122 (S122) control and $\Omega baiH$ mutant colonized SPF mice harbor a highly complex gut microbiota and analysis of the correlation between DCA and the relative abundances of gut bacterial taxonomic groups using data from healthy human stools. Figs. 19A-19B: The Chao 1 index (**Fig. 19A**) and Shannon index (**Fig. 19B**) of the fecal microbiota of the SPF mice colonized with the control (**Con**) and the $\Omega baiH$ mutant (**Mut**). Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**). **Figs. 19C-19D:** Relative abundances of *Bacteroidetes* and *Proteobacteria* in the stool microbiome of the control and $\Omega baiH$ mutant colonized SPF mice. $\Omega baiH$ mutant (**Mut**) colonized mice harbor significantly higher abundances of *Bacteroidetes* and lower abundances of *Proteobacteria* compared to the control group (**Con**). Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**). **Fig. 19E:** Relative abundances of gut microbial taxa (at the family level) in the stool microbiome of the control and $\Omega baiH$ mutant colonized SPF mice. Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**). **Fig. 19F:** Observed OTUs rarefaction curves of 16s rRNA gene sequencing of fecal samples from SPF mice colonized with *Faecalicatena contorta*

S122 (S122) control (**Con**) and *ΩbaiH* mutant (**Mut**). This Fig. shows that the Con and Mut mice harbor a highly complex gut microbiota. The depth of our 16s rRNA sequencing is enough to cover the breadth of gut bacterial taxa groups in the control and mutant colonized mice. **Fig. 19G:** Rank-abundance curves of 16s rRNA gene sequencing of fecal samples from SPF mice colonized with *Faecalicatena contorta* S122 (S122) control (**Con**) and *ΩbaiH* mutant (**Mut**). (The X-axis is the OTU rank in descending order, and the Y-axis is the relative abundance of the OTU). **Figs. 19H-19I:** In nonIBD human stools, the fecal DCA level is positively associated with the relative abundance of microbes in the *Erysipelotrichaceae* family and negatively associated with that of the *B. eggerthii* species.

[0043] Figs. 20A-20J. *baiH* modulates colon inflammation in the context of complex gut microbiota. **Fig. 20A:** Macroscopic observations of colon length of SPF mice colonized with *Faecalicatena contorta* S122 (S122) control (**Con**) and *ΩbaiH* mutant (**Mut**) post-DSS treatment. **Fig. 20B:** Macroscopic observations of colon length of germ-free mice colonized with *Faecalicatena contorta* S122 (S122) control (**Con**) and *ΩbaiH* mutant (**Mut**) post-DSS treatment. **Fig. 20C:** Quantification of fecal lipocalin-2 (LCN-2) in SPF mice colonized with *Faecalicatena contorta* S122 (S122) control (**Con**) and *ΩbaiH* mutant (**Mut**) post-DSS treatment on the day-5, 6, 7, and day-sac, the amount of lipocalin-2 was significantly higher in the control group. Data are shown as mean ± SEM. Student's T-test was performed, and the asterisk indicates *p*-value < 0.05 (*) or < 0.01 (**). **Fig. 20D:** Quantification of fecal lipocalin-2 (LCN-2) in germ-free mice colonized with *Faecalicatena contorta* S122 (S122) control (**Con**) and *ΩbaiH* mutant (**Mut**) post-DSS treatment on the day-sac, no significant difference was found. Data are shown as mean ± SEM. Student's T-test was performed, n.s.: not statistically significant. **Fig. 20E:** Comparison of colonic expression of inflammatory genes in SPF mice colonized with *Faecalicatena contorta* S122 (S122) control (**Con**) and *ΩbaiH* mutant (**Mut**) post-DSS treatment. mRNA expression of inflammatory genes was normalized to Hprt1 and shown as the fold change relative to the mutant group. Colonic expression of inflammatory genes was significantly higher in the control group. Data are shown as mean ± SEM. Student's T-test was performed, and the asterisk indicates *p*-value < 0.05 (*) or < 0.01 (**). **Fig. 20F:** Density of intestinal colonization of SPF mice supplemented with low dose antibiotics (15 µg/mL thiamphenicol and 10 µg/mL erythromycin in drinking water) (n = 4 per group) by the S122 control (**Con**) and *ΩbaiH* mutant (**Mut**) (Fig. 6A). The colony-forming units (CFU) were

calculated before DSS treatment and on day 3 and day 6 of DSS treatment. Student's T-test was performed. n.s.: not statistically significant. **Fig. 20G:** Relative abundances (by 16s rRNA sequencing) of *Erysipelotrichaceae* in the stool microbiome of the S122 control (**Con**) and *ΩbaiH* mutant (**Mut**) colonized SPF mice (n=4 per group) at day 3 and day 6 of DSS treatment (**Fig. 6A**). Student's T-test was performed. The asterisk indicates p -value < 0.01 (**). n.s.: not statistically significant. **Fig. 20H:** Fecal DCA levels of the SPF mice (grey bar), and the SPF mice supplemented with low dose antibiotics (n = 4 per group) colonized with the S122 control (**Con**) strain (before and at day 3 of DSS treatment) (**Fig. 6A**). Student's T-test was performed. n.s.: not statistically significant. **Fig. 20I:** The relative abundances of S122 in the control and mutant group (n = 4 per group) of the DSS germ-free experiment (**Fig. 6B**) were assessed by 16s rRNA sequencing and were comparable. Student's T-test was performed. n.s.: not statistically significant. **Fig. 20J:** Targeted metabolomics analyses of the stool bile acid (BA) compositions of the S122 control (**Con**) and *ΩbaiH* mutant (**Mut**) colonized gnotobiotic mice (n = 4 per group) at day 3 after the DSS treatment (**Fig. 6F**). Student's T-test was performed. The asterisk indicates p -value < 0.01 (**). n.s.: not statistically significant, n.d.: not detected.

[0044] Fig. 21. Growth curve of *Bacteroides* and *Erysipelotrichaceae* microbes. Growth curve of *Bacteroides* (*Bacteroides fragilis* 3_1_12 (*Bac1*) and *Bacteroides vulgatus* ATCC 8482 (*Bac2*)) and *Erysipelotrichaceae* (*Clostridium ramosum* ATCC 25554 (*Ery1*), *Erysipelatoclostridium ramosum* strain 113-1 (*Ery2*), *Clostridium ramosum* DSM 24812 (*Ery3*), *Clostridium ramosum* DSM 1402 (*Ery4*), *Clostridium innocuum* 6_1_30 (*Ery5*), *Clostridium innocuum* DSM 22910 (*Ery6*), and *Holdemania filiformis* DSM 12042 (*Ery7*)) were measured in 500 μM of DCA, 3-oxoDCA, CA and 7-oxoCA) with DMSO as control. Optical densities at 600 nm (OD₆₀₀) were recorded every 30min until the cultures reached the stationary phase. Bacterial growth curves were performed in triplicate, with each biological replicate deriving from a single colony.

[0045] Figs. 22A-22L. The *baiH*-mediated microbiota composition shift exacerbates DSS-Induced colitis in gnotobiotic mice. **Fig. 22A:** Density of intestinal colonization (assessed by CFU) of gnotobiotic mice colonized with *Faecalicatena contorta* S122 (**S122**) control (**Con**) and *ΩbaiH* mutant (**Mut**), and total intestinal bacterial load in the 10-member community as shown in **Fig. 7D**. We calculated colony-forming units in fecal pellets to estimate the density of

intestinal colonization. n.s.: not statistically significant. **Fig. 22B:** Density of intestinal colonization (assessed by CFU) of gnotobiotic mice colonized with *Faecalicatena contorta* S122 (S122) control (Con) and $\Omega baiH$ mutant (Mut), and total intestinal bacterial load in the 3-member community as shown in **Fig. 22G**. We calculated colony-forming units in fecal pellets to estimate the density of intestinal colonization. n.s.: not statistically significant. **Figs. 22C-22D:** Metabolomics analyses of fecal bile acids in feces of two *Bacteroides* (bac) microbes and seven *Erysipelotrichaceae* (Ery) microbes + *Faecalicatena contorta* S122 control (Con) / (S122 $\Omega baiH$ mutant) (Mut) co-colonized germ-free mice before DSS treatment (**Fig. 22C**) and on day 3 of DSS treatment (**Fig. 22D**). Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**), n.s.: not statistically significant, n.d.: not detected. **Figs. 22E-22F:** Metabolomics analyses of bile acids in feces of two *Bacteroides* (bac) microbes + *Faecalicatena contorta* S122 control (Con) / (S122 $\Omega baiH$ mutant) (Mut) co-colonized germ-free mice before DSS treatment (**Fig. 22E**) and on day 3 of DSS treatment (**Fig. 22F**). Data are shown as mean $\pm$ SEM. Student's T-test was performed, and the asterisk indicates p -value < 0.05 (*) or < 0.01 (**), n.s.: not statistically significant, n.d.: not detected. **Figs. 22G-22L:** DSS-induced murine colitis model was applied to the gnotobiotic mice colonized with a synthetic consortium consisting of the genetically tagged S122 control (Con) or $\Omega baiH$ mutant (Mut) along with 2 *Bacteroides* (Bac) microbes only. Mice were colonized with the consortium for at least two weeks, followed by 2.5% DSS for 9 days. The disease state was monitored by weight loss (**Fig. 22H**), hematoxylin and eosin (H&E) staining of the distal colon (**Fig. 22I**), colon shortening (**Fig. 22J**), fecal lipocalin-2 (**Fig. 22K**), and daily hematochezia score (**Fig. 22L**). Data shown in **Figs. 22H-22L** are representations of $n = 4$ mice per group replicated in two independent experiments. In **Fig. 22H**, % of starting weight was calculated by normalizing weights at sacrifice to starting weight. In **Figs. 22J** and **22K**, colon length and LCN2 data were analyzed using unpaired two-tailed Student's T-test. In **Figs. 22H** and **22L**, % of starting weight and hematochezia score data were analyzed using Two-way ANOVA followed by the Bonferroni post hoc test ($n=4$). Data are shown as mean $\pm$ SEM. The asterisk indicates p -value < 0.05 (*), < 0.01 (**) or < 0.001 (***). The numbering of the strains corresponds to the strain information shown in **Fig. 22**.

[0046] **Fig. 23. Culture conditions of all the gut commensals screened in this study.**

[0047] Fig. 24. Optimized factors for introducing plasmid DNA into non-model gut microbes.

[0048] Fig. 25. 91 gut microbes and the corresponding compatible plasmids after multifactorial optimization.

[0049] Fig. 26. Genes targeted in gram-positive Firmicutes/Clostridia strains and strains from other phyla.

[0050] Fig. 27. Genes targeted in gram-negative Bacteroidia strains and strains from other phyla.

[0051] Fig. 28. Vectors for the construction of mutants in gram+ and gram- strains.

[0052] Fig. 29. Bacterial strains used in this study.

[0053] Fig. 30. Primers and gRNA sequences used in this study (SEQ ID NOs: 23-287 in order of appearance).

[0054] Fig. 31. Recipes for all the culture media used in this study and their ingredient details.

[0055] Fig. 32. Taxonomy abundances (%) of 16s rRNA gene sequencing of fecal samples from SPF mice colonized with *Faecalicatena contorta* S122 (S122) control (Con) and Ω baiH mutant (Mut).

[0056] Fig. 33. Number of reads and taxonomy of each OTU of 16s rRNA gene sequencing of fecal samples from SPF mice colonized with *Faecalicatena contorta* S122 (S122) control (Con) and Ω baiH mutant (Mut).

[0057] Fig. 34. Effective statistics of 16s rRNA gene sequencing of fecal samples from SPF mice colonized with *Faecalicatena contorta* S122 (S122) control (Con) and Ω baiH mutant (Mut).

[0058] Fig. 35. Nomenclature of all the pGM vectors designed in this study.

[0059] Fig. 36. Putative RM sites (SEQ ID NOs: 288-309 in order of appearance) reduced in the sequence optimization.

DETAILED DESCRIPTION

[0060] It is to be appreciated that certain aspects, modes, embodiments, variations and features of the present methods are described below in various levels of detail in order to provide a substantial understanding of the present technology.

[0061] In practicing the present methods, many conventional techniques in molecular biology, protein biochemistry, cell biology, immunology, microbiology and recombinant DNA are used. *See, e.g.*, Sambrook and Russell eds. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edition; the series Ausubel *et al.* eds. (2007) *Current Protocols in Molecular Biology*; the series *Methods in Enzymology* (Academic Press, Inc., N.Y.); MacPherson *et al.* (1991) *PCR 1: A Practical Approach* (IRL Press at Oxford University Press); MacPherson *et al.* (1995) *PCR 2: A Practical Approach*; Harlow and Lane eds. (1999) *Antibodies, A Laboratory Manual*; Freshney (2005) *Culture of Animal Cells: A Manual of Basic Technique*, 5th edition; Gait ed. (1984) *Oligonucleotide Synthesis*; U.S. Patent No. 4,683,195; Hames and Higgins eds. (1984) *Nucleic Acid Hybridization*; Anderson (1999) *Nucleic Acid Hybridization*; Hames and Higgins eds. (1984) *Transcription and Translation; Immobilized Cells and Enzymes* (IRL Press (1986)); Perbal (1984) *A Practical Guide to Molecular Cloning*; Miller and Calos eds. (1987) *Gene Transfer Vectors for Mammalian Cells* (Cold Spring Harbor Laboratory); Makrides ed. (2003) *Gene Transfer and Expression in Mammalian Cells*; Mayer and Walker eds. (1987) *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); and Herzenberg *et al.* eds (1996) *Weir's Handbook of Experimental Immunology*. Methods to detect and measure levels of polypeptide gene expression products (*i.e.*, gene translation level) are well-known in the art and include the use of polypeptide detection methods such as antibody detection and quantification techniques. (*See also*, Strachan & Read, *Human Molecular Genetics*, Second Edition. (John Wiley and Sons, Inc., NY, 1999)).

[0062] Disclosed herein is a genetic manipulation (GM) pipeline to identify gene transfer methodology and build a genetic tool for non-model human gut commensals on a large scale (201 gut isolates from >140 species in five phyla) (**Fig. 1**). This pipeline efficiently identified the gene transfer methods for 91 non-model gut bacterial isolates (72 species), including 81 previously untransformed microbes, and built their tools for targeted gene manipulation (**Fig. 1**). Of note, gut *Firmicutes/Clostridia* comprises one of the most abundant bacterial groups in

healthy human guts, yet its genetic manipulation is largely unexplored (Waller et al., 2017). Via a multifactorial optimization of their conjugation/transformation conditions, the present disclosure identified the gene transfer methods for 38 non-model *gut Clostridia*, and set up CRISPRi or Group II intron-based genetic tools in 27 of them. The Examples herein demonstrated the utility of these toolsets by modulating short-chain fatty acids (SCFAs) and secondary bile acids *in vitro* and in the context of host colonization. As a proof of principle, one *Clostridia* specific pathway-bile acid 7 α dihydroxylation was selected for further functional investigation. By genetically tagging the *Clostridia* commensal, the *bai* gene in a complex microbiome was manipulated. Provided herein is evidence that the *bai* gene significantly impacts host gut microbiome and bile acid composition and mediates colon inflammation in a complex microbiome.

[0063] The pipeline described here and the related findings represent the first large-scale identification of gene transfer methodology for non-model gut bacterial isolates. This screen greatly expands the manipulatable genes/pathways coded by the gut microbiota. For instance, microbiota pathways encoded by the gut microbes that previously had no tractable genetic tools, like that for butyrate or bile acid 7 α -dehydroxylation, were identified in the library of genetically targetable commensals described herein and manipulated. This library of targetable gut isolates and their genetic tools serve as a starting point for precisely controlling microbiome molecular output and interrogating their effects on host biology. The GM pipeline efficiently identifies gene transfer methods for gut bacterial isolates and develops their gene manipulation tools without prior knowledge of their genome sequence. Both features suggest its application as a useful technology to delineate the genetics for non-model gut *Firmicutes/Clostridia* commensals.

Definitions

[0064] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this technology belongs. As used in this specification and the appended claims, the singular forms “a”, “an” and “the” include plural referents unless the content clearly dictates otherwise. For example, reference to “a cell” includes a combination of two or more cells, and the like. Generally, the nomenclature used herein and the laboratory procedures in cell culture, molecular

genetics, organic chemistry, analytical chemistry and nucleic acid chemistry and hybridization described below are those well-known and commonly employed in the art.

[0065] As used herein, the term “about” in reference to a number is generally taken to include numbers that fall within a range of 1%, 5%, or 10% in either direction (greater than or less than) of the number unless otherwise stated or otherwise evident from the context (except where such number would be less than 0% or exceed 100% of a possible value).

[0066] As used herein, the terms “amplify” or “amplification” with respect to nucleic acid sequences, refer to methods that increase the representation of a population of nucleic acid sequences in a sample. Nucleic acid amplification methods are well known to the skilled artisan and include ligase chain reaction (LCR), ligase detection reaction (LDR), ligation followed by Q-replicative amplification, PCR, primer extension, strand displacement amplification (SDA), hyperbranched strand displacement amplification, multiple displacement amplification (MDA), nucleic acid strand-based amplification (NASBA), two-step multiplexed amplifications, rolling circle amplification (RCA), recombinase- polymerase amplification (RPA)(TwistDx, Cambridge, UK), transcription mediated amplification, signal mediated amplification of RNA technology, loop-mediated isothermal amplification of DNA, helicase-dependent amplification, single primer isothermal amplification, and self- sustained sequence replication (3SR), including multiplex versions or combinations thereof. Copies of a particular nucleic acid sequence generated *in vitro* in an amplification reaction are called “amplicons” or “amplification products.”

[0067] The term "Cas9" or "Cas9 nuclease" refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof (*e.g.*, a protein comprising an active, inactive, or partially active DNA cleavage domain of Cas9, and/or the gRNA binding domain of Cas9). A Cas9 nuclease is also referred to sometimes as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. CRISPR is an adaptive immune system that provides protection against mobile genetic elements (viruses, transposable elements and conjugative plasmids). CRISPR clusters contain spacers, sequences complementary to antecedent mobile elements, and target invading nucleic acids. CRISPR clusters are transcribed and processed into CRISPR RNA (crRNA). In type II CRISPR systems correct processing of pre-crRNA requires a trans-encoded small RNA (tracrRNA), endogenous ribonuclease 3 (rnc)

and a Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3-aided processing of pre-crRNA. Subsequently, Cas9/crRNA/tracrRNA endonucleolytically cleaves linear or circular dsDNA target complementary to the spacer. The target strand not complementary to crRNA is first cut endonucleolytically, then trimmed 3'-5' exonucleolytically. In nature, DNA-binding and cleavage typically requires protein and both RNAs. However, single guide RNAs ("sgRNA", or simply "gNRA") can be engineered so as to incorporate aspects of both the crRNA and tracrRNA into a single RNA species. See, *e.g.*, Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. *Science* 337:816-821(2012), the entire contents of which is hereby incorporated by reference. Cas9 recognizes a short motif in the CRISPR repeat sequences (the PAM or protospacer adjacent motif) to help distinguish self versus non-self. Cas9 nuclease sequences and structures are well known to those of skill in the art (see, *e.g.*, "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*." Ferretti *et al.*, J. J., McShan W. M., Ajdic D. J., Savic D. J., Savic G., Lyon K., Primeaux C., Sezate S., Suvorov A. N., Kenton S., Lai H. S., Lin S. P., Qian Y., Jia H. G., Najar F. Z., Ren Q., Zhu H., Song L., White J., Yuan X., Clifton S. W., Roe B. A., McLaughlin R. E., *Proc. Natl. Acad. Sci. U.S.A.* 98:4658-4663(2001); "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Deltcheva E., Chylinski K., Sharma C. M., Gonzales K., Chao Y., Pirzada Z. A., Eckert M. R., Vogel J., Charpentier E., *Nature* 471:602-607(2011); and "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A., Charpentier E. *Science* 337:816-821(2012), the entire contents of each of which are incorporated herein by reference). Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. In some embodiments, a Cas9 nuclease has an inactive (*e.g.*, an inactivated) DNA cleavage domain, that is, the Cas9 is a nickase.

[0068] A nuclease-defective Cas9 protein may interchangeably be referred to as a "dCas9" protein (for nuclease-"dead" Cas9). Methods for generating a Cas9 protein (or a fragment thereof) having an inactive DNA cleavage domain are known (See, *e.g.*, Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression" (2013) *Cell*. 28; 152(5):1173-83, the entire contents of each of which are incorporated herein by reference). For example, the DNA cleavage domain of Cas9 is known to include two subdomains, the HNH nuclease subdomain and the RuvC1

subdomain. The HNH subdomain cleaves the strand complementary to the gRNA, whereas the RuvC1 subdomain cleaves the non-complementary strand. Mutations within these subdomains can silence the nuclease activity of Cas9. For example, the mutations D10A and H840A completely inactivate the nuclease activity of *S. pyogenes* Cas9 (Jinek *et al.*, *Science*. 337:816-821(2012); Qi *et al.*, *Cell*. 28; 152(5):1173-83 (2013)). In some embodiments, proteins comprising fragments of Cas9 are provided. For example, in some embodiments, a protein comprises one or two Cas9 domains: (1) the gRNA binding domain of Cas9; or (2) the DNA cleavage domain of Cas9.

[0069] The terms “complementary” or “complementarity” as used herein with reference to polynucleotides (*i.e.*, a sequence of nucleotides such as an oligonucleotide or a target nucleic acid) refer to the base-pairing rules. The complement of a nucleic acid sequence as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5' end of one sequence is paired with the 3' end of the other, is in “antiparallel association.” For example, the sequence “5'-A-G-T-3'” is complementary to the sequence “3'-T-C-A-5'.” Certain bases not commonly found in naturally-occurring nucleic acids may be included in the nucleic acids described herein. These include, for example, inosine, 7-deazaguanine, Locked Nucleic Acids (LNA), and Peptide Nucleic Acids (PNA). Complementarity need not be perfect; stable duplexes may contain mismatched base pairs, degenerative, or unmatched bases. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength and incidence of mismatched base pairs. A complement sequence can also be an RNA sequence complementary to the DNA sequence or its complement sequence, and can also be a cDNA.

[0070] As used herein, “conjugation” refers to the temporary direct contact between two bacterial cells leading to an exchange of genetic material (DNA). This exchange is unidirectional, *i.e.* one bacterial cell is the donor of DNA and the other is the recipient. In this way, genes are transferred laterally amongst existing bacterial as opposed to vertical gene transfer in which genes are passed on to offspring. Conjugation is a convenient means for transferring genetic material to bacteria.

[0071] “Cpf1 protein,” as used herein, refers to a Cpf1 wild-type protein derived from Class 2 Type V CRISPR-Cpf1 systems, modifications of Cpf1 proteins, variants of Cpf1 proteins, Cpf1 orthologs, and combinations thereof. Cpf1 proteins include, but not limited to, *Francisella novicida* (UniProtKB—A0Q7Q2 (CPF1_FRATN)), *Lachnospiraceae bacterium* (UniProtKB—A0A182DWE3 (A0A182DWE3_9FIRM)), and *Acidaminococcus* sp. (UniProtKB—U2UMQ6 (CPF1_ACISB)). Cpf1 is the signature protein characteristic for Class 2 Type V CRISPR systems. Cpf1 homologs can be identified using sequence similarity search methods known to one skilled in the art. “dCpf1,” as used herein, refers to variants of Cpf1 protein that are nuclease-deactivated Cpf1 proteins, also termed “catalytically inactive Cpf1 protein,” or “enzymatically inactive Cpf1.”

[0072] As used herein, “expression” includes one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

[0073] As used herein, an “expression control sequence” refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operably linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term “control sequences” is intended to encompass, at a minimum, any component whose presence is essential for expression, and can also encompass an additional component whose presence is advantageous, for example, leader sequences.

[0074] “Gene” as used herein refers to a DNA sequence that comprises regulatory and coding sequences necessary for the production of an RNA, which may have a non-coding function (*e.g.*, a ribosomal or transfer RNA) or which may include a polypeptide or a polypeptide precursor. The RNA or polypeptide may be encoded by a full length coding sequence or by any portion of the coding sequence so long as the desired activity or function is retained. Although a sequence of the nucleic acids may be shown in the form of DNA, a person of ordinary skill in the art recognizes that the corresponding RNA sequence will have a similar sequence with the thymine being replaced by uracil, i.e., “T” is replaced with “U.”

[0075] As used herein, the term “genome” refers to the whole hereditary information of an organism that is encoded in the DNA (or RNA for certain viral species) including both coding and non-coding sequences. In various embodiments, the term may include the chromosomal DNA of an organism and/or DNA that is contained in an organelle such as, for example, the mitochondria or chloroplasts and/or extrachromosomal plasmid and/or artificial chromosome.

[0076] As used herein, the term “group II intron” refers to a class of bacterial retrotransposons that insert site-specifically into DNA target sites by a mechanism termed “retrohoming” in which the excised intron RNA reverse splices into a DNA strand and is reverse transcribed by the intron-encoded protein (a reverse transcriptase). Retrohoming is mediated by a ribonucleoprotein particle that contains the intron-encoded protein and excised intron RNA, with target specificity determined largely by base pairing of the intron RNA to the DNA target sequence. This feature enabled the development of mobile group II introns into bacterial gene targeting vectors (“targetrons”) with programmable target specificity.

[0077] The term “guide sequence” refers to the portion of a crRNA or guide RNA (gRNA) that is responsible for hybridizing with the target DNA.

[0078] As used herein, a “heterologous nucleic acid sequence” is any nucleic acid sequence placed at a location where it does not normally occur. A heterologous nucleic acid sequence may comprise a sequence that does not naturally occur in a cell, or it may comprise only sequences naturally found in the cell, but placed at a non-normally occurring location in the cell. In some embodiments, the heterologous nucleic acid sequence is not an endogenous sequence. In certain embodiments, the heterologous nucleic acid sequence is an endogenous sequence that

is derived from a different cell. In other embodiments, the heterologous nucleic acid sequence is a sequence that occurs naturally in a cell but is then relocated to another site where it does not naturally occur, rendering it a heterologous sequence at that new site.

[0079] “Homology” or “identity” or “similarity” refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. A polynucleotide or polynucleotide region (or a polypeptide or polypeptide region) has a certain percentage (for example, at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99%) of “sequence identity” to another sequence means that, when aligned, that percentage of bases (or amino acids) are the same in comparing the two sequences. This alignment and the percent homology or sequence identity can be determined using software programs known in the art. In some embodiments, default parameters are used for alignment. One alignment program is BLAST, using default parameters. In particular, programs are BLASTN and BLASTP, using the following default parameters: Genetic code=standard; filter=none; strand=both; cutoff=60; expect=10; Matrix=BLOSUM62; Descriptions=50 sequences; sort by =HIGH SCORE; Databases=non-redundant, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+SwissProtein+SPupdate+PIR. Details of these programs can be found at the National Center for Biotechnology Information. Biologically equivalent polynucleotides are those having the specified percent homology and encoding a polypeptide having the same or similar biological activity. Two sequences are deemed “unrelated” or “non-homologous” if they share less than 40% identity, or less than 25% identity, with each other.

[0080] As used herein, the phrase “homologous recombination” refers to the process in which nucleic acid molecules with similar nucleotide sequences associate and exchange nucleotide strands. A nucleotide sequence of a first nucleic acid molecule that is effective for engaging in homologous recombination at a predefined position of a second nucleic acid molecule can therefore have a nucleotide sequence that facilitates the exchange of nucleotide strands between the first nucleic acid molecule and a defined position of the second nucleic acid molecule. Thus,

the first nucleic acid can generally have a nucleotide sequence that is sufficiently complementary to a portion of the second nucleic acid molecule to promote nucleotide base pairing.

Homologous recombination requires homologous sequences in the two recombining partner nucleic acids but does not require any specific sequences. Homologous recombination can be used to introduce a heterologous nucleic acid and/or mutations into the host genome. Such systems typically rely on sequence flanking the heterologous nucleic acid to be expressed that has enough homology with a target sequence within the host cell genome that recombination between the vector nucleic acid and the target nucleic acid takes place, causing the delivered nucleic acid to be integrated into the host genome. These systems and the methods necessary to promote homologous recombination are known to those of skill in the art.

[0081] The term “hybridize” as used herein refers to a process where two substantially complementary nucleic acid strands (at least about 65% complementary over a stretch of at least 14 to 25 nucleotides, at least about 75%, or at least about 90% complementary) anneal to each other under appropriately stringent conditions to form a duplex or heteroduplex through formation of hydrogen bonds between complementary base pairs. Hybridizations are typically and preferably conducted with probe-length nucleic acid molecules, preferably 15-100 nucleotides in length, more preferably 18-50 nucleotides in length. Nucleic acid hybridization techniques are well known in the art. *See, e.g.,* Sambrook, *et al.*, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Press, Plainview, N.Y. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is influenced by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, and the thermal melting point (T_m) of the formed hybrid. Those skilled in the art understand how to estimate and adjust the stringency of hybridization conditions such that sequences having at least a desired level of complementarity will stably hybridize, while those having lower complementarity will not. For examples of hybridization conditions and parameters, *see, e.g.,* Sambrook, *et al.*, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Press, Plainview, N.Y.; Ausubel, F. M. *et al.* 1994, *Current Protocols in Molecular Biology*, John Wiley & Sons, Secaucus, N.J. In some embodiments, specific hybridization occurs under stringent hybridization conditions. An oligonucleotide or polynucleotide (*e.g.*, a probe or a primer) that is specific for a target nucleic acid will “hybridize” to the target nucleic acid under suitable conditions.

[0082] As used herein, the terms “individual”, “patient”, or “subject” are used interchangeably and refer to an individual organism, a vertebrate, a mammal, or a human. In a preferred embodiment, the individual, patient or subject is a human.

[0083] As used herein, “microbiome” refers to the collective genetic content of the communities of microbes that live in and on the human body, both sustainably and transiently, including eukaryotes, fungi, archaea, bacteria, and viruses (including bacterial viruses (i.e., phage)), wherein “genetic content” includes genomic DNA, RNA such as micro RNA and ribosomal RNA, the epigenome, plasmids, and all other types of genetic information. As used herein, the term “gut microbiome” refers to the collective genetic content of the communities of microbes present in the gastrointestinal tract (GIT).

[0084] As used herein, “microbiota” refers to the collective microbes that live in and on the human body, both sustainably and transiently, including eukaryotes, fungi, archaea, bacteria, and viruses (including bacterial viruses (i.e., phage)). “Gut microbiota” as used herein refers to the totality of the microbes present in the GIT, including eukaryotes, fungi, archaea, bacteria, and viruses (including bacterial viruses (i.e., phage)).

[0085] As used herein, “oligonucleotide” refers to a molecule that has a sequence of nucleic acid bases on a backbone comprised mainly of identical monomer units at defined intervals. The bases are arranged on the backbone in such a way that they can bind with a nucleic acid having a sequence of bases that are complementary to the bases of the oligonucleotide. The most common oligonucleotides have a backbone of sugar phosphate units. A distinction may be made between oligodeoxyribonucleotides that do not have a hydroxyl group at the 2' position and oligoribonucleotides that have a hydroxyl group at the 2' position. Oligonucleotides may also include derivatives, in which the hydrogen of the hydroxyl group is replaced with organic groups, *e.g.*, an allyl group. Oligonucleotides of the method which function as primers or probes are generally at least about 10-15 nucleotides long and more preferably at least about 15 to 25 nucleotides long, although shorter or longer oligonucleotides may be used in the method. The exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including, for example, chemical synthesis, DNA replication, restriction endonuclease digestion of plasmids or

phage DNA, reverse transcription, PCR, or a combination thereof. The oligonucleotide may be modified *e.g.*, by addition of a methyl group, a biotin or digoxigenin moiety, a fluorescent tag or by using radioactive nucleotides.

[0086] As used herein, “operably linked” means that expression control sequences are positioned relative to a nucleic acid of interest to initiate, regulate or otherwise control transcription of the nucleic acid of interest. In some embodiments, transcription of a polynucleotide operably linked to an expression control element (*e.g.*, a promoter) is controlled, regulated, or influenced by the expression control element.

[0087] As used herein, the term “polynucleotide” or “nucleic acid” means any RNA or DNA, which may be unmodified or modified RNA or DNA. Polynucleotides include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, RNA that is mixture of single- and double-stranded regions, and hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, polynucleotide refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons.

[0088] A “protospacer sequence” refers to the target double stranded DNA and specifically to the portion of the target DNA (*e.g.*, target region in the genome (*e.g.*, the genome of the target bacterium)) that is fully or substantially complementary (and hybridizes) to a guide sequence of a CRISPR RNA (crRNA). In the case of Type I and II CRISPR-Cas systems, the protospacer sequence is directly flanked by a PAM.

[0089] The term “protospacer adjacent motif” (or PAM) as used herein, refers to a 2-6 base pair DNA sequence that flanks the DNA region targeted for cleavage by the CRISPR system, such as CRISPR-Cas9. The PAM is required for a Cas nuclease to cut and is generally found 3-4 nucleotides downstream from the cut site. The PAM specificity may be a function of the DNA-binding specificity of the Cas nuclease protein.

[0090] As used herein, the term “primer” refers to an oligonucleotide, which is capable of acting as a point of initiation of nucleic acid sequence synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a target nucleic acid strand is induced, *i.e.*, in the presence of different nucleotide triphosphates and a polymerase in an appropriate buffer (“buffer” includes pH, ionic strength, cofactors *etc.*) and at a suitable temperature. One or more of the nucleotides of the primer can be modified for instance by addition of a methyl group, a biotin or digoxigenin moiety, a fluorescent tag or by using radioactive nucleotides. A primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being substantially complementary to the strand. The term primer as used herein includes all forms of primers that may be synthesized including peptide nucleic acid primers, locked nucleic acid primers, phosphorothioate modified primers, labeled primers, and the like. The term “forward primer” as used herein means a primer that anneals to the anti-sense strand of dsDNA. A “reverse primer” anneals to the sense-strand of dsDNA.

[0091] As used herein, “primer pair” refers to a forward and reverse primer pair (*i.e.*, a left and right primer pair) that can be used together to amplify a given region of a nucleic acid of interest.

[0092] The term “promoter” as used herein refers to any sequence that regulates the expression of a coding sequence, such as a gene. Promoters may be constitutive, inducible, repressible, or tissue-specific, for example. A “promoter” is a control sequence that is a region of a polynucleotide sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at which regulatory proteins and molecules may bind such as RNA polymerase and other transcription factors.

[0093] As used herein, the term “recombinant” when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the material is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant)

form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

[0094] As used herein, an endogenous nucleic acid sequence in the cell of an organism (or the encoded protein product of that sequence) is deemed “recombinant” herein if a heterologous sequence is placed adjacent to the endogenous nucleic acid sequence, such that the expression of this endogenous nucleic acid sequence is altered. In this context, a heterologous sequence is a sequence that is not naturally adjacent to the endogenous nucleic acid sequence, whether or not the heterologous sequence is itself endogenous to the organism (originating from the same organism or progeny thereof) or exogenous (originating from a different organism or progeny thereof). By way of example, a promoter sequence can be substituted (*e.g.*, by homologous recombination) for the native promoter of a gene in the cell of an organism, such that this gene has an altered expression pattern. This gene would be “recombinant” because it is separated from at least some of the sequences that naturally flank it. A nucleic acid is also considered “recombinant” if it contains any modifications that do not naturally occur in the corresponding nucleic acid in a cell. For instance, an endogenous coding sequence is considered “recombinant” if it contains an insertion, deletion or a point mutation introduced artificially, *e.g.*, by human intervention. A “recombinant nucleic acid” also includes a nucleic acid integrated into a host cell chromosome at a heterologous site and a nucleic acid construct present as an episome.

[0095] As used herein, the term “replication origins”, “origins of replications” or “*rep origins*” refers to a unique DNA sequence of a replicon at which DNA replication is initiated and proceeds bidirectionally or unidirectionally. It contains the sites where the first separation of the complementary strands occurs, a primer RNA is synthesized, and the switch from primer RNA to DNA synthesis takes place.

[0096] As used herein, a “reporter gene” refers to a polynucleotide sequence encoding a gene product (*e.g.*, polypeptide) that can generate, under appropriate conditions, a detectable signal that allows detection of the presence and/or quantity of the gene product. Reporter genes are often used as an indication of whether a certain gene has been introduced into or expressed in the host cell or organism. Examples of commonly used reporters include: antibiotic resistance genes, fluorescent proteins, auxotrophic selection modules, β -galactosidase (encoded by the

bacterial gene lacZ), luciferase (from lightning bugs), chloramphenicol acetyltransferase (CAT; from bacteria), GUS (β -glucuronidase; commonly used in plants) and green fluorescent protein (GFP; from jelly fish). Reporters or selection modules can be selectable or screenable.

[0097] The term “seed region” refers to the RNA sequence responsible for initial complexation between a target DNA sequence and CRISPR gRNA/nuclease complex. Mismatches between the seed region and a target DNA sequence have a stronger effect on target site recognition and cleavage than the remainder of the crRNA/sgRNA sequence. In some embodiments, a single mismatch in the seed region of a crRNA/gRNA can render a CRISPR complex inactive at that binding site. In some embodiments, the seed regions for Cas9 endonucleases are located along the last -12 nts of the 3' portion of the guide sequence, which correspond (hybridize) to the portion of the protospacer target sequence that is adjacent to the PAM. In some embodiments, the seed regions for Cpf1 endonucleases are located along the first -5 nts of the 5' portion of the guide sequence, which correspond (hybridize) to the portion of the protospacer target sequence adjacent to the PAM.

[0098] As used herein, “selection marker” refers to a gene that confers a trait suitable for artificial selection. Typically host cells expressing the selectable selection marker is protected from a selective agent that is toxic or inhibitory to cell growth. Examples of commonly used selective markers include antibiotic resistance genes. A screenable selection marker (*e.g.*, gfp, lacZ) generally allows researchers to distinguish between wanted cells (expressing the selection module) and unwanted cells (not expressing the selection module or expressing at insufficient level).

[0099] The term “stringent hybridization conditions” as used herein refers to hybridization conditions at least as stringent as the following: hybridization in 50% formamide, 5xSSC, 50 mM NaH₂PO₄, pH 6.8, 0.5% SDS, 0.1 mg/mL sonicated salmon sperm DNA, and 5x Denhart's solution at 42° C. overnight; washing with 2x SSC, 0.1% SDS at 45° C; and washing with 0.2x SSC, 0.1% SDS at 45° C. In another example, stringent hybridization conditions should not allow for hybridization of two nucleic acids which differ over a stretch of 20 contiguous nucleotides by more than two bases.

[0100] As used herein, "16S ribosomal RNA" or "16S rRNA", is a component of the prokaryotic ribosome 30S subunit. The 16S rRNA gene is the DNA sequence corresponding to rRNA encoding bacteria, which exists in the genome of all bacteria. 16S rRNA is highly conserved and specific, and the gene sequence is long enough (about 1,500 base pairs) for informatics purposes. 16S rRNA sequences are used for phylogenetic reconstruction as they are generally highly conserved, but contain specific hypervariable regions that harbor sufficient nucleotide diversity to differentiate genera and species of most bacteria.

[0101] As used herein, a "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid," which generally refers to a circular double stranded DNA loop into which additional DNA segments may be ligated, but also includes linear double-stranded molecules such as those resulting from amplification by the polymerase chain reaction (PCR) or from treatment of a circular plasmid with a restriction enzyme. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, vectors having an origin of replication which functions in the host cell). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and are thereby replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply "expression vectors").

CRISPR-Cas Systems

[0102] CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR-associated (cas) endonucleases were originally discovered as adaptive immunity systems evolved by bacteria and archaea to protect against viral and plasmid invasion. Naturally occurring CRISPR/Cas systems in bacteria are composed of one or more Cas genes and one or more CRISPR arrays consisting of short palindromic repeats of base sequences separated by genome-targeting sequences acquired from previously encountered viruses and plasmids (called spacers). (Wiedenheft, B., et al., *Nature* 482: 331 (2012); Bhaya, D., et al., *Annu. Rev. Genet.* 45: 231 (2014); and Terms, M.P., et. al, *Curr. Opin. Microbiol.* 14: 321 (2011)). Bacteria and archaea

possessing one or more CRISPR loci respond to viral or plasmid challenge by integrating short fragments of foreign sequence (protospacers) into the host chromosome at the proximal end of the CRISPR array. Transcription of CRISPR loci generates a library of CRISPR-derived RNAs (crRNAs) containing sequences complementary to previously encountered invading nucleic acids (Haurwitz, R.E., et al., *Science* 329: 1355 (2012); Gesner, E.M., et al., *Nat. Struct. Mol. Biol.* 18: 688 (2001); Jinek, M., et al., *Science* 337: 816-21 (2012)). Target recognition by crRNAs occurs through complementary base pairing with target DNA, which directs cleavage of foreign sequences by means of Cas proteins. (Jinek et al., *Science* 337: 816-821 (2012)).

[0103] There are at least five main CRISPR system types (Type I, II, III, IV and V) and at least 16 distinct subtypes (Makarova, K.S., et al., *Nat. Rev. Microbiol.* 13: 722-736 (2015)). CRISPR systems are also classified based on their effector proteins. Class 1 systems possess multi-subunit crRNA-effector complexes, whereas in class 2 systems all functions of the effector complex are carried out by a single protein (e.g., Cas9 or Cpf1). As used herein, “CRISPR enzyme”, “Cas protein” and “CRISPR-Cas protein” refer to CRISPR-associated proteins (Cas) including, but not limited to Class 1 Type I CRISPR-associated proteins, Class 1 Type III CRISPR-associated proteins, and Class 1 Type IV CRISPR-associated proteins, Class 2 Type II CRISPR-associated proteins, Class 2 Type V CRISPR-associated proteins, and Class 2 Type VI CRISPR-associated proteins. The Cas protein of the present technology can be selected from the group consisting of Cas9, dCas9, Cpf1, dCpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, and Csf4.

[0104] In some embodiments, the present disclosure teaches using type II and/or type V single-subunit effector systems. Thus, in some embodiments, the present disclosure teaches using class 2 CRISPR systems. Class 2 Cas proteins include Cas9 proteins, Cas9-like proteins encoded by Cas9 orthologs, Cas9-like synthetic proteins, Cpf1 proteins, proteins encoded by Cpf1 orthologs, Cpf1-like synthetic proteins, C2c1 proteins, C2c2 proteins, C2c3 proteins, and variants and modifications thereof. In some embodiments, Cas proteins are Class 2 CRISPR-associated proteins, for example one or more Class 2 Type II CRISPR-associated proteins, such as Cas9, one or more Class 2 Type V CRISPR-associated proteins, such as Cpf1, and one or more Class

2 Type VI CRISPR-associated proteins, such as C2c2. In preferred embodiments, Cas proteins are one or more Class 2 Type II CRISPR-associated proteins, such as Cas9, and one or more Class 2 Type V CRISPR-associated proteins, such as Cpf1. Typically, for use in aspects of the present technology, a Cas protein is capable of interacting with one or more cognate polynucleotides (most typically RNA) to form a nucleoprotein complex (most typically, a ribonucleoprotein complex).

[0105] CRISPR-Cas nucleases and associated RNAs can be repurposed to edit the genomes in bacteria, yeast and human cells. These techniques all rely on the use of a Cas nucleases to introduce double strand breaks at specific loci.

[0106] In addition to gene editing, CRISPR-Cas has been further exploited for CRISPR activation (CRISPRa) and CRISPR interference (CRISPRi) using nuclease-deactivated Cas proteins. CRISPRa and CRISPRi utilize nuclease-deactivated Cas proteins (*e.g.*, dCas9, dCpf1) that cannot generate a double strand, but instead target genomic regions resulting in RNA-directed transcriptional control. CRISPRi utilizes nuclease-deactivated Cas proteins that complexes with gRNA to target promoter regions for transcriptional repression, or knockdown, of the gene. CRISPRa employs nuclease-deactivated Cas proteins fused to different transcriptional activation domains, which can be directed to promoter regions by either standard gRNA or special gRNAs that recruit additional transcriptional activation domains to upregulate expression of the target gene.

CRISPR/Cas9

[0107] In some embodiments, the present disclosure provides gene editing methods using a Type II CRISPR system. In some embodiments, the Type II CRISPR system uses the Cas9 enzyme. Type II systems rely on a i) single endonuclease protein, ii) a tracrRNA, and iii) a crRNA where a ~20-nucleotide (nt) portion of the 5' end of crRNA is complementary to a target nucleic acid. The region of a crRNA strand that is complementary to its target DNA protospacer is hereby referred to as "guide sequence." In some embodiments, the tracrRNA and crRNA components of a Type II system can be replaced by a single-guide RNA (sgRNA)

[0108] Cas9 endonucleases produce blunt end DNA breaks and are recruited to target DNA by a combination of a crRNA and a tracrRNA oligos, which tether the endonuclease via complementary hybridization of the RNA CRISPR complex. DNA recognition by the crRNA/endonuclease complex requires additional complementary base-pairing with a protospacer adjacent motif (PAM) (*e.g.*, 5'-NGG-3') located in a 3' portion of the target DNA, downstream from the target protospacer. (Jinek, M., et al., *Science* 337: 816-821 (2012)). In some embodiments, the PAM motif recognized by a Cas9 varies for different Cas9 proteins.

[0109] In some embodiments, one skilled in the art can appreciate that the Cas9 disclosed herein can be any variant derived or isolated from any source. In other embodiments, the Cas9 peptide of the present disclosure can include one or more of the mutations described in the literature, including but not limited to the functional mutations described in: Fonfara et al., *Nucleic Acids Res.* 42(4):2577-2590 (2014); Nishimasu H. et al., *Cell* 156(5): 935-949 (2014); Jinek M. et al., *Science* 337:816-821 (2012); and Jinek M. et al., *Science* 343 (6176): 1247997 (2014); see also U.S. Pat. App. No. 13/842,859, filed March 15, 2013, which is hereby incorporated by reference; further, see ET.S. Pat. Nos. 8,697,359; 8,771,945; 8,795,965; 8,865,406; 8,871,445; 8,889,356; 8,895,308; 8,906,616; 8,932,814; 8,945,839; 8,993,233; and 8,999,641, which are all hereby incorporated by reference. Thus, in some embodiments, the systems and methods disclosed herein can be used with the wild type Cas9 protein having double-stranded nuclease activity, Cas9 mutants that act as single stranded nickases, or other mutants with modified nuclease activity.

[0110] The present disclosure further envisions the use of catalytically inactivated Cas9 mutants, or dCas9. A non-limiting list of mutations that reduce or eliminate nuclease in Cas9 includes: D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, or A987, or a mutation in a corresponding location in a Cas9 homologue or ortholog. The mutation(s) can include substitution with any natural (*e.g.*, alanine) or non-natural amino acid, or deletion. An exemplary nuclease defective dCas9 protein is Cas9D10A&H840A (Jinek, et al., *Science* 337: 816-821 (2012); Qi, et al., *Cell* 152(5): 1173-1183 (2013)).

CRISPR/CpfI

[0111] In other embodiments, the present disclosure teaches methods of gene editing using a Type V CRISPR system. In some embodiments, the present disclosure teaches methods of using CRISPR from *Prevotella* and *Francisella* 1 (Cpf1).

[0112] The Cpf1 CRISPR systems of the present disclosure comprise i) a single endonuclease protein, and ii) a crRNA, wherein a portion of the 3' end of crRNA contains the guide sequence complementary to a target nucleic acid. In this system, the Cpf1 nuclease is directly recruited to the target DNA by the crRNA. In some embodiments, guide sequences for Cpf1 must be at least 12nt, 13nt, 14nt, 15nt, or 16nt in order to achieve detectable DNA cleavage, and a minimum of 14nt, 15nt, 16nt, 17nt, or 18nt to achieve efficient DNA cleavage.

[0113] The Cpf1 systems of the present disclosure differ from Cas9 in a variety of ways. First, unlike Cas9, Cpf1 does not require a separate tracrRNA for cleavage. In some embodiments, Cpf1 crRNAs can be as short as about 42-44 bases long— of which 23-25 nt is guide sequence and 19 nt is the constitutive direct repeat sequence. In contrast, the combined Cas9 tracrRNA and crRNA synthetic sequences can be about 100 bases long. In some embodiments, the present disclosure will refer to a crRNA for Cpf1 as a “guide RNA.”

[0114] Second, Cpf1 prefers a “TTN” PAM motif that is located 5' upstream of its target. This is in contrast to the “NGG” PAM motifs located on the 3' of the target DNA for Cas9 systems. In some embodiments, the uracil base immediately preceding the guide sequence cannot be substituted (Zetsche, B., et al., *Cell* 163: 759-771 (2015), which is hereby incorporated by reference in its entirety for all purposes).

[0115] Third, the cut sites for Cpf1 are staggered by about 3-5 bases, which create “sticky ends” (Kim D., et al., *Nat Biotechnol.* 34(8): 863-868 (2016)). These sticky ends with ~3-5 nt overhangs are thought to facilitate NHEJ-mediated-ligation, and improve gene editing of DNA fragments with matching ends. The cut sites are in the 3' end of the target DNA, distal to the 5' end where the PAM is. The cut positions usually follow the 18th base on the non-hybridized strand and the corresponding 23rd base on the complementary strand hybridized to the crRNA.

[0116] Fourth, in Cpf1 complexes, the “seed” region is located within the first 5 nt of the guide sequence. Cpf1 crRNA seed regions are highly sensitive to mutations, and even single base

substitutions in this region can drastically reduce cleavage activity (see Zetsche B., et al., *Cell* 163: 759- 771 (2015)). Critically, unlike the Cas9 CRISPR target, the cleavage sites and the seed region of Cpf1 systems do not overlap. Additional guidance on designing Cpf1 crRNA targeting oligos is available on (Zetsche B., et al., *Cell* 163: 759-771 (2015)).

[0117] One skilled in the art will appreciate that the Cpf1 disclosed herein can be any variant derived or isolated from any source. The present disclosure further envisions the use of catalytically inactivated Cpf1 mutants. Thus in some embodiments, the present disclosure teaches dCpf1 mutants. In some embodiments, the dCpf1 of the present disclosure comprises: ddCpf1 (Zhang et al., *Cell Discov.* 3: 17018 (2017); *Francisella novicida* (UniProtKB— A0Q7Q2 (CPF 1 FRATN)), *Lachnospiraceae* bacterium (UniProtKB— A0A182DWE3 (A0A182DWE3 9FIRM)), and *Acidaminococcus* sp. (UniProtKB— U2UMQ6 (CPF1 ACISB)). In preferred embodiments, the dCpf1 of the present disclosure is generated by mutating the catalytic domain AsCpf1, for example, dCpf1 having a D908A mutation, as described by Yamano, T., et al., *Cell* 165: 949-962 (2016), which is incorporated herein by reference in its entirety.

Expression Vectors for Genetically Modifying Gram-negative Commensal Human Gut Bacteria

[0118] In one aspect, the present disclosure provides a bacterial expression vector comprising (a) a nucleic acid encoding a target gene that is conserved in a plurality of human gut commensal gram-negative bacterial species and (b) a heterologous nucleic acid encoding a selectable marker, wherein the selectable marker is an antibiotic resistance gene or an auxotrophic marker. Examples of target genes that are largely conserved in human gut commensal gram-negative bacterial species include, but are not limited to 16s rRNA, 23s rRNA, *mmdA*, *RokA* (Clucokinase gene), and ABC transporter genes. Additionally or alternatively, in some embodiments, the bacterial expression vector further comprises at least one open reading frame encoding a bioluminescent protein, a chemiluminescent protein, a fluorescent protein, a CRISPR enzyme, a Group II intron-encoded protein, at least one sgRNA, at least one Group II intron, or any combination thereof.

[0119] In some embodiments, the target gene is a chimeric sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the 16S

rRNA gene sequence of at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 *Bacteroidia* (e.g., *Prevotella* and *Bacteroides*) microbes. In other embodiments, the target gene is a chimeric sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the 23S rRNA gene sequence of at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 *Bacteroidia* (e.g., *Prevotella* and *Bacteroides*) microbes.

[0120] In some embodiments, the target gene is a chimeric sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the mmdA gene sequence of at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 *Bacteroidia* (e.g., *Prevotella* and *Bacteroides*) microbes. In other embodiments, the target gene is a chimeric sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to the Roka gene sequence of at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 *Bacteroidia* (e.g., *Prevotella* and *Bacteroides*) microbes. In certain embodiments, the target gene is a chimeric sequence that is at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identical to an ABC transporter gene sequence of at least 5, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 *Bacteroidia* (e.g., *Prevotella* and *Bacteroides*) microbes.

[0121] A non-limiting example of a chimeric 16S rRNA sequence is:

CGAATTCCTGCAGCCCGGGTGGGGATGCGTTCCATTAGGTAGTTGGCGGGGTAACG
 GCCCACCAAGCCTACGATGGATAGGGGTCTGAGAGGAAGGTCCCCACATTGGAA
 CTGAGACACGGTCCAACTCCTACGGGAGGCAGCAGTGAGGAATATTGGTCAATGG
 GCGAGAGCCTGAACCAGCCAAGTAGCGTGAAGGATGACTGCCCTATGGGTTGTAAA
 CTTCTTTTATATGGGAATAAAGTTCGGTACGTGTGGGATTTTGTATGTACCATATGAA
 TAAGGATCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGATCCGAGCGTT
 ATCCGGATTTATTGGGTTTAAAGGGAGCGTAGGTGGATCGTTAAGTCAGTTGTGAAA
 GTTTGCGGCTCAACCGTAAAATTGCAGTTGATACTGGAGGTCTTGAGTACAGTAGAG
 GTAGGCGGAATTCGTGGTGTAGCGGTGAAATGCTTAGATATCACGAAGAAGTCCGA
 TTGCGAAGGCAGCTTACTAGACTGCAACTGACACTGATGCTCGAAAGTGTGGGTATC
 AACAGGATTAGATACCCTGGTAGTCCACACAGTAAACGATGAATACTCGCTGTTTG
 CGATATACAGTAAGCGGCCAAGCGAAAGCATTAAAGTATTCCACCTGGGGAGTACGC
 CGGCAACGGTGAAACTCAAAGGAATTGACGGGGGCGCCGACAAAGCGGAGGAACAT
 GTGGTTTAATTCGATGATACGCGAGGAACCTTACCCGGGCTTAAATTGCTGAATTGG
 AACAGCTTAGCCGCAAGGCAGATGTGAAGGTGCTGCATGGTTGTCGTCAGCTCGT
 GCCGTGAGGTGTCGGCTTAAGTGCCATAACGAGCGCAACCCTTATCTTTAGTTACTA
 ACAGGTCATGCTGAGGACTCTAGAGAGACTGCCGTCGTAAGATGTGAGGAAGGTGG
 GGATGACGTCAAATCAGCACGGCCCTTACGTCCGGGGCTACACACGTGTTACAATG
 GGGGGTACAGAAGGCAGCTACCTGGCGACAGGATGCTAATCCCCAAAACCTCTCTC
 AGTTCGGATCGGAGTCTGCAACCCGACTCCGTGAAGCTGGATTCGCTAGTAATCGCG
 CATCAGCCACGGCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCAAGC
 CATGAAAGCCGGGGGTACCTGAAGTACGTGGATCCACTAGTTCTAGAGC (SEQ ID
 NO: 11)

[0122] Additionally or alternatively, in some embodiments, the bacterial expression vector comprises the nucleic acid sequence of SEQ ID NO: 310 (provided below):

1 acgcgttatg agtaaacttg gtctgacagt taccaatgct taatcagtga ggcacctatc
 61 tcagcatct gtctatttcg ttcattcata gttgcctgac tccccgtcgt gtagataact
 121 acgatacggg agggcttacc atctggcccc agtgctgcaa tgataccgcg agaccacgc
 181 tcaccggctc cagatttate agcaataaac cagccagccg gaagggccga gcgcagaagt
 241 ggtcctgcaa cttatccgc ctccatccag tctattaatt gttgccggga agctagagta
 301 agtagttcgc cagttaatag ttgcgcaac gttgttgcca ttgctacagg catcgtggtg
 361 tcacgctcgt cgtttggtat ggcttcattc agctccggtt cccaacgata aaggcgagtt

421 acatgatccc ccatgttggtg caaaaaagcg gttagctcct tcggtcctcc gatcgttgct
 481 agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt
 541 actgtcatgc catccgtaag atgcttttct gtgactgggt agtactcaac caagtcattc
 601 tgagaatagt gtatgcggcg accgagttgc tcttgcccgg cgtcaatacg ggataatacc
 661 gcgccacata gcagaacttt aaaagtgtc atcattggaa aacgttcttc ggggcgaaaa
 721 ctctcaagga tcttaccgct gttgagatcc agttcgatgt aaccactcg tgcaccaac
 781 tgatcttcag catcttttac tttaccagc gtttctgggt gagcaaaaac aggaaggcaa
 841 aatgccgcaa aaaagggaat aaggcgaca cggaaatgtt gaatactcat actcttcctt
 901 tttcaatcat ggccgcggga taaaagtcg gggattgggt aaaaaagg tgttctctc
 961 ttaagagaa atatcgtttt gctaaacagt tgatattgag gtatcattt atcgtaaaag
 1021 acattttgc tacaacattg ctgacggaa atcaacaaat ttagcattt tgtaaaaaag
 1081 tcgctatata atttggtgaa ttggagtta tttcatattt ttgcatcccg aagagttct
 1141 cttaaagaga gaaacatctt ttgcatacct ttccgaccg aattttatg tcgtaaaagag
 1201 gggctttgca ggggggtggac tcagaaagat gagaatagat gactattgta gttgaaacac
 1261 atagaaagtt gctgatatac agaccgatac gcatatcggg atgaaccatg agtacgttct
 1321 ttttcaaaa aacataaata ttgaaaaga gatgcaataa attaaggaga ggtataatg
 1381 gtatttgaaa aaattgataa aaatagttgg aacagaaaag agtattttga ccactacttt
 1441 gcaagtgtac ctgtaccta cagcatgacc gtaaaagtg atatacaca aataaaggaa
 1501 aagggaatga aactataatc tgcaatgctt tattatattg caatgattgt aaaccgcat
 1561 tcagagtta ggacggcaat caatcaagat ggtgaattgg ggatatatga tgagatgata
 1621 ccaagctata caatatttca caatgatact gaaacatttt ccagcctttg gactgagtgt
 1681 aagtctgact taaatcatt ttagcagat tatgaaagt atacgcaacg gtatggaaac
 1741 aatcatagaa tggaggaaa gccaatgct ccggaaaaca ttttaatgt atctatgata
 1801 ccgtggtaaa ccttcgatgg cttaatctg aatttgaga aaggatatga ttatttgatt
 1861 cctattttta ctatggggaa atattataaa gaagataaca aaattatact tccttggca
 1921 attcaagttc atcagcagt atgtgacgga tttcacattt gccgtttgt aaacgaattg
 1981 caggaaatga taaatagta accaataggc cacatgcaac tgtaaatgtt tacgcgtcct
 2041 cggtagcgtc tcttcacaa cagtctcggg ttctgtact atcacaggt catcttctcc
 2101 cgcgaattcc tgcagcccg gtggggatgc gttccattag gtagttggcg gggtaacggc
 2161 ccaccaagcc tacgatggat aggggttctg agaggaaagt cccacattt ggaactgaga
 2221 cacgttccaa actctacgg gaggcagcag tgaggaatat tggatcaatgg gcgagagcct
 2281 gaaccagcca agtagcgtga aggatgactg ccctatgggt tgtaaaactt tttatawgg
 2341 gaataaagtt cgtgacgtgt gggattttgt atgtaccata tgaataagga tcggctaact
 2401 ccgtgccagc agccgcggt atacggagga tccgagcgtt atccggattt attgggttta
 2461 aaggagcgt aggtgatcg ttaagtcagt tgtgaaagt tgcggctcaa ccgtaaaatt
 2521 gcagttgata ctggaggtct tgagtacagt agaggtaggc ggaattcgtg gtgtagcgg
 2581 gaaatgctta gatatacga agaactccga ttgcgaaggc agcttactag actgcaactg
 2641 acactgatgc tcgaaagtgt gggatcaaaa caggattaga taccctggta gtccacacag
 2701 taaacgatga atactcgtg ttgcatat acagtaagcg gccaaagcga agcattaagt
 2761 attccactg gggagtacgc cggcaacggt gaaactcaa ggaattgacg ggggccgca
 2821 caagcggagg aacatgtggt ttaattcgat gatacgcgag gaaccttacc cgggcttaaa
 2881 ttgctgaatt ggaaacagct tagccgcaag gcagatgtga aggtgctgca tgggtgctg
 2941 cagctcgtgc cgtgaggtgt cggcttaagt gccataacga gcgcaaccct tatcttagt
 3001 tactaacagg tcatgctgag gactctagag agactgccg cgttaagatgt gaggaagggt
 3061 gggatgacgt caaatcagca cggcccttac gtccggggct acacacgtgt tacaatgggg
 3121 ggtacagaag gcagctacct ggcgacagga tgctaattcc aaaaacctct ctgattcgg

3181 atcggagtct gcaacccgac tccgtgaagc tggattcgct agtaatcgcg catcagccac
 3241 ggcgcggtga atacgttccc gggccttgta cacaccgccc gtcaagccat gaaagccggg
 3301 ggtacctgaa gtacgtggat ccactagttc tagagccgag tcgacggat cgataagctt
 3361 gatatcgaat tctgcagcc cgggggatcc actagtctta gagcggccgc caccgcggtg
 3421 gaggggaatt cccatgtcag ccgttaagtg ttctgtgtc actcaaaatt gctttgagag
 3481 gctctaaggg cttctcagt cgttacatcc ctggctgtt gtccacaacc gttaaacctt
 3541 aaaagcttta aaagccttat atattcttt tttcttata aaacttaaaa ccttagaggc
 3601 tattaagt gctgatttat attaatttta ttgtcaaac atgagagctt agtacgtgaa
 3661 acatgagagc ttagtacgtt agccatgaga gcttagtacg ttagccatga gggtttagt
 3721 cgtaaacaat gagagcttag tacgttaaac atgagagctt agtacgtgaa acatgagagc
 3781 ttagtacgta ctatcaacag gttgaactgc tgatcttcag atcctctacg cgggacgcat
 3841 cgtggccgga tcaattccgt ttccgctgc ataaccctgc ttcggggtca ttatagcgt
 3901 ttttcggtat tatccatcct ttttcgcag atatacagga ttttgccaaa gggttcgtg
 3961 agactttctt tggtgtatcc aacggcgta gccgggcagg ataggtgaag tagggccacc
 4021 cgcgagcggg ttttcttct tcaactgtcc ttattcgac ctggcggtgc tcaacgggaa
 4081 tctgtctctg cgaggctggc cggctaccgc cggcgtaaca gatgaggga agcggtatggc
 4141 tgatgaaacc aagccaacca ggaagggcag cccacctatc acggaattga tccccctga
 4201 attg (SEQ ID NO: 310)

[0123] In any and all embodiments of the bacterial expression vectors disclosed herein, the antibiotic resistance gene is selected from the group consisting of *catP*, *ermB*, *aad9*, *tetA*, and *ampR*, or the auxotrophic marker is *pyrG*, or *pyrF*.

[0124] In any of the preceding embodiments of the bacterial expression vectors disclosed herein, the CRISPR enzyme is selected from the group consisting of Cas9, dCas9, Cpf1, dCpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, and Csf4.

[0125] Examples of fluorescent proteins include, but are not limited to, GFP, YFP, CFP, RFP, TagBFP, Azurite, EBFP2, mKalama1, Sirius, Sapphire, T-Sapphire, ECFP, Cerulean, SCFP3A, mTurquoise, monomeric Midoriishi-Cyan, TagCFP, mTFP1, EGFP, Emerald, Superfolder GFP, Monomeric Azami Green, TagGFP2, mUKG, mWasabi, EYFP, Citrine, Venus, SYFP2, TagYFP, Monomeric Kusabira-Orange, mKOκ, mKO2, mOrange, mOrange2, mRaspberry, mCherry, dsRed, mStrawberry, mTangerine, tdTomato, TagRFP, TagRFP-T, mApple, mRuby, mPlum, HcRed-Tandem, mKate2, mNeptune, NirFP, TagRFP657, IFP1.4, iRFP, mKeima Red, LSS-mKate1, LSS-mKate2, PA-GFP, PAmCherry1, PAmTagRFP, Kaede (green), Kaede (red),

KikGR1 (green), KikGR1 (red), PS-CFP2, PS-CFP2, mEos2 (green), mEos2 (red), PSmOrange, or Dronpa. Examples of chemiluminescent proteins include, but are not limited to, β -galactosidase, horseradish peroxidase (HRP), or alkaline phosphatase. Examples of bioluminescent protein include, but are not limited to, Aequorin, firefly luciferase, *Renilla* luciferase, red luciferase, luxAB, or nanoluciferase.

[0126] In any and all embodiments of the bacterial expression vectors disclosed herein, the at least one sgRNA specifically hybridizes with a heterologous or endogenous target gene expressed in a gut bacterial host cell, and/or wherein the at least one sgRNA and/or the CRISPR enzyme is operably linked to a constitutive promoter or a conditional promoter. Additionally or alternatively, in some embodiments of the bacterial expression vectors disclosed herein, the at least one Group II intron specifically targets a heterologous or endogenous gene expressed in a gut bacterial host cell, and/or wherein the at least one Group II intron and/or the Group II intron-encoded protein is operably linked to a constitutive promoter or a conditional promoter.

[0127] In another aspect, the present disclosure provides an engineered gram-negative human gut bacterial cell comprising any of the preceding embodiments of the bacterial expression vector described herein, wherein the engineered gram-negative human gut bacterial cell is derived from a family selected from the group consisting of *Enterobacteriaceae*, *Bacteroidaceae*, *Tannerellaceae*, and *Prevotellaceae*. In some embodiments, the engineered gram-negative human gut bacterial cell is derived from *Bacteroides cellulosilyticus*, *Bacteroides cellulosilyticus*, *Bacteroides dorei*, *Bacteroides eggerthii*, *Bacteroides finegoldii*, *Bacteroides fragilis*, *Bacteroides intestinalis*, *Bacteroides nordii*, *Bacteroides oleiciplenus*, *Bacteroides ovatus*, *Bacteroides salyersiae*, *Bacteroides sp.*, *Bacteroides thetaiotaomicron*, *Bacteroides uniformis*, *Bacteroides vulgatus*, *Bacteroides xylanisolvens*, *Parabacteroides faecis*, *Parabacteroides merdae*, or *Prevotella bivia*.

Expression Vectors for Genetically Modifying Gram-positive Commensal Human Gut Bacteria

[0128] Also disclosed herein are bacterial expression vectors comprising a gram-positive bacteria replication origin that are useful for genetically modifying a plurality of human gut commensal gram-positive bacterial species. Examples of suitable gram-positive bacteria replication origin sequences include:

[0129] pBP1 (SEQ ID NO: 1)

ggcgcgccgttctgaatccttagctaattggtcaacaggtactatgacgaagatagcacctggataagtctgtaattgattctaaggcattt
aatgaagacgtgtatataaaatgtgctaattgaaaaagaaaatgcgttaaaagagcctaaaatgagttcaaatggtttgaaattgattggtagttt
aatataatatttttctattggctatctcgatacctatagaatcttctgttcactttgttttgaaatataaaaaggggcttttagccctttttttaa
actcggaggagtttcttcttctgatactatacgtactattttcgatttgacttcattgtcaattaagctagtaaaatcaatgggtaaaaacaaa
aaacttgcattttctacctaagtaattataatttaagtgtcgagtttaaaagtataatttaccaggaaaggagcaagtttttaataaggaaaaattt
ttccttttaaaattctatttgcgttatatgactaattataatcaaaaaatgaaaataacaagaggtaaaaactgcttagagaaatgtactgataaa
aaaagaaaaaatcctagatttacgtacatacatagcacctttaactactaagaaaaatattgaaaggacttccactgtggagattattgtttatgtt
gagtgatgcagacttagaacatttttaattacataaaggtaattttgcggtaatagattttgtccaatgtgtagttggcgacttgcttgaaggata
gttagaaatatctattctatggagcatttaagaaaagaagaaaataaagagtttatattttaactcttacaactccaaatgtaaaaagttagat
cttaattattctattaaacaatataataaatcttttaaaaaattaatggagcgtaaggaagtaaggatataactaaagggtatataagaaaattaga
agtaacttaccaaaaggaaaaatacataacaaaggatttatggaaaaataaaaaagattattatcaaaaaaaggacttgaaattggtgattta
gaacctaattttgatacttataatcctcattttcatgtatttgcagtttaataaaagttattttacagataaaaatttatataaatcgagaaagat
ggttggaaattatggaagtttgctactaaggatgattctataactcaagttgatgttagaaaagcaaaaattaatgattataaagaggtttacgaac
ttcgaaataattcagctaaagacactgattatttaatatcgaggccagttttgaaatttttataaagcattaaaaggcaagcaggtattagtttt
agtggatttttaaaagatgcacacaaattgtacaagcaaggaaaacttgatgtttataaaaagaaagatgaaattaatatgtctatatagtttt
ataattggtgcaaaaaacaatatgaaaaactagaataagggaacttacggaagatgaaaaagaagaattaatcaagtttaatatagatgaa
atagaaatagattaaagttaactatactttatatatatatgattaaaaaataaaaaacaacagcctattaggttggtgttttttttttatttttataattt
tttaatttttagtttttagttcttttttaaaataagtttcagcctcttttcaatatttttaaaagaggagttttgcatgaattgccttttttctaacagact
taggaaatattttacagtatcttcttgcgccggtgattttggaactcataacttactaatttataattatttttcttttaattgtaacagttgcaaa
agaagctgaacctgttcttcaactagtttatcatcttcaatataatattcttgacctatagataaatatattttattatattttactttttctgaatc
tattttttataatcataaaaagttttaccacaaaagaaggtgtactccttctgggtccaacatattttttactatattatctaaataattttgggaac
tggtgttgaatttgattaatcgacaaccagttatacttaaggaattataactataaaaatatataggattatcttttaaaatttcattattggcctcc
ttttattaaatttatgttaccataaaaaggacataacgggaatatgtagaataatttttaattgtagacaaaattttacataaatataagaaaggaag
tggttgttaattttatagcaaaactatcaaaaattagggggataaaaatttatgaaaaaagggtttcgatgttattttatgtttaactttaatagtttg
tggtttatttacaattcggccggcc

[0130] pCB102 (SEQ ID NO: 2)

ggcgcgccgcattattttttgaacaattgacaattcatttctattttttattaagtgatagtcaaaaggcataacagtgtgtaatagaaagaaat
ttacagaaaagaaaattatagaatttagtatgattaattatactcatttatgaatgtttaattgaatacaaaaaaaatacttggttatgtattcaattac

gggttaaaatatagacaaggtgaaaaatttaaaaaaaataagtcctcagctcttatattaagctaccaacttagtatataagccaaaactta
aatgtgctaccaacacatcaagccgttagagaactctatctatagcaatatttcaaatgtaccgacatacaagagaaacattaactatatattc
aatttatgagattatcttaacagatataaatgtaaattgcaataagtaagatttagaagtttatagcctttgtgtattggaagcagtagcgaaggc
ttttttatttgataaaaattagaagtattttttttcataattaatttatgaaaatgaaaggggtgagcaaagtgcagaggaaagcagtagctt
atcaataacaaggtattagcaatatcattattgacttttagcagtaaacattatgacttttatagtgcttgtagctaagtagtacgaaagggggag
ctttaaaagctccttgggaatcacatagaattcataaattaatttatgaaaagaagggcgctatatgaaaactgtaaaaattgcaaagagttattaa
agatactgaaatatgcaaaatacattcgttgatgattcatgataaacagtagcaacctattgcagtaatacaatgagcaagatgtttacataa
agggaaagtccaatgtattaattgttcaaagatgaaccgatatggatgggtgtgccataaaaatgagatgtttacagaggaagaacagaaaa
agaacgtacatgcattaaatattatgcaaggagctttaaaaagctcatgtaaagaagagtaaaaagaaaaataattttatttaatttaatt
gagagtgcgacacagtagtgcactaaaaatatactgtggtgtagttagccgatacaaaaggatagtcactgcattttcataatacatcttat
gttatgattatgtgtcgggtgggacttcacgacgaaaaccacataaaaaaagagttcggggtaggggtaagcatagttgaggcaactaaac
aatcaagctaggatatgcagtagcagaccgtaaggctgtttaggtgtgtgtaatacatagctattaagatgtaaaaatacggataccaat
gaagggaaaagtataatttttgatgtattgtttgttcattctatgggcaaactacgtccaaagccgttccaaatctgctaaaaagtatatccttt
ctaaatcaaagtcaagtagtaaatcataaataaagtttaattttgaagttattatgatattatgttttctattaaaaataaattaagtatatagaatag
ttaataatagtatatacttaattgtgataagtgctgacagtgacagaaaggatgattgttatggattataagcggccggcc

[0131] pCD6 (SEQ ID NO: 3)

ggcgcgcccgccttaagtctaaaaattaggggagatgtaaggatttgggaaaaatagaagatgttataatcataaatatggtattcgtaggct
taaagtcaaaaaggaggtgaaatataaatagatttttagctaaattaagtaagaaataggaggagatttattgaacaaaaattagaaaaacca
tttgatataagagagagtagctatttgactggatatgatgttgaaattttacaaaaatagagtagaacaagcaatatatgtttatgttgggagta
gttgatcatataacatgagagctagaagtagtaaatggagataccatataagaacaaataaagtctatatgttgtaacattaaaaattttatc
ataacttggaattgtttataaaatggaattaaagttgtcagataatattattaatgataagctatactatagcaatatagcagagttgaagaatttg
aaacactagaaaaagctagagaggtagaagtagtattataagtcaatatcaatttttagattctataaatcacatgttaaaacaaaaataattt
attgagtaataaggatagtggttaaacataactaaaaatggaaatacaaaattatttgaaagtaaaaaataacatagaaaaacataagaac
aagccaataatgagataccatatcaactgtcaattcaatacagatggaagtgtaaaagtattacacaggagttgaaccaatattggaattaa
acaaaaaaataaccctaagccgaccaagcagagtagtttttaaaataatatttaagataacaacaaatgagataatactactagacaatgaca
actcaactaccaattgagtttatggagctaccaactccaatatcggcttaactgattaagtagtctgtagtattataataatattgctatcaattttagc
atcttaacaatattattatataactaagctaaaatttcaatagttgtaaaagtgattagtagcaataagtatattttaatgtagtgtagtctcttaa
aaaactagataaggagataataaatatatggaacaattagattcaaaatataagttgaaaaatttctaatggcagtagtttagagatggtagg
acaaggaaataatcttattgataatgaatatgtagagtagtttcaaaataataaaagtaataagtaaacaaactgagagaagaatttaaga
atagtagtaaaacaacttttttaaaatatagatgatagtagaatttaccttcgcaaaaaatatttattatgaaaatacatttttaacctatgtacta

ctgatggaaaagcaggaaccaatgaaaacttaataaataagatatgcattaggatttgattttgacaaaaagaattaggacaagggtttaattat
aaagatataattaattttactaagataggattacattatcatatcctagttgtagtggaatggattccatgtttatgtgctaattaataaaacta
ataacattaagtttagtatcagaagttacaaatacattaataaataaattgggtgcagataaacaagcaaatttatctactcaagtattaagagtac
cttatacatataatattaaaaatactactaaacaagtaaaaaataacaccaagacaaaaatatatagatatgacatagaaaagttagctaaa
aaatattgcaaagatgtaaaaacagtaggtaataactaatacaaaaatatattagatagtaagctaccaaatgtatagtagatattttaaaaaatg
gtagtaaaagatggacataaaaacctagatttgcaaaaaatagttgtgactttaagattgaggaataaaagtttaagtcaagtaatatccgttgct
agagaatggaactatatatcacaaaatagtcctttcaaatagtgagctagaatatcaagtcaagtatatgtatgagaaacttaaacggttaatttt
gggtgtactgggtgtgagtttaatagtgattgttggaataaaatagaatcagattttatatatagtgatgaagatactttgttcaatatgccacataa
gcactcaaaggatttgaaatataagaataggaaaggggttaaaataatgactggtaaatcaattgtttatctataatgtgttacttaacaataaaga
tagagaattaacatagacgatataatggagctgataacctataaacgtaagaagaaagttaaaaacattgttatgagtgaagacattaag
agaaacattaaaagaacttcaacataatgattatattacaaaaacaaaaggtgttacaagctaggaataaaagatacatacaatgtaaaaga
agtttagatgtaatatagataaacaataactattagttactttgttaccatggcagtaatttggggaataatttcaactgaagaattaagattatata
ctcatagagatataagcaagattttattgggtcaaagatgataaaataaaaggaaatatattaagaattaatcaagaggaattagcaaaagattt
ggagtaacacagcaaagaatttcaaatatgatagaatctttattagatactaaaatttttagatgtatgggaaactaaaataaatgatagaggattt
atgtactatacatatagattaaacaagtagatttttgataggattagaattgattttctagtcctattttatgcaaaaaactaattaataaaatttct
tttgtaaaataattgtacgagaatttgcaaaaaaaaatggcatcaaagtattgaaattaagccgttttaaaatttcttttgtaaaataattctac
atatatatgtagtatatatatatgttttttagagaatgtataactagaatatagagctagaatatagagaatgtataactagaatatagagctaga
atatagagaatgtataactagaatatagagctagaatatagagaatgtataactagaatatagagctagaatatagagaatgtataactagaat
atagagctagaatatagagaatgtataactagaatatagagctagaatatagagaatgtataactagaatatagagctagaatcctaattagta
gggtgcttttttaaaacaagttaaaaatcaaaaatagttattagtaagcattggaaatgctagattctaaaatagaaaagtaaaaattgggtgcacta
tctaaacttatctatatcgctttttccgtcgtttgttctctagttacgatacaggggatgcttatattgagttatagtactaatcagtgcttaata
gttaataaaattatagttaccatagtttagtaactatgatgtatgttagttagaaacttgcatcgcggccggcc

[0132] pIM13 (SEQ ID NO: 4)

ggcgcgccgcattcacttctttctatataaatatgagcgaagcgaataagcgtcggaaaagcagcaaaaagttccttttctgtgttgagcat
gggggttcaggggtgcagtatctgacgtcaatgccgagcgaagcgagccgaagggtagcatttacgttagataacccctgatatgctc
cgacgctttatagaaaagaagattcaactaggtaaaatcttaatataggttgagatgataaggttataaggaattgtttgttctaattttcact
cattttgttctaatttctttaacaaatgttcttttttttagaacagttatgatatagttagaatgtttaaaataaggagtgagaaaaagatgaaaga
aagatatggaacagtcataaaggctctcagaggctcatagacgaagaaagtgagaaagtcataaggttagacaagttataccgtaaacaa
acgtctggttaacttcgtaaaggcatatatagtgaattaataagtatgttagatatgattggcggaaaaaaactaaaatcgtaactataccta
gataatgtccacttaagtaacaatacaatgatagctacaacaagagaatagcaaaagctacaggaacaagtcataaacagtaataacaac

acttaaaatcttagaagaaggaaatattataaaaagaaaaactggagtattaatgttaaaccctgaactactaatgagaggcgacgacccaaaa
acaaaaatacctcttactcgaatttgggaactttgagcaagaggcaaatgaaatagattgacctccaataacaccacgtagttattgggaggt
caatctatgaaatgcgattaagggccggcc

[0133] pMU102/ Cthem-based rep origin (SEQ ID NO: 5)

ggcgcgccctgattctgtggataaccgtattaccgctttgagtgagctgataccgctcgccgcagccgaacgaccgagcgagcgagt
cagtgagcgaggaagcggaagagcgcccaatacgcaaaccgctctccccgcgcttgccgattcattaatgcagtatgggaacaaa
atattgcgtatgcgactgtcacatggacgagaaaaccctcacatgcatttaggagtggttctatgcgcctagagggttttcgtgcgtcag
catgagcgatcggaagaaaagaagaatacttttcgcttcaagatgtttgcagcgtgatcgagaacttcgtgagcaaaagaaaagcaaagag
gaaaaatcgcatgatttggagcgataagaaaaagcactcgaatgagtgctttttgcgtttgagcgtagcgaaaaacgagttctttctattct
tgatacatatagaaataacgtcatttttattttagttgctgaaaggtgcgttgaaagtgttggtatgtatgtgttttaaagtattgaaaacccttaaatt
ggttgacagaaaaacccatctgttaaagttataagtaccaaacaataactaaatagatgggggtttctttaattatgtgtcctaatagt
agcatttattcagatgaaaaatcaagggttttagtggaagacaaaaagtgaaaaagtgagaccatggagagaaaaagaaaatcgctaattg
tgattactttgaacttctgcataattctgaattaaaaaggctgaaagagtaaaagattgtgctgaaatattagagtataaacaatacgtgaaac
aggcgaaagaaagttgtatcgagtggtgtttgtaaatccaggctttgtccaatgtgcaactggaggagagcaatgaaacatggcattcagtc
acaaaagggtgttgctgaagttataaacaagaagccaacagttcggttggtgttctcacattaacagttaaaaatgtttatgatggcgaagaatta
aataagagttgtcagatatggctcaaggatttcgccgaatgatgcaataaaaaaataataaaaaatcttggtggtttatgcgtgcaacggaa
gtgacaataaataaagataattcttataatcagcacatgcattgtatgttggtatgttggaaccaactttttaagaatacagaaaaactacgtg
aatcaaaaacaatggattcaattttgaaaaaggcaatgaaattagactatgatccaatgtaaaagttcaaatgattcgaccgaaaaataaat
ataaatcggatatacaatcggaattgacgaaactgcaaatatcctgtaaaggatagcgattttatgaccgatgatgaagaaaagaatttgaa
acgtttgtctgatttggaggaaggtttacaccgtaaaagggttaatctcctatggtggtttgttaaaagaatacataaaaaataaaccttgatga
cacagaagaaggcgatttgattcatacagatgatgacgaaaaagccgatgaagatggattttctattattgcaatgtggaattgggaacggaa
aaattattttataaagagtagttcaacaacgggattgacttttaaaaaggattgattctaatagaagaagcagacaagtaagcctcctaatt
cacttagataaaaaatttaggaggcataatcaatgaacggccggcc

[0134] pAMβ1 (SEQ ID NO: 6)

ggcgcgccctcacgttaagggttttggctatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagttttaaatcaatct
aaagtatatatgagtaaaccttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgc
ctgactccccgtctgtagataactacgatacgggaggggttaccatctggccccagtgctgcaatgataccgcgagaccacgctcaccg
gtccagatttatcagcaataaaccagccagccggaaggggccgagcgcagaagtgttcctatttaatacatttgactagcaataactaaciaa
caagacacacacacaaaaatcaaaaattcactacttttagttaaaaaccacgtaaccacaagaactaatccaatccatgtaatcgggttcttc

aaatatttctccaagattttcctcctctaatatgctcaacttaaatgacctattcaataaatctattatgctgctaaatagtttataggacaaataagta
tactctaatagacctataaaagatagaaaattaaaaaatcaagtgttcgcttcgctctcactgcccctcgacgttttagtagcctttccctcacttcg
ttcagccaagccaactaaaagtttcgggctactctcctctctcccctaataattaataaaatcttactctgtatattctgctaatacttcgcta
aacagcaaagaaaaacaaacacgtatcatagatataaatgtaatggcatagtgccgggtttatttcagcctgtatcatagctaacaacatcg
agttgtgtgtccgttttagggcggttctgctagctgtttaagctcttgatgaatgtatgctctaagtcaaaagaatttgcagcgcctttatatag
ctttcttttctcttttttactttaatgatcgatagcaacaatgatttaacactagcaagttgaatgccaccatttctcctggtttaactttaagaaa
atttctgattcgccttcagtaccttcagcaatttatctaagtccgttcaggaatgcctagcacttctctaactcttttttggtcgtcactaaataag
gcttgatacatcgctttttcgctaataataagccattaaactctttccattctgacaaatgaacacgttgacgttcgcttctttttctgaatttaa
accacccttgacggacaaataaatctttactggttaaatcacttgatacccaagctttgcaaagaatggtaatgtattccctattagccccttgat
agtttctgaataggcacttctaacaattttgattacttcttttctctaagggttgatctaactgattattaaactcaaacatattatattcgcacgttt
cgattgaatagcctgaactaaagtaggctaaagagagggtaaacatgacgttattacgccctattaaacccttttctcctgaaaatttcgttcgt
gcaataagagattaaaccagggttcactctactgtttttgcttctgtaccgcttaaaaccggttagacttgaacgagtaaagcccttattatctgtt
tgtttgaaagaccaatcttgccattctttgaaagaataacggtaattaggatcaaaaaattctacattgtccgttcttggtatgcgagcaatacca
aaatgattacacgttagatcaactggcaaagactttccaaaataattctcggatattttgcgaaattattttggctgctttgacagatttaaattctgat
tttgaagtcacatagactggcggtttctaaaacaaaatatgcttgataacctttatcagatttgataatcatagtaggcataaaacctaaatcaatag
cgggttgtaaaatatcgcttgctgaaatagtttcttttgcggtgtgaatatcaaaaatcaataaagaaggtattgatttgcttaaatgttttcagaatg
tccttctgtgtatgaacggttttcgtctgcatacggttcataacgataaacgttgggtgtccaatgtgtaaatgtatcttgattttctgaatcgcttc
ctcggaagtcagaacaacaccacgaccgccaatcatgcttgattttgagcgatacgcaaaaatagccccttgcttttacctggcttggtagtg
attgagcgaattttactatttttaattgtactttaacaagccgtcatgaagcacagtctacaacaaaagggatattcattcagctgttctccttt
cctataaatcctataaaataggttggtttaattaacttggtttgcttttcttcaactgtttcaatattgcatgttttgaaaaagatttttctttataagt
caattttttccactaatcgaataaattattttgtattttctattaacttatataataatcttccccctccgaagaaaaatacttatctgattttgtttctaa
gtagatatttctcttttctaactctttctaaacgtttctagtgtatagatatttgctaattttcttatctccaataaactatttttatataagttttacattcat
catgattcatacaaactccacttctataaatgaatacaaaaaagcaatcaaacgatttccgattgattgcttaacaattcttaaatcagtagct
tagatacttgaaaactctctgatttccctatataatgatagtagcgggtatataccgtcttcaacaaagttaattaataacttctacgaggggaaga
gttcatctgactaactgataagcggttggtttggcaatcttatcgggctatgcatttataaaatgtcgtcaaacattttataaatgtgtcatggctctttt
ttcgtttctattcagttcgttgtttcgttatctagtagtaccgcttttaaaaaaaataagcaacgatttcgtgcattattcacacgaagtcattgcttt
tttcttctccatttctaaatccaatgttactgttctgattctgtttctggttctggttctggttgcatttgggattaaatccactactagcgttgagtt
agtttaactttgcaattttgttctagtgttttatgggttgatctgattttcctgggcccggcc

[0135] pWV01 (SEQ ID NO: 7)

ggcgcgccgcagcgaagatgtgtctgtagattatgaaagccgatgactgaatgaaataaagcgcgccctctatttcggttgag
gaggctcaagggagtatgagggaatgaaattccctcatgggttgattttaaaaattgcttgcaatttgcgagcggtagcgctggaaaattt
tgaaaaaatttggatttggaaaaaatggggggaaggaagcgaatttgcctccgtactacgacccccattaagtccgagtgccaatt
tttgtgcaaaaacgctctatcccaactggctcaagggttaaggggttttcaatcgccaacgaatcgccaacgttttcgccaacgtttttata
aatctatatttaagtagctttattgtgttttatgattacaaagtatacactaactttataaaattatttgattggagtttttaaatggtgatttcagaa
tcgaaaaaagagttatgatttctctgacaaaagagcaagataaaaaattaacagatatggcgaaacaaaaaggttttcaaatctgcggttg
cggcgttagctatagaagaatatgcaagaaaggaatcagaacaaaaaaataagcgaagctcgcgtttttagaaggatacaggttttcgt
actgttttgataaggaattatcatggctattaaaaatactaaagctagaaatttggattttattatctgactcaattcctaattgattggaa
agaaaaattagagagtttgggcgtatctatggctgtcagtcctttacagatatggacgaaaaaaaagataagatacatggaatagtagta
tgttatacgaaatggaaagcactataaaaaaccacactatcacgttatatatattgcacgaaatcctgtaacaatagaaagcgttaggaacaag
attaagcgaaaattgggaatagttcagttgctcatgttgagatacttgattatatcaaaggtcatatgaatttgactcatgaatcaaaggac
gctattgctaagaataaacatatatacgacaaaaaagatatttgaacattaatgattttgatattgaccgctatataacacttgatgaaagccaaa
aaagagaattgaagaatttacttttagatatagtgatgactataatttggtaaatacaaaaagattaatggctttattcgcttaggggagcgga
gtttggaattttaatacgaatgatgtaaagatatgtttcaacaaactctagcgccttagattatggttgagggaattatcagtggtgatata
gagcaagttatgcaaaggttctgatgtgaaacgggggaaataaaatgacaaacaaagaaaaagagttatttgctgaaaatgaggaattaa
aaaaagaaattaaggacttaaaagagcgtattgaaagatacagagaaatggaagtgaattaagtacaacaatagattttatgagaggaggg
attattgaataaataaaaagccccctgacgaaagtcgaaggggttttatttggttgatgttgcgattaatagcaatacaaggccggcc

[0136] pMB1 (SEQ ID NO: 8)

ggcgcgccgcgcgggcctcagcctgcggaacgcgcagcggacgcccagcggtcagacggctcagaaacgtccgtgagtgccctcc
acgcgccgaacaggtcagggaggtcgcgcatacgtgagcggcgtggagaagcgggtgaaggccgtccagcggcttttcgtgcagg
atgtgctgggctgggttcagccgacgcttcgtgggctgaaatatctgacttggtcccgcgtatttgttactgtacaaatagatgtatgctgt
agccatgtccgatgagtattcgcagccgacgcttgagctgtcgcgcacgttcgaaggctggtggctgccgaacgcccgtgtgctgcga
cgacgactactcccggtgcaccgcaggagccgcgcccagcgcgtcaaatgcaagcacatcgaggcgaaccccgccgctggtgaa
cacgatcgtggtggacatcgacgacggaacgccaaggcgtggccctgtgggagcacgaggcatcggccgaactggatcgcgga
gaacccggccaacgggcacgctcacgaggctgggtgctcacctttccggtgccagaaccgatctggcgcgtctcaagccgttgaagc
tcctgcacgccaccacggagggtgctgcccgtcctgcgacggggacatgggctattcgggacttctgatgaagaaccccgagcatccg
gcgtggggcgtcgacatcatcgagtgggacacctacgacctggaacagctcgtgcagtcgctccaggaacacggggacatccgccc
tcagctggaagcgcaccaagcgcgcccgcagcaggggtgggacgcaactgcacgctcttcgacaaggcccgacgctcgcctacc
gctacgttgcggcggctgccgaccgttcggaggccagcagcgaggcattgcgcctatacgtgcgtgcacctgccacgaactcaacgtct
cgctgttccccgatccgctgcacgcgctgaggtcgaggacatcgccaagagcatccacaaatggatcgtcaccgcagccgcatgtgg

cgcgacgggtgccattgccaacgcagccacattcatcgccatccaatccgcacgaggacacaaacacgggtgagaacaaatatcagcaggt
catgaaggaggcactggaatggtaaggacgactttgaggaagaagcgcccggtgtctgcacgtgaattagctgaagcatacggcgctctcc
acgcgcaccattcagagctgggtggcaatgaagcgcgaggattggattgatgaacaagccgctatgcgcgaagcagtcgctcatatcac
gatgacgagggccatacatggccgcagaccgaggcatttaacatgagccagggtgccgtgcgtcaacgctgctacagggctcgcaa
ggagcgcgaggacgagggcgaggagaaatcgaagcatctacccggcgagattccactgttcgactgacgctaacgctgtcccaaacgc
gaacgcagcacctccctcgcttgcggcttttctcttccatcgcccttcggcactcggtgtgtctccagcgccgcagggcgcgaggg
ctgcggccggcc

[0137] pIP404 (SEQ ID NO: 9)

ggcgccccgaagaacgtttccaatgatgagcacttttaattaaaaatgaagttttaaacttcatttttaatttaaaaaatgaagttttat
caaaaaattttccaataatcccactctaagccacaaacacgcctataaaatcccgcttaatcccactttgagacacatgtaattactttacg
ccctagtatagtgataatttttacattcaatgccacgcaaaaaataaaggggcactataaaaagttccttcggaactaactaaagtaaaaa
attatctttacaacctccccaaaaaaagaacaggtacaaagtacctataatacaagcgtaaaaaaatgagggtaaaaataaaaaataaa
aaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaa
aatataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaa
aaaataaaaaatatttttttaagtttgaaaaaatttttatattataatctttgaagaaaagaataaaaaatgagcctttataaaagccc
attttttcatatacgaatatgacgttctaattgttttattggtacttctaacttagagtaatttcttttttaagccttttcttaagggttttattt
tttcttaatacatttaattcctctttttgttgcctttccttagcttttaattgctcttgataattttttacctctaattttctctctcttatattcctttta
gaaattattattgcatatattttgttcttcttctgaatttctaataactctataagagtttcattcttatacttataattgcttattttatctaaataacatct
ttcagcacttctagtgtcttataacttctttcacttaaatgttgcctaaacatactattaagttctaaaacatcatttaagccttctcaatgtcttct
gtaaagctacaaagataatatctatataaaaaataataagctctctgtgccttttaaatcatattctcttagttcacaaagttttattatgtcttgatt
cttcataatataaacttcttctataaatataattttttgcttggtctacccttttctttcatatggtttaattcaggtaaaaatccattttgtattt
ctcttaagtcataaatatattcgtactcatctaataattgactactgttttgatttagagtttatacttctggaactcttaattctggttgcatctaa
ggcttgctctatctgctcaaagtattttaattgattatataaatattctgaaccgctttccataatggtaatgctttactaggtactgcatttattatcc
atattaaatcacctccttccactatctattacatagtttggtataggaatactttgattaaaaataattcttttctaagtcattaataacctggctttagt
tttgccagttttataataatccaagtctataaacagtgtatttaactcttttatattttctaactgcctacacggcttataaaagggtatttagagttatata
gatattttcatcactcatatctaaatcttttaattcagcgatttatagtgccattggctatatccctttttatctataacgctcctgggtatccacccttta
cttctactatgaatattatctatagttctttttattcagctttaatgcgtttctcacttattcacctccccttctgtaaaactaagaaaattatatcatatt
ttcaataatttataactattcttaaaccttaataaaaaatagagtaagtcccaattgaaacttaactctattttttatgttttaattattttttattaaaa
tattttaaactaaattaatgattcttttaatttttactatttcattccataatataattactataatttttacaataatatttcttcatttgtaatttagat
gatttactaatttttagtttttatattaaataaataatgtataattttatataaaaaatcaaaggagcttataaattatgattatttccaaagatactaaag

atttaatttttcaattttaacaatactttttgtaatatattgtttaaatttaattgtattttttcatataataaagccgttgaagtaaaccaatccattttcct
tatgatgggcccggcc

[0138] In some embodiments, the bacterial expression vectors of the present technology comprise a gram-positive bacteria replication origin comprising a sequence selected from among:

[0139] pBP1 (*C. botulinum*) (SEQ ID NO: 311)

[0140] gaatggcgaatggcgctagcataaaaaataagaagcctgcatttgcaggcttctattttatggcgcgccgttctgaatccttagcta
atgggtcaacaggttaactatgacgaagatagcacctggataagtctgtaattggattctaaggcatttaataagacgtgtatataaaatgtgct
aatgaaaaagaaaatgcgttaaaagagcctaaaatgagtcaaatggtttgaaattgattggtagtttaatttaatatatttttctattggctatct
cgatacctatagaatcttctgttcacttttgttttgaaatataaaaaggggcttttagcccccttttttaaaactccggaggagtttcttcattctg
atactatacgttaactattttcgatttgacttcattgtcaattaagctagtaaaatcaatggtaaaaaacaaaaactgcatttttctacctagtaatt
tataattttaagtgtcgagtttaaaagtataatttaccaggaaaggagcaagtttttaataaggaaaaattttccttttaaaattctatttcgttatat
gactaattataatcaaaaaaatgaaaataacaagaggtaaaaactgcttagagaaatgtactgataaaaaaagaaaaatcctagatttacg
tcatacatagcacctttaactactaagaaaaatattgaaaggacttccactgtggagattattgtttatgttgagtgtgcagacttagaacattt
taaattacataaaggtaatttttgcggtaatagattttgtccaatgtgtagtggcgacttgcttgaaggatagtttagaaatatctattcttatgga
gcatttaagaaaagaagaaaaataagagtttatatttttaactcttacaactccaaatgtaaaaagttatgatcttaattattctattaaacaatataa
taaactcttttaaaaaattaatggagcgttaagggaagttaaggatataactaaaggttatataagaaaattagaagtaacttacaaaaggaaaaat
acatacaaaaggatttatggaaaataaaaaaagattattatcaaaaaaaggacttgaaattgggtgatttagaacctaattttgatacttataatcc
tcattttcatgtagtatttgcagttaataaaagtattttacagataaaaaattattatataaatcgagaaagatggttggaattatggaagtttgctact
aaggatgattctataactcaagttgatgttagaaaagcaaaaattaatgattataaagaggtttacgaacttgcgaaatattcagctaaagacac
tgattatttaatatcgaggccagttttgaaatttttataaagcattaaaaggcaagcaggtatttagtttttagtgatttttaagatgcacacaa
attgtacaagcaaggaaaacttgatgtttataaaaagaaagatgaaattaaatgtctatatagtttattataattgggtgcaaaaaacaatatgaa
aaaactagaataagggaacttacggaagatgaaaaagaagaattaaatcaagatttaatatagatgaaatagaaatagattaaagtgaactata
ctttatatatatatgattaaaaaaataaaaaacaacagcctattaggttggtgttttttattttcttattaattttttaatttttagtttttagtctttttaa
ataagtttcagcctcttttcaatatttttaagaaggagatttgcataaattgccttttttctaacagacttaggaaatattttaacagtatcttcttg
cgccgggtgattttggaacttcataactactaatttataattatttttctttttaattgtaacagttgcaaaaagaagctgaacctgttccttcaacta
gtttatcatcttcaatataatattcttgacctatatagtataaatatatttttattatatttttactttttctgaatctattattttataatcataaaagttttac
caccaaaagaaggttgactccttctggtccaacataatttttactatattatctaaataattttgggaactgggtgtgtaatttgattaatcgaaca
accagttatacttaaagggaattataactataaaaaatatataggattatcttttaatttcattattggcctcctttttattaaatttatgtaccataaaa

aggacataacgggaatatgtagaatTTTTaatgtagacaaaattttacataaatataaagaaaggaagtgtttgtttaaattttatagcaaactat
 caaaaattagggggataaaaaatttatgaaaaaaaggttttcgatgtatttttatgtttaactttaatagtttggtttatttacaattcggccggcc
agtgggcaagttg

[0141] pCB102 (*C. butyricum*) (SEQ ID NO: 312)

[0142] gaatggcgaatggcgctagcataaaaaataagaagcctgcatttgcaggcttctattttatggcgcgccgccattatttttgaaca
 attgacaattcatttctatttttattaagttagatgtagcaaaaggcataacagtgctgaatagaaagaaatttacagaaaagaaattatagaattta
 gtatgattaattatactcatttatgaatgtttaattgaatacaaaaaaaaaaatacttgttatgtattcaattacgggttaaaatagacaagttgaaaa
 atttaataaaaaaataagtcctcagctcttatattaagctaccaacttagtatataagccaaaacttaaatgtgctaccaacacatcaagccgtt
 agagaactctatctatagcaatatttcaaatgtaccgacatacaagagaaacattaactatatattcaatttatgagattatcttaacagatataa
 atgtaaattgcaataagtaagatttagaagtttatagcctttgtgtattggaagcagtagcgaagggctttttatttgataaaaattagaagtattt
 tatttttcataattaatttatgaaaatgaaaggggtgagcaaaagtacagaggaaagcagtagcttatcaataacaaggtatttagcaatattc
 attattgacttttagcagtaaacatttatgactttttagtgcttagctaaagtagtagcgaagggggagctttaaaaagctccttggaatacatag
 aattcataaattaatttatgaaaagaagggcgctatatgaaaacttgtaaaaattgcaaagagtttattaaagatactgaaatatgcaaatacatt
 cgttgatgattcatgataaaacagtagcaacctattgcagtaatacaatgagtcagatgtttacataaagggaagtcgaatgtattaattgtt
 caaagatgaaccgatatggatgggtgtgccataaaaatgagatgtttacagaggaagaacagaaaaagaacgtacatgcattaatatattatg
 caaggagctttaaaaaagctcatgtaaaagaagagtaaaaagaaaaataatttatttattaatttaattatgagagtgccgacacagtagtcact
 aaaaaatatactgtgggtgtagtgagccgatacaaaaggatagtcactcgcattttcataatacatcttatgttatgattatgtgtcgggtgggactt
 cacgacgaaaaccacaataaaaaaagagttcggggtaggggttaagcatagttgaggcaactaaacaatcaagctaggatatgcagtagc
 agaccgtaagggtcgttgttttaggtgtgttgaatacatagctattatgaataaacgataccaatgaagggaaggtataatttttgat
 gtagtttgtttgttcatctatgggcaaaactacgtccaaagccgtttccaaatctgctaaaaagtagtatcctttctaaatcaagtcagtagtaa
 tcataaataaagtttaattttgaagttattatgatattatgttttctattaaaataaattaagtagtatagaaatagtttaataatagtagtatatacttaattgtgat
 aagtgtctgacagtgtagcagaaaggatgattgttatggattataagcggccggccagtgaggcaagttg

[0143] pCD6 (*C. difficile*) (SEQ ID NO: 313)

[0144] gaatggcgaatggcgctagcataaaaaataagaagcctgcatttgcaggcttctattttatggcgcgccccgcccttaagtctaaaa
 attaggggagatgtaaggatttgggaaaaatagaagatgttataatcataaatatggatttcgtaggcttaaagtcaaaaaggaggtgaaatat
 aaatagatttttagctaaattaagtaagaaataggaggagatttattgaacaaaaattagaaaaaccattttagtatataagagagagtagcatttg
 actggatatgatgtgaaattttacaaaaatagtagttagaacaagcaatatatgtttatgttgggagtagttgtgcatataacatgagagctaga
 agtagtaaatggagataccatataagaacaaataaagtagtatgttgaacattaaaaattttatataacttggaattgtttataaaatgga
 attaaagttgtcagataatattattaatgataagctatactatagcaatatagcagagtttgaagaatttgaaacactagaaaaagctagagaggt

agaaagtactattataagtcaatatcaatttttagattctataatcacatgttaaaacaaaaataattttattgagtaataaggatagtggttaaa
cataactaaaaatggaaatacaaatttttgaaagtaaaaaataatacatagaaaaacataagaacaagccaataatgagataccatatcaa
ctgtcaattcaatacagatggaagtgtcaaaagtattacacaggagtttgaaccaatattggaattaaacaaaaaaataccctaagccgacc
aagcagagtatttttaaaataatatttaagataacaacaaaatgagataatactactagacaatgacaactcaactaccaattgagtttatggag
ctaccaactccaatatcggctaaactgattaagtatctgtagtatatataataattgctatcaatttttagcatcttaacaatattattatacatactaa
gctaaaattattcaatagttgtaaaagttgattagtcataagtatataattatgtagtgttatctttaaaaaactagataaggagataataaata
tatggaacaattagattcaaaatataagttgaaaaaatttctaattggcagtattttagagatggataggacaaggaaataatcttattgataatga
atatgttagagtatttcaaaataataaaagtaatagtaaacaaactgaactcggagaagaatttaagaatataagtaaaacaacttttttaaaat
atagatgatatagtagaatttacccttcgcaaaaaatatttattatgaaaatacatttttaacctatgtactactgatggaaaagcaggaaccaatg
aaaacttaataaatagatatgcattaggatttgattttgacaaaaagaattaggacaaggtttaattataaagataataaatttatttactaaga
taggattacattatcatatcctagttgatagtggaaatggattccatgtttatgtgctaattaataaaactaataacattaagttagtagatcagaagtta
caaatacattaataataaattgggtgcagataaacagcaaatatttactactcaagtattaagagtagccttatacatataatattaaaaatactac
taacaagtaaaaaataatacaccaagacaaaaatatatatagatatgacatagaaaagtttagctaaaaaatattgcaaagatgtaaaaacagt
aggtaactaatacaaaaatatattagatagtaagctaccaaattgtatagtagatattttaaaaaatggtagtaaaagatggacataaaaacct
agatttgcaaaaaatagttgtgactttaagattgaggaataaaaagtttaagtcaagtaatatccgttgctagagaatggaactatatacacaaaa
tagtctttcaaatagtgagctagaatatcaagtcaagtatatgtatgagaaacttaaaacggttaattttgggtgtactgggtgtgagtttaatagtg
attgttggaataaaaatagaatcagattttatatatagtgatgaagatactttgttcaatatgccacataagcactcaaggatttgaatataagaa
taggaaaggggttaaaataatgactggtaatacaattgtttatctataatgtgttacttaacaataaagatagagaattaaacatagacgatataat
ggagctgataacctataaacgtaagaagaaagttaaaaacattgttatgagtgaagacattaagagaaacattaaaagaacttcaacataa
tgattatattacaaaaacaaaagggtgtacaaagctaggaataaaaagatacatacaatgtaaaagaagttagatgtaatatagataacaatata
ctattagttactttgttaccatggcagtaatttggggaataatttcaactgaagaattaagattatatactcacatgagatataagcaagatttattg
gtcaaagatgataaaataaaaaggaaatatattaagaattaatcaagaggaattagcaaaagatttaggagtaacacagcaaagaatttcaaat
atgataagaatctttattagatactaaaatttttagatgtatgggaaactaaaataaatgatagaggatttatgtactatacatatagattaaacaagta
gatttttgataggattagaattgattttctagtcctattttatgcaaaaaactaattaataaaaaatttcttttggtaaaataattgtacgagaattgca
aaaaaaaaatggcatcaaagtattgaaattaagccgttttaaaaaatttcttttggtaaaataattctacatatatatgtagtatatatatatgttttt
agagaatgtataactagaatatagagctagaatatagagaatgtataactagaatatagagctagaatatagagaatgtataactagaatatag
agctagaatatagagaatgtataactagaatatagagctagaatatagagaatgtataactagaatatagagctagaatatagagaatgtataa
ctagaatatagagctagaatatagagaatgtataactagaatatagagctagaatcctaattagtaggtgcttttttaaaacaagttaaaaatcaa
aaatagtagtaagcattggaaatgctagattctaaaatagaaaagtaaaaaattgggtgcactatctaaactatctatatcgctttttccgtcgt
ttggttctctagttacgatacaggggatgcttatattgagttatagtactaatcagtgcttaatatagttaataaaattatagttaccatagtttagt
aactatgatgtatgttagttagaaacttgcatctcggccggccagtgggcaagttg

[0145] pIM13 (*B. subtilis*) (SEQ ID NO: 314)

[0146] gaatggcgaatggcgctagcataaaaaataagaagcctgcatttgcaggcttctatttttatggcgcgccgcattcacttctttctat
ataaatatgagcgaagcgaataagcgtcggaaaagcagcaaaaagtttcttttgcgttgaggcatgggggttcaggggtgcagtatct
gacgtcaatgccgagcgaagcgaagcgaaggtagcatttacgttagataacccctgatatgctccgacgctttatagaaaagaagat
tcaactaggtaaaatcttaatataggttgagatgataaggtttataaggaattgtttgttctaatttttactcattttgttctaatttctttaacaaatg
ttcttttttttagaacagttatgataggttagaatagtttaaaataaggagtgagaaaaagatgaaagaaagatatggaacagtctataaagg
ctctcagaggctcatagacgaagaaagtggaagtcataaggttagacaagttataaccgtaaacaaacgtctgtaacttcgtaaaggcat
atatagtgcaattaataagtatgttagatatgattggcggaaaaaaacttaaaatcgtaactatatcctagataatgtccacttaagtaacaatac
aatgatagctacaacaagagaaatagcaaaagctacaggaacaagtcacaaacagtaataacaacacttaaaatcttagaagaaggaaat
attataaaaagaaaaactggagtattaatgttaaacctgaactactaatgagaggcgacgacaaaaacaaaaatacctcttactcgaatttg
ggaactttgagcaagaggcaaatgaaatagattgacctccaataacaccacgtagttattgggaggtcaatctatgaaatgcgattaaggg
ccggccagtgggcaagttg

[0147] Cthem-based rep origin (*C. thermocellum*) (SEQ ID NO: 315)

[0148] gaatggcgaatggcgctagcataaaaaataagaagcctgcatttgcaggcttctatttttatggcgcgcccctgattctgtggataa
ccgtattaccgcctttgagtgagctgataccgctcggcgagccgaacgaccgagcgcagcagtcagtgagcgaagcgggaagag
cgccaatacgcgaacgcctctccccgcgcttgccgattcattaatgcagtatgggaaacaaaatattgcgtatgcgactgttcacatgg
acgagaaaacccctcacatgcatttaggagttgttctatgcgcctagagggtttctgtcgtcagcatgagcgtcggaagaaaagaaga
atactttttcgttcaagatgtttgcagcgtgatcgagaacttcgtgagcaaaagaaagcaagaggaaaaaatcgcatgatttgagcgtat
aagaaaaagcactcgaatgagtgcttttttgcgtttgagcgtagcgaaaaacgagttcttctattcttgatacatatagaaataacgtcatttt
atttagttgctgaaaggtgcgttgaaagtgttggtatgtatgttttaagattgaaaaccccttaaaattggttcacagaaaaacccatctgtt
aaagtataagtgaccaaacaataactaaatagatgggggtttcttttaattattatgtgtcctaatagtagcatttattcagatgaaaaatcaagg
gttttagtgacaagacaaaaagtggaagtgagaccatggagagaaaagaaatcgctaattgttgattactttgaacttctgcatattcttg
aatttaaaaggctgaaagagtaaaagattgtgctgaaatattagagtataacaaaatcgtaaacaggcgaaagaaagttgtatcgagtg
ggtttgtaaatccaggctttgtccaatgtgcaactggaggagcaatgaaacatggcattcagtcacaaaagggtgtgtgtaagtatttaa
acaaaagccaacagttcgttggtgtttctcacattaacagttaaaaatgtttatgatggcgaagaattaataagagttgtcagatatggctca
aggatttcgccgaatgatgcaatataaaaaaattaataaaaatctgttggtttatgcgtgcaacggagtgacaataataataaagataatc
ttataatcagcatgcatgtattggtatgtgtggaaccaactatttttaagaatacagaaaactacgtgaatcaaaaacaatggattcaatttg
gaaaaaggcaatgaaattagactatgatccaatgtaaaagttaaattgattcgaccgaaaaataataataatcggatatacaatcggaatt
gacgaaactgcaaaatcctgtaaaggatacggattttatgaccgatgatgaagaaaagaatttgaaacgtttgtctgatttgaggaagggt

tacaccgtaaaaggftaatctctatgggtggtttgttaaagaatacataaaaaattaaccttgatgacacagaagaaggcgatttgattcat
 acagatgatgacgaaaaagccgatgaagatggattttctattattgcaatgtggaattgggaacggaaaaatttttattaaagagtagttcaa
 caaacgggattgacttttaaaaaaggattgattctaatagaagaagcagacaagtaagcctcctaaattcactttagataaaaaatttaggaggc
 atatcaaatgaacggccggccagtgggcaagttg

[0149] pAM β 1-based (*E. faecalis*) (SEQ ID NO: 316)

[0150] gaatggcgaatggcgctagcataaaaaataagaagcctgcatttgaggcttcttattttatggcgcgccctcacgttaagggatttt
 ggtcatgagattatcaaaaaggatcttcacctagatccttttaattaaaaatgaagttttaaataaatctaaagtatatatgagtaaacttggtctg
 acagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatcatagttgccctgactccccgtcgtgtagataacta
 cgatacgggagggccttaccatctggccccagtgtgcaatgataccgcgagaccacgctcaccggctccagatttatcagcaataaacca
 gccagccggaaggccgagcgcagaagtggcctatttaatacatttgactagcaataactaacaacaagacacacacacacaaaaatcaa
 aaattcactacttttagttaaaaaccacgtaaccacaagaactaatccaatccatgtaatcgggttcttcaaataatttctccaagatttctcctct
 aatatgctcaacttaaatgacctattcaataaatctattatgctgctaaatagtttataggacaaataagtatactctaatagacctataaaagataga
 aaattaaaaaatcaagtgttcgcttcgctctcactgccccctcgacgttttagtagcctttccctcacttcgttcagccaagccaactaaaagttt
 cgggctactctctccttctccccctaataattaataaaatcttactctgtatatttctgctaatacttcgctaacagcaaagaaaaaacaacac
 gtatcatagatataaatgtaatggcatagtgccgggtttatttcagcctgtatcatagctaaacaatcgagttgtgtgtccgttttagggcggttct
 gtagcttgtttaaagtctcttgaatgaatgtatgctctaagcaaaagaatttgcagcgctttatagctttcttttcttttttactttaatga
 tcgatagcaacaatgatttaactagcaagttgaatgccaccatttctcctgggttaatacttaagaaaatttctgattcgccctcagttaccttc
 agcaatttatctaattgtccgttcaggaatgcctagcacttctctaattcttttttggtcgtcactaaataaggcctgtatacatcgctttttcgctaa
 tataagccattaaatcttcttccattctgacaaatgaacacgttgacgttcgcttcttttttctgaatttaaacacccttgacggacaaataaat
 ctttactgggttaaatcacttgatacccaagccttgcaagaatggtaatgtattccctatttagcccccttgatagttttctgaataggcacttctaaca
 attttgattacttcttttcttaagggtgatctaategattattaaactcaaacataattatctgcacgtttcgattgaatagcctgaactaaagta
 ggctaaagagagggtaaacatgacgttattacgccctattaaaccctttctcctgaaaattcgtttcgtgcaataagagattaaaccagggtt
 catctacttgtttttgcttctgtaccgcttaaaaccgttagacttgaacgagtaaagcccttattatctgtttgttgaagaccaatcttgccattc
 tttgaaagaataacggtaattagatcaaaaaattctacattgtccgttcttggtatgcgagcaataacaaaatgattacacgttagatcaactgg
 caaagactttccaaaatattctcgatattttgcgaaattatttggctgcttgacagatttaattctgattttgaagtcacatagactggcggttct
 aaaacaaaatatgcttgataacctttatcagatttgataatcatagtaggcataaaacctaaatcaatagcgggtgttaaaatacgttgcgtgaa
 atagtttcttttgcgtgtgaatatcaaaatcaataaagaaggatttgatttgcgttaaattgttttcagaatgtccttcgtgtatgaacgggttctgc
 tgcatacgttccataacgataaacgttgggtgtccaatgtgtaaatgtatcttgattttcttgaatcgcttctcgggaagtcagaacaacaccacg
 accgccaatcatgcttgattttgagcgatacgcaaaaatagcccccttgctttacctggcttggtagtgattgagcgaattttactatttttaattt
 gtactttaacaagccgtcatgaagcacagtttctacaacaaaagggatattcattcagctgttctccttctataaatctataaaatagggtgttt

aattaacttgggttgccttttcaactgtttcaatattgcatgtttgaaaaagatttttctttataagtcaatttttccactaatcgaataaatta
 tttgttattttctattaacttatatatataatcttccccctccgaagaaaaatacttatctgattttgttctaagtagatatttcttttctaactctttctt
 aaacgtttctagtgtatagatatttgcataatttcttatctccaataaactatttttatataagttttacattcatcatgattcatacaaaactccacctct
 ataatgaatacaaaaaaagcaatcaaacgatttccgattgattgcttaacaattcttaattcagtagcttagatacttgaaaactctctgatttcc
 ctatataatgatagtacgggttatataccgtcttcaacaaagttaataaataacttcttacgagggagagttcatctgactaactgataagcgtt
 ggtttggcaatcttatcgggctatgcatttataaaatgtcgtcaaacattttataaatgtgcatggctctttttcgtttctattcagttcgttgttctgtt
 atatctagtataccgcttttaaaaaaaataagcaacgatttctgtcattattcacacgaagtcattgctttttcttctccatttctaaatccaatgtt
 acttgttctgattctgttctgttctgttctgttggctcatttgggattaaatccactactagcgttgagtagttaactttgcaatttgttctagtgtt
 ttatggttggatctgattttctgggcccggccagtgggcaagtgtg

[0151] pWV01-based (*L. lactis*) (SEQ ID NO: 317)

[0152] gaatggcgaatggcgctagcataaaaaataagaagcctgcatttgcaggcttctattttatggcgcgccgcagcgaagatgtgtg
 ctggttagattatgaaagccgatgactgaatgaaataataagcgcagcgccttctatttctgggtggaggaggctcaaggagatgaggggaa
 tgaaattccctcatgggtttgattttaaattgcttgcaattttgccgagcggtagcgctggaaaattttgaaaaaatttggaatttggaaaaa
 aatgggggggaaaggaagcgaattttgcttccgtactacgacccccattaagtccgagtgccaatttttgccaaaaacgctctatcccaa
 ctggctcaagggttaagggttttcaatcgccaacgaatcgccaacgttttcgccaacgtttttataaatctatatattaagtagctttattgtgtt
 tttatgattacaaagtatacactaactttataaaattatttgattggagtttttaaatggtgatttcagaatcgaaaaaagagttatgatttctctg
 acaaaagagcaagataaaaaattaacagatatggcgaaacaaaaagggttttcaaatctgcggttgcggcggttagctatagaagaatatgc
 aagaaaggaatcagaacaaaaaaataagcgaaagctcgcgtttttagaaggatacgagtttctgctacttgttttgataaggttaattatatca
 tggctattaaaaataactaaagctagaaattttggatttttattatctctgactcaattcctaattggtgaaagaaaaattagagagtttggcgcta
 tctatggctgtcagtcctttacacgatatggacgaaaaaaaagataaagatacatggaatagtagtgatttatacgaatggaaagcactata
 aaaaaccacactatcacgttatatatattgcacgaaatcctgtaacaatagaaaagcgttaggaacaagattaagcgaaaattgggggaatagt
 cagttgctcatgttgagatacttgattatatcaaagggtcatatgaattttgactcatgaatcaaaggacgctattgctaagaataaacatatata
 cgacaaaaaagatattttgaacattaatgattttgatattgaccgctatataacacttgatgaaagccaaaaaagagaattgaagaatttactttta
 gatatagtggatgactataatttggtaatacaaaagatttaattggcctttattcgccttaggggagcggagtttgaattttaatacgaatgatg
 taaaagatattgtttcaaaaactctagcgccttttagattatgggttgagggcaattatcagtggtgatatagagcaagttatgcaaagggttctga
 tgctgaaacgggggaaataaaatgacaaacaaagaaaaagagttatttgcgtaaagaggaattaaaaaagaaattaaggacttaaaag
 agcgtattgaaagatacagagaaatggaagttgaattaagtacaacaatagatttattgagaggagggttattgaataaataaaagccccct
 gacgaaagtcgaagggggttttattttggtttgatgttgcgattaatagcaatacaaggcccggccagtgggcaagtgtg

[0153] pMB1 (*B. longum*) (SEQ ID NO: 318)

[0154] gaatggcgaatggcgctagcataaaaaataagaagcctgcatttgaggcttcttattttatggcgcgccgcccggggcctcagc
ctgcggaacgcgcagcggacgccgacggctcagacggctcagaaacgtccgtgagtggtccacgcggccgaacaggtcagggag
gctcgcgcatacgtgagcggcgtggagaagcggctgaaggccgtccagcggctttctgtcaggatgtgctgggctgggttcagccgac
gcttcgtgggctgaaatatctgacttgggtcccgctatttgttactgtacaaatacgtatgtgtagccatgtccgatgagtattcgag
ccgacgcttgagctgtcgcgcacgttcgaaggctgggtggctgccgaaccccgcgcgtgtgctgcgacgacgactactcccggtgcaccg
caggagccgcgcgacgcgctcaaatgcaagcacatcgaggcgaaccccgcgcgtgggtgaacacgatcgtgggtggacatcgacga
cggaacgccaaggcgtggccctgtgggagcacgagggcatgcggccgaactggatcgcggagaacccggccaacgggcacgctc
acgcgggctgggtgctcacctttccggtgcccagaaccgatctggcgcgtctcaagccgttgaagctcctgcacgccaccacggaggga
ctgcgcgcgtcctgcgacggggacatgggctattcgggacttctgatgaagaaccccgagcatccggcgtgggctcggacatcatga
gtgggacacctacgacctggaacagctcgtgcagtcgctccaggaacacggggacatgccgcccgtcagctggaagcgcaccaagcgc
gcccgcacgcaggggctgggacgcaactgcacgctcttcgacaaggccgcacgctcgcctaccgctacgttgcggcggctgccgacc
gttcggaggccagcagcagggcattgcgcctatacgtgcgtgcacctgccacgaactcaacgtctcgtgttccccgatccgctgcacgc
gcgtgaggtcgaggacatcgccaagagcatccacaaatggatcgtcaccgcagccgatgtggcgcgacgggtgccattgccaacgca
gccacattcatgccatccaatccgcacgaggacacaaacacgggtgagaacaaatatcagcaggtcatgaaggaggcactggaatggtg
aggacgactttgaggaagaagcgcgggtgtctgcacgtgaattagctgaagcatagggcgtctccacgcgcaccattcagagctgggtg
gcaatgaagcgcgaggattggattgatgaacaagccgctatgcgcgaagcagtcgctcatatcacgatgacgagggccatacatggcc
gcagaccgcgagcatttcaacatgagccagggtgccgtgcgtcaacgctgtacagggctcgcaaggagcgcgaggacgagggcgc
ggagaaatcgaagcatctaccggcgagattccactgttcgactgacgctaaacgttgtcccaaacgcgaacgcagcacctccctgcctt
gcggcttttctcttccatcggccttcggcactcgggtgttgcctccagcgcgcagggcgcgagggtgcggccggccagtgggcaa
gttg

[0155] pIP404-based (*C. perfringens*) (SEQ ID NO: 319)

[0156] gaatggcgaatggcgctagcataaaaaataagaagcctgcatttgaggcttcttattttatggcgcgccccgaagaacgtttcca
atgatgagcacttttaattaaaaatgaagttttaaaacttcatttttaatttaaaaaatgaagttttatcaaaaaatttccaataatccactct
aagccacaaacacgccctataaaatcccgctttaatccactttgagacacatgtaataattactttacgccctagtatagtataatttttacattc
aatgccacgcaaaaaataaaggggcactataataaaagtcttcggaactaactaaagtaaaaaattatctttacaacctccccaaaaaaa
agaacaggtacaaagtaccctataatacaagcgtaaaaaatgagggtaaaaataaaaaataaaaaataaaaaataaaaaataaaaa
aaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaata
aaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaataaaaaat
gaaaaaattttttatattataatctttgaagaaaagaatataaaaaatgagcctttataaaagccattttttcatatacgtaatatgacgttct
aatgtttttattggtacttctaacattagagtaatttcttttttaagcccttttcttaagggttttttttttcttaatacatttaattcctctttttgt

tgctttccttagctttaattgctcttgataatTTTTTaccctctaattttctcttcttattccttttagaaattattgtcatatattttgtcttc
 ttctgtaatttctaataactctataagagtttcattcttatacttatattgcttattttatctaaataacatcttcagcacttctagttgctctataactct
 ctttcacttaaatgttgctctaaacatactattaagtctaaaacatcatttaaatgccttctcaatgtcttctgtaaagctacaaagataatatctatataa
 aaataatataagctctctgtgtccttttaaatacatattctttaggtcacaaagtttattatgtcttgattctccataataaaactctttctataaa
 tataatttattttgcttggtctacccttttccattatgggtttaattcaggtaaaaatccattttgtatttctttaaagtcataaatatattcgactcat
 ctaatatattgactactgttttgatttagagttatacttctggaactcttaattctgggtgcatctaaggctgtctatctgctccaaagtatttta
 attgattatataaatattctgaaccgctttccataatggtaatgcttactaggtactgcatttattatccatataaatacattccttccactatcta
 ttacatagtttggtataggaatactttgattaaaataattctttctaagtcattaataacctggctttagtttgccagttttataataatccaagtctat
 aaacagtgtatttaactctttatattttctaatacgctacacggcttataaaaggatttagagttatatagatatttcatcactcatatctaaatcttt
 aattcagcgtatttatagtgccattggctatatcctttttatctataacgctcctgggtatccacccttacttctactatgaatattatctatatagttct
 tttattcagctttaatgcgtttctcacttattcacctcccctctgtaaaactaagaaaattatcatattttcaataattattaactattcttaactctt
 aataaaaaatagagtaagtcccaattgaaacttaattctttttatgttttaatttattttttataaaatattttaaactaaattaaatgattctttta
 atttttactatttcattccataatataattactataattatttacaataatattttcttcatttgtaatttagatgattactaattttagttttatatattaaat
 aattaatgtataatttatataaaaaatcaaaggagcttataaattatgattatttccaaagataactaaagatttaatttttcaattttaacaatactttt
 gtaattatgtttaatttaattgtattttttcatataataaagccgtgaagtaaccaatccatttcttatgatgggcccggccagtgggcgaag
ttg

[0157] In one aspect, the present disclosure provides a bacterial expression vector comprising (a) a gram-positive bacteria replication origin comprising a sequence selected from the group consisting of SEQ ID NOs: 1-9 or 311-319, (b) a heterologous nucleic acid encoding a selectable marker that is an antibiotic resistance gene or an auxotrophic marker, and (c) at least one open reading frame, wherein the at least one open reading frame encodes a bioluminescent protein, a chemiluminescent protein, a fluorescent protein, a CRISPR enzyme, a Group II intron-encoded protein, at least one sgRNA, at least one Group II intron, or any combination thereof. The bacterial expression vector of the present technology may further comprise one or more bacterial conjugation transfer genes and/or an *E. coli* replication origin. Examples of bacterial conjugation transfer genes include *traJ* and *oriT*, and examples of *E. coli* replication origin include *colEI*, *pBR*, and *R6K*. Additionally or alternatively, in some embodiments, the one or more bacterial conjugation transfer genes, the gram-positive bacteria replication origin, and the heterologous nucleic acid encoding the selectable marker are codon optimized. Additionally or alternatively, in some embodiments, the at least one sgRNA or the at least one Group II intron

targets one or more genes selected from among 16S rRNA, *porA*, *bcat*, *croA*, *baiA2*, *baiCD*, *baiF*, *baiH*, *baiB*, *baiE*, *baiG* and *bail*.

[0158] In any and all embodiments of the bacterial expression vectors disclosed herein, the antibiotic resistance gene is selected from the group consisting of *catP*, *ermB*, *aad9*, *tetA*, and *ampR*, or the auxotrophic marker is *pyrG*, or *pyrF*.

[0159] In any of the preceding embodiments of the bacterial expression vectors disclosed herein, the CRISPR enzyme is selected from the group consisting of Cas9, dCas9, Cpf1, dCpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, and Csf4.

[0160] Examples of fluorescent proteins include, but are not limited to, GFP, YFP, CFP, RFP, TagBFP, Azurite, EBFP2, mKalama1, Sirius, Sapphire, T-Sapphire, ECFP, Cerulean, SCFP3A, mTurquoise, monomeric Midoriishi-Cyan, TagCFP, mTFP1, EGFP, Emerald, Superfolder GFP, Monomeric Azami Green, TagGFP2, mUKG, mWasabi, EYFP, Citrine, Venus, SYFP2, TagYFP, Monomeric Kusabira-Orange, mKOk, mKO2, mOrange, mOrange2, mRaspberry, mCherry, dsRed, mStrawberry, mTangerine, tdTomato, TagRFP, TagRFP-T, mApple, mRuby, mPlum, HcRed-Tandem, mKate2, mNeptune, NirFP, TagRFP657, IFP1.4, iRFP, mKeima Red, LSS-mKate1, LSS-mKate2, PA-GFP, PAmCherry1, PAmTagRFP, Kaede (green), Kaede (red), KikGR1 (green), KikGR1 (red), PS-CFP2, PS-CFP2, mEos2 (green), mEos2 (red), PSmOrange, or Dronpa. Examples of chemiluminescent proteins include, but are not limited to, β -galactosidase, horseradish peroxidase (HRP), or alkaline phosphatase. Examples of bioluminescent protein include, but are not limited to, Aequorin, firefly luciferase, *Renilla* luciferase, red luciferase, luxAB, or nanoluciferase.

[0161] In any and all embodiments of the bacterial expression vectors disclosed herein, the at least one sgRNA specifically hybridizes with a heterologous or endogenous target gene expressed in a gut bacterial host cell, and/or wherein the at least one sgRNA and/or the CRISPR enzyme is operably linked to a constitutive promoter or a conditional promoter. Additionally or alternatively, in some embodiments of the bacterial expression vectors disclosed herein, the at

least one Group II intron specifically targets a heterologous or endogenous gene expressed in a gut bacterial host cell, and/or wherein the at least one Group II intron and/or the Group II intron-encoded protein is operably linked to a constitutive promoter or a conditional promoter.

[0162] In another aspect, the present disclosure provides an engineered gram-positive human gut bacterial cell comprising any of the preceding embodiments of the bacterial expression vector disclosed herein, wherein the engineered gram-positive human gut bacterial cell is derived from a family selected from the group consisting of *Clostridiaceae*, *Lachnospiraceae*, *Eubacteriaceae*, *Erysipelotrichaceae*, *Enterococcaceae*, and *Bifidobacteriaceae*. In some embodiments, the engineered gram-positive human gut bacterial cell is derived from *Blautia hydrogenotrophica*, *Blautia luti*, *Blautia sp.*, *Blautia wexlerae*, *Clostridium bolteae*, *Clostridium innocuum*, *Clostridium paraputrificum*, *Clostridium saccharolyticum*, *Clostridium senegalense*, *Clostridium sp.*, *Clostridium sporogenes*, *Clostridium symbiosum*, *Eubacterium limosum*, *Eubacterium maltosivorans*, *Eubacterium ramulus*, *Eubacterium sp.*, *Roseburia inulinivorans*, *Bifidobacterium catenulatum*, *Enterococcus faecium*, *Escherichia fergusonii*, *Roseburia inulinivorans*, or *Bifidobacterium catenulatum*.

Methods for Genetically Modifying Commensal Human Gut Bacteria

[0163] In one aspect, the present disclosure provides a method for modifying a gram-negative human gut bacteria cell genome comprising transferring at least one gram-negative specific bacterial expression vector described herein into a gram-negative human gut bacteria cell via conjugation. In some embodiments, the at least one bacterial expression vector is integrated into the genome of the gram-negative human gut bacteria cell.

[0164] In another aspect, the present disclosure provides a method for genetically modifying a gram-positive human gut bacteria cell comprising transferring two or more distinct bacterial expression vectors into a gram-positive human gut bacteria cell simultaneously via conjugation, wherein each of the two or more distinct bacterial expression vectors comprise: (a) a gram-positive bacteria replication origin comprising a sequence selected from the group consisting of SEQ ID NOs: 1-9 or 311-319, (b) a heterologous nucleic acid encoding a selectable marker that is an antibiotic resistance gene or an auxotrophic marker, and (c) at least one open reading frame, wherein the at least one open reading frame encodes a bioluminescent protein, a

chemiluminescent protein, a fluorescent protein, a CRISPR enzyme, a Group II intron-encoded protein, at least one sgRNA, at least one Group II intron, or any combination thereof. The antibiotic resistance gene or the auxotrophic marker of each of the two or more distinct bacterial expression vectors may be independently selected from the group consisting of *catP*, *ermB*, *aad9*, *tetA*, *ampR*, *pyrG*, and *pyrF*.

[0165] In some embodiments, each of the two or more distinct bacterial expression vectors further comprise one or more bacterial conjugation transfer genes and/or an *E. coli* replication origin, optionally wherein the one or more bacterial conjugation transfer genes are selected from the group consisting of *traJ*, and *oriT* and/or the *E. coli* replication origin is selected from the group consisting of *colEI*, *pBR*, and *R6K*.

[0166] Additionally or alternatively, in some embodiments, the CRISPR enzyme of each of the two or more distinct bacterial expression vectors is independently selected from the group consisting of Cas9, dCas9, Cpf1, dCpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, and Csf4.

[0167] Additionally or alternatively, in some embodiments of the methods disclosed herein, the fluorescent protein of each of the two or more distinct bacterial expression vectors is independently selected from the group consisting of GFP, YFP, CFP, RFP, TagBFP, Azurite, EBFP2, mKalama1, Sirius, Sapphire, T-Sapphire, ECFP, Cerulean, SCFP3A, mTurquoise, monomeric Midoriishi-Cyan, TagCFP, mTFP1, EGFP, Emerald, Superfolder GFP, Monomeric Azami Green, TagGFP2, mUKG, mWasabi, EYFP, Citrine, Venus, SYFP2, TagYFP, Monomeric Kusabira-Orange, mKOk, mKO2, mOrange, mOrange2, mRaspberry, mCherry, dsRed, mStrawberry, mTangerine, tdTomato, TagRFP, TagRFP-T, mApple, mRuby, mPlum, HcRed-Tandem, mKate2, mNeptune, NirFP, TagRFP657, IFP1.4, iRFP, mKeima Red, LSS-mKate1, LSS-mKate2, PA-GFP, PAmCherry1, PATagRFP, Kaede (green), Kaede (red), KikGR1 (green), KikGR1 (red), PS-CFP2, PS-CFP2, mEos2 (green), mEos2 (red), PSmOrange, and Dronpa. Additionally or alternatively, in certain embodiments of the methods disclosed herein, the chemiluminescent protein of each of the two or more distinct bacterial expression

vectors is independently β -galactosidase, horseradish peroxidase (HRP), or alkaline phosphatase. Additionally or alternatively, in some embodiments of the methods of the present technology, the bioluminescent protein of each of the two or more distinct bacterial expression vectors is independently Aequorin, firefly luciferase, *Renilla* luciferase, red luciferase, luxAB, or nanoluciferase.

[0168] In any and all embodiments of the methods disclosed herein, the at least one sgRNA sequence of the two or more distinct bacterial expression vectors specifically hybridizes with a heterologous or endogenous target gene expressed in a gut bacterial host cell, and/or wherein the at least one sgRNA and/or the CRISPR enzyme is operably linked to a constitutive promoter or a conditional promoter. Additionally or alternatively, in some embodiments of the methods disclosed herein, the at least one Group II intron of the two or more distinct bacterial expression vectors specifically targets a heterologous or endogenous gene expressed in a gut bacterial host cell, and/or wherein the at least one Group II intron and/or the Group II intron-encoded protein is operably linked to a constitutive promoter or a conditional promoter. In some embodiments, three or four distinct bacterial expression vectors are simultaneously transferred into a gram-positive human gut bacteria cell simultaneously via conjugation.

[0169] In any and all embodiments of the methods disclosed herein, the gram-negative or gram-positive human gut bacteria cell is isolated from a colonic mucosa-enriched lavage sample, a fecal sample, a rectal swab, or an intestinal sample obtained from a human subject.

[0170] Also disclosed herein are engineered human gut bacterial cells generated by any and all embodiments of the methods of the present technology. Additionally or alternatively, in some embodiments of the methods disclosed herein, the engineered human gut bacterial cells are generated using at least two, at least three, at least four, at least five, at least six, at least eight, at least ten, or at least twelve or more primers and/or gRNAs of any one of SEQ ID NOs: 23-287.

Kits of the Present Technology

[0171] Also provided herein are kits comprising any and all embodiments of the bacterial expression vectors of the present technology and instructions for using the bacterial expression vectors to genetically modify human gut bacteria. The kits may further comprise one or more primers and/or gRNAs comprising the sequence of any one of SEQ ID NOs: 23-287.

[0172] In some embodiments, the kits further comprise buffers, enzymes having polymerase activity, enzymes having polymerase activity and lacking 5'→3' exonuclease activity or both 5'→3' and 3'→5' exonuclease activity, CRISPR enzymes, enzyme cofactors such as magnesium or manganese, salts, chain extension nucleotides such as deoxynucleoside triphosphates (dNTPs), modified dNTPs, nuclease-resistant dNTPs or labeled dNTPs, necessary to carry out an assay or reaction, such as amplification and/or engineering alterations (*e.g.*, knock-in or knock-out alterations) in target nucleic acid sequences corresponding to specific human gut bacterial genes disclosed herein.

[0173] In one embodiment, the kits of the present technology further comprise a positive control nucleic acid sequence and a negative control nucleic acid sequence to ensure the integrity of the assay during experimental runs. A kit may further contain a means for comparing the levels and/or activity of one or more of the preselected set of human gut bacterial genes described herein in a sample obtained from a subject with a reference nucleic acid sample (*e.g.*, from a control sample or isolated culture). The kit may also comprise instructions for use, software for automated analysis, containers, packages such as packaging intended for commercial sale and the like.

[0174] The kits of the present technology can also include other necessary reagents to perform any of the NGS techniques disclosed herein. For example, the kit may further comprise one or more of: adapter sequences, barcode sequences, reaction tubes, ligases, ligase buffers, wash buffers and/or reagents, hybridization buffers and/or reagents, labeling buffers and/or reagents, and detection means. The buffers and/or reagents are usually optimized for the particular amplification/detection technique for which the kit is intended. Protocols for using these buffers and reagents for performing different steps of the procedure may also be included in the kit.

[0175] The kits of the present technology may include components that are used to prepare nucleic acids from a colonic mucosa-enriched lavage sample, a fecal sample, a rectal swab, or an intestinal sample obtained from a human subject for the subsequent amplification and/or detection of engineered alterations (*e.g.*, knock-in or knock-out alterations) in target nucleic acid sequences corresponding to specific human gut bacterial genes disclosed herein. Such sample preparation components can be used to produce nucleic acid extracts from tissue samples. The

test samples used in the above-described methods will vary based on factors such as the assay format, nature of the detection method, and the specific tissues, cells or extracts used as the test sample to be assayed. Methods of extracting nucleic acids from samples are well known in the art and can be readily adapted to obtain a sample that is compatible with the system utilized. Automated sample preparation systems for extracting nucleic acids from a test sample are commercially available, *e.g.*, Roche Molecular Systems' COBAS AmpliPrep System, Qiagen's BioRobot 9600, and Applied Biosystems' PRISM™ 6700 sample preparation system.

EXAMPLES

[0176] The present technology is further illustrated by the following Examples, which should not be construed as limiting in any way.

[0177] Hundreds of microbiota genes are associated with host biology/disease. Unraveling the causal contribution of a microbiota gene to host biology remains difficult because many are encoded by non-model gut commensals and not genetically targetable. A general approach to identify their gene transfer methodology and build their gene manipulation tools would enable mechanistic dissections of their impact on host physiology.

[0178] We developed a pipeline that identifies the gene transfer methods for 91 non-model microbes spanning > 70 species and 5 phyla, and we demonstrated the utility of their genetic tools by modulating microbiome-derived short-chain fatty acids and bile acids *in vitro* and in the host. In a proof-of-principle study, by deleting a commensal gene for bile acid synthesis in a complex microbiome, we discover an unprecedented role of this gene in regulating colon inflammation. This technology will enable genetically engineering the non-model gut microbiome and facilitate mechanistic dissection of microbiota-host interactions.

[0179] The pipeline disclosed herein would not only facilitate dissection of the effect of microbiota on the associated treatments but would also enable genetic engineering of the gut microbiome, as a whole, for improved therapeutics.

Example 1: Materials and Methods

[0180] Screen the culture conditions of Gram-positive Clostridia strains

[0181] The culture was incubated in an anaerobic chamber at 37 °C under an atmosphere of 5% CO₂, 7.5% H₂, 87.5% N₂. To pre-reduce, the plates were left in the chamber overnight before being used, and the liquid medium was left in the chamber with a loosened cap for at least 48 hrs before inoculation. Firstly, we screened the culture conditions of the agar plate for the Gram-positive *Clostridia* strains. Strains were restreaked (from original glycerol stock or medium suspension of freeze-dried powder) onto pre-reduced TSAB (Tryptic Soy Agar + blood) plates (**Fig. 31**) or BHIB (Brain Heart Infusion Agar + blood) plates (**Fig. 31**). Then, those that can grow on either TSAB or BHIB plates were sub-cultured into 1 mL pre-reduced liquid medium: TYGB (**Fig. 31**), Mega (**Fig. 31**), Chopped Meat Medium (CMM) (**Fig. 31**), and Reinforced Clostridial Medium (RCM) (BD 218081), strains that can grow in any one of the four liquid cultures were subjected to the antibiotics test.

[0182] *Antibiotic test of Gram-positive Clostridia strains*

[0183] We tested the antibiotic resistance of 109 *Clostridia* microbes. To find the antibiotic and its optimal concentration that suppresses the growth of conjugation donor *E. coli*, the *Clostridia* strains were restreaked on TSAB or BHIB plates supplemented with 250 µg/mL D-cycloserine or 200 µg/mL gentamicin (to suppress the growth of conjugation donor *E. coli* CA434 during conjugation), or with 200 µg/mL kanamycin (to suppress the growth of conjugation donor *E. coli* HB101/pRK24). Both *E. coli* have been shown to successfully transform exogenous genomic DNA into *Clostridium* bacteria like *C. sporogenes* or *C. acetobutylicum* in previous studies (Cañadas et al., 2019; Guo et al., 2019; Heap et al., 2007). We found that 92 out of 109 strains are resistant to either D-cycloserine, gentamicin, or kanamycin. We next screened these 92 microbes against TSAB or BHIB plates with 15 µg/mL thiamphenicol. We found that they are all sensitive to thiamphenicol, so the thiamphenicol resistant gene can be exploited as a universal marker to select transconjugants that can uptake and stably maintain extracellular plasmid DNA. Further, the minimum inhibitory concentrations (MICs) of thiamphenicol of the 92 *Clostridia* candidates were tested with TSAB or BHIB plates containing thiamphenicol at different concentrations (**Fig. 24**).

[0184] *Vector assembly for Clostridia GM screening*

[0185] (i) Expand the replication origins (*rep oris*)

[0186] We first amplified the RP4 *oriT* component from the pExchange vector using primers R6K_F + R6K_R, and the amplified PCR product was Gibson assembled with the backbone amplified from pMTL82151 using primers pmtl+RP4 oriT_F and pmtl+RP4 oriT_R. The assembled vector was then double-digested with AscI and FseI and used as a backbone to fuse with nine replication origins (**Fig. 35**) that are PCR amplified from a synthetic DNA fragment (Twist Bioscience) or Addgene vectors. This gave a series of vectors pGM-ABCM, BBCM, CBCM, DBCM, EBCM, FBCM, GBCM, HBCM, and IBCM that would be used in the following mixed-conjugation experiment to identify *Clostridia* gut microbes that uptake and stably maintain exogenous DNA (**Fig. 9**). All this set of shuttle plasmids and the plasmids described in the latter sections are verified by restriction digestion and Sanger sequencing of their core functional components.

[0187] (ii) Sequence optimization

[0188] We further sequence-optimized the set of *Clostridia* conjugation plasmids by **1**) codon-optimizing the coding sequences (CDSs) of *catP*, *traJ*, and *Clostridial rep oris* to reduce their putative *Clostridial* Type II-RM sites (**Figs. 35** and **36**), and **2**) replacing the *catP* promoter Ppmtl-catP with Pfdx (the most potent promoter identified via a promoter screen). In brief, we searched the REBASE database and found 23 cutting sites that are most often recognized by the Type II-RM of *Clostridia* bacteria (including the solventogenic *Clostridium* genus) (**Fig. 36**). We then codon-optimized the CDSs of *catP*, *traJ*, and *rep oris* to reduce the number of these restriction sites by at least half (**Fig. 35**). Of note, the promoter and terminator of the CDS like *catP* or *traJ* and some highly repetitive motifs in the *rep oris* were left untouched. These sequences play a key role in regulating the functions of *catP* and *rep ori*, and any mutation (or nucleotide switch) could potentially cause dysfunction of *catP* or *rep ori* and lead to unsuccessful transformation. These set of plasmids are labeled with 'seq-opt'. Please refer to **Fig. 25** for the *Clostridia* that uptakes this set of vectors.

[0189] *Testing if Clostridium sporogenes ATCC 15579 can uptake multiple replication origins (plasmids) in one conjugation using mixed-conjugation strategies*

[0190] We did a preliminary test to assess if a model gut commensal *C. sporogenes* ATCC 15579 can uptake plasmids with a compatible replication from three *E. coli* conjugation donors

in one conjugation. (**Fig. 10A**). We inoculated three *E. coli* HB101/pRK24 donors harboring three different vectors pMTL82254 (rep ori: pBP1; antibiotic: erythromycin), pMTL83353 (rep ori: pCB102; antibiotic: spectinomycin), and pMTL84151 (rep ori: pCD6; antibiotic: thiamphenicol) (Heap et al., 2009), respectively. *C. sporogenes* ATCC 15579 was inoculated in 1 mL TYGC liquid broth and grown anaerobically at 37 °C for 12~18 hrs. The three *E. coli* donors were inoculated into LB liquid broth supplemented with the corresponding antibiotics (erythromycin: 250 µg/mL; spectinomycin: 100 µg/mL; chloramphenicol: 25 µg/mL) and shaken at 220 rpm for overnight. The next day, 700 µL of each *E. coli* culture were mixed and centrifuged at 1500 x g for 2 min. The cell pellet was washed with 1.5 mL PBS (pH 7.4) and centrifuged again at 1500 x g for 2 min. The PBS supernatant was removed, and the cell pellet was transferred on ice into the anaerobic chamber. The cell pellet was mixed with 300 µL of the overnight *C. sporogenes* culture, and a 35 µL cell mixture was dotted on pre-reduced TYG agar plates. After 48 hrs, the cell dots were scraped using a sterile inoculation loop and resuspended in 300 µL pre-reduced PBS (pH 7.4) buffer. 50 µL of the cell suspension was plated onto three TYG agar plates that were supplemented with D-cycloserine (250 µg/mL) + erythromycin (10 µg/mL, to select for pMTL82254), or spectinomycin (500 µg/mL, to select for pMTL83353), or thiamphenicol (15 µg/mL, to select for pMTL84151).

[0191] *Identify gene transfer methods for non-model Clostridia gut commensals that uptake and maintain exogenous plasmid DNA*

[0192] (i) Mixed-conjugation strategies for *Clostridia* gut commensals

[0193] A series of plasmids (**Fig. 9**) harboring different *rep oris* but the same antibiotic marker *catP* (against thiamphenicol) were transformed into chemical competent *E. coli* CA434, *E. coli* HB101/pRK24 or *E. coli* T7 Express harboring R702 or pRK24 (plasmid R702 or pRK24 was transferred from *E. coli* CA434 or *E. coli* HB101/pRK24 into T7 Express via conjugation) (Woods et al., 2019). We used the aforementioned mixed-conjugation strategies to identify the compatible *rep ori* for each *Clostridia* microbe of interest. For the *Clostridia* microbes resistant to D-cycloserine (250 µg/mL) or gentamicin (200 µg/mL), we would use *E. coli* CA434 as the conjugation donor. For the microbes that are not resistant to D-cycloserine (250 µg/mL) but resistant to kanamycin (200 µg/mL), we used *E. coli* HB101/pRK24 as their conjugation donors.

[0194] We began the conjugation by restreaking the target *Clostridia* microbe on a pre-reduced TSAB or BHIB agar plate. After 24-48 hrs, a single colony was inoculated in 1 mL of liquid broth (Mega/RCM/CMM) that supports its growth (mostly Mega, see **Figs. 23** and **31**) in an anaerobic chamber at 37 °C under an atmosphere consisting of 5% CO₂, 7.5% H₂, 87.5% N₂. On the same day, *E. coli* strains, each has a different *rep ori* (as described in the main text and **Fig. 9**), were inoculated into 6 mL of LB supplemented with tetracycline (15 µg/mL) and chloramphenicol (25 µg/mL) and shaken aerobically at 37 °C for 12-18 hrs (overnight). The next day, these *E. coli* donors were separated into three groups, including group I: pGM-ABCM, BBCM, and CBCM; group II: pGM-DBCM, EBCM, and FBCM; and group III: pGM-GBCM, HBCM, IBCM, and a negative *rep ori*-less control. For conjugating one *Clostridia* microbe, a 1.0 mL culture of each *E. coli* within the same group was mixed and centrifuged at 1500 x g for 2 min. The culture supernatant was discarded, and the cell pellet was gently washed with 500 µL PBS buffer (pH = 7.4). The PBS supernatant was then removed after centrifugation at 1500 x g for 2 min, and the cell pellet was transferred on ice into the anaerobic chamber. Next, the cell pellet (a total of three cell pellets) was mixed gently with 300 µL overnight culture of the targeting *Clostridia* microbe, and a 35 µL cell mixture was dotted on pre-reduced TSAB or BHIB agar plates. After 48 hrs, the cell dots were scraped using a sterile inoculation loop and resuspended in 300 µL pre-reduced PBS (pH 7.4) buffer. 100 µL of the cell suspension was plated on TSAB or BHIB plate supplemented with 15 µg/mL thiamphenicol (or MIC, see **Fig. 24**) and 250 µg/mL D-cycloserine (if *E. coli* CA434 is the conjugation donor), or 200 µg/mL kanamycin (if *E. coli* HB101/pRK24 is the conjugation donor). Colonies typically appeared after 36-48 hrs. Four colonies were picked and restreaked onto TSAB or BHIB plates with the same antibiotics to isolate single colonies.

[0195] Attempts to expand the number of either the conjugation donor *E. coli* or the recipient *Clostridia* were made, for instance, conjugating 5 or more *E. coli* to 2 *Clostridia* in one conjugation. We obtained some transconjugants, but this set-up decreases the conjugation efficiency and makes the followed-up diagnostic PCR (to identify which *rep ori* gets uptaken) more complicated and less efficient. All the working or non-working conjugations have been repeated at least three times in our experiment.

[0196] (ii) Electroporation of *Clostridia* microbes

[0197] To make electroporation competent cells, *Clostridia* microbe was first streaked on a pre-reduced TSAB or BHIB agar plate. After 24-48 hrs, a single colony was inoculated in 1 mL of pre-reduced liquid broth (Mega/RCM/CMM) that supports its growth (mostly Mega, see **Figs. 23** and **31**) and incubated in an anaerobic chamber at 37 °C overnight, then 1 mL of the seed culture was inoculated in 45 mL of liquid broth supplied with 0.4 M sucrose and 0.625% or 1.25% glycine (see **Fig. 24**). When the culture attained an OD600 of 0.6-0.8, culture was chilled on ice for at least 10 min (from this time point, all manipulations were performed at 4°C using an ice-bath and pre-chilled buffer). Cells were harvested by centrifugation at 8000×g and 4°C for 10 min. The resulting cell pellet was washed twice with 10 mL of pre-reduced, filter-sterilized SMP buffer (270 mM sucrose, 1 mM MgCl₂, and 5 mM sodium phosphate, pH 6.9). Following centrifugation, the final cell pellet was resuspended in 1.8 mL SMP buffer.

[0198] Plasmids harboring different replication origins were extracted and purified from *E. coli* CA434 using Plasmid Midiprep Kit (Zymo Research). Plasmid was pre-methylated using CpG (M. SssI) and GpC (M. CviPI) methyltransferases following the manufacturer's protocol (by NEB). After DNA purification, plasmids were separated into three groups, including group I: pGM-ABC1M_seq-opt, BBCM, and CBC1M_seq-opt; group II: pGM-DBC1M_seq-opt, EBC1M_seq-opt, and FBC1M_seq-opt; and group III: pGM-GBC1M_seq-opt, HBC1M_seq-opt, IBCM.

[0199] All the experimental procedures described below are carried out in an anaerobic chamber under an atmosphere consisting of 5% CO₂, 7.5% H₂, 87.5% N₂. Each group of plasmid mixtures (containing 2 µg of each plasmid) were added into 600 µL electroporation competent cell and mixed gently by flicking, and the DNA-cell mixture was transferred to a pre-chilled electroporation cuvette (4mm, Fisher Scientific). After incubated on ice for at least 10 min, a single exponential decay pulse was applied under anaerobic condition using an ECM 630 Electroporation System (BTX) set at 2.0 kV, 25 µF, and 400 Ω. Immediately following pulse delivery, 900 µL of liquid broth containing 0.2 M sucrose was added into the electroporation cuvette, and the entire suspension was transferred to 400 µL of the same medium. The cell suspension was recovered at 37 °C overnight, then 200 µL of the recovery culture was plated onto TSAB or BHIB agar plates with 15 µg/mL thiamphenicol (or MIC, see **Fig. 24**). Colonies typically appeared after 36-48 hrs.

[0200] Eight colonies were picked and restreaked onto TSAB or BHIB plates with the same antibiotics to isolate single colonies. The isolated single colony was cultivated in 3 mL liquid broth supplemented with the same antibiotics. The genomic DNA was isolated from the resulting cell material using the Quick DNA fungal/bacterial kit (Zymo Research). Then multiplex diagnostic PCR was conducted to assess which plasmid was incorporated by the recipient *Clostridia* microbe. PCR products of *rep oris* were purified and verified by sanger sequencing. Additionally, we confirmed that the colonies we picked and restreaked are the target *Clostridia* strain by amplifying the 16s rRNA region of the colony using primers 16s_27F + 16s_1391R, and the PCR product was purified and sent for Sanger sequencing using primer 16s_1391R.

[0201] *Diagnostic PCR and sanger sequencing to verify the plasmids uptaken by the Clostridia strains*

[0202] The isolated single colony was cultivated in 3 mL Mega/RCM/CMM broth supplemented with the corresponding antibiotics 250 µg/mL D-cycloserine (or 200 µg/mL kanamycin) + 15 µg/mL thiamphenicol (or MIC, see Fig. 24; For electroporation, colonies were plated on the plates with only thiamphenicol). The genomic DNA was isolated from the resulting cell material using the Quick DNA fungal/bacterial kit (Zymo Research). Then we performed multiplex diagnostic PCR to assess which plasmid was uptaken by the conjugation recipient *Clostridia* microbe. For the mixed-conjugation with group I, primers pMTL_*laz*_diag_F (universal forward primer) + pGM-ABCM_rep_R_1500bp + pGM-BBCM_rep_R_1000bp + pGM-CBCM_rep_R_2000bp (for 15 µL PCR reaction, the amount of the four primers is: 0.75 µL, 0.3 µL, 0.3 µL and 0.3 µL (10 µM)) were used for diagnostic PCR. We would see a 1.5 kb (or 1.0 kb, or 2.0 kb) PCR band if pGM-ABCM (or BBCM, or CBCM) is uptaken by the *Clostridia* microbe (Figs. 11A-11B). In the meantime, we confirmed that the colonies we picked and restreaked are the target *Clostridia* strain but not the *E. coli* that escapes the antibiotics. We amplified the 16s rRNA region of the colony using primers 16s_27F + 16s_1391R, and the PCR product was purified and sent for Sanger sequencing using primer 16s_1391R.

[0203] *Validating the mixed-conjugation result by conjugating the E. coli donor that harbors the identified plasmid(s) to the targeted Clostridia microbe*

[0204] We next did the single strain conjugation (one *E. coli* donor to one *Clostridia* recipient) to validate that the PCR-identified plasmid(s) can indeed be transformed into the targeted *Clostridia* microbe. A single colony of targeted *Clostridia* strain was inoculated in a 1 mL Mega (or RCM or CMM) broth in an anaerobic chamber at 37 °C under an atmosphere consisting of 5% CO₂, 7.5% H₂, 87.5% N₂. The conjugation donor *E. coli* (CA434 or HB101/pRK24) harboring the PCR-identified plasmid was inoculated into 6 mL of LB supplemented with tetracycline (15 µg/mL) and chloramphenicol (25 µg/mL) and shaken aerobically at 37 °C for 12-18 hrs (overnight). After 12-18 hrs, 1.5 mL of the *E. coli* culture was centrifuged at 1500 x g for 2 min. The supernatant was discarded, and the cell pellet was washed with 500 µL PBS buffer (pH = 7.4). The PBS supernatant was then removed after centrifugation at 1500 x g for 2 min, and the cell pellet was transferred on ice into the anaerobic chamber. Next, the cell pellet was mixed gently with a 300 µL overnight culture of the targeting *Clostridia* microbe, and a 35 µL cell mixture was dotted on pre-reduced TSAB or BHIB agar plates. After 48 hrs, the cell dots were scraped using a sterile inoculation loop and resuspended in 300 µL pre-reduced PBS (pH 7.4) buffer. 100 µL of the cell suspension was plated on TSAB or BHIB plate supplemented with 15 µg/mL thiamphenicol (or MICs, see **Fig. 24**) and 250 µg/mL D-cycloserine (if *E. coli* CA434 is the conjugation donor), or 200 µg/mL kanamycin (if *E. coli* HB101/pRK24 is the conjugation donor). Colonies typically appeared after 36-48 hrs. Four colonies were picked and restreaked onto TSAB or BHIB plates with the same antibiotics to isolate single colonies. The isolated single colonies will be cultured in 1 mL of pre-reduced Mega (or RCM or CMM) with the same antibiotics, and the glycerol stock will be prepared using the culture.

[0205] Developing and testing a CRISPRi-dCpf1 lacZα system for *Clostridia* GM

[0206] *Vector assembly for utilizing dCpf1 to suppress the lacZα transcription in Clostridia strains*

[0207] We followed a previously reported literature to mutate the aspartic acid (D) catalytic site at 908 position of the Cpf1 amino acid sequence to alanine (A) to get the deactivated Cpf1 (dCpf1) (Tang et al., 2017). We amplified the Cpf1 coding sequence (CDS) from the vector pDEST-hisMBP-AsCpf1-EC (Hur et al., 2016) using primers 83153_AsCpf-1_XbaI_F + dAsCpf-1_D908A_R and dAsCpf-1_D908A_F + 83153_AsCpf-1_XhoI_R. The two fragments

were assembled via fusion PCR using primers 83153_AsCpf-1_XbaI_F+83153_AsCpf-1_XhoI_R. The purified PCR product and plasmid pMTL83153 were double-digested with XbaI/XhoI and ligated together using Instant Sticky-end Ligase (NEB), yielding plasmid pGM-BBCD. Then, we amplified the *rep ori* fragments from plasmid pGM-ABCM using primers pMTL_rep_origin_F and pMTL_rep_origin_R. The purified PCR products were then Gibson assembled with the pGM-BBCD backbone amplified using primers pMTL_dCpf1_backbone_F and pMTL_dCpf1_backbone_R to give plasmid pGM-ABCD (**Fig. 9**). The *rep ori* fragments from plasmids pGM-CBCM, DBCM, EBCM, FBCM, GBCM, HBCM, and IBCM were amplified using primers pMTL_rep_origin_F and pMTL_rep_origin_R. The purified PCR products were then Gibson assembled with the pGM-BBCD backbone amplified using primers pMTL_dCpf1_backbone_F and pMTL_dCpf1_backbone_R, yielding a new set of vectors pGM-CBCD, DBCD, EBCD, FBCD, GBCD, HBCD, and IBCD (including the aforementioned pGM-ABCD) (**Fig. 9**). Each of these vectors carries a *Clostridia*-specific *rep ori* and the coding sequence of dCpf1 driven by a strong and constitutive promoter Pfdx (Heap et al., 2009).

[0208] We next assembled this set of plasmids with the *lacZα*. One fragment (see details below) includes the gRNA promoter, the *lacZα* promoter, the *lacZα* coding sequence and was amplified from the plasmid pMTL82254_ *lacZα* (obtained from from pMTL82254) using primers *lacZα*_dCpf1_F and *lacZα*_dCpf1_R. The sequence is shown below:

[0209] *cctgcaggccaacacatcaagc*TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCCGA
GACGGCGCAACGCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTT
TACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAAT
TTCACACAGGAAACAGCTATGACCATGATTACGGATTCACTGGCCGTCGTTTTACAACG
TCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGACGACATCCCCCTTTC
GCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAG
*CCTGAATGGCGAATGGTAATAGT*CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAA
AGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCTGGTGAACGCTCTC*cctgacgtctcctacgtaggc*
ggccgc (SEQ ID NO: 13)

[0210] The lowercase italicized sequences are restriction sites of SbfI and NotI, respectively. The underlined sequence is the gRNA promoter PJ23119. The bold sequence is the *lacZα*

promoter. The italicized uppercase sequence is the coding sequence of *lacZα*. The double underlined sequence is the *lacZα* terminator. This fragment and the plasmid pGM-ABCD were digested with SbfI/NotI and ligated together using Instant Sticky-end Ligase (NEB), yielding plasmid pGM-ABCL. Then the *rep ori* fragments from plasmids pGM-BBCM, CBCM, DBCM, EBCM, FBCM, GBCM, HBCM and IBCM were amplified using primers pMTL_rep_origin_F and pMTL_rep_origin_R. The purified PCR products were then Gibson assembled with the backbone amplified from vector pGM-ABCL using primers pMTL_dCpf1_backbone_F and pMTL_dCpf1_backbone_R, yielding a whole set of plasmids that carry the CRISPRi-dCpf1 machinery and the *lacZα* reporter gene (**Fig. 9**).

[0211] The gRNA fragment targeting the promoter region and CDS of *lacZα* was introduced into the set of plasmids harboring dCpf-1 and *lacZα*. First, we used primers dCpf1-*lacZα*_gRNA_F_V6_R1 and gRNA_Cas9_Cpf1_R and a synthetic fragment (gBlocks, IDT) containing the terminator region as a template to amplify a PCR product that has one direct repeat sequence and gRNA fused with the terminator. Next, this PCR product was purified and used as the template for the second PCR, using primers dCpf1-*lacZα*_gRNA_F_V6_R2 and dCpf1_ *lacZα*_gRNA_Gib_R, to get this gRNA fragment. The sequence of this fragment is shown below:

gtcctaggtataatgctagcTAATTTCTACTCTTGTAGATACACAGGAAACAGCTATGACTAAT
 TTCTACTCTTGTAGATCAACGTCGTGACTGGGAAAACCTAATTTCTACTCTTGT
GATAGGAGAATAGAAAGAAGAAAATTCTTTCTAAAGGCTGAATTCTCTGTTTAATTTTGAGA
GACCATCTCTCAAAAATTGAACTTCTCAATAAAAAATTGAGAAGTAGCTGACCATCACAAAA
TCGTAGATTTTGGATGTCTAGCTATGTTCTTTGAAAATTGCACAGTGAATAAGTAAAGCTAA
AGGTATATAAAAAATCCTTTGTAAGAATACAATTTGCAAAGTGACAGAGGAAAGCgagacggc
 gcaacgcaattaatg (SEQ ID NO: 14)

[0212] The lowercase sequences are homologous to the sequence in pGM-ABCL. The bold sequences are the dCpf1 direct repeat sequence. The double underlined sequences are two gRNA targeting both the promoter region and the template strand of *lacZα*. The italicized sequence is a terminator region obtained from the *Cs* 16s rRNA gene (*CLOSPO_00916*).

[0213] This gRNA fragment was then Gibson assembled with the backbone amplified from pGM-ABCL using primers dCpf1-*lacZα*_backbone_F and 8x151 without *lacZα*_RN to get pGM-ABCF. The previously PCR amplified replication origin fragments were then Gibson assembled with the backbone amplified from pGM-ABCF using primers pMTL_dCpf1_backbone_F and pMTL_dCpf1_backbone_R. This generated a whole set of plasmids carrying nine different *rep* *oris*, dCpf1, *lacZα*, and the *lacZα* targeting gRNA (**Fig. 9**). We found that this set of plasmids were not uptaken by 2 microbes during our first round of conjugation. Because the same set of vectors without the *lacZα* targeting gRNA can be successfully transformed into these two microbes, we thus changed the gRNA sequence (double underlined) in the *lacZα* targeting gRNA fragment, and the new gRNA sequence is TGCTTCCGGCTCGTATGTTG (SEQ ID NO: 17) and ACACAGGAAACAGCTATGAC (SEQ ID NO: 18). This set of plasmids were then used to generate the gRNA-only control plasmids (**Fig. 2B**) by excising the dCpf-1 CDSs using PCR amplification (primers 8x153_lacza_dcpf-1_gRNA_No_dcpf-1_F and 8x153_lacza_dcpf-1_gRNA_No_dcpf-1_R) followed by T4 ligation.

[0214] *Perform GM screen in Clostridia microbes using the CRISPRi-dCpf1 lacZα system*

[0215] Using the Gram-positive strain *Clostridium bolteae* DSM 29485 (S74) as an example, pGM-ABCL and pGM-ABCF were transformed into chemically competent *E. coli* CA434, respectively. *E. coli* CA434 harboring pGM-ABCL and pGM-ABCF were conjugated to *Clostridium bolteae* DSM 29485. The transconjugants were picked and restreaked onto TSAB supplemented with D-cycloserine (250 µg/mL) + thiamphenicol (15 µg/mL) (or MICs, see **Fig. 24**). Then, we cultivated three isolated single colonies in 5 mL Mega liquid broth supplemented with 15 µg/mL thiamphenicol (or MICs, see **Fig. 24**) for 36 hrs, extracted the RNA using Quick RNA fungal/bacterial kit (Zymo Research), and performed qPCR to quantify the relative expression of *lacZα* after normalizing to 16s rRNA gene, using primers dCpf1-*lacZα*_qPCR_F and dCpf1-*lacZα*_qPCR_R for *lacZα* gene and S74_16s_qPCR_F and S74_16s_qPCR_R for the control 16s rRNA (**Fig. 30**). All the non-working conjugations have been repeated at least twice in our experiment.

[0216] *Developing and testing a 16s-tron strategy for Clostridia GM*

[0217] *Assemble vectors to test a Group II intron targeting a conserved target site in the Clostridia 16s rRNA genes*

[0218] We first performed multiple sequence alignment using 16s rRNAs of *Clostridia* that can uptake plasmids and identified a highly conserved target site of Group II intron. Then we used the Intron targeting and design tool on the Clostron website (<http://www.clostron.com/clostron2.php>) to design the Group II introns targeting the conserved 16s sequence. The 16s-targeting intron was amplified using primers EBS universal primer + WBJ_16s_tgt_685_IBSN + WBJ_16s_tgt_685_EBS1d + WBJ_16s_tgt_685_EBS2, and the purified PCR product was then Gibson assembled with backbone that amplified from the plasmid pGM-BCAR-001 using primers pMTL007C-E2_F and pMTL007C-E2_R to get the plasmid pGM-BCAQ. The *rep ori* fragments from plasmids pGM-ABCM, CBCM, DBCM, EBCM, FBCM, GBCM, HBCM, and IBCM were amplified using primers pMTL_rep_origin_F and pMTL_rep_origin_R. The purified PCR products were then Gibson assembled with the pGM-BCAQ backbone amplified using primers pMTL_dCpf1_backbone_F and clostron_rep_origin_backbone_R, yielding a new set of vectors pGM-ACAQ, CCAQ, DCAQ, ECAQ, FCAQ, GCAQ, HCAQ, ICAQ (whose conjugation-selection marker is *catP*, and retrotransposition-activated marker (RAM) is *ermB*) (**Fig. 9**). Then, we changed the RAM in plasmid pGM-FCAQ from *ermB* (antibiotic: erythromycin) into *aad9* (antibiotic: spectinomycin) by assembling antibiotic marker *aad9* amplified using primers clostron_Spec_F + clostron_Spec_R with the backbone of pGM-FCAQ amplified using primers Csp-316s_marker_F + Csp-316s_marker_R to get plasmid pGM-FCBQ (whose conjugation-selection marker is *catP*, and RAM is *aad9*) (**Fig. 9**), and then the replication origin of plasmid pGM-FCBQ was replaced to get a new set of vectors pGM-ACBQ, BCBQ, CCBQ, DCBQ, ECBQ, GCBQ, HCBQ, ICBQ.

[0219] *Introduce the assembled 16s-tron vectors into Clostridia and select the RAM integrated mutants*

[0220] Using the strain *Blautia luti* DSM 14534 (S54) as an example. The assembled vectors pGM-FCAQ was transformed into chemically competent *E. coli* CA434. Then *E. coli* CA434 harboring plasmid pGM-FCAQ was conjugated to S54. The transconjugants were picked and

restreaked onto TSAB supplemented with D-cycloserine (250 µg/mL) + thiamphenicol (15 µg/mL). Then, we cultivated three single colonies into 1 mL Mega supplied with 15 µg/mL thiamphenicol and 250 µg/mL D-cycloserine. After 24-36 hrs, 50 µL of cultures were spread onto TSAB plates supplemented with 250 µg/mL D-cycloserine and 10 µg/mL erythromycin. The transconjugants typically appeared after 36-48 hrs. Eight colonies were picked to inoculate 3 mL Mega supplemented with 250 µg/mL D-cycloserine and 10 µg/mL erythromycin. After 24-36 hrs, we extracted their genomic DNA using Quick DNA fungal/bacterial kit (Zymo Research) and performed diagnostic PCR using primers 16s_tron_diagR_v4 + 16s_1391R + 16s_1391R_3to5 (with 16s_tron_diagR_v4 binding the integrated intron part and 16s_1391R + 16s_1391R_3to5 binding the target 16s site, only colonies that undergo RAM integration will have the band of ~2.5kb) (**Figs. 13A and 13B**).

[0221] Identifying the gene transfer methods for Gram-negative Bacteroidia and building their gene insertion tools

[0222] *Screen the culture conditions of Gram-negative Bacteroidia strains*

[0223] Strains were restreaked (from original glycerol stock or medium suspension of freeze-dried powder) onto pre-reduced TSAB (Tryptic Soy Agar + blood) plates (**Fig. 31**) or BHIB (Brain Heart Infusion Agar + blood) plates (**Figs. 24 and 31**) to screen the culture conditions of the agar plate for these Gram-negative *Bacteroidia* strains. Then, those that can grow on either TSAB or BHIB plates were sub-cultured into pre-reduced liquid medium (see **Figs. 24 and 31**): TYGB (**Fig. 31**), Mega (**Fig. 31**), Chopped Meat Medium (CMM) (**Fig. 31**), and Reinforced Clostridial Medium (RCM) (BD 218081). Strains that can of *Bacteroidia* strains in our library can grow on a TSAB plate and in the TYGB or Mega liquid medium (see **Figs. 24 and 31**).

[0224] *Antibiotics test of target Gram-negative Bacteroidia strains*

[0225] We tested the antibiotic resistance of 66 *Bacteroidia* (*Prevotella* and *Bacteroides*) microbes. To find the antibiotic and its optimal concentration that suppresses the growth of conjugation donor *E. coli*, the *Bacteroidia* strains were restreaked on TSAB plates supplemented with 200 µg/mL gentamycin or 250 µg/mL D-cycloserine. We found that all of them are resistant to either gentamycin or D-cycloserine. We next screened these microbes against TSAB plates

with 15 µg/mL thiamphenicol (or MICs, see **Fig. 24**). We expect that they are sensitive to thiamphenicol, so the thiamphenicol resistant gene can be a universal marker to select transconjugants whose genome has been integrated by the suicide vector after the conjugation. All *Bacteroidia* strains tested are not resistant against thiamphenicol, and they were selected as candidates for the GM screening.

[0226] *Vector assembly for Bacteroidia GM screening*

[0227] We first amplified ~1 kb fragment of the 16s rRNA gene of *Bacteroides theta* VPI-5482 (*Bt*) and *Bacteroides ovatus* ATCC8483 (*Bo*) using primers BO_16S_F1+BO_16S_R2, BO_16S_F3N+BO_16S_R4 (two fragments, fused using fusion PCR, **Fig. 30**). The fragment was assembled with the pExchange vector to get the pGM vectors pGM-NAEM-001 and pGM-NAEM-002 for testing whether the assembled vector will integrate into the 16s rRNA loci of *Bt* and *Bo* (**Fig. 28**).

[0228] To generate the chimeric 16s rRNA sequence (chi-16s) for the *Bacteroidia* GM screening, we first performed multiple sequence alignment using 16s rRNAs of *Prevotella* (and *Bacteroides*) and synthesized ~1 kb fragments containing the nucleotides that are conserved in at least 50% of the aligned 16s sequences for both *Prevotella* and *Bacteroides*. Then, the synthetic *Bacteroides* chi-16s was amplified using primers CJG_syn16s_F and CJG_syn16s_R. The purified PCR product was then Gibson assembled with the backbone amplified from the vector pExchange using primers R6K_F and Erm_R to get the plasmid pGM-NAEB (**Fig. 9**). Because multiple target strains are resistant against erythromycin but not against thiamphenicol, we replaced the antibiotic marker *ermB* with *catP* to use thiamphenicol as a universal selective antibiotic for the *Bacteroidia* GM screen. The *catP* coding sequence was amplified from the vector pGM-ABCM using the primers pMTL_cat_F and pMTL_cat_R. The purified PCR product was then Gibson assembled with the backbone amplified from pGM-NAEB using the primers pEx_Erm_change_F and pEx_Erm_change_R to give pGM-NAC2B (**Fig. 9**). The suicide plasmid pGM-NAC2B was used for the *Bacteroides* GMS screen. Likewise, synthesized ~1 kb chi-16s for *Prevotella* amplified using primers Pre16s_F and R6K_F_RC. The purified PCR product was then Gibson assembled with the backbone amplified from the plasmid pGM-NAC2B using primers R6K_F and Erm_R to get the plasmid pGM-NAC2P (**Fig. 9**).

[0229] *Introducing suicide vectors into the Bacteroidia commensals by E. coli conjugation*

[0230] We introduced the suicide vectors pGM-NAC2P/B into the target *Prevotella/Bacteroides* using *E. coli* conjugation following the previously published protocol (Martens et al., 2008). A single colony of the target commensal was inoculated in 3 mL TYGB broth and cultured in an anaerobic chamber at 37 °C. The *E. coli* S17 harboring the pGM-NAC2P/B vector was inoculated in the LB broth supplemented with carbenicillin (100 µg/mL) grown at 37 °C with aerobic shaking at 220 rpm. After ~12-16 hrs, when the OD600 of *E. coli* S17 reached 0.8-1.0, 6mL of *E. coli* S17 culture was centrifuged at 1500 x g for 2 min. The supernatant was discarded, and the cell pellet was washed twice with 3 mL PBS buffer (pH = 7.4). The washed *E. coli* S17 cell pellet was resuspended in 3 mL overnight culture of the target *Bacteroidia* strain and gently mixed by pipetting. The mixture was filtered through a 0.2 µm filter. The filtered liquor was discarded, and the filter with the mixture of donor and recipient cells was placed onto the surface of a pre-reduced TSAB plate. The plate was incubated in a 37 °C incubator aerobically.

[0231] After incubation aerobically at 37 °C for 24 hrs, the filter was soaked in 2 mL of pre-reduced TYGB medium. The cell on the filter was resuspended into the medium by gentle vortexing. The mixture was then transferred into the anaerobic chamber, and 100 µL was plate onto a pre-reduced TSAB plate + 200 µg/mL gentamycin + 15 µg/mL thiamphenicol (or MICs, see **Fig. 24**). Colonies of the target strain typically appeared after 36-48 hrs. Four colonies were picked and restreaked on a pre-reduced TSAB plate + 200 µg/mL gentamycin + 15 µg/mL thiamphenicol (or MICs, see **Fig. 24**) to isolate single colonies. All the working or non-working conjugations have been repeated at least twice in our experiment.

[0232] *Diagnostic PCR and sequencing to verify the single crossover integration of pGM-NAC2P/B*

[0233] The isolated single colony was inoculated in 3 mL TYBG broth supplemented with 200 µg/mL gentamycin + 15 µg/mL thiamphenicol (or MICs, see **Fig. 24**). After 12 hrs, we extracted their genomic DNA using Quick DNA fungal/bacterial kit (Zymo Research) and performed diagnostic PCR using primers 16s_27F and R6K_R to verify the single crossover integration of pGM-NAC2P/B at their 16s rRNA loci. (**Fig. 13C**) We would see a ~2.5 kb PCR band in the

transconjugants, one of whose chromosomal 16s rRNA loci was integrated by pGM-NAC2P/B. The 2.5 kb PCR product was purified using DNA Clean & Concentrator kit (Zymo Research) and sent for sequencing using primer R6K_F_RC. The sequencing results showed that the partial sequence of the 2.5 kb fragment came from the synthetic chi-16s in pGM-NAC2P/B, and the partial sequence of the 16s rRNA gene of the target strain, suggesting a single crossover of pGM-NAC2P/B into one of its 16s rRNA loci. (**Fig. 13D**) If the single crossover takes place at the 5' of the synthetic chi-16s, we would see that most of the resulting sequence will be the synthetic chi-16s. If the single crossover takes place at the 3' end of the synthetic 16s or if the chi-16s is highly similar to the 16s rRNA of the targeting microbes, most of the resulting sequence will be the original 16s (**Fig. 13D**).

[0234] Identifying the gene transfer methods for microbes of other phyla and building their gene insertion tools

[0235] *Culture conditions of candidate strains of other phyla*

[0236] In addition to strains mentioned above in phyla of *Firmicutes* and *Bacteroidetes*, we also applied our pipeline to screen a batch of microbes of other phyla, including *Fusobacteria* (8 *Fusobacterium*), *Proteobacteria* (8 *Desulfovibrio*, 6 *Klebsiella*, 10 *Proteus*, and 3 clinical isolates) and one *Actinobacteria* S201 (**Fig. 22**). Strains were restreaked from the medium suspension of freeze-dried powder onto agar plates and then sub-cultured into liquid broth recommended by DSMZ, all *Fusobacterium* can grow on Columbia agar with 5% blood (CBAB) and in Columbia Broth (CB) or CMM anaerobically, all *Desulfovibrio* can grow on *Desulfovibrio* (postgate) medium + 1.5% agar and in *Desulfovibrio* (postgate) medium anaerobically, and all *Klebsiella*, *Proteus* and, 3 clinical isolates can grow on LB agar and in LB broth aerobically (for *Proteus*, 3% agar was added in LB to avoid swarm) (**Fig. 22**).

[0237] *Antibiotics test of target strains of other phyla*

[0238] To find the antibiotic and its optimal concentration that suppresses the growth of donor *E. coli* in conjugation (conjugation was applied for *Desulfovibrio*, *Proteus*, and 3 clinical isolates), the strains were restreaked on corresponding agar plates supplemented with 250 µg/mL D-cycloserine or 30 µg/mL kanamycin. We found that all of them are resistant to either D-

cycloserine or kanamycin (see **Fig. 24**). To find the antibiotic and its optimal concentration that selects the growth of recipient strains in conjugation and electroporation (electroporation was applied for *Fusobacterium*, *Klebsiella*, and one *Proteus*), we screened all these microbes against agar plates with different concentrations of thiamphenicol (for *Fusobacterium*), chloramphenicol (for *Desulfovibrio*, *Proteus* and clinical isolates), carbenicillin (for clinical isolates), and kanamycin (for *klebsiella*) (see **Fig. 24**). We expect that they are sensitive to the antibiotics tested, so that the corresponding resistant gene can be used as a marker to select transconjugants whose genome has been integrated by the suicide vector after conjugation or electroporation. Antibiotics and concentrations for each candidate strain are listed in **Fig. 24**.

[0239] *Introducing suicide vectors into the candidate microbes of other phyla by conjugation and electroporation*

[0240] For *Fusobacterium*, the suicide vector pGM-NACO2 was introduced into target microbes via electroporation. A single colony of the target *Fusobacterium* was inoculated in 1 mL liquid broth and cultured in an anaerobic chamber at 37 °C overnight. Then the 1 mL seed culture was inoculated into 45 mL of the same liquid broth and incubated at 37 °C till the OD600 reached ~1.2, culture was chilled on ice for at least 10 min (from this time point, all manipulations were performed at 4°C using an ice-bath and pre-chilled buffer.), then the cell was harvested by centrifugation at 8000×g and 4°C for 10 min, the resulting cell pellet was washed twice with 25 mL of pre-reduced, filter-sterilized water. Following centrifugation, the final cell pellet was resuspended in 1 mL of 10% (v/v) cold glycerol. Then 2 µg of plasmid pGM-NACO2 was added into 100 µL electroporation competent cell and mixed gently by flicking, and the DNA-cell mixture was transferred to a pre-chilled electroporation cuvette (1mm, Fisher Scientific). After incubated on ice for at least 10 min, a single exponential decay pulse was applied under anaerobic condition using an ECM 630 Electroporation System (BTX) set at 2.5 kV, 25 µF, and 200 Ω. Immediately following pulse delivery, 1 mL of liquid broth supporting growth was added into the electroporation cuvette, and the entire suspension was recovered at 37 °C for 3 hrs, then 200 µL of the recovery culture was plated onto CBAB agar plates containing thiamphenicol (with different recipient supplied with their corresponding MICs). Colonies typically appeared after 48-72 hrs.

[0241] Similarly, for *Klebsiella* and one *Proteus*, the suicide vector pGM-NACO3 and pGM-NACO4 were also introduced into target strains via electroporation. A single colony of the target *Klebsiella* and *Proteus* was inoculated in 1 mL LB and cultured aerobically at 37 °C overnight. Then the 1 mL seed culture was inoculated into 45 mL LB and incubated at 37 °C till the OD600 reached ~0.6, culture was chilled on ice for at least 10 min (from this time point, all manipulations were performed at 4°C using an ice-bath and pre-chilled buffer.), then the cell was harvested by centrifugation at 5500 rpm and 4°C for 10 min, the resulting cell pellet was washed with 25 mL of pre-reduced, filter-sterilized water and 2 mL of 10% (v/v) cold glycerol for twice. Following centrifugation, the final cell pellet was resuspended in 1 mL of 10% (v/v) cold glycerol. Then 2 µg of plasmid pGM-NACO3 and pGM-NACO4 was added into 70 µL electroporation competent cell and mixed gently by flicking, and the DNA-cell mixture was transferred to a pre-chilled electroporation cuvette (1mm, Fisher Scientific). After incubated on ice for at least 10 min, a single exponential decay pulse was applied using an ECM 630 Electroporation System (BTX) set at 2.5 kV, 25 µF, and 200 Ω. Immediately following pulse delivery, 500 µL of LB was added into the electroporation cuvette, and the entire suspension was recovered at 37 °C for 1 hr, then 100 µL of the recovery culture was plated onto LB agar plates containing selective antibiotics 30 µg/mL kanamycin or thiamphenicol. Colonies typically appeared after 36-48 hrs.

[0242] For *Desulfovibrio* and clinical isolates, the suicide vector pGM-NACO1 and pGM-NACO5,6,7 were introduced into target strains via conjugation, and we also applied conjugation to transfer plasmid pGM-NACO4 for two *Proteus*. A single colony of target strain was inoculated in 1 mL of liquid broth that supports its growth (Figs. 23 and 31). On the same day, *E. coli* S17 containing the corresponding suicide vector was inoculated into 6 mL of LB supplemented with 25 µg/mL chloramphenicol and shaken aerobically at 37 °C for 12-18 hrs (overnight). The next day, 1.5 mL of S17 donor was centrifuged at 1500 x g for 2 min. The culture supernatant was discarded, and the cell pellet was gently washed with 500 µL PBS buffer (pH = 7.4). The PBS supernatant was then removed after centrifugation at 1500 x g for 2 min, and the cell pellet was mixed gently with 300 µL overnight culture of the targeting microbe. A 35 µL cell mixture was dotted on agar plates. After 48 hrs, the cell dots were scraped using a sterile inoculation loop and resuspended in 300 µL pre-reduced PBS (pH 7.4) buffer. 100 µL of the cell suspension was plated on agar plate supplemented with 250 µg/mL D-cycloserine (for

Proteus and clinical isolates) or 30 µg/mL kanamycin (for *Desulfovibrio*) plus chloramphenicol (with their corresponding MICs). Colonies typically appeared after 48-72 hrs.

[0243] *Diagnostic PCR and sequencing to verify the single crossover integration of pGM plasmids*

[0244] After electroporation or conjugation plating, at least eight colonies were picked and restreaked onto agar plates with the same antibiotics used for plating. The isolated single colony was cultivated in 3 mL liquid broth supplemented with the same antibiotics. The genomic DNA was isolated from the resulting cell material using the Quick DNA fungal/bacterial kit (Zymo Research). Diagnostic PCR was performed using primers 16s_27F and R6K_R to verify the single crossover integration of suicide plasmids at their 16s rRNA loci. (Fig. 13C) We would see a ~2.5 kb PCR band in the transconjugants, one of whose chromosomal 16s rRNA loci was integrated. The 2.5 kb PCR product was purified using DNA Clean & Concentrator kit (Zymo Research) and sent for sequencing using primer pEx_diag_R.

[0245] Modulating *Clostridia* *bcat* expression and microbiome-derived metabolites using gene manipulation tools developed via the GM pipeline.

[0246] *Targeted suppression of BCAA aminotransferase (bcat) gene and modulate butyrate production in non-model gut Clostridia*

[0247] (i) Vector assembly for utilizing dCpf1 to suppress the *bcat* and *croA* transcription or utilizing Group II intron to deplete *croA* in *Clostridia* strains

[0248] The design of targeting gRNA for targeting *bcat* and *croA* in genome-sequenced *Clostridia* strains is about the same as that of *lacZα*. We used Golden Gate Ligation or Gibson assembly to introduce the targeting gRNA into the dCpf1 harboring plasmids.

[0249] The sequence of targeting gRNA that is introduced by Golden Gate ligation is shown as below:

tcgtctcctagcTAATTTCTACTCTTGTAGATCTATATGCCGACGGACAAGCTAATTTCTA
CTCTTGTAGATAGGATTAATACGATTATAATTAATTTCTACTCTTGTAGATAGGAG
AATAGAAAGAAGAAAATTCTTTCTAAAGGCTGAATTCCTCTGTTTAATTTTGAGAGACCATTCT

*CTCAAAATTGAAACTTCTCAATAAAAAATTGAGAAGTAGCTGACCATCACAAAATCGTAGATT
TTGGATGTCTAGCTATGTTCTTTGAAAATTGCACAGTGAATAAGTAAAGCTAAAGGTATATA
AAAATCCTTTGTAAGAATACAATTTGCAAAGTGACAGAGGAAAGCtacgggagacgg* (SEQ ID
NO: 15)

[0250] The lowercase sequences are Esp3I restriction sites. The bold sequences are dCpf1 direct repeat sequences. The underlined sequences are duplex gRNA targeting the *bcat* or *croA* gene. The italicized sequence is a terminator region obtained from the *Cs* 16s rRNA gene (*CLOSPO_00916*). Take *Clostridium bolteae* ATCC BAA-613 (S72) as an example. First, we used primers gRNA_S72_CGC65_03110_dCpf1_round1 and gRNA_Cas9_Cpf1_R and a synthetic fragment (gBlocks, IDT) containing the terminator region as a template to amplify a PCR product that has one direct repeat sequence and gRNA fused with the terminator. Next, this PCR product was purified and used as the template for the second PCR, using primers gRNA_S72_CGC65_03110_dCpf1_round2 and gRNA_dCas9_R, to get this gRNA fragment. The purified PCR product was ligated using Esp3I and T4 ligase with pGM-FBCL to give pGM-FBCD-010 (Figs. 28 and S8).

[0251] The sequence of targeting gRNA that is introduced by Gibson assembly is shown as below:

gtcctaggtataatgctagc**TAATTTCTACTCTTGTAGATTACCTCATAGCTACCCTTCACTAAT
TTCTACTCTTGTAGATTATAATGGTGATATGAAAAC**TAATTTCTACTCTTGTAGAT****
*AGGAGAATAGAAAGAAGAAAATTCTTTCTAAAGGCTGAATTCTCTGTTTAATTTGAGAGAC
CATTCTCTCAAAATTGAACTTCTCAATAAAAAATTGAGAAGTAGCTGACCATCACAAAATCG
TAGATTTTGGATGTCTAGCTATGTTCTTTGAAAATTGCACAGTGAATAAGTAAAGCTAAAGG
TATATAAAATCCTTTGTAAGAATACAATTTGCAAAGTGACAGAGGAAAGCctgacgtctcctacg*
tagg (SEQ ID NO: 16)

[0252] The lowercase sequences are homologous to regions in pGM-xBCL. The boldface sequences are dCpf1 direct repeat sequences. The underlined sequences are duplex gRNAs targeting *bcat* or *croA* gene. The italicized sequence is a terminator region obtained from the *Cs* 16s rRNA gene (*CLOSPO_00916*). To assemble the dCpf1 targeting vector for *bcat* in *C. senegalense* DSM 25507 (S100), we first used primers gRNA_S100_BCAA

aminotransferase_dCpf1_round1 and *gRNA_Cas9_Cpf1_R* and a synthetic fragment (gBlocks, IDT) containing the terminator region as a template to amplify a PCR product that has one direct repeat sequence and one gRNA fused with the terminator. Next, this PCR product was purified and used as the template for the second PCR, using primers *gRNA_S100_BCAA_aminotransferase_dCpf1_round2* and *dCpf1_gRNA_Gib_R*, to get the above gRNA fragment. The purified PCR product was then Gibson-assembled with the backbone amplified from vector pGM-ABCL using primers 8x151 without *lacZα_FN* and 8x151 without *lacZα_RN*, yielding plasmid pGM-ABCD-013 (**Figs. 28 and S8**).

[0253] To deplete *croA* in *Clostridia* strains utilizing Group II intron, we used the Intron targeting and design tool on the ClosTron website (<http://www.clostron.com/clostron2.php>) to design the Group II introns targeting the *croA* gene. The *croA*-targeting intron was amplified using primers EBS universal primer + S115_cro_123_IBSN+ S115_cro_123_EBS1d+ S115_cro_123_EBS2, and the purified PCR product was then Gibson assembled with the backbone that amplified from the plasmid pGM-FCAQ using primers pMTL007C-E2_F and pMTL007C-E2_R to get the *croA*-targeting ClosTron plasmid pGM-FCAR-003. Then plasmid pGM-FCAR-003 was introduced into S115 following the aforementioned conjugation procedure (see **Example 1** and **Fig. 24**). After conjugation, the S115 colonies harboring pGM-FCAR-003 appeared on the TSAB plate supplemented with 9 µg/mL thiamphenicol and 200 µg/mL gentamycin. Next, four colonies were restreaked onto a TSAB plate with the same antibiotics to isolate a single colony. The single colonies were inoculated into 1 mL Mega supplied with 9 µg/mL thiamphenicol and 200 µg/mL gentamycin. After 24-36 hrs, 50 µL of cultures were spread onto TSAB plates supplemented with 200 µg/mL gentamycin and 10 µg/mL erythromycin. The integrated colonies typically appeared after 48-72 hrs. Eight colonies were picked to inoculate 3 mL Mega supplemented with 200 µg/mL gentamycin and 10 µg/mL erythromycin. After 24-36 hrs, we extracted their genomic DNA using Quick DNA fungal/bacterial kit (Zymo Research) and performed diagnostic PCR using primers S115_cro_DiagF and S115_cro_DiagR.

[0254] The butyrate production was evaluated by glucose assay with PBS washed cell of control and *croA* mutant, 3 mL of culture was first centrifuged at 1500 x g for 3 min. The cell pellet was washed twice with 1 mL PBS (pH 7.4) and centrifuged again at 1500 x g for 3 min. The PBS

supernatant was removed, and the cell pellet was resuspended with 500 μ L PBS and then glucose was added to the concentration of 5 mM. The mixture was incubated anaerobically at 37 °C for 1 h. The PBS suspension was subjected to SCFAs derivatization and LCMS measurement.

[0255] (ii). Transform the assembled vectors into the *Clostridia* microbes via *E. coli* conjugation

[0256] We use the strain *Clostridium bolteae* ATCC BAA-613 (S72) as an example. The assembled vectors pGM-FBCD and pGM-FBCD-010 were transformed into chemically competent *E. coli* CA434, respectively. *E. coli* CA434 harboring pGM-FBCD and pGM-FBCD-010 were conjugated to S72. The transconjugants were picked and restreaked onto TSAB supplemented with D-cycloserine (250 μ g/mL) + thiamphenicol (15 μ g/mL). Then, we cultivated three isolated single colonies in 5 mL Mega liquid broth supplemented with 15 μ g/mL thiamphenicol for 36 hrs, extracted the RNA using Quick RNA fungal/bacterial kit (Zymo Research), and performed qPCR to quantify the relative expression of *lacZ α* after normalizing to 16s rRNA gene, using primers S72_CGC65_03110_qPCR_F and S72_CGC65_03110_qPCR_R for *bcat* and S72_16s_qPCR_F and S72_16s_qPCR_R for the control 16s rRNA (Fig. 30).

[0257] Genetic manipulation of *Bacteroidia* strains to deplete propionate production

[0258] (i). Assemble pGM vectors to generate *Bacteroidia* mutants that abolish propionate

[0259] Take *Bacteroides* sp. 1_1_6 (strain 25, abbreviated as S25) as an example, a ~1kb fragment of gene *BSIG_3264* that encodes a methylmalonate mutase (*mmdA*), was amplified from S25 genomic DNA using primers S25_BSIG_3264_mmdA_pEX_F and S25_BSIG_3264_mmdA_pEX_R. The purified PCR product was Gibson assembled with the backbone amplified from the vector pGM-NAC2B using primers R6K_F and Erm_R to give pGM-NACM-003 (Fig. 28).

[0260] (ii). Introducing propionate deletion vector pGM-NACM-003 into S25 via *E. coli* conjugation

[0261] We used the same protocol above to introduce pGM-NACM-003 into S25 via *E. coli* conjugation. About 48 hrs after plating the conjugation cell mixture, we picked four colonies and restreaked them on a pre-reduced TSAB plate + 200 μ g/mL gentamycin + 15 μ g/mL

thiamphenicol to isolate single colonies. A single colony was inoculated in 3 mL TYBG broth supplemented with 200 µg/mL gentamycin + 15 µg/mL thiamphenicol, and we extracted the bacterial genomic DNA using Quick DNA fungal/bacterial kit (Zymo Research). We performed the diagnostic PCR using primers S25_*BSIG_3264_mmdA*_diagF (for *Bacteroides* sp. 1_1_6 (S25)) and R6K_R to verify the single crossover integration of pGM-NACM-003 into *mmdA*. There was a ~2.0 kb PCR band in the colonies whose *mmdA* gene was inserted and mutated by pGM-NACM-003.

[0262] *Targeted suppression of porA gene in C. sporogenes ATCC 15579 (S107)*

[0263] (i) Vector assembly

[0264] To introduce the gRNA targeting the metabolic gene *porA* responsible for the branched short-chain fatty acid synthesis (Guo et al., 2019), we used primers gRNA_clo02083_z2 and gRNA_dCas9_R and a synthetic fragment (gBlocks, IDT) containing the terminator region as a template to amplify a PCR product that has two direct repeat sequences and gRNA fused with the terminator. Next, this PCR product was purified and ligated using Esp3I and T4 ligase with pGM-ABCL to give pGM-ABCD-006 (**Figs. 28 and S8**).

tcgtctcctagc**TAATTTCTACTCTTG**TAGATATAAGAATGCCTTACAAGTCTTAATTTCT
ACTCTTGTAGATAGGAGAATAGAAAGAAGAAAATTCTTTCTAAAGGCTGAATTCTCTGTT
TAATTTTGAGAGACCATTCTCTCAAAATTGAACTTCTCAATAAAAAATTGAGAAGTAGCTGA
CCATCACAAAATCGTAGATTTTGATGTCTAGCTATGTTCTTTGAAAATTGCACAGTGAATA
AGTAAAGCTAAAGGTATATAAAAAATCCTTTGTAAGAATACAATTTGCAAAGTGACAGAGGA
 AAGCtacgggagacgg (SEQ ID NO: 12)

[0265] The lowercase sequences are Esp3I restriction sites. The boldface sequence is the dCpfI direct repeat sequence. The underlined sequences are the gRNA targeting the promoter region of the *porA* metabolic gene cluster. The italicized sequence is a 16s rRNA terminator region obtained from the *Cs* 16s rRNA gene (*CLOSPO_00916*).

[0266] (ii). Introduce the vectors pGM-ABCD and pGM-ABCD-006 into *C. sporogenes* ATCC 15579 and quantify their production of branched short-chain fatty acid

[0267] We used the same protocol as described herein to introduce the vectors pGM-ABCD (control) and pGM-ABCD-006 (*porA* transcription repression mutant) into *C. sporogenes* ATCC 15579. For each conjugation, we cultivated three isolated single colonies in 5 mL TYGC liquid broth supplemented with 15 µg/mL thiamphenicol for 36 hrs. We extracted RNA from 5 mL of liquid culture using Quick RNA fungal/bacterial kit (Zymo Research). We quantified the relative expression of *porA* in the control strain and its transcription repression mutant. To quantify the production of branched short-chain fatty acid, 10 µL supernatant of both the control strain and the *porA* transcription repression mutant was derivatized and subject to LC-qTOF analysis.

[0268] Genetic manipulation of *baiH* gene in the *bai* operon of *Faecalicatena contorta* S122 (S122)

[0269] *Screening 27 genetically targetable Clostridia strains that 7 α -dehydroxylate primary bile acid cholic acid (CA) and chenodeoxycholic acid (CDCA)*

[0270] To identify if there are any 7 α -dehydroxylating bacteria in the group of 27 genetically targetable *Clostridia* commensals characterized via the GM pipeline, we restreaked the bacteria on the TSAB or BHIB agar, and a single colony of each strain was cultivated in 1mL liquid medium supplemented with 100 µM CA and 100 µM CDCA. After 48 hrs, 1ml of the culture was centrifuged at 15000 g for 20 min, and the supernatant was subjected to LC-MS analysis to examine if CA and CDCA were 7 α -dehydroxylated to DCA and LCA (see **Example 1** for detailed information).

[0271] *Whole-genome sequencing of *Faecalicatena contorta* S122 (S122)*

[0272] The biosafety level 1 *Faecalicatena contorta* S122 was isolated from healthy human stool. We cultivated a single colony of *Faecalicatena contorta* S122 (S122) in 3 mL Mega liquid broth for 24 hrs and extracted the genomic DNA using Quick DNA fungal/bacterial kit (Zymo Research). The S122 genomic DNA was sent for whole genome sequencing (BGI). The raw sequencing reads were filtered (for quality control), and *de novo* assembled (Geneious). The assembled contig in fasta format was further annotated using Prokka (v1.12) (Seemann, 2014). To locate the *bai* operon in the S122 genome, we performed a tblastn search of each *bai* gene

annotated in the genome of *C. scindens* ATCC 35704 and identified a cluster of nine genes as a candidate *bai* operon in the S122 genome (**Fig. 4A**).

[0273] Vector assembly for utilizing Group II intron to disrupt *baiH* in *Faecalicatena contorta* S122 (S122)

[0274] We used the Intron targeting and design tool on the ClosTron website (<http://www.clostron.com/clostron2.php>) to design the Group II introns targeting the S122 *baiH* gene. The *baiH*-targeting intron was amplified using primers EBS universal primer + WBJ_BaiH_tgt_645_IBSN + WBJ_BaiH_tgt_645_EBS1d + WBJ_BaiH_tgt_645_EBS2, and the purified PCR product was then Gibson assembled with the backbone that amplified from the plasmid pGM-FCAQ using primers pMTL007C-E2_F and pMTL007C-E2_R to get the plasmid pGM-FCAR-002.

[0275] To introduce thiamphenicol resistance to S122, we generated a plasmid pGM-FCFQ by replacing the original conjugation-selection marker *catP* marker of pGM-FCAQ with *aad9-ampR*, and retrotransposition-activated marker (RAM) is changed from *ermB* to *catP*. (**Fig. 9**) The antibiotic marker *aad9-ampR* was amplified using primers *aad9_carb_007C2_thiam_F* and *aad9_carb_007C2_thiam_R*, and the purified PCR product was then Gibson assembled with the backbone that amplified from the plasmid pGM-FCAQ using primers *pmtl_007C2_thiam_marker_F* and *pMTL007C_Clostron_87_87_Erm_ItrA_R* to get the plasmid pGM-FCDQ. And the antibiotic marker *catP* was then amplified using primers *clostron_Thiam_F* and *clostron_Thiam_R*, and the purified PCR product was then Gibson assembled with the backbone that amplified from the plasmid pGM-FCDQ using primers *Csp-316s_marker_F* and *Csp-316s_marker_R* to get the plasmid pGM-FCFQ.

[0276] *Genetic disruption of baiH in Faecalicatena contorta S122 (S122)*

[0277] To disrupt the *baiH* gene, *baiH*-targeting plasmid pGM-FCAR-002 was first introduced into S122 following the aforementioned conjugation procedure (see **Example 1** and **Fig. 24**). After conjugation, the S122 colonies harboring pGM-FCAR-002 appeared on the TSAB plate supplemented with 15 µg/mL thiamphenicol and 250 µg/mL D-cycloserine. Next, four colonies were restreaked onto a TSAB plate with the same antibiotics to isolate a single colony. The

single colonies were inoculated into 1 mL Mega supplied with 15 µg/mL thiamphenicol and 250 µg/mL D-cycloserine. After 24-36 hrs, 50 µL of cultures were spread onto TSAB plates supplemented with 250 µg/mL D-cycloserine and 10 µg/mL erythromycin. The integrated colonies typically appeared after 36-48 hrs. Eight colonies were picked to inoculate 3 mL Mega supplemented with 250 µg/mL D-cycloserine and 10 µg/mL erythromycin. After 24-36 hrs, we extracted their genomic DNA using Quick DNA fungal/bacterial kit (Zymo Research) and performed diagnostic PCR using primers WBJ_BaiH_tgt_DiagF and WBJ_BaiH_tgt_DiagR (Fig. 17B).

[0278] To confer the *baiH* mutant strain with thiamphenicol resistance, the plasmid pGM-FCFQ was first introduced into S122+pGM-FCAR-002 following the aforementioned conjugation procedure (Fig. 24). After conjugation, the S122+pGM-FCAR-002 colonies harboring pGM-FCFQ appeared on the TSAB plate supplemented with 10 µg/mL erythromycin, 300 µg/mL spectinomycin, and 250 µg/mL D-cycloserine. Next, four colonies were restreaked onto a TSAB plate with the same antibiotics to isolate single colony. The single colonies were inoculated into 1 mL Mega supplied with 10 µg/mL erythromycin, 10 µg/mL spectinomycin, and 250 µg/mL D-cycloserine. After 24-36 hrs, 50 µL of cultures were spread onto TSAB plates supplemented with 250 µg/mL D-cycloserine, 15 µg/mL thiamphenicol, and 10 µg/mL erythromycin. The integrated colonies typically appeared after 36-48 hrs. Colonies were picked to inoculate Mega supplemented with 250 µg/mL D-cycloserine, 15 µg/mL thiamphenicol, and 10 µg/mL erythromycin.

[0279] Likewise, we constructed the S122 control strain with both erythromycin and thiamphenicol resistance using 16s-targeting plasmid pGM-FCAQ and pGM-FCFQ. Both Group II introns in these two plasmids are targeting the 16s rRNA genes, we have validated that the engineered strains with thiamphenicol and erythromycin resistance still carry at least one copy of intact 16s rRNA gene in their genomes by diagnostic PCR.

[0280] Quantification of microbiome-derived SCFAs, BSCFAs, and bile acids using LC-MS

[0281] Quantification of isovalerate, propionate, butyrate, and bile acids production in the culture supernatant of the control and mutant strains

[0282] A single colony of control or mutant strain was used to inoculate 1 mL pre-reduced liquid medium (TYGB or Mega), if needed, supplemented with 200 µg/mL gentamycin + 15 µg/mL thiamphenicol for *Bacteroidia* strains, or 250 µg/mL D-cycloserine + 15 µg/mL thiamphenicol for *Clostridia* strains, or for bile acids measurement in *Faecalicatena contorta* S122 (S122), Mega with 250 µg/mL D-cycloserine + 10 µg/mL erythromycin was used. To pre-reduce, the liquid medium was left in the chamber with a loosened cap for at least 48 hrs before inoculation. The culture was incubated in an anaerobic chamber at 37 °C under an atmosphere consisting of 5% CO₂, 7.5% H₂, 87.5% N₂.

[0283] The cultures were incubated for 48 hrs in the anaerobic chamber. For the quantification of isovalerate, propionate, and butyrate, a 10 µL aliquot of the culture was mixed with 190 µL of short-chain fatty acids (SCFAs) derivatization solution (1 mM 2,2'-dipyridyl disulfide, 1 mM triphenylphosphine, and 1 mM 2-hydrazinoquinoline dissolved in acetonitrile) (Lu et al., 2013). The resulting mixture was vortexed and incubated at 60 °C for 1 hr. The mixture was centrifuged at 21000 x g for 20 min, and the supernatant was analyzed using an Agilent 1290 LC system coupled to an Agilent 6530 quadrupole time-of-flight (QTOF) mass spectrometer with a 130Å, 1.7 µm, 2.1 mm × 100 mm ACQUITY UPLC BEH C18 column (Waters). We used the following solvent system: A: H₂O with 0.1% formic acid; B: Methanol with 0.1% formic acid. 1 µL of each sample was injected, and the flow rate was 0.35 mL/min with a column temperature of 40 °C. The gradient for HPLC-MS analysis was: 0-6.0 min, 99.5%-70.0% A; 6.0-9.0 min, 70.0%-2.0% A; 9.0-9.4 min, 2.0% A; 9.4-9.6min, 2.0%-99.5% A. Peaks were assigned by comparison with authentic standards and relative analyte concentrations were quantified by comparing their peak areas with those of internal standards.

[0284] For bile acids detection and quantification, 100 µL of culture was centrifuged at 21000 x g for 20 min, and the supernatant was analyzed using an Agilent 1290 LC system coupled to an Agilent 6530 quadrupole time-of-flight (QTOF) mass spectrometer with a 1.7 µm, 2.1 mm × 100 mm Kinetex C18 column (Phenomenex). We used the following solvent system: A: H₂O with 0.05% formic acid; B: Acetone with 0.05% formic acid. 1 µL of each sample was injected, and the flow rate was 0.35 mL/min with a column temperature of 40 °C. 0-1 min, 75% A; 1-25 min, 75%-25% A; 25-26 min, 25%-0% A; 26-30 min, 0% A; 30-32 min 0%-75% A. Peaks were

assigned by comparison with authentic standards. Their concentrations were calculated using the standard curve and normalized to the fecal/cecal weight.

[0285] *Quantification of isovalerate, propionate, butyrate, and bile acids in mouse biological samples*

[0286] For the quantification of isovalerate, propionate, and butyrate, we made standard curves of isovalerate, propionate, and butyrate based on the Area Under Curve (AUC) of true chemical standards at different concentrations. A ~ 10mg fecal samples (or cecal samples) were resuspended in 50 μ L of 50% MeOH (in H₂O) and vortexed for 10 min (some beads were added to disperse the fecal/cecal material). Then the mixture was spun down, and 10 μ L of supernatant was mixed with 190 μ L short-chain fatty acids (SCFAs) derivatization solution. The resulting mixture was vortexed and incubated at 60 °C for 1 hr, then the mixture was centrifuged at 21000 x g for 20 min, and the supernatant was analyzed by LC-MS. The method and column for LC-MS are the same as described above. The concentrations of isovalerate, propionate, and butyrate were calculated using the standard curve and normalized to the fecal/cecal weight.

[0287] For the detection of bile acids, A ~ 10mg fecal samples were resuspended in 100 μ L of 50% MeOH (in H₂O) and vortexed for 10 min (some beads were added to disperse the fecal material). Then the mixture was centrifuged at 21000 x g for 20 min, and the supernatant was analyzed by LC-MS. The method and column for LC-MS are the same as described above. Their concentrations were calculated using the standard curve and normalized to the fecal/cecal weight.

[0288] *Growth curve of bacteria*

[0289] *Bacteroides (Bacteroides fragilis 3_1_12 (Bac1) and Bacteroides vulgatus ATCC 8482 (Bac2)) and Erysipelotrichaceae (Clostridium ramosum ATCC 25554 (Ery1), Erysipelatoclostridium ramosum strain 113-1 (Ery2), Clostridium ramosum DSM 24812 (Ery3), Clostridium ramosum DSM 1402 (Ery4), HM-173 Clostridium innocuum 6_1_30 (Ery5), Clostridium innocuum DSM 22910 (Ery6) and Holdemania filiformis DSM 12042 (Ery7))* were streaked from a glycerol stock onto TSAB agar plates and incubated anaerobically for ~ 24 h at 37 °C. Three colonies were inoculated into 1 ml of Mega broth and were anaerobically cultured

for overnight at 37 °C. Cells were diluted 1,000-fold into Mega broth to reach late-log phase. Then 5 µL of the culture was resuspended in 145 µL broth, loaded into a 96-well plate, and incubated anaerobically at 37 °C in Multiskan Sky Microplate Spectrophotometer (Thermo Fisher Scientific). Four bile acids (CA, 7-oxoCA, DCA, and 3-oxoDCA, 500 µM each) were tested with their solvent DMSO as control. Optical densities at 600 nm (OD₆₀₀) were recorded every 60 min until the cultures reached the stationary phase. Bacterial growth curves were performed in triplicate with each biological replicate deriving from a single colony.

[0290] Quantitative PCR (qPCR)

[0291] *qPCR of dCpfI targeting genes*

[0292] Three isolated single colonies of control or mutant strain were used to inoculate 5 mL pre-reduced Mega medium supplemented with 15 µg/mL thiamphenicol. The cultures were incubated for 36 hrs in the anaerobic chamber. Following incubation, the cultures were centrifuged at 1500 x g for 5 min, and the supernatant was discarded. RNA was extracted from the resulting bacterial pellet using Quick RNA fungal/bacterial kit (Zymo Research) following the manufacturer's protocol. Reverse transcription of extracted RNA into cDNA was performed using PrimeScript™ RT Reagent Kit (TaKaRa) following the manufacturer's protocol. Real-time quantitative PCR (qPCR) was performed on cDNA using SYBR green chemistry (Applied Biosystems). Reactions were run on a real-time quantitative PCR system (ABI 7500; Applied Biosystems). Samples were normalized to 16s rRNA of each strain.

[0293] *qPCR for the comparison of Erysipelotrichaceae relative fold change between groups*

[0294] For the qPCR of *Erysipelotrichaceae* abundance in liquid culture, overnight culture of two *Bacteroides* (*Bacteroides fragilis* 3_1_12 (*Bac1*) and *Bacteroides vulgatus* ATCC 8482 (*Bac2*)) and seven *Erysipelotrichaceae* (*Clostridium ramosum* ATCC 25554 (*Ery1*), *Erysipelatoclostridium ramosum* strain 113-1 (*Ery2*), *Clostridium ramosum* DSM 24812 (*Ery3*), *Clostridium ramosum* DSM 1402 (*Ery4*), HM-173 *Clostridium innocuum* 6_1_30 (*Ery5*), *Clostridium innocuum* DSM 22910 (*Ery6*) and *Holdemania filiformis* DSM 12042 (*Ery7*)) in Mega (~1 x 10⁷ CFU) were inoculated into Mega with different concentrations of DCA (0 µM, 250 µM, 500 µM), or co-inoculated together with S122 control or S122 *ΩbaiH* mutant into Mega

with 500 μ M CA. After incubation for 24h, gDNA was extracted using Quick RNA fungal/bacterial kit (Zymo Research) and qPCR was performed using primers Bac_Erysi_16s_qPCR_F-2 + Bac_Erysi_16s_qPCR_R-2 to amplify total 16s of both *Bacteroides* and *Erysipelotrichaceae* as reference, and primers Erysi_16s_qPCR_F + Erysi_16s_qPCR_R to amplify *Erysipelotrichaceae*-specific 16s for the comparison of *Erysipelotrichaceae* abundance between groups.

[0295] For the qPCR of *Erysipelotrichaceae* relative fold change in fecal samples, gDNA in fecal samples was extracted using QIAamp Fast DNA Stool Mini Kit (Cat. # 51604), and qPCR was performed using primers Bac_Erysi_16s_qPCR_F-2 + Bac_Erysi_16s_qPCR_R-2 to amplify total 16s of both *Bacteroides* and *Erysipelotrichaceae* as a reference, and primers Erysi_16s_qPCR_F + Erysi_16s_qPCR_R to amplify *Erysipelotrichaceae*-specific 16s for the comparison of *Erysipelotrichaceae* relative fold change between groups.

[0296] Colonize germ-free and SPF mice with the control and mutant bacteria

[0297] Germ-free mouse experiments were performed on gnotobiotic Swiss Webster or C57BL/6 mice, which were bred within sterile vinyl isolators and maintained at the gnotobiotic facility at Weill Cornell Medicine. SPF mice on a C57BL/6 background were purchased from the Jackson Laboratory and were bred and maintained in specific-pathogen-free facilities at Weill Cornell Medicine. Sex- and age-matched mice between 8 and 14 weeks of age were used for experiments if not otherwise indicated (n = 4 or 5 per group).

[0298] For mono-colonization in germ-free mice, taking *Eubacterium maltosivorans* DSM 105863 control (S117+pGM-FBCD) and its SCFAs mutant (S117+pGM-FBCD-020) as an example, a 200 μ L portion of their overnight RCM culture ($\sim 1 \times 10^7$ CFU) were mono-colonized with germ-free mice (n = 4 per group) via oral gavage. The germ-free mice were maintained on standard chow and water containing minimal thiamphenicol (15 μ g/mL). Successful colonization was determined by colony-forming unit (CFU) counting. After two weeks of colonization, mice were euthanized humanely by CO₂ asphyxiation. Blood was collected by cardiac puncture, and serum was prepared using microtainer serum separator tubes obtained from Becton Dickinson (Cat. # 365967). The urine, cecal contents, and feces were collected and snap-frozen in liquid nitrogen and stored at -80°C until use.

[0299] For co-colonization of *Faecalicatena contorta* S122 (S122) control or its *baiH* mutant together with *Bacteroides* sp. 1_1_6 (S25) in germ-free mice, three verified transconjugants were restreaked and subcultured in Mega broth, and 1 mL of their overnight Mega culture ($\sim 1 \times 10^7$ CFU) were mixed and co-colonized with germ-free mice ($n = 5$ per group) via oral gavage (300 μ L per mouse). The germ-free mice were maintained on standard chow and water supplemented with 15 μ g/ml thiamphenicol. Successful colonization was determined by the quantitative PCR (qPCR) of 16s gene of S122 and S25 (data not shown) and 16s rRNA sequencing.

[0300] For co-colonization of *Faecalicatena contorta* S122 (S122) control or its *baiH* mutant together with two *Bacteroides* (3-member community) mentioned in **Fig. 22G**, and co-colonization of *Faecalicatena contorta* S122 (S122) control or its *baiH* mutant together with two *Bacteroides* (*Bac1-2*) and seven *Erysipelotrichaceae* (*Ery1-7*) (10-member community) in **Fig. 7D**, 1 mL of their overnight Mega culture ($\sim 1 \times 10^7$ CFU) were mixed and co-colonized with germ-free mice ($n = 4$ per group) via oral gavage (300 μ L per mouse). The germ-free mice were maintained on standard chow, and cholic acid sodium salt (5mM for the 10-member community and 0.5 mM for the 3-member community) was supplied in water to facilitate S122 colonization and ensure both gnotobiotic mice settings have comparable gut bile acid profiles. Successful colonization of S122 was determined by the Colony-forming unit (CFU) and LCMS. After 14 days, mice were administered with DSS for 8 or 9 days. After DSS was removed and mice were recovered with regular drinking water for 1 or 2 days, mice were euthanized humanely by CO₂ asphyxiation. Blood was collected by cardiac puncture, and serum was prepared using microtainer serum separator tubes obtained from Becton Dickinson (Cat. # 365967). Colon length was measured, proximal colon/distal colon/ileum tissue samples were collected for histology, and colon/ileum tissue samples were collected for qPCR. The urine, cecal contents, ileal content, and feces were collected and snap-frozen in liquid nitrogen and stored at -80°C until use.

[0301] For co-colonization of *Faecalicatena contorta* S122 (S122) together with 55 other genetically targetable strains identified in this study, 1 mL of their overnight Mega/RCM/CMM culture ($\sim 1 \times 10^7$ CFU) were mixed and co-colonized with germ-free mice ($n = 5$ per group) via oral gavage (300 μ L per mouse). The germ-free mice were maintained on standard chow. Successful colonization of S122 was determined by LCMS, and colonization of other strains was

confirmed by 16s rRNA sequencing. After two weeks of colonization, mice were euthanized humanely by CO₂ asphyxiation. Blood was collected by cardiac puncture, and serum was prepared using microtainer serum separator tubes obtained from Becton Dickinson (Cat. # 365967). The urine, cecal contents, and feces were collected and snap-frozen in liquid nitrogen and stored at –80 °C until use.

[0302] For the colonization of *Faecalicatena contorta* S122 (S122) control or its *baiH* mutant in SPF mice, three verified transconjugants were restreaked and subcultured in Mega broth, and a 300 µL portion of their overnight Mega culture ($\sim 1 \times 10^7$ CFU) were colonized with SPF mice (n = 4 per group) via oral gavage, twice per day for 3 days in a row. The SPF mice were maintained on standard chow (Lab Diet 5053) and water containing thiamphenicol (15 µg/mL) and erythromycin (10 µg/mL). Successful colonization was determined by colony-forming unit (CFU) counting. After 14 days, mice were administered with DSS for 8 days. After DSS was removed and mice were recovered with regular drinking water (water containing 15 µg/mL thiamphenicol and 10 µg/mL erythromycin) for 3 days, mice were euthanized humanely by CO₂ asphyxiation. Blood was collected by cardiac puncture, and serum was prepared using microtainer serum separator tubes obtained from Becton Dickinson (Cat. # 365967). Colon length was measured, proximal colon/distal colon/ileum tissue samples were collected for histology, and colon/ileum tissue samples were collected for qPCR. The urine, cecal contents, ileal content, and feces were collected and snap-frozen in liquid nitrogen and stored at –80 °C until use.

[0303] Colony-forming unit (CFU) quantification of mouse fecal samples

[0304] For CFU of mono-colonized germ-free mice, a ~5 mg fecal material was resuspended in 200 µL pre-reduced Gibco™ phosphate-buffered saline buffer, pH 7.4. A 10-fold serial dilution (to 10⁻⁴) was made in the same buffer on a 96-well plate, and 50 µL from each well was plated on pre-reduced TSAB agar and was incubated anaerobically at 37 °C. After 24 hrs, colonies will appear, and the CFU of fecal samples from control and mutant strains colonized germ-free mice is calculated after normalizing to fecal weight.

[0305] For CFU of SPF mice colonized with *Faecalicatena contorta* S122 (S122) control (S122+ pGM-FCAQ + pGM-FCFQ) or its *baiH* mutant (S122 + pGM-FCAR-002 + pGM-

FCFQ, S122 *ΩbaiH* mutant), a ~5 mg fecal material was resuspended in 200 µL pre-reduced Gibco™ phosphate-buffered saline buffer, pH 7.4. A 10-fold serial dilution (to 10⁻⁴) was made in the same buffer on a 96-well plate and 50 µL from each well was plated on pre-reduced TSAB agar supplemented with 250 µg/mL D-cycloserine + 15 µg/mL thiamphenicol + 10 µg/mL erythromycin and was incubated anaerobically at 37 °C. After 24 hrs, colonies will appear and colonies were inoculated in 3 mL Mega broth supplemented with 250 µg/mL D-cycloserine + 15 µg/mL thiamphenicol + 10 µg/mL erythromycin. After 12 hrs, we extracted their genomic DNA using Quick DNA fungal/bacterial kit (Zymo Research). We performed diagnostic PCR using primers WBJ_BaiH_tgt_DiagF and WBJ_BaiH_tgt_DiagR to verify that colonies on plates were the control and the *baiH* mutant strain of S122. The CFU of fecal samples from control and mutant strains colonized SPF mice is calculated after normalizing to fecal weight.

[0306] Isolation of gut bacterial strains from collected fecal samples

[0307] Fecal samples (from human or mouse) were suspended in PBS (1:10 w/v), the suspension was then restreaked on TSAB/BHIB plates and incubated in an anaerobic chamber at 37 °C under an atmosphere consisting of 5% CO₂, 7.5% H₂, 87.5% N₂, in the meantime, the suspension was also restreaked on LB plates and incubated aerobically at 37 °C. Colonies typically appeared after 24-36 hrs, the isolated single colonies were inoculated in 3 mL Mega/RCM/CMM/TYBG or LB broth. After ~12 hrs, we extracted their genomic DNA using Quick DNA fungal/bacterial kit (Zymo Research), and amplified the 16s rRNA region of the colony using primers 16s_27F + 16s_1391R. The PCR product was purified and sent for Sanger sequencing using primer 16s_1391R to identify the species of the isolated strains.

[0308] DSS treatment of GF and SPF mice

[0309] *DSS administration*

[0310] Dextran sulfate sodium salt (DSS) of colitis-grade with an average MW of 36,000-50,000 Da (MP Biomedicals) was added to drinking water at day 0. DSS was administered until substantial inflammation was induced as evidenced by significant weight loss. For the GF mice experiment in **Fig. 6F**, DSS was used at a concentration of 2% (in water supplemented with 15 µg/ml thiamphenicol) for 7 days, and for SPF mice experiment in **Fig. 6A**, DSS was used at a

concentration of 2.5% (in water supplemented with 15 µg/ml thiamphenicol and 10 µg/ml erythromycin) for 8 days. After DSS treatment, DSS was removed from the drinking water, GF mice in **Fig. 6F** were recovered with regular water (with 15 µg/ml thiamphenicol) for 3 days and SPF mice in **Fig. 6A** were recovered with water containing 15 µg/mL thiamphenicol and 10 µg/mL erythromycin for 3 days. For the experiment in **Fig. 7D** and **Fig. 22G**, DSS was used at a concentration of 2.5 % for 9 days. After DSS treatment, DSS was removed from the drinking water, mice in **Fig. 7D** and **Fig. 22G** were recovered with regular water for 1 or 2 days. Throughout DSS treatment and recovery, mice were weighed daily at the same time of day at indicated time points, and feces were collected daily at the same time points. Mice were then euthanized by CO₂ asphyxiation. Blood was collected by cardiac puncture, and serum was prepared using microtainer serum separator tubes obtained from Becton Dickinson (Cat. # 365967). Colon length was measured, proximal colon/distal colon/ileum tissue samples were collected for histology, and colon/ileum tissue samples were collected for qPCR. The urine, cecal contents, ileal content, and feces were collected and snap-frozen in liquid nitrogen and stored at -80 °C until use.

[0311] *Quantitative PCR (qPCR) of colonic inflammatory genes expression post DSS-treatment*

[0312] For the qPCR of inflammatory genes in the colon, colonic samples were homogenized, and then RNA was extracted from the resulting homogenate using Quick RNA fungal/bacterial kit (Zymo Research) following the manufacturer's protocol. Reverse transcription of extracted RNA into cDNA was performed using PrimeScript™ RT Reagent Kit (TaKaRa) following the manufacturer's protocol. Real-time quantitative PCR (qPCR) was performed on cDNA using SYBR green chemistry (Applied Biosystems). Reactions were run on a real-time quantitative PCR system (ABI 7500; Applied Biosystems). Samples were normalized to *Hprt1*.

[0313] *Quantification of fecal lipocalin-2 (LCN-2) by ELISA*

[0314] For fecal lipocalin-2 quantification, fecal samples were collected and suspended in PBS containing 1% Bovine Serum Albumin (1g/100mL) to a final concentration of 100 mg/mL and vortexed for 20 min to get a homogenous fecal suspension. These samples were then centrifuged for 10 min at 14 000 g and 4°C to remove aggregates, and the resulting supernatant was collected. Afterward, according to the manufacturer's instructions, a sandwich ELISA was

performed following appropriate dilution using mouse lipocalin-2/NGAL DuoSet ELISA (R & R&D Systems).

[0315] *Assessment of fecal hematochezia score*

[0316] Fecal samples were collected daily at the same time of day at indicated time points and subjected to Hemocult II SENSE Dispensapak Plus kit (Becton Dickinson) to assess hematochezia scores following the manufacturer's instructions.

[0317] *Histology*

[0318] Distal colon sections were obtained and fixed in 10% neutral buffered formalin overnight at room temperature and then were transferred to 70% ethanol. Then sections were paraffin-embedded, sectioned, and stained with hematoxylin and eosin by IDEXX BioAnalytics company. Blinded histological evaluation was conducted on a scale of 1-3 or 4 for the following histologic parameters: area involved (0-4), erosion/ulceration (0-4), follicles (0-3), edema (0-3), fibrosis (0-3), crypt loss (0-4), granulocytes (0-3), mononuclear cells (0-3), and crypt damage/apoptosis (0-4). Scores were accumulated to give a total score of inflammation.

[0319] *16s rRNA gene sequencing of fecal samples*

[0320] For the 16s rRNA gene sequencing of fecal samples from SPF mice colonized with *Faecalibacterium prausnitzii* S122 (S122) control (**Con**) and *ΔbaiH* mutant (**Mut**), gDNA in fecal samples was extracted using QIAamp Fast DNA Stool Mini Kit (Cat. # 51604), and the concentration of double-stranded gDNA in the extracted gDNA was measured using Quant-iT™ dsDNA Assay Kit, high sensitivity (Cat. # Q33120). Then gDNA was normalized to 20 ng/μL and sent for 16s rRNA gene sequencing.

[0321] Next generation sequencing library preparations and Illumina MiSeq sequencing were conducted at GENEWIZ, Inc. (Suzhou, China). DNA samples were quantified using a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA). 30-50 ng DNA was used to generate amplicons using a MetaVx™ Library Preparation kit (GENEWIZ, Inc., South Plainfield, NJ, USA).

[0322] V3, V4, and V5 hypervariable regions of prokaryotic 16s rDNA were selected for generating amplicons and following taxonomy analysis. GENEWIZ designed a panel of

proprietary primers aimed at relatively conserved regions bordering the V3, V4, and V5 hypervariable regions of bacteria 16s rDNA. The v3 and v4 regions were amplified using forward primers containing the sequence "CCTACGRRBGCASCAGKVRVGAAT" (SEQ ID NO: 19) and reverse primers containing the sequence "GGACTACNVGGGTWTCTAATCC" (SEQ ID NO: 20). The v4 and v5 regions were amplified using forward primers containing the sequence "GTGYCAGCMGCCGCGGTAA" (SEQ ID NO: 21) and reverse primers containing the sequence "CTTGTGCGGKCCCCCGYCAATTC" (SEQ ID NO: 22). 1st round PCR products were used as templates for 2nd round amplicon enrichment PCR. At the same time, indexed adapters were added to the ends of the 16s rDNA amplicons to generate indexed libraries ready for downstream NGS sequencing on Illumina Miseq.

[0323] DNA libraries were validated by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified by Qubit 2.0 Fluorometer. DNA libraries were multiplexed and loaded on an Illumina MiSeq instrument according to manufacturer's instructions (Illumina, San Diego, CA, USA). Sequencing was performed using a 2x300/250 paired-end (PE) configuration.

[0324] The QIIME data analysis package was used for 16s rRNA data analysis. The forward and reverse reads were joined and assigned to samples based on barcode and truncated by cutting off the barcode and primer sequence. Quality filtering on joined sequences was performed and sequence which did not fulfill the following criteria were discarded: sequence length <200bp, no ambiguous bases, mean quality score ≥ 20 . Then the sequences were compared with the reference database (RDP Gold database) using UCHIME algorithm to detect chimeric sequence, then the chimeric sequences were removed (**Fig. 34**).

[0325] The effective sequences were used in the final analysis. Sequences were grouped into operational taxonomic units (OTUs) using the clustering program VSEARCH (1.9.6) against the Silva 119 database pre-clustered at 97% sequence identity. The Ribosomal Database Program (RDP) classifier was used to assign taxonomic category to all OTUs at a confidence threshold of 0.8. The RDP classifier uses the Silva 119 database which has taxonomic categories predicted to the species level (**Figs. 32 and 33**).

[0326] Sequences were rarefied prior to the calculation of alpha and beta diversity statistics. Alpha diversity indexes were calculated in QIIME from rarefied samples using for diversity the

Shannon index, for richness the Chao1 index. Beta diversity was calculated using principal coordinate analysis (PCoA) performed.

[0327] *Bioinformatics analyses performed in this study*

[0328] *Phylogenetic analyses of the 91 non-model gut commensals that are genetically targetable*

[0329] To construct the phylogenetic tree in **Fig. 1**, we extracted their 16s rRNA sequences from their sequenced genomes. If the bacteria have no sequenced genome information, their 16s rRNA sequences were downloaded from the Silva database (Joseph et al., 2018) or from the Sanger sequencing result of the PCR product amplified from their extracted genomic DNA. These 16s rRNA sequences were aligned using Clustal Omega, and the aligned sequences were used to construct a phylogenetic tree (neighbor-joining) with a bootstrap test of 5000 using Mega11 (Kumar et al., 2016).

[0330] *Analyses of the prevalence and relative abundances of Clostridia commensals harboring the bai operon in human stool datasets*

[0331] The publicly available 16s rRNA sequencing reads were downloaded and mapped to the 16s rRNA sequences of five *Clostridia* commensals, including *Faecalicatena contorta* S122 (S122), *Clostridium hylemonae* DSM15053, *Clostridium hiranonis* DSM 13275, *Clostridium scindens* ATCC35704, and *Dorea* sp. D27 using Geneious. We used a very stringent setting, and only reads with >95% quality and minimal 100% overlap identity will be mapped to their 16s rRNA sequences. The prevalence of the strain and their closely related microbes is calculated by dividing the number of stool samples with at least one mapped read by the total number of stool samples. Their relative abundances were calculated by dividing the total mapped reads by the stool sample's total reads. Similarly, the relative abundances of S122 control or mutant strain shown in **Fig. 4E** were analyzed and calculated by dividing the total mapped high-quality reads by the total nonchimeric reads of the mouse stool samples.

[0332] *Metatranscriptomic analyses of the S122 bai operon*

[0333] To determine if the S122 *bai* operon is actively transcribed under the condition of host colonization, we built a local DNA sequence database consisting of all the *bai* operons identified so far (**Figs. 16A-16C**) metatranscriptomic analyzes. We mapped metatranscriptomic reads from the stool of healthy human subjects (David et al., 2014) to this database using Bowtie 2 (local, high sensitivity); representative mapping results are shown in **Fig. 4A**.

[0334] *Analyses of the correlation between human fecal DCA level and specific taxonomic groups of gut microbes and fecal calprotectin*

[0335] The metabolomics data, fecal calprotectin data, and relative abundances of specific taxonomic groups of gut microbes whose relative abundances are affected by *baiH* depletion in our experiment were downloaded and extracted from the iHMP-IBD website (<https://ibdmdb.org/>). Longitudinal data of the same participant were Z-transformed (with a mean of 0 and an SD of 1). The correlation between **1)** fecal DCA and fecal calprotectin level, and **2)** fecal DCA and specific taxonomic groups of gut microbes whose abundances have been affected by *baiH* depletion, were assessed using Pearson correlation, with a pre-specific alpha level of 0.05 to assess statistical significance. A correlation of 0.2 or higher or of -0.2 or lower was considered moderate. In addition, a linear mixed model with random intercept was used to assess the association between fecal DCA and fecal calprotectin/gut microbe relative abundances, accounting for the longitudinal measurements obtained from the same individual. Z-transformed values (dots) and the fitted values based on the linear mixed model (line) were presented in **Figs. 5H and 19A-19I**.

Example 2: An overview of the GM pipeline

[0336] The overall GM workflow is summarized in **Fig. 1A**. There are three challenges toward building such a pipeline. First, there is no previously reported antibiotic marker that universally functions in most of the non-model microbes. By assessing the antibiotic resistance and testing different conjugation donors to introduce antibiotic markers into 201 gut isolates, we found that one chloramphenicol resistant marker operates in majority of the 91 transformed microbes (**Fig. 1B**). Second, *Firmicutes/Clostridia* microbes are highly abundant in healthy human guts, yet the genetic manipulation of this physiologically important, host-associated bacterial group remains largely unexplored (Waller et al., 2017). Lacking their genetic tools greatly limits mechanistic

dissection of the effects of *Firmicutes/Clostridia* genes on host biology. By optimizing multiple factors, we managed to identify gene transfer methods for 38 non-model gut *Clostridia* microbes (**Fig. 1**). Third, when we started building the pipeline, the genomes of many isolates had not been sequenced, posing a considerable roadblock in establishing their targeted gene manipulation tools on a large scale. To overcome this hurdle, we incorporated CRISPRi and a *lacZ* α transcription reporter or developed strategies to genetically target the bacterial 16S rRNA gene (**Fig. 1A**). For consistency, we consider a gut microbe as genetically targetable if exogenous DNA (shuttle or suicide plasmids) can be repeatedly introduced into the microbe *in vitro*. A genetic manipulation tool is established if targeted manipulation of its gene/gene expression is achieved in the microbes of interest.

Example 3: Selection of gut microbes and screening their culture conditions and antibiotic resistance

[0337] We prioritized *Firmicutes* and *Bacteroidetes* microbes that dominate healthy human guts (Cho and Blaser, 2012), but many (like *Clostridia* and *Prevotella*) do not have gene transfer methods and tractable genetic tools. We diversified the screened pool by selecting commensals from a variety of genera/species (**Fig. 22**). We first identified the culture conditions supporting the growth of these gut microbes (**Figs. 1A and 8, Figs. 23 and 31**). Next, we screened these microbes against a collection of antibiotics to identify the following (**Figs. 1A and 8, Fig. 24**): **1**) the MIC of an antibiotic to which they are susceptible, allowing its resistance gene to be used as a universal selective marker, and **2**) an antibiotic which they are resistant to but is active against *E. coli*, enabling suppression of *E. coli* growth after conjugation. For **1**), we determined the MIC of thiamphenicol that inhibits the growth of almost all the tested microbes (**Fig. 24**), and for **2**), almost all the *Clostridia* are resistant to D-cycloserine or kanamycin, and all the *Prevotella* to gentamycin or D-cycloserine (**Fig. 24**).

Example 4: A multifactorial optimization to identify gene transfer methodology for non-model *Clostridia*

[0338] Multiple reasons, including incompatible origins of replication (*rep oris*) and antibiotic markers, host endogenous defense systems, and very inefficient homologous recombination (HR), cause the genetics of gut *Firmicutes/Clostridia* commensals to be poorly investigated compared to its counterpart *Bacteroides*. (Waller et al., 2017b). Therefore, herein we have performed a multifactorial optimization of the transformation/conjugation parameters to identify

gene transfer conditions for previously untransformed gut *Firmicutes/Clostridia commensals* (**Fig. 2A**):

[0339] First, because our initial attempt to introduce the four most-used *Clostridium rep ori*s (Heap et al., 2009) into several gut *Clostridia* like *C. bolteae* were unsuccessful, we expanded the repertoire of *Clostridia rep ori*s and developed a mixed-conjugation strategy to introduce compatible *rep ori* into non-model gut *Clostridia* (**Figs. 2A, 10A-10B, and 11A-11B**, see **Example 1** for more details). Second, we utilized a universal *catP* marker regulated by a potent constitutive promoter, P_{pmtl-catP} or P_{fdx-cs} (identified via a promoter library screen in multiple non-model *Clostridia*, data not shown), to confer antibiotic resistance during conjugation/transformation (see **Example 1** for more details). This effort significantly reduced the workload of marker-switching when applying the pipeline to a large number of previously non-targetable *Clostridial* microbes (**Fig. 2A**). Third, we attempted different approaches, including utilization of an *E. coli* methylase-free 'sExpress' conjugation donor (Woods et al., 2019), decreasing restriction-modification (RM) recognition sites (Mermelstein et al., 1992; Purdy et al., 2002; Yang et al., 2016), and/or pre-methylate transforming DNA (Jennert et al., 2000; Pyne et al., 2014), to reduce the effect of *Clostridia* host defense systems during conjugation/transformation (**Fig. 2A**, see **Example 1** for more details). Last, several other parameters are optimized in this study, including conjugation time length, conjugation donor/recipient ratio, different conjugative plasmids, etc. (**Fig. 2A**, see **Example 1** and **Fig. 24** for more details). These optimized parameters are summarized in the **Fig. 24**, and detailed protocols for conjugation/electroporation are reported in **Example 1**.

[0340] These concerted efforts allowed us to identify gene transfer conditions for 38 *Clostridia* commensals (of 27 species) out of 92 *Clostridia* (of 66 species) tested (an overall 41.3 % success rate) (**Figs. 2A and 11A-11B**), suggesting the possibility of developing associated gene manipulation systems. As may be anticipated, multiple factors need to be optimized simultaneously to successfully introduce plasmids into previously untransformed *Clostridia* (**Fig. 24**). For instance, introducing plasmids into S71 *C. barlette* (that harbors a putative Type-IV RM system) requires a compatible *Clostridia rep ori*, a functional *catP* marker driven by a strong promoter (and plating on plates supplemented with thiamphenicol at MIC), an *E. coli* 'sExpress' donor that does not methylate plasmid DNA and harbors R702 conjugative plasmid, and

combination with other optimized parameters such as conjugation time and conjugation antibiotics detailed in **Fig. 24**. Interestingly, some *Clostridia* accept different *rep oris* even if they are closely related (*e.g.*, *C. bolteae* isolates, see **Fig. 25**), demonstrating the necessity of expanding the collection of *Clostridia* rep origins.

Example 5: Testing CRISPRi-dCpf1 system in multiple *Clostridia* commensals

[0341] The following critical step toward developing a *Clostridia* GM pipeline is identifying a genetic manipulation tool that enables targeted gene manipulation in most *Clostridia*. As with Cas9-initiated cutting and dCas9-mediated interference, CRISPR-based systems have been recently applied to *C. sporogenes* (Cañadas et al., 2019; Guo et al., 2019) and *C. difficile* (McAllister et al., 2017). We prioritized the CRISPRi-dCpf1 (deactivated Cpf1) system (Hong et al., 2018; Hur et al., 2016; Kim et al., 2017; Tang et al., 2017; Zetsche et al., 2015; Zhang et al., 2017) mainly because the dCpf1 does not initiate a DNA double-strand break, and the dCpf1 plasmids showed less toxicity and higher conjugation efficiency than Cas9 or Cpf1. In comparison, our preliminary test found that the double-stranded cut by Cas9/Cpf1 is lethal to many *Clostridia* because of their very inefficient HR. The CRISPRi-dCpf1 system incorporates a catalytically dead dCpf1 and a guide RNA (gRNA) repurposed for gene regulation in bacteria. During regulation, the dCpf1/crRNA complex binds to the template strand of a target gene and blocks the transcription elongation, thus suppressing gene expression (Kim et al., 2017; Zhang et al., 2017).

[0342] To test CRISPRi-dCpf1 in the genetically targetable *Clostridia*, we assembled the *dCpf1* and *lacZα* (as a transcription reporter) with the pGM plasmids harboring the nine rep origins (**Figs. 2B (left) and 9**, see **Example 1** for more details). *LacZα* was selected because of its small size (~300 bp) and robust expression in multiple *Clostridia*. We designed a duplex gRNA targeting both the promoter and the template strand of *lacZα* (**Figs. 2B (right), and 12A-12B**). We found that dCpf1 leads to efficient knockdowns (~3 to over 100 fold) of *lacZα* transcription in 25 *Clostridia* (**Figs. 2B and 12A-12B, Fig. 26**). Several tested *Clostridia* could not take in this set of vectors, probably because the conjugation efficiency is greatly diminished due to this vector's large size (>10kb) (Guo et al., 2019; Zhang et al., 2018). Altogether our data suggest that the CRISPRi-dCpf1 system regulates gene transcription in almost all the *Clostridia* that uptake extracellular plasmid DNA.

Example 6: A strategy targeting bacterial 16s rRNA genes to generate targeted gene insertion tools in non-model gut commensals

[0343] Besides CRISPRi, a targeted gene insertion tool will also facilitate studying the molecular functions of *Clostridia* genes. Over half of the 38 targetable *Clostridia* are not genome sequenced. We considered whether targeting their universally conserved DNA sequences (as 'an archery target') could enable selective genetic insertion of a *Clostridia* gene without prior knowledge of its genome sequence. However, highly conserved genes are generally functionally essential (Isenbarger et al., 2008), and a genetic mutation to these genes could be lethal. To find such a target, we interrogated the 16s rRNA gene that has been used to assess microbiome diversity and construct bacterial phylogeny. We believe that the 16s rRNA gene is an optimal target for two reasons: **1)** a microbe usually has multiple copies, such that the disruption of one will not be lethal; **2)** it is highly conserved among bacteria (Isenbarger et al., 2008). The same set of 16s-targeting vectors can be applied to different bacteria, thus significantly saving time and effort in sequencing and cloning. One example of a *Clostridia* 16s rRNA is provided below:

[0344] caggaaacagctatgacctgagtgggcgacgggtgagtaacgcgtgggtaacctgcctcatcacagggggataacagttggaa
acggatgctaataaccgataagaccacagcaccgcatggtgcgggggtaaaaactccggcggtatgagatggaccgcgtctgattagct
agttggtgaggtaacggcccaccaaggcgacgatcagtagccgacctgagaggggtgaccggccacattgggactgagacacggcccaa
actcctacgggagggcagcagtggtgggaatattgcacaatggcgaaagcctgatgcagcgacgccgcgtgagtgaagaaggatttcggtt
gtaaagctctttatcaggaagaaaatgacggtacctgactaagaagccccggctaactacgtgccagcagccgcggttaatacgtagg
ggcaagcgttatccggttactgggcgtaaaggagcgtaggcggaagctgatgtgaaagccggggctcaaccgcgggactgctt
ggaaactgtgagtgcaggagaggttaagtgaattcctagtgtagcgggtgaaatgcgtagatattaggaggaacaccagtggcgaaggcg
gcttactggactgtaactgacgctgaggctcgaaagcgtggggagcaaacaggattagataccctggtagtccacgccgtaaacgatgaat
actaggtgtygggagcccttcggtgccgcagctaacgcagtaagtattccgcctggggagtagcttcgaagaatgaaactcaaaggaatt
gacggggggcccgcaacgggtggagcatgtggttaattcgaagcaacgcgaagaaccttaccaggtcttgacatccatctgaccgaga
gatggggccttccttcgggcaggggagacaggtggtgcatggtgtcgtcagctcgtgtcgtgagatgttgggttaagtcccgaacgag
cgcaacccttatcyttagttgccagcattaagctgggcactctaggagactgccggggataacccggaggaaggtggggatgacgtcaa
atcatcatgccccttatgacctgggctacacacgtgctacaatggcgtaacaaagggaagcgagaccgcgaggccgagcaaatcccaa
aaagtctcagttcggttagtctgcaactcgactacatgaagctggaatcgctagtaatcgcgatcagaatgccgcggtgaatacgttcc
cgggccttgtagacaccgcccgtcacaccatgggagtcagtaacgcccgaagtcggtgacctaaccaaggaggagctgccgaaggtg
ggachgatgactggggtgaagtcgtaacaaggtagccgtatcggaaggtgcggctggatcacctctttctaaggaatacaaatccggccg
gccag (SEQ ID NO: 10)

[0345] Group II intron-directed mutagenesis systems, such as Targetron (Zhong et al., 2003) or Clostron (Heap et al., 2007), utilize base-pairing (between the excised intron lariat RNA and the target site DNA) for DNA target recognition to direct the site-specific insertion of a retrotransposition-activated selectable marker (RAM) into the targeted DNA loci. The RAM itself is interrupted by a self-splicing group I intron and only confers the corresponding antibiotic resistance after splicing out the group I intron and successful insertion into the *Clostridial* chromosome (Heap et al., 2007; Zhong et al., 2003). We proposed that a Group II intron targeting the 16s gene will likely integrate into the 16s loci of multiple *Clostridia*. To test this assumption, we aligned their 16s rRNA genes (from the HMP reference genomes (Turnbaugh et al., 2007)) and identified one potential, highly conserved target site of Group II intron (**Fig. 2C**). We then introduced the 16s-targeting Group II intron (16s-tron), along with their compatible rep origins and antibiotic RAM, into 19 targetable *Clostridia* (**Figs. 2C, 9, and 13A**). The RAM provides antibiotic resistance only upon integration into the *Clostridia* chromosome. We found 16 *Clostridia* whose chromosomes were targetedly inserted by the 16s-tron (**Figs. 2C and 13, Fig. 26, and Example 1** for detailed information), suggesting that this strategy is efficient in developing gene insertion tools for many *Clostridia*.

[0346] We tested whether a similar strategy can be applied to non-model Gram-negative gut commensals. We prioritized *Prevotella* microbes because there are limited genetic tools available for this genus (Li et al., 2021). Gram-negative bacteria, in general, have more efficient HR. We synthesized a chimeric 16s (Chi-16s) sequence with high homology to the *Prevotella* 16s rRNA genes (**Fig. 2D**). We introduced the Chi-16s with a suicide vector into 21 human-associated *Prevotella* isolates (Kraal et al., 2014) (**Fig. 22**). We found 7 *Prevotella* whose 16s loci were inserted by pGM-NAC₂P (**Fig. 2D, Figs. 24 and 27, and Example 1** for detailed information). The Chi-16s strategy was also applied to 45 *Bacteroides/Parabacteroides* microbes (some with genetic tools (Bencivenga-Barry et al., 2020; García-Bayona and Comstock, 2019; Salyers et al., 1999; Taketani et al., 2020)), and 35 gut-associated Gram-negative microbes from other phyla (Fung et al., 2016), leading us to identify the gene transfer methods for 41 of them (**Figs. 1C, Fig. 22, S2, S3, and S5 and Example 1** for detailed information). These data demonstrate that the HR-based Chi-16s strategy efficiently identifies their gene transfer methods and generates gene insertion tools in non-model *Prevotella* and *Bacteroides*, and gut-associated microbes from other phyla.

Example 7: Constructing mutants to modulate *Clostridia* gene transcription and microbiome metabolites production

[0347] To demonstrate the utility of genetic tools developed in this study, we selected a widely distributed gene *bcat* and modulated its expression in 12 *Clostridia*. (**Fig. 3A (top)**). The BCAT protein deaminates branched-chain amino acids into their keto acid form (Hur et al., 2016). (**Figs. 3A (bottom), 14, and Figs. 26 and 28**). A duplex *bcat*-targeting gRNA along with dCpf1 was introduced into 12 *Clostridia*, and *bcat* transcription was repressed in all the mutants with the dCpf1+gRNAs, compared to control with only dCpf1 (**Fig. 3A (bottom)**).

[0348] We next sought to utilize these gene insertion tools to modulate the production of microbiome-derived metabolites *in vitro* and in the context of host colonization. We selected short-chain fatty acids (SCFAs) propionate and butyrate, as well as branched-SCFAs (BSCFAs), because of their vital role in maintaining host immune homeostasis and metabolic health (Blander et al., 2017; Cani et al., 2019; Rooks and Garrett, 2016). We first identified several gut commensals as abundant producers of the corresponding metabolites by analyzing the SCFA profiles of our targetable commensals. Next, we generated a series of mutants that reduce their production *in vitro* by targeting the corresponding metabolic genes. For propionate, we deleted the methylmalonate mutase (*mmdA*) genes of three *Bacteroides* microbes that convert methylmalonate to propionate (**Figs. 3B and 15A-15C, and Figs. 27 and 28**) (Fischbach and Sonnenburg, 2011) (Reichardt et al., 2014). For butyrate, we targeted the crotonase gene (*croA*) (Vital et al., 2014) and used either dCpf1 to downregulate its expression or group II intron to knock out the gene (**Figs. 3C and 15A-15C, and Figs. 26 and 28**). For BSCFAs, we applied the dCpf1 tool to suppress *porA* expression in *C. sporogenes* (Guo et al., 2019). For all the mutants we generated, we found that the *in vitro* production of the corresponding metabolites is significantly reduced compared to the control, and their levels in the mono-colonized mice are also much less than that of the control (**Figs. 3D and 15A-15C, and Figs. 26, 27, and 28**). Taken together, these data show that we can utilize the genetic tools developed via the GM pipeline to modulate microbiome-derived metabolites *in vitro* and in the context of host colonization, suggesting their potential in systematically linking microbiota genes with their responsible metabolites and associated host biology.

Example 8: A case study of *Clostridia*-specific bile acid 7 α -dehydroxylation

[0349] We sought to use these genetic tools to study the effect of one microbiota gene on host biology. We selected the *bai* operon for 7 α -dehydroxylating of CA (cholate)/CDCA (chenodeoxycholate) to DCA (deoxycholate)/LCA (lithocholate) for follow-up studies. Three reasons motivated us to choose this pathway (**Figs. 4, 16A, and 17A**): **1)** Interestingly, we found one commensal S122 that efficiently converts CA(**1**)/CDCA to DCA(**3**)/LCA (**Figs. 4A and 4B**). Previous works have stepwise elucidated the chemistry and enzymology of 7 α -dehydroxylation (Funabashi et al., 2020; Ridlon et al., 2006, 2016). However, a key impediment to investigating *bai* operon biology is that all the identified *bai*-coding *Clostridia* (**Fig. 16A**) have no published gene transfer methods and tractable genetic tools (Ridlon et al., 2016). **2)** DCA/LCA and their derivatives dominate the host secondary bile acid pool (Arab et al., 2017). **3)** Amphipathic bile acids have intriguing biological activities: they inhibit the growth of enteric pathogens (Buffie et al., 2015), regulate mucosal immunity (Chen et al., 2019; Fiorucci et al., 2018), and promote liver cancer (Yoshimoto et al., 2013).

[0350] We sequenced S122 and identified its *bai* operon (**Fig. 4A**, see **Example 1** for detailed information). Our bioinformatic analyses revealed three unique features of S122: **1)** The strain is widely distributed among the healthy human population in two independent cohorts (41.30% (Lloyd-Price et al., 2019), 85.98% (Yatsunenko et al., 2012)). **2)** Like other 7 α -dehydroxylating *Clostridia*, S122 has a low intestinal abundance (~0.016%), but its *bai* operon is actively transcribed under conditions of host colonization (**Fig. 4A**). **3)** S122 and its close relatives are more prevalent and abundant than *C. hiranonis* or *C. scindens* (**Figs. 16B-16C**), indicating they play a significant role in regulating gut 7 α -dehydroxylating activity.

[0351] To manipulate the *bai* pathway *in vivo*, we generated a *baiH* insertion mutant (S122 Ω *baiH*) (**Figs. 4B and 17B**, see **Example 1** for detailed information). The *baiH* gene encodes an oxidoreductase that reduces 3-oxo-4,5-6,7-didehydro-DCA (**2**) to 3-oxo-4,5-didehydro-DCA (**Figs. 4B, 17A, and 18**) (Funabashi et al., 2020; Kang et al., 2008). The Ω *baiH* mutant depleted DCA and accumulated the intermediate (**2**) and 7-oxo CA *in vitro* (**Figs. 4B and 17A**). Unexpectedly, our attempt to efficiently mono-associate GF mice with S122 to induce *in vivo* DCA production proved unsuccessful. Instead, we found S122 can stably co-colonize the GF mice with S25, and knocking out *baiH* eliminates gut 7 α -dehydroxylating activity: The control

accumulated ~12 pmol/mg DCA (**Figs. 4C and 4D**), while the mutant abolished DCA but accumulated 7-oxo-CA (**Figs. 4D and 17C**). Moreover, S122 can stably co-colonize GF mice with 55 other genetically targetable microbes identified in this study (**Fig. 17D**). The relative abundance of S122 is low in both cases (**Fig. 4C and 17D**), but robust CA to DCA conversion can be detected, suggesting that S122 is a highly active 7 α -dehydroxylating bacterium in the host.

Example 9: The *baiH* gene has significant effects on the host bile acid pool and microbiota composition

[0352] This finding motivates us to knock out *baiH* in complex microbiota, like that of Specific Pathogen Free (SPF) mice. Manipulating microbiota genes in a complex microbiome provides a direct readout of their effects on microbial composition, which can be critical to explaining its impact on host biology. Unlike GF mice, the GI tract of SPF mice already harbors a complex microbiome with robust 7 α -dehydroxylating activity, leaving a limited niche for S122 to occupy. To overcome this challenge, we genetically tagged the control and Ω *baiH* mutant with a thiamphenicol-resistant marker. We supplemented their drinking water with thiamphenicol (15 μ g/ml) and erythromycin (10 μ g/ml) at very low concentrations for two reasons: **1**) to facilitate the colonization of the tagged strains that are resistant to these two antibiotics, and **2**) to eliminate the background 7 α -dehydroxylating activity conferred by the existing *bai*-coding *Clostridia*. This strategy led us to stably colonize the SPF mice with S122 control and the Ω *baiH* mutant at about the same level with comparable total bacterial load (**Figs. 5A, and 5B**) for at least 4 weeks. Because supplemented antibiotics minimally accumulate in the feces (~5 pmol/mg for thiamphenicol and not detectable for erythromycin), they do not reduce the total bacterial load compared to the SPF mice (**Fig. 17E**). Additionally, their effect on the gut microbiome is also controlled under this experimental setting.

[0353] To examine whether *baiH* deletion affects gut bile acid composition and the microbiome, we performed metabolomics and 16s rRNA sequencing analyses on stool samples (**Figs. 5 and 19A-19I**). Principal coordinate analysis (PCoA) demonstrated that stool samples are clustered by genotype (**Fig. 5C**). We drew two observations from these data: First, the control and Ω *baiH* colonized mice have different intestinal bile acid pools. Both groups have similar levels of conjugated bile acids like TCA and TCDCA (**Fig. 5D**), indicating *baiH* depletion does not

significantly modify microbiome bile salt hydrolyzing activity. DCA and its derivatives (such as isoDCA and 3-oxo DCA) are accumulated in the control group at levels comparable to host physiological levels. In contrast, the mutant group has higher levels of CA and its derivatives, including 7-oxo CA and UCA. (**Fig. 5D**)

[0354] Second, knocking out *baiH* modifies host gut microbiome composition. Both the control and mutant groups harbor a highly complex stool microbiota, and their overall phylum composition was maintained (**Fig. 5E**). The control group has a lower abundance of *Bacteroidetes*, higher *Proteobacteria*, and a significantly elevated *Erysipelotrichaceae* (**Figs. 5F, 19C, and 19D**). This compositional shift has been associated with worsened intestinal inflammation (Kaakoush, 2015; Kaser et al., 2010; Palm et al., 2014). A total of 56 operational taxonomic units (OTUs) were differentially abundant between groups, and they belong predominantly to the *Bacteroidia*, *Betaproteobacteria*, and *Erysipelotria* (**Fig. 5G**). Of note, the control has significantly more *Erysipelotrichaceae* that have high IgA coating and are associated with exacerbated colon inflammation (**Figs. 5F and 5G**) (Kaakoush, 2015; Palm et al., 2014). Aligned with our findings in the SPF mice, a higher stool DCA is positively associated with *Erysipelotrichaceae* abundance (**Fig. 19H**) and fecal calprotectin (a marker for the level of intestinal inflammation) (**Fig. 5H**) in nonIBD human stools. However, this correlation is not observed in patients with ulcerative colitis or Crohn's disease whose gut microbiota are usually structurally altered and whose gut 7 α -dehydroxylation activity is disrupted because of an exaggerated immune response (Lloyd-Price et al., 2019). These data indicate a potential modulatory role of *bai* operon in human gut microbiota and the onset of intestinal inflammation.

Example 10: Assessing the effect of *baiH* on intestinal inflammation

[0355] Because knocking out *baiH* shifts the gut microbiome to a less proinflammatory state (Kaser et al., 2010), we assessed whether *baiH* regulates intestinal inflammatory responses in a dextran sodium sulfate (DSS)-induced murine colitis model. The control and Ω *baiH* mutant colonized SPF mice were given drinking water with DSS (**Figs. 6A and 6F**). We found that both the S122 control and Ω *baiH* mutant strains stably colonized the mice (**Figs. 13F and 13I**), and the control group still has significantly higher *Erysipelotrichaceae* during DSS treatment (**Fig. 13G**). As colonic inflammation progressed, *baiH* indeed played a modulatory role in intestinal inflammation: The control lost more weight and experienced more severe inflammation as shown

by enhanced colonic pathology, shorter colon lengths, increased fecal lipocalin-2 levels, higher hematochezia score, and upregulation of inflammatory genes (**Figs. 6B-6E, 20A, 20C, and 20E**). Interestingly, the same DSS treatment successfully triggered an inflammation response in the GF C57BL/6 mice co-colonized with the control or Ω *baiH* mutant and S25, but knocking out *baiH* has no notable effect on intestinal inflammation (**Figs. 6G-6J, 20B, and 20D**). Taken together, these data indicate that *baiH*-mediated inflammatory responses are microbiota dependent, and *baiH* depletion in a complex microbiota reshapes host bile acid profiles and presets gut microbiome composition to a more protective state against DSS-induced colitis. More importantly, using a combination of microbiome genetics, metabolomics, and colitis mouse models, we demonstrate how a single commensal gene of a low intestinal abundance may significantly impact host biology by reshaping bile acid metabolism and the gut microbiota ecosystem.

Example 11: The *baiH*-mediated microbiota composition shift exacerbates DSS-Induced colitis in gnotobiotic mice

[0356] Motivated by our findings that *baiH* mediates colon inflammation in a complex microbiome, we proceeded to examine if microbiota composition shift induced by *baiH* deletion (**Figs. 5F and 20G**) could be related to the different intestinal inflammatory responses in the SPF mice under DSS treatment. First, we determined whether *Erysipelotrichaceae* expansion is *baiH*-dependent. Indeed, *Erysipelotrichaceae* isolates are more resistant to high concentrations of DCA and 3-oxo DCA compared to *Bacteroides* (**Figs. 7A and 21**). In a 10-member synthetic consortium we prepared *in vitro*, *Erysipelotrichaceae* also expands in the presence of *baiH* and its product DCA (**Figs. 7B and 7C**).

[0357] Next, we asked whether *baiH* drives *Erysipelotrichaceae* expansion *in vivo*, and whether this microbiota composition shift affects colon inflammatory responses in the DSS colitis model. We colonized two groups of germ-free C57BL6/N mice with the same 10-member synthetic consortium (S122 control or Ω *baiH* mutant with 7 *Erysipelotrichaceae* and 2 *Bacteroides*, **Figs. 7B and 7D**) and applied the DSS treatment two weeks post colonization (**Fig. 7D**). As expected, *baiH* also drives the expansion of *Erysipelotrichaceae* in the context of host colonization (**Fig. 7E**). The control group has exacerbated colon inflammation in this gnotobiotic setting as evaluated by severe weight loss (**Fig. 7F**), enhanced colonic pathology (**Fig. 7G**), shorter colon

lengths (**Fig. 7H**), increased fecal lipocalin-2 levels (**Fig. 7I**), and higher hematochezia score (**Fig. 7J**). The same DSS treatment also induced a robust inflammation response in the GF C57BL/6 mice co-colonized with the S122 control or $\Omega baiH$ mutant with only the two *Bacteroides* (three-member, **Fig. 22G**), however depleting *baiH* in this gnotobiotic setting has no notable effect on intestinal inflammation (**Figs. 22H to 22L**). The S122 control and $\Omega baiH$ mutant colonized the mice at comparable levels under both gnotobiotic settings (10-member vs. 3 member) (**Figs. 22A and 22B**), and their fecal bile acid profiles are comparable (**Figs. 22C to 22F**), suggesting that the different intestinal inflammatory response observed in the 10-member consortium colonized mice is more likely due to the expansion of *Erysipelotrichaceae* driven by *baiH*. Altogether, these data indicate that a *baiH*-mediated microbiota composition shift could exacerbate DSS-Induced colitis in the gnotobiotic mice, and the similar shift observed in the SPF mice (**Figs. 5F and 20G**) could be potentially related to the different intestinal inflammatory responses induced by *baiH* depletion in a complex microbiota. Of note, members of the synthetic consortium were selected based on the information we obtained by depleting *baiH* in a highly diverse microbiome, demonstrating the usefulness and necessity of studying the function of a microbiota gene in the background of a complex microbiota.

EQUIVALENTS

[0358] The present technology is not to be limited in terms of the particular embodiments described in this application, which are intended as single illustrations of individual aspects of the present technology. Many modifications and variations of this present technology can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. Functionally equivalent methods and apparatuses within the scope of the present technology, in addition to those enumerated herein, will be apparent to those skilled in the art from the foregoing descriptions. Such modifications and variations are intended to fall within the scope of the present technology. It is to be understood that this present technology is not limited to particular methods, reagents, compounds compositions or biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

[0359] In addition, where features or aspects of the disclosure are described in terms of Markush groups, those skilled in the art will recognize that the disclosure is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[0360] As will be understood by one skilled in the art, for any and all purposes, particularly in terms of providing a written description, all ranges disclosed herein also encompass any and all possible subranges and combinations of subranges thereof. Any listed range can be easily recognized as sufficiently describing and enabling the same range being broken down into at least equal halves, thirds, quarters, fifths, tenths, *etc.* As a non-limiting example, each range discussed herein can be readily broken down into a lower third, middle third and upper third, *etc.* As will also be understood by one skilled in the art all language such as “up to,” “at least,” “greater than,” “less than,” and the like, include the number recited and refer to ranges which can be subsequently broken down into subranges as discussed above. Finally, as will be understood by one skilled in the art, a range includes each individual member. Thus, for example, a group having 1-3 cells refers to groups having 1, 2, or 3 cells. Similarly, a group having 1-5 cells refers to groups having 1, 2, 3, 4, or 5 cells, and so forth.

[0361] All patents, patent applications, provisional applications, and publications referred to or cited herein are incorporated by reference in their entirety, including all figures and tables, to the extent they are not inconsistent with the explicit teachings of this specification.

REFERENCES

1. Arab, J.P., Karpen, S.J., Dawson, P.A., Arrese, M., and Trauner, M. (2017). Bile acids and nonalcoholic fatty liver disease: Molecular insights and therapeutic perspectives. *Hepatology* 65, 350–362.
2. Bencivenga-Barry, N.A., Lim, B., Herrera, C.M., Stephen Trent, M., and Goodman, A.L. (2020). Genetic manipulation of wild human gut bacteroides. *Journal of Bacteriology* 202.
3. Blander, J.M., Longman, R.S., Iliev, I.D., Sonnenberg, G.F., and Artis, D. (2017). Regulation of inflammation by microbiota interactions with the host. *Nature Immunology* 18, 851–860.

4. Buffie, C.G., Bucci, V., Stein, R.R., McKenney, P.T., Ling, L., Gobourne, A., No, D., Liu, H., Kinnebrew, M., Viale, A., et al. (2015). Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature* *517*, 205–208.
5. Campbell, C., McKenney, P.T., Konstantinovskiy, D., Isaeva, O.I., Schizas, M., Verter, J., Mai, C., Jin, W.B., Guo, C.J., Violante, S., et al. (2020). Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. *Nature* *581*, 475–479.
6. Cañadas, I.C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., and Minton, N.P. (2019). RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*. *ACS Synthetic Biology* *8*, 1379–1390.
7. Cani, P.D., Van Hul, M., Lefort, C., Depommier, C., Rastelli, M., and Everard, A. (2019). Microbial regulation of organismal energy homeostasis. *Nature Metabolism* *1*, 34–46.
8. Chen, M.L., Takeda, K., and Sundrud, M.S. (2019). Emerging roles of bile acids in mucosal immunity and inflammation. *Mucosal Immunology* *12*, 851–861.
9. Cho, I., and Blaser, M.J. (2012). The human microbiome: At the interface of health and disease. *Nature Reviews Genetics* *13*, 260–270.
10. Fiorucci, S., Biagioli, M., Zampella, A., and Distrutti, E. (2018). Bile acids activated receptors regulate innate immunity. *Frontiers in Immunology* *9*, 1.
11. Fischbach, M.A., and Sonnenburg, J.L. (2011). Eating for two: How metabolism establishes interspecies interactions in the gut. *Cell Host and Microbe* *10*, 336–347.
12. Funabashi, M., Grove, T.L., Wang, M., Varma, Y., McFadden, M.E., Brown, L.C., Guo, C., Higginbottom, S., Almo, S.C., and Fischbach, M.A. (2020). A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* *582*, 566–570.
13. Fung, T.C., Bessman, N.J., Hepworth, M.R., Kumar, N., Shibata, N., Kobuley, D., Wang, K., Ziegler, C.G.K., Goc, J., Shima, T., et al. (2016). Lymphoid-Tissue-Resident Commensal Bacteria Promote Members of the IL-10 Cytokine Family to Establish Mutualism. *Immunity* *44*, 634–646.

14. García-Bayona, L., and Comstock, L.E. (2019). Streamlined genetic manipulation of diverse bacteroides and parabacteroides isolates from the human gut microbiota. *MBio* 10.
15. Guo, C.-J., Allen, B.M., Hiam, K.J., Dodd, D., Van Treuren, W., Higginbottom, S., Nagashima, K., Fischer, C.R., Sonnenburg, J.L., Spitzer, M.H., et al. (2019). Depletion of microbiome-derived molecules in the host using *Clostridium* genetics. *Science* 366, eaav1282.
16. Hang, S., Paik, D., Yao, L., Kim, E., Jamma, T., Lu, J., Ha, S., Nelson, B.N., Kelly, S.P., Wu, L., et al. (2019). Bile acid metabolites control TH17 and Treg cell differentiation. *Nature* 576, 143–148.
17. Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P., and Minton, N.P. (2007). The ClosTron: A universal gene knockout system for the genus *Clostridium*. *Journal of Microbiological Methods* 70, 452–464.
18. Helmink, B.A., Khan, M.A.W., Hermann, A., Gopalakrishnan, V., and Wargo, J.A. (2019). The microbiome, cancer, and cancer therapy. *Nature Medicine* 25, 377–388.
19. Hong, W., Zhang, J., Cui, G., Wang, L., and Wang, Y. (2018). Multiplexed CRISPR-Cpf1-Mediated Genome Editing in *Clostridium difficile* toward the Understanding of Pathogenesis of *C. difficile* Infection. *ACS Synthetic Biology* 7, 1588–1600.
20. Hur, J.K., Kim, K., Been, K.W., Baek, G., Ye, S., Hur, J.W., Ryu, S.M., Lee, Y.S., and Kim, J.S. (2016). Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nature Biotechnology* 34, 807–808.
21. Isenbarger, T.A., Carr, C.E., Johnson, S.S., Finney, M., Church, G.M., Gilbert, W., Zuber, M.T., and Ruvkun, G. (2008). The most conserved genome segments for life detection on earth and other planets. *Origins of Life and Evolution of Biospheres* 38, 517–533.
22. Jennert, K.C.B., Tardif, C., Young, D.I., and Young, M. (2000). Gene transfer to *Clostridium cellulolyticum* ATCC 35319. *Microbiology* 146, 3071–3080.

23. Johnston, C.D., Cotton, S.L., Rittling, S.R., Starr, J.R., Borisy, G.G., Dewhirst, F.E., and Lemon, K.P. (2019). Systematic evasion of the restriction-modification barrier in bacteria. *Proceedings of the National Academy of Sciences* *116*, 11454–11459.
24. Kaakoush, N.O. (2015). Insights into the role of Erysipelotrichaceae in the human host. *Frontiers in Cellular and Infection Microbiology* *5*, 1–4.
25. Kang, D.-J., Ridlon, J.M., Moore, D.R. 2nd, Barnes, S., and Hylemon, P.B. (2008). *Clostridium scindens* baiCD and baiH genes encode stereo-specific 7 α /7 β -hydroxy-3-oxo- Δ^4 -cholenoic acid oxidoreductases. *Biochimica et Biophysica Acta* *1781*, 16–25.
26. Kaser, A., Zeissig, S., and Blumberg, R.S. (2010). Inflammatory bowel disease. *Annual Review of Immunology* *28*, 573–621.
27. Kim, S.K., Kim, H., Ahn, W.C., Park, K.H., Woo, E.J., Lee, D.H., and Lee, S.G. (2017). Efficient Transcriptional Gene Repression by Type V-A CRISPR-Cpf1 from *Eubacterium eligens*. *ACS Synthetic Biology* *6*, 1273–1282.
28. Klompe, S.E., Vo, P.L.H., Halpin-Healy, T.S., and Sternberg, S.H. (2019). Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* *571*, 219–225.
29. Kraal, L., Abubucker, S., Kota, K., Fischbach, M.A., and Mitreva, M. (2014). The prevalence of species and strains in the human microbiome: A resource for experimental efforts. *PLoS ONE* *9*.
30. Li, J., Gálvez, E.J.C., Amend, L., Almasi, É., Iljazovic, A., Lesker, T.R., Bielecka, A.A., and Strowig, T. (2021). A versatile genetic toolbox for *Prevotella copri*; enables studying polysaccharide utilization systems. *BioRxiv* 2021.03.19.436125.
31. Lim, B., Zimmermann, M., Barry, N.A., and Goodman, A.L. (2017). Engineered Regulatory Systems Modulate Gene Expression of Human Commensals in the Gut. *Cell* *169*, 547-558.e15.

32. Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* *569*, 655–662.
33. McAllister, K.N., Bouillaut, L., Kahn, J.N., Self, W.T., and Sorg, J.A. (2017). Using CRISPR-Cas9-mediated genome editing to generate *C. difficile* mutants defective in selenoproteins synthesis. *Scientific Reports* *7*, 1–12.
34. Mermelstein, L.D., Welker, N.E., Bennett, G.N., and Papoutsakis, E.T. (1992). Expression of cloned homologous fermentative genes in *clostridium acetobutylicum* ATCC 824. *Bio/Technology* *10*, 190.
35. Mimee, M., Tucker, A.C., Voigt, C.A., and Lu, T.K. (2015). Programming a Human Commensal Bacterium, *Bacteroides thetaiotaomicron*, to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell Systems* *1*, 62–71.
36. Palm, N.W., De Zoete, M.R., Cullen, T.W., Barry, N.A., Stefanowski, J., Hao, L., Degnan, P.H., Hu, J., Peter, I., Zhang, W., et al. (2014). Immunoglobulin A coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell* *158*, 1000–1010.
37. Purdy, D., O'Keeffe, T.A.T., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., and Minton, N.P. (2002). Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. *Molecular Microbiology* *46*, 439–452.
38. Pyne, M.E., Bruder, M., Moo-Young, M., Chung, D.A., and Chou, C.P. (2014). Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. *Biotechnology Advances* *32*, 623–641.
39. Reichardt, N., Duncan, S.H., Young, P., Belenguer, A., McWilliam Leitch, C., Scott, K.P., Flint, H.J., and Louis, P. (2014). Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME Journal* *8*, 1323–1335.
40. Ridlon, J.M., Kang, D.J., and Hylemon, P.B. (2006). Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research* *47*, 241–259.

41. Ridlon, J.M., Harris, S.C., Bhowmik, S., Kang, D.J., and Hylemon, P.B. (2016). Consequences of bile salt biotransformations by intestinal bacteria. *Gut Microbes* 7, 22–39.
42. Rooks, M.G., and Garrett, W.S. (2016). Gut microbiota, metabolites and host immunity. *Nature Reviews Immunology* 16, 341–352.
43. Roy, S., and Trinchieri, G. (2017). Microbiota: A key orchestrator of cancer therapy. *Nature Reviews Cancer* 17, 271–285.
44. Salyers, A.A., Shoemaker, N., Cooper, A., D'Elia, J., and Shipman, J.A. (1999). 8 Genetic Methods for *Bacteroides* Species. *Methods in Microbiology* 29, 229–249.
45. Sinha, S.R., Haileselassie, Y., Nguyen, L.P., Tropini, C., Wang, M., Becker, L.S., Sim, D., Jarr, K., Spear, E.T., Singh, G., et al. (2020). Dysbiosis-Induced Secondary Bile Acid Deficiency Promotes Intestinal Inflammation. *Cell Host and Microbe* 27, 659-670.e5.
46. Song, X., Sun, X., Oh, S.F., Wu, M., Zhang, Y., Zheng, W., Geva-Zatorsky, N., Jupp, R., Mathis, D., Benoist, C., et al. (2020). Microbial bile acid metabolites modulate gut ROR γ ⁺ regulatory T cell homeostasis. *Nature* 577, 410–415.
47. Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E. v., and Zhang, F. (2019). RNA-guided DNA insertion with CRISPR-associated transposases. *Science* 364, 48–53.
48. Taketani, M., Zhang, J., Zhang, S., Triassi, A.J., Huang, Y.J., Griffith, L.G., and Voigt, C.A. (2020). Genetic circuit design automation for the gut resident species *Bacteroides thetaiotaomicron*. *Nature Biotechnology* 1–8.
49. Tang, X., Lowder, L.G., Zhang, T., Malzahn, A.A., Zheng, X., Voytas, D.F., Zhong, Z., Chen, Y., Ren, Q., Li, Q., et al. (2017). A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nature Plants* 2017 3:3 3, 1–5.
50. Thomas, A.M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., Beghini, F., Manara, S., Karcher, N., Pozzi, C., et al. (2019). Metagenomic analysis of colorectal cancer

datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nature Medicine* 25, 667–678.

51. Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R., and Gordon, J.I. (2007). The Human Microbiome Project. *Nature* 449, 804–810.

52. Vital, M., Howe, A., and Tiedje, J. (2014). Revealing the Bacterial Synthesis Pathways by Analyzing (Meta) Genomic Data. *MBio* 5, 1–11.

53. Vo, P.L.H., Ronda, C., Klompe, S.E., Chen, E.E., Acree, C., Wang, H.H., and Sternberg, S.H. (2021). CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nature Biotechnology* 39, 480–489.

54. Waller, M.C., Bober, J.R., Nair, N.U., and Beisel, C.L. (2017a). Toward a genetic tool development pipeline for host-associated bacteria. *Current Opinion in Microbiology* 38, 156–164.

55. Waller, M.C., Bober, J.R., Nair, N.U., and Beisel, C.L. (2017b). Toward a genetic tool development pipeline for host-associated bacteria. *Current Opinion in Microbiology* 38, 156–164.

56. Wang, J., Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60.

57. Whitaker, W.R., Shepherd, E.S., and Sonnenburg, J.L. (2017). Tunable Expression Tools Enable Single-Cell Strain Distinction in the Gut Microbiome. *Cell* 169, 538–546.e12.

58. Wirbel, J., Pyl, P.T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J.S., Voigt, A.Y., Palleja, A., Ponnudurai, R., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nature Medicine* 25, 679–689.

59. Woods, C., Humphreys, C.M., Rodrigues, R.M., Ingle, P., Rowe, P., Henstra, A.M., Köpke, M., Simpson, S.D., Winzer, K., and Minton, N.P. (2019). A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe* 59, 184–191.
60. Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., Watanabe, H., Masuda, K., Nishimoto, Y., Kubo, M., et al. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nature Medicine* 25, 968–976.
61. Yang, X., Xu, M., and Yang, S.T. (2016). Restriction modification system analysis and development of in vivo methylation for the transformation of *Clostridium cellulovorans*. *Applied Microbiology and Biotechnology* 100, 2289–2299.
62. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222–227.
63. Yoshimoto, S., Loo, T.M., Atarashi, K., Kanda, H., Sato, S., Oyadomari, S., Iwakura, Y., Oshima, K., Morita, H., Hattori, M., et al. (2013). Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* 499, 97–101.
64. Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A., et al. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* 163, 759–771.
65. Zhang, J., Zong, W., Hong, W., Zhang, Z.T., and Wang, Y. (2018). Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. *Metabolic Engineering* 47, 49–59.
66. Zhang, X., Wang, J., Cheng, Q., Zheng, X., Zhao, G., and Wang, J. (2017). Multiplex gene regulation by CRISPR-ddCpf1. *Cell Discovery* 2017 3:1 3, 1–9.
67. Zhao, S., Gong, Z., Du, X., Tian, C., Wang, L., Zhou, J., Xu, C., Chen, Y., Cai, W., and Wu, J. (2018). Deoxycholic acid-mediated sphingosine-1-phosphate receptor 2 signaling

exacerbates DSS-induced colitis through promoting cathepsin b release. *Journal of Immunology Research* 2018.

68. Zhong, J., Karberg, M., and Lambowitz, A.M. (2003). Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Research* 31, 1656–1664.

69. Zhou, W., Sailani, M.R., Contrepois, K., Zhou, Y., Ahadi, S., Leopold, S.R., Zhang, M.J., Rao, V., Avina, M., Mishra, T., et al. (2019). Longitudinal multi-omics of host–microbe dynamics in prediabetes. *Nature* 569, 663–671.

70. Cañadas, I.C., Groothuis, D., Zygouropoulou, M., Rodrigues, R., and Minton, N.P. (2019). RiboCas: A Universal CRISPR-Based Editing Tool for *Clostridium*. *ACS Synthetic Biology* 8, 1379–1390.

71. David, L.A., Maurice, C.F., Carmody, R.N., Gootenberg, D.B., Button, J.E., Wolfe, B.E., Ling, A. V, Devlin, A.S., Varma, Y., Fischbach, M.A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563.

72. Funabashi, M., Grove, T.L., Wang, M., Varma, Y., McFadden, M.E., Brown, L.C., Guo, C., Higginbottom, S., Almo, S.C., and Fischbach, M.A. (2020). A metabolic pathway for bile acid dehydroxylation by the gut microbiome. *Nature* 582, 566–570.

73. Guo, C.-J., Allen, B.M., Hiam, K.J., Dodd, D., Van Treuren, W., Higginbottom, S., Nagashima, K., Fischer, C.R., Sonnenburg, J.L., Spitzer, M.H., et al. (2019). Depletion of microbiome-derived molecules in the host using *Clostridium* genetics. *Science* 366, eaav1282.

74. Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P., and Minton, N.P. (2007). The ClosTron: A universal gene knock-out system for the genus *Clostridium*. *Journal of Microbiological Methods* 70, 452–464.

75. Heap, J.T., Pennington, O.J., Cartman, S.T., and Minton, N.P. (2009). A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods* 78, 79–85.

76. Hur, J.K., Kim, K., Been, K.W., Baek, G., Ye, S., Hur, J.W., Ryu, S.M., Lee, Y.S., and Kim, J.S. (2016). Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nature Biotechnology* 34, 807–808.

77. Lu, Y., Yao, D., and Chen, C. (2013). 2-Hydrazinoquinoline as a Derivatization Agent for LC-MS-Based Metabolomic Investigation of Diabetic Ketoacidosis. *Metabolites* 3, 993–1010.

78. Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008). Mucosal Glycan Foraging Enhances Fitness and Transmission of a Saccharolytic Human Gut Bacterial Symbiont. *Cell Host and Microbe* 4, 447–457.

79. Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069.

80. Tang, X., Lowder, L.G., Zhang, T., Malzahn, A.A., Zheng, X., Voytas, D.F., Zhong, Z., Chen, Y., Ren, Q., Li, Q., et al. (2017). A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nature Plants* 3, 17018.

81. Woods, C., Humphreys, C.M., Rodrigues, R.M., Ingle, P., Rowe, P., Henstra, A.M., Köpke, M., Simpson, S.D., Winzer, K., and Minton, N.P. (2019). A novel conjugal donor strain for improved DNA transfer into *Clostridium* spp. *Anaerobe* 59, 184–191.

82. Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222–227.

CLAIMS

1. A bacterial expression vector comprising (a) a nucleic acid encoding a target gene that is conserved in a plurality of human gut commensal gram-negative bacterial species and (b) a heterologous nucleic acid encoding a selectable marker, wherein the selectable marker is an antibiotic resistance gene or an auxotrophic marker, and optionally wherein the target gene is selected from the group consisting of 16s rRNA, 23s rRNA, mmdA, RokA (Clucokinase gene), and an ABC transporter gene.
2. The bacterial expression vector of claim 1, wherein the 16s rRNA comprises the nucleic acid sequence of SEQ ID NO: 11.
3. The bacterial expression vector of claim 1 or 2, wherein the bacterial expression vector comprises the nucleic acid sequence of SEQ ID NO: 310.
4. The bacterial expression vector of any one of claims 1-3, further comprising at least one open reading frame encoding a bioluminescent protein, a chemiluminescent protein, a fluorescent protein, a CRISPR enzyme, a Group II intron-encoded protein, at least one sgRNA, at least one Group II intron, or any combination thereof.
5. A bacterial expression vector comprising (a) a gram-positive bacteria replication origin comprising a sequence selected from the group consisting of SEQ ID NOs: 1-9 or 311-319, (b) a heterologous nucleic acid encoding a selectable marker that is an antibiotic resistance gene or an auxotrophic marker, and (c) at least one open reading frame, wherein the at least one open reading frame encodes a bioluminescent protein, a chemiluminescent protein, a fluorescent protein, a CRISPR enzyme, a Group II intron-encoded protein, at least one sgRNA, at least one Group II intron, or any combination thereof.
6. The bacterial expression vector of claim 5, wherein the at least one sgRNA or the at least one Group II intron targets one or more genes selected from among 16S rRNA, *porA*, *bcat*, *croA*, *baiA2*, *baiCD*, *baiF*, *baiH*, *baiB*, *baiE*, *baiG* and *bail*.
7. The bacterial expression vector of any one of claims 5-6, further comprising one or more bacterial conjugation transfer genes and/or an *E. coli* replication origin, optionally wherein the

one or more bacterial conjugation transfer genes are selected from the group consisting of *traJ*, and *oriT* and/or the *E. coli* replication origin is selected from the group consisting of *colE1*, *pBR*, and *R6K*.

8. The bacterial expression vector of any one of claims 5-7, wherein the one or more bacterial conjugation transfer genes, the gram-positive bacteria replication origin, and the heterologous nucleic acid encoding the selectable marker are codon optimized.

9. The bacterial expression vector of any one of claims 1-8, wherein the antibiotic resistance gene is selected from the group consisting of *catP*, *ermB*, *aad9*, *tetA*, and *ampR*, or wherein the auxotrophic marker is *pyrG*, or *pyrF*.

10. The bacterial expression vector of any one of claims 4-9, wherein the CRISPR enzyme is selected from the group consisting of Cas9, dCas9, Cpf1, dCpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, and Csf4.

11. The bacterial expression vector of any one of claims 4-10, wherein the fluorescent protein is GFP, YFP, CFP, RFP, TagBFP, Azurite, EBFP2, mKalamal, Sirius, Sapphire, T-Sapphire, ECFP, Cerulean, SFCFP3A, mTurquoise, monomeric Midoriishi-Cyan, TagCFP, mTFP1, EGFP, Emerald, Superfolder GFP, Monomeric Azami Green, TagGFP2, mUKG, mWasabi, EYFP, Citrine, Venus, SYFP2, TagYFP, Monomeric Kusabira-Orange, mKOκ, mKO2, mOrange, mOrange2, mRaspberry, mCherry, dsRed, mStrawberry, mTangerine, tdTomato, TagRFP, TagRFP-T, mApple, mRuby, mPlum, HcRed-Tandem, mKate2, mNeptune, NirFP, TagRFP657, IFP1.4, iRFP, mKeima Red, LSS-mKate1, LSS-mKate2, PA-GFP, PAmCherry1, PAmTagRFP, Kaede (green), Kaede (red), KikGR1 (green), KikGR1 (red), PS-CFP2, PS-CFP2, mEos2 (green), mEos2 (red), PSmOrange, or Dronpa.

12. The bacterial expression vector of any one of claims 4-11, wherein the chemiluminescent protein is β -galactosidase, horseradish peroxidase (HRP), or alkaline phosphatase.

13. The bacterial expression vector of any one of claims 4-12, wherein the bioluminescent protein is Aequorin, firefly luciferase, *Renilla* luciferase, red luciferase, luxAB, or nanoluciferase.

14. The bacterial expression vector of any one of claims 4-13, wherein the at least one sgRNA specifically hybridizes with a heterologous or endogenous target gene expressed in a gut bacterial host cell, and/or wherein the at least one sgRNA and/or the CRISPR enzyme is operably linked to a constitutive promoter or a conditional promoter.

15. The bacterial expression vector of any one of claims 4-14, wherein the at least one Group II intron specifically targets a heterologous or endogenous gene expressed in a gut bacterial host cell, and/or wherein the at least one Group II intron and/or the Group II intron-encoded protein is operably linked to a constitutive promoter or a conditional promoter.

16. An engineered gram-negative human gut bacterial cell comprising the bacterial expression vector of any one of claims 1-4 or 9-15, wherein the engineered gram-negative human gut bacterial cell is derived from a family selected from the group consisting of *Enterobacteriaceae*, *Bacteroidaceae*, *Tannerellaceae*, and *Prevotellaceae*.

17. The engineered gram-negative human gut bacterial cell of claim 16, wherein the engineered gram-negative human gut bacterial cell is derived from *Bacteroides cellulosilyticus*, *Bacteroides cellulosilyticus*, *Bacteroides dorei*, *Bacteroides eggerthii*, *Bacteroides finegoldii*, *Bacteroides fragilis*, *Bacteroides intestinalis*, *Bacteroides nordii*, *Bacteroides oleiciplenus*, *Bacteroides ovatus*, *Bacteroides salyersiae*, *Bacteroides sp.*, *Bacteroides thetaiotaomicron*, *Bacteroides uniformis*, *Bacteroides vulgatus*, *Bacteroides xylanisolvens*, *Parabacteroides faecis*, *Parabacteroides merdae*, or *Prevotella bivia*.

18. An engineered gram-positive human gut bacterial cell comprising the bacterial expression vector of any one of claims 5-15, wherein the engineered gram-positive human gut bacterial cell is derived from a family selected from the group consisting of *Clostridiaceae*, *Lachnospiraceae*, *Eubacteriaceae*, *Erysipelotrichaceae*, *Enterococcaceae*, and *Bifidobacteriaceae*.

19. The engineered gram-positive human gut bacterial cell of claim 18, wherein the engineered gram-positive human gut bacterial cell is derived from *Blautia hydrogenotrophica*,

Blautia luti, *Blautia* sp., *Blautia wexlerae*, *Clostridium bolteae*, *Clostridium innocuum*, *Clostridium paraputrificum*, *Clostridium saccharolyticum*, *Clostridium senegalense*, *Clostridium* sp., *Clostridium sporogenes*, *Clostridium symbiosum*, *Eubacterium limosum*, *Eubacterium maltosivorans*, *Eubacterium ramulus*, *Eubacterium* sp., *Roseburia inulinivorans*, *Bifidobacterium catenulatum*, *Enterococcus faecium*, *Escherichia fergusonii*, *Roseburia inulinivorans*, or *Bifidobacterium catenulatum*.

20. A kit comprising the bacterial expression vector of any one of claims 1-15 and instructions for using the bacterial expression vector to genetically modify human gut bacteria.

21. The kit of claim 16, further comprising one or more primers and/or gRNAs comprising the sequence of any one of SEQ ID NOs: 23-287.

22. A method for modifying a gram-negative human gut bacteria cell genome comprising transferring at least one bacterial expression vector of any one of claims 1-4 or 9-15 into a gram-negative human gut bacteria cell via conjugation.

23. The method of claim 22, wherein the at least one bacterial expression vector is integrated into the genome of the gram-negative human gut bacteria cell.

24. A method for genetically modifying a gram-positive human gut bacteria cell comprising transferring two or more distinct bacterial expression vectors into a gram-positive human gut bacteria cell simultaneously via conjugation, wherein each of the two or more distinct bacterial expression vectors comprise:

(a) a gram-positive bacteria replication origin comprising a sequence selected from the group consisting of SEQ ID NOs: 1-9 or 311-319,

(b) a heterologous nucleic acid encoding a selectable marker that is an antibiotic resistance gene or an auxotrophic marker, and

(c) at least one open reading frame, wherein the at least one open reading frame encodes a bioluminescent protein, a chemiluminescent protein, a fluorescent protein, a CRISPR enzyme, a Group II intron-encoded protein, at least one sgRNA, at least one Group II intron, or any combination thereof.

25. The method of claim 24, wherein each of the two or more distinct bacterial expression vectors further comprise one or more bacterial conjugation transfer genes and/or an *E. coli* replication origin, optionally wherein the one or more bacterial conjugation transfer genes are selected from the group consisting of *traJ*, and *oriT* and/or the *E. coli* replication origin is selected from the group consisting of *colE1*, *pBR*, and *R6K*.

26. The method of claim 24 or 25, wherein the antibiotic resistance gene or the auxotrophic marker of each of the two or more distinct bacterial expression vectors is independently selected from the group consisting of *catP*, *ermB*, *aad9*, *tetA*, *ampR*, *pyrG*, and *pyrF*.

27. The method of any one of claims 24-26, wherein the CRISPR enzyme of each of the two or more distinct bacterial expression vectors is independently selected from the group consisting of Cas9, dCas9, Cpf1, dCpf1, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, and Csf4.

28. The method of any one of claims 24-27, wherein the fluorescent protein of each of the two or more distinct bacterial expression vectors is independently selected from the group consisting of GFP, YFP, CFP, RFP, TagBFP, Azurite, EBFP2, mKalama1, Sirius, Sapphire, T-Sapphire, ECFP, Cerulean, SCFP3A, mTurquoise, monomeric Midoriishi-Cyan, TagCFP, mTFP1, EGFP, Emerald, Superfolder GFP, Monomeric Azami Green, TagGFP2, mUKG, mWasabi, EYFP, Citrine, Venus, SYFP2, TagYFP, Monomeric Kusabira-Orange, mKOκ, mKO2, mOrange, mOrange2, mRaspberry, mCherry, dsRed, mStrawberry, mTangerine, tdTomato, TagRFP, TagRFP-T, mApple, mRuby, mPlum, HcRed-Tandem, mKate2, mNeptune, NirFP, TagRFP657, IFP1.4, iRFP, mKeima Red, LSS-mKate1, LSS-mKate2, PA-GFP, PAmCherry1, PAtagRFP, Kaede (green), Kaede (red), KikGR1 (green), KikGR1 (red), PS-CFP2, PS-CFP2, mEos2 (green), mEos2 (red), PSmOrange, and Dronpa.

29. The method of any one of claims 24-28, wherein the chemiluminescent protein of each of the two or more distinct bacterial expression vectors is independently β -galactosidase, horseradish peroxidase (HRP), or alkaline phosphatase.

30. The method of any one of claims 24-29, wherein the bioluminescent protein of each of the two or more distinct bacterial expression vectors is independently Aequorin, firefly luciferase, *Renilla* luciferase, red luciferase, luxAB, or nanoluciferase.

31. The method of any one of claims 24-30, wherein the at least one sgRNA sequence of the two or more distinct bacterial expression vectors specifically hybridizes with a heterologous or endogenous target gene expressed in a gut bacterial host cell, and/or wherein the at least one sgRNA and/or the CRISPR enzyme is operably linked to a constitutive promoter or a conditional promoter.

32. The method of any one of claims 24-31, wherein the at least one Group II intron of the two or more distinct bacterial expression vectors specifically targets a heterologous or endogenous gene expressed in a gut bacterial host cell, and/or wherein the at least one Group II intron and/or the Group II intron-encoded protein is operably linked to a constitutive promoter or a conditional promoter.

33. The method of any one of claims 24-32, wherein three or four distinct bacterial expression vectors are simultaneously transferred into a gram-positive human gut bacteria cell simultaneously via conjugation.

34. The method of any one of claims 22-33, wherein the gram-negative or gram-positive human gut bacteria cell is isolated from a colonic mucosa-enriched lavage sample, a fecal sample, a rectal swab, or an intestinal sample obtained from a human subject.

35. An engineered human gut bacterial cell generated by the method of any one of claims 22-34.

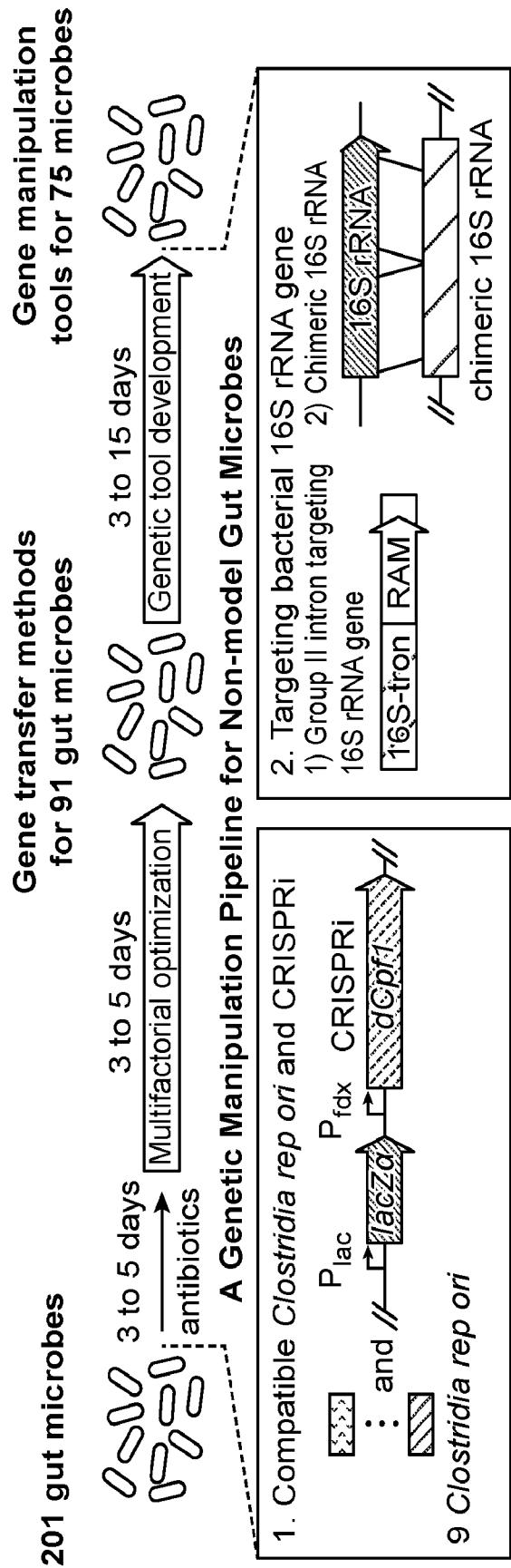


FIG. 1A

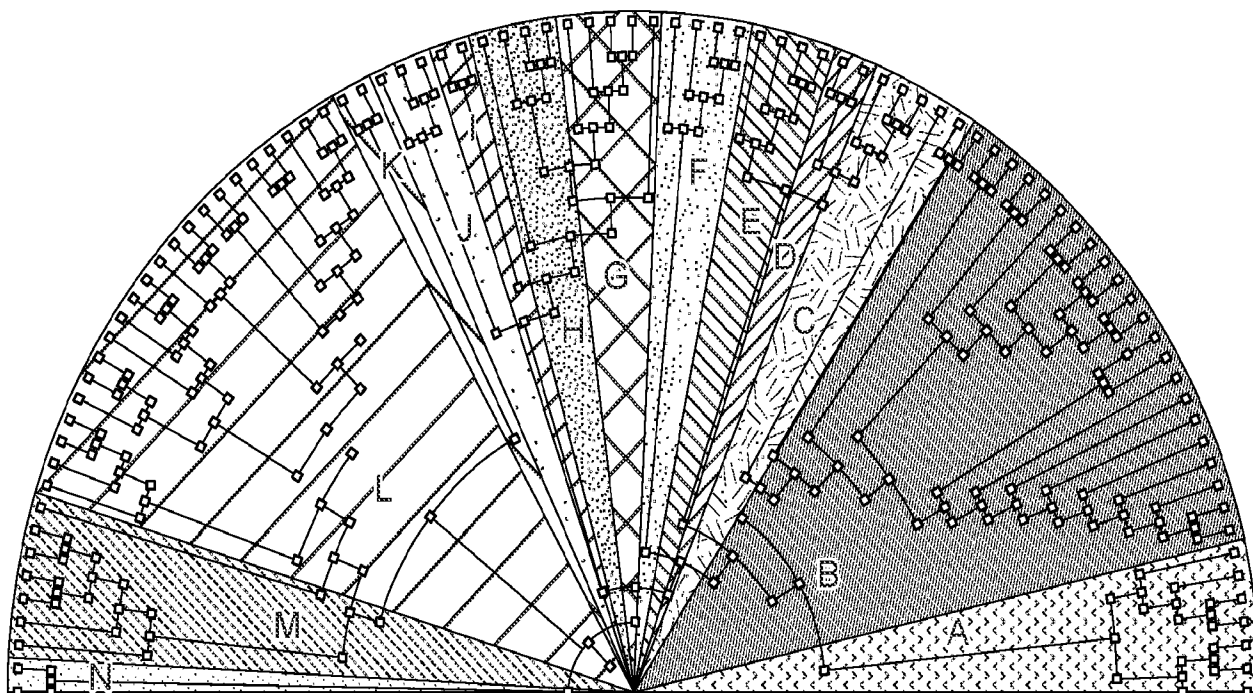


FIG. 1B

Color code	Phylum	Class	Family	Genus	Species	Strains
A	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Clostridiaceae</i>	1	6	7
B	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Lachnospiraceae</i>	5	14	22
C	<i>Firmicutes</i>	<i>Clostridia</i>	<i>Eubacteriaceae</i>	2	4	5
D	<i>Firmicutes</i>	<i>Bacilli</i>	<i>Enterococcaceae</i>	1	2	2
E	<i>Firmicutes</i>	<i>Erysipelotrichia</i>	<i>Erysipelotrichaceae</i>	1	3	4
F	<i>Fusobacteria</i>	<i>Fusobacteriia</i>	<i>Fusobacteriaceae</i>	1	3	4
G	<i>Proteobacteria</i>	<i>γ-Proteobacteria</i>	<i>Enterobacteriaceae</i>	2	2	5
H	<i>Proteobacteria</i>	<i>γ-Proteobacteria</i>	<i>Morganell/Pseudomona</i>	2	3	4
I	<i>Proteobacteria</i>	<i>γ-Proteobacteria</i>	<i>Alcaligenaceae</i>	2	2	2
J	<i>Proteobacteria</i>	<i>δ/ε-Proteobacteria</i>	<i>Desulfovibrionaceae</i>	2	3	3
K	<i>Bacteroidetes</i>	<i>Bacteroidia</i>	<i>Tannerellaceae</i>	1	2	2
L	<i>Bacteroidetes</i>	<i>Bacteroidia</i>	<i>Bacteroidaceae</i>	2	20	22
M	<i>Bacteroidetes</i>	<i>Bacteroidia</i>	<i>Prevotellaceae</i>	1	6	7
N	<i>Actinobacteria</i>	<i>Actinobacteria</i>	<i>Bifido/Corynebacteriaceae</i>	2	2	2

FIG. 1C

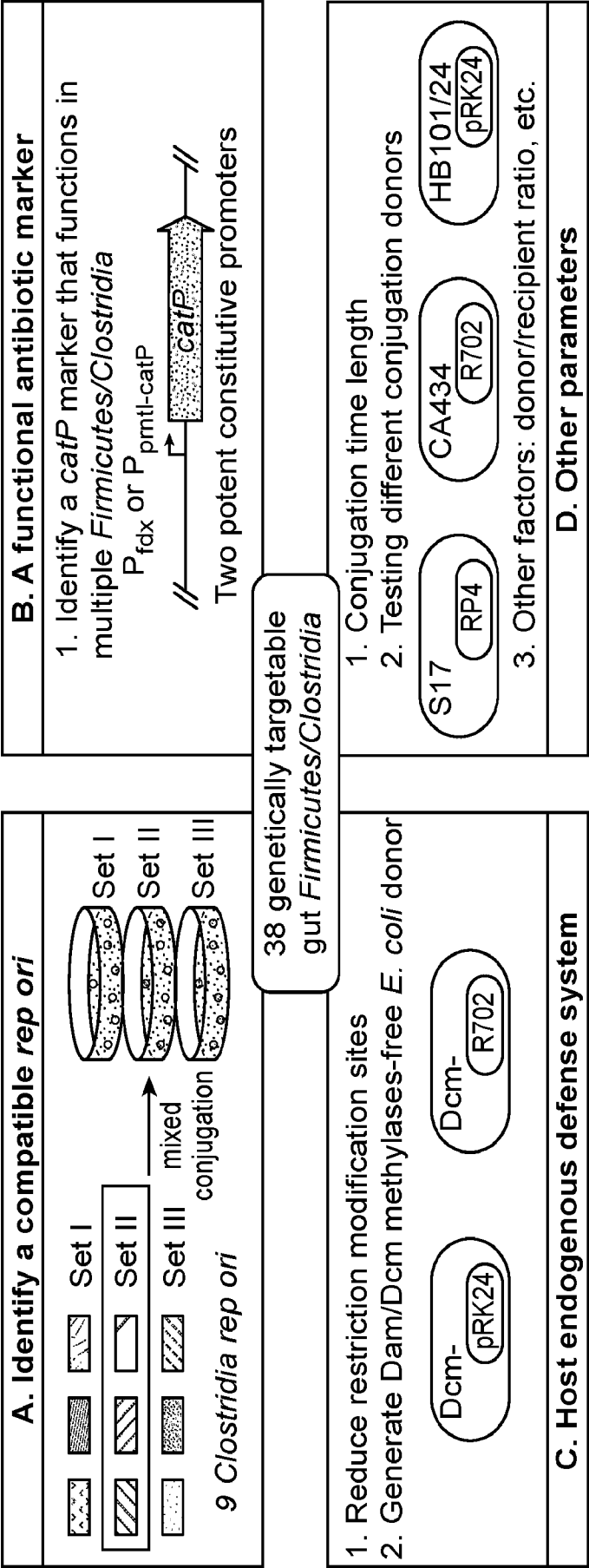


FIG. 2A

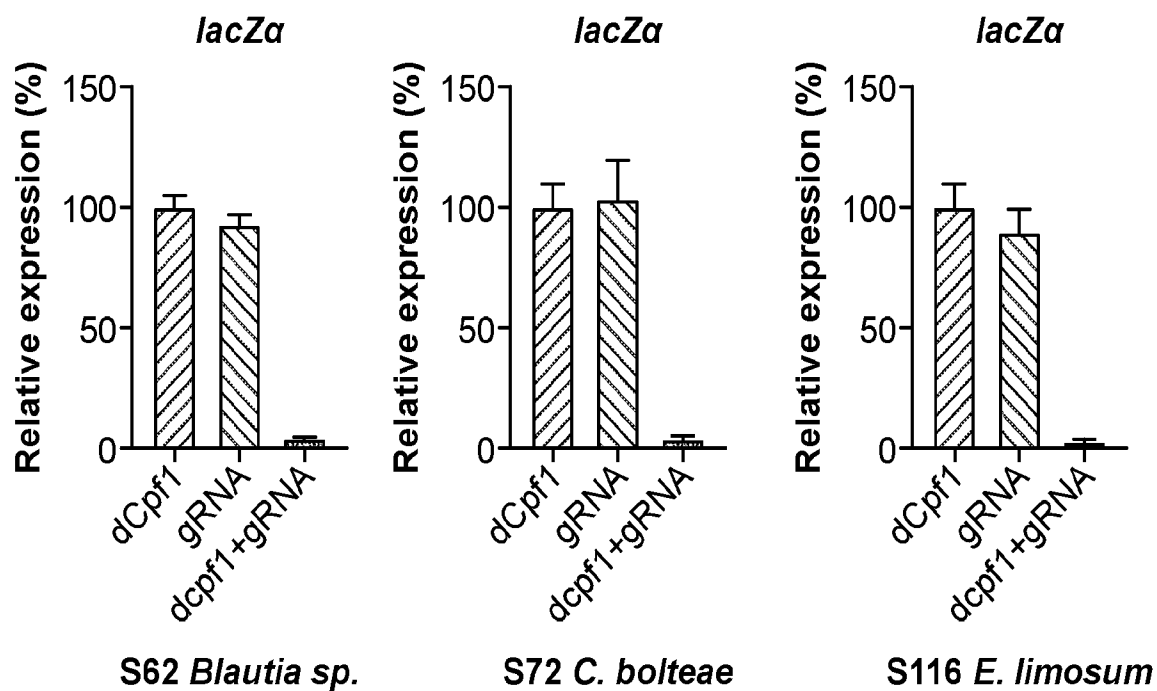
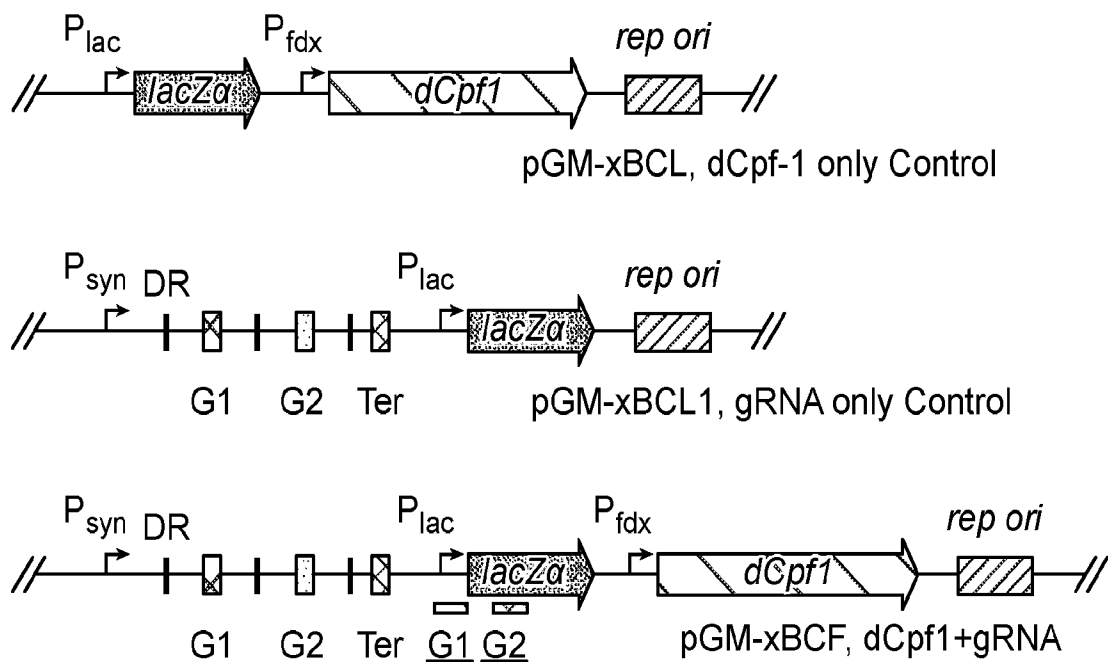


FIG. 2B

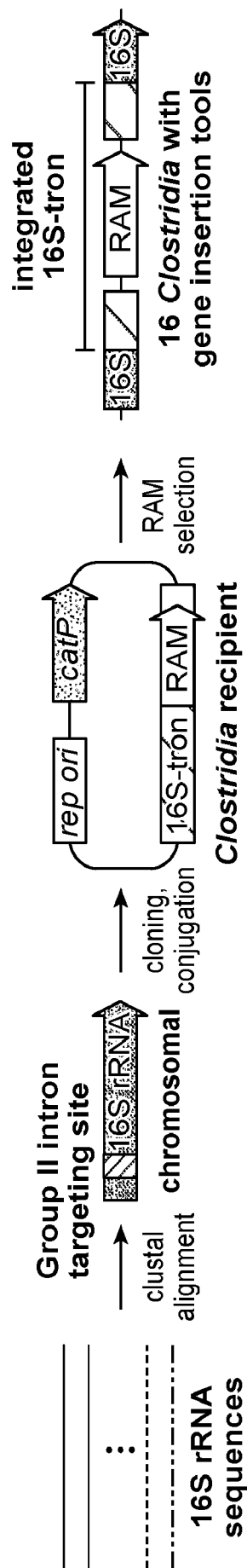


FIG. 2C

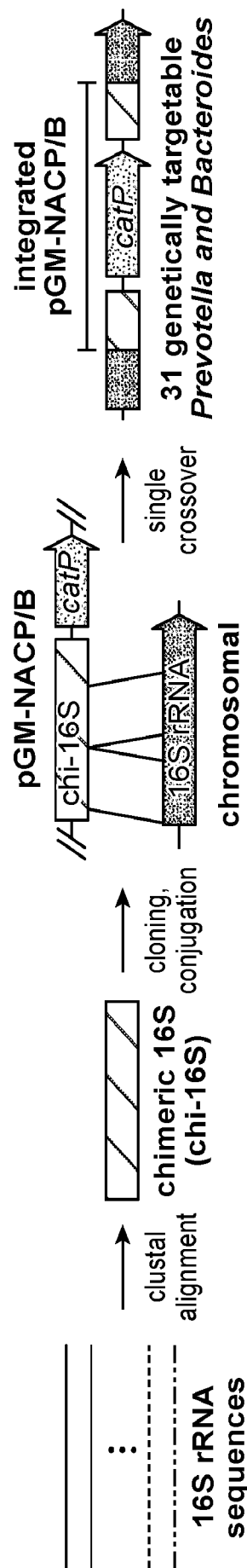


FIG. 2D

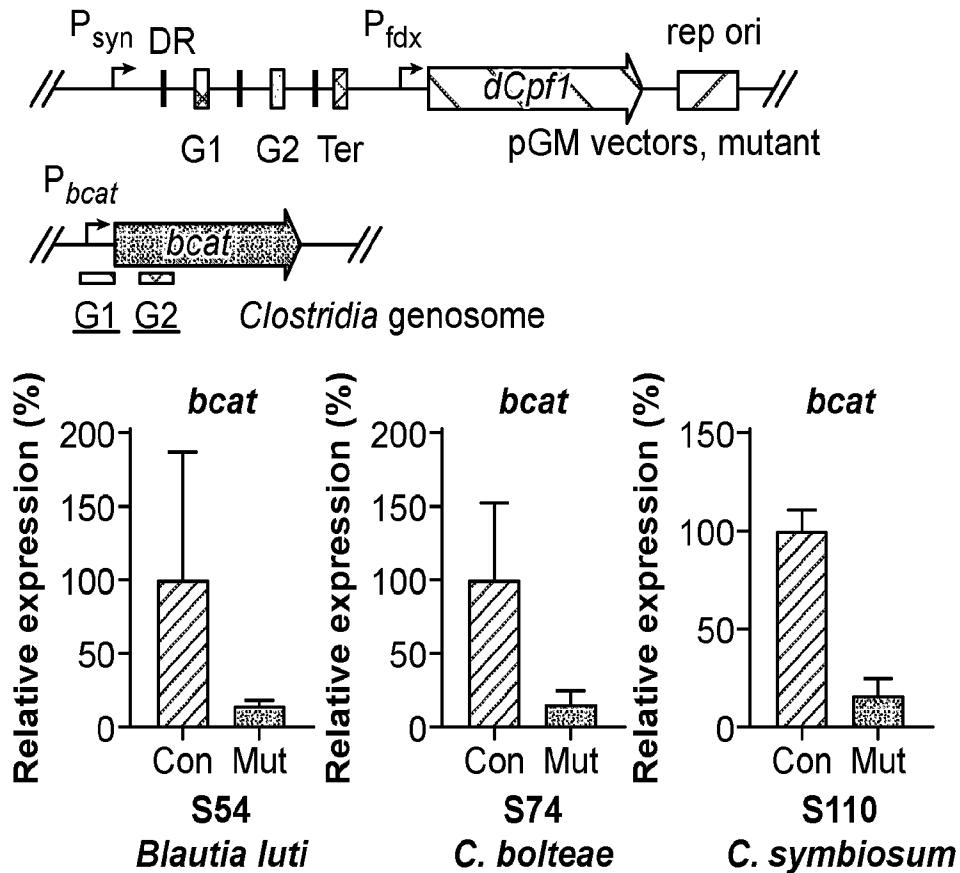


FIG. 3A

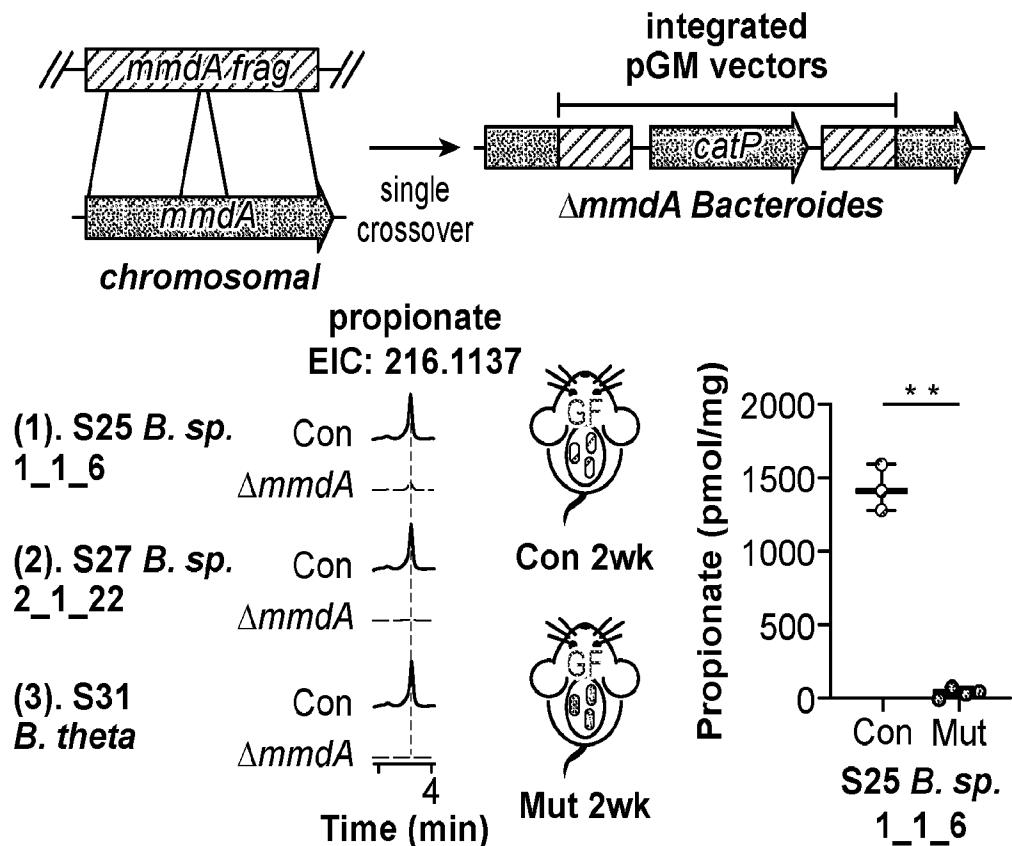


FIG. 3B

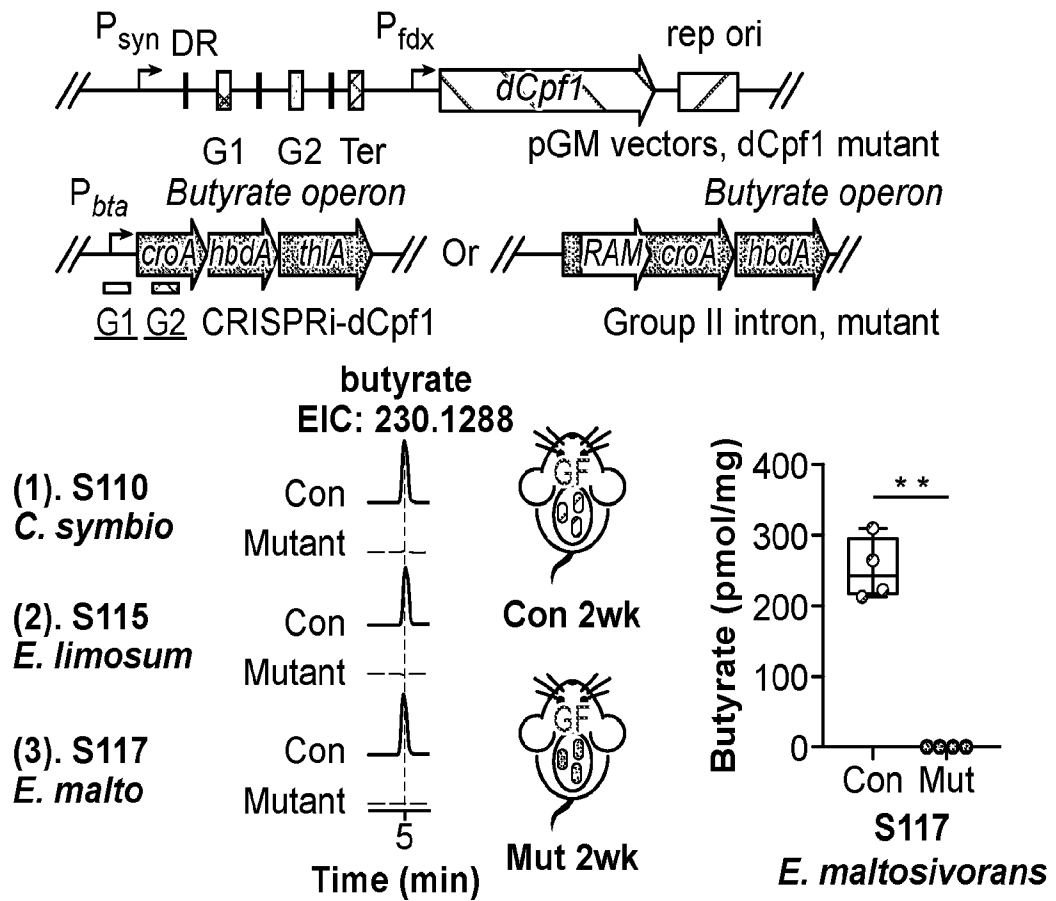


FIG. 3C

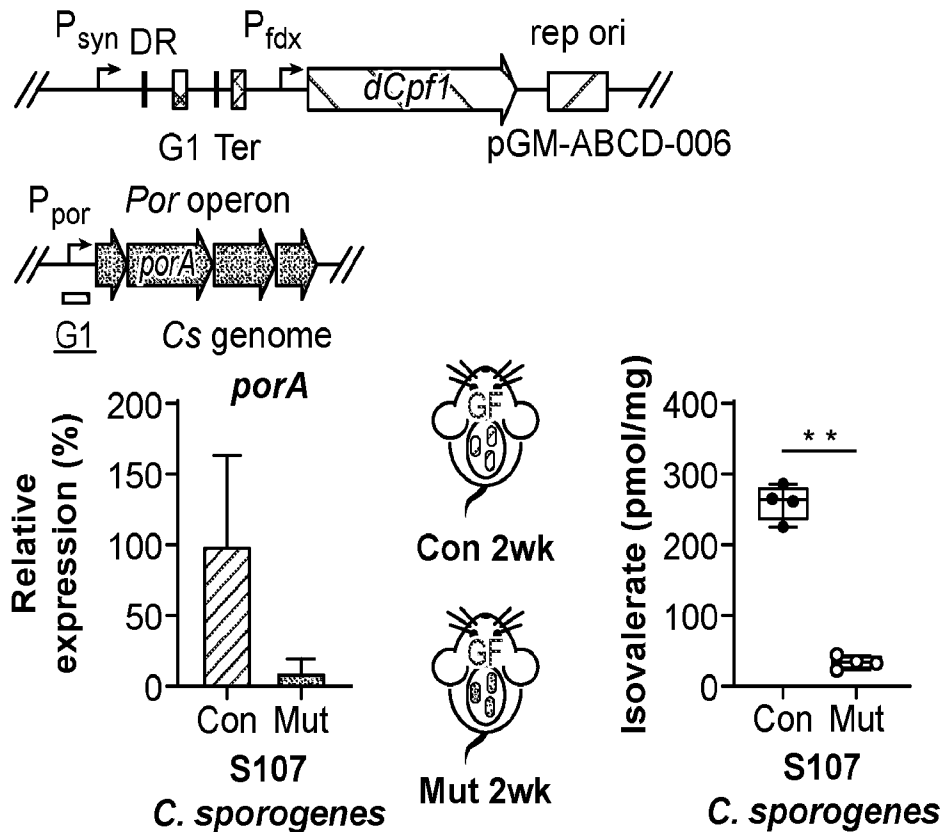


FIG. 3D

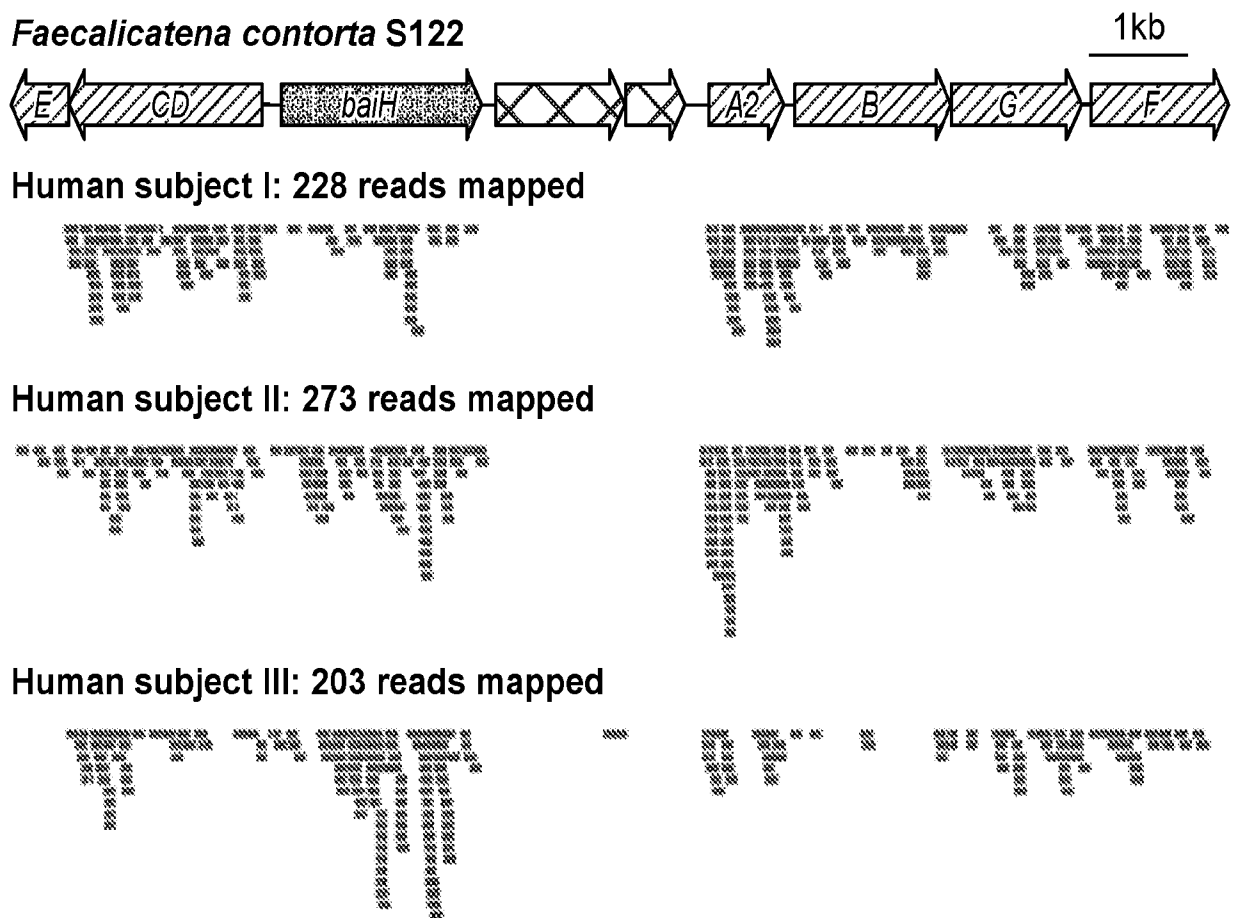
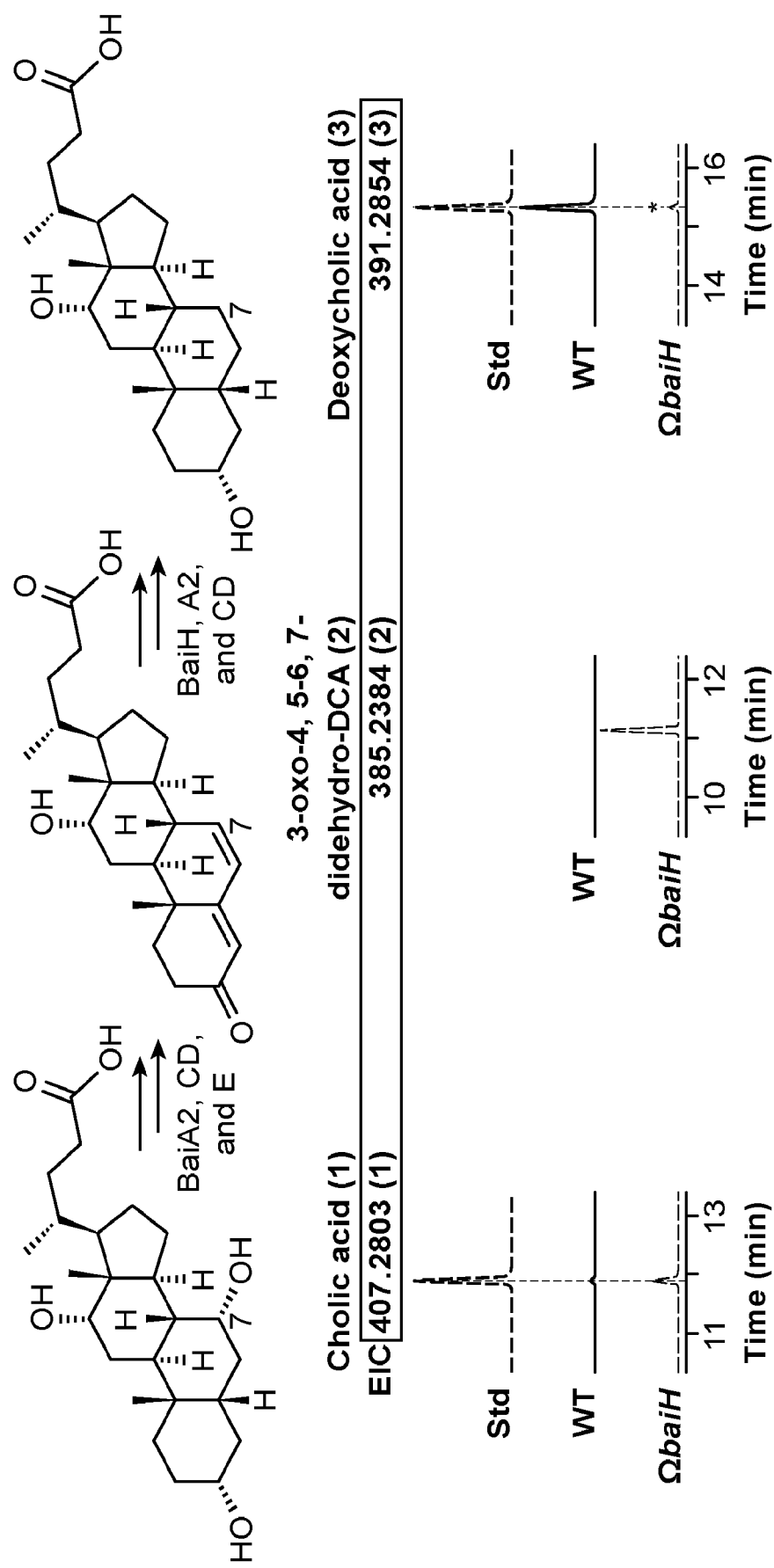


FIG. 4A



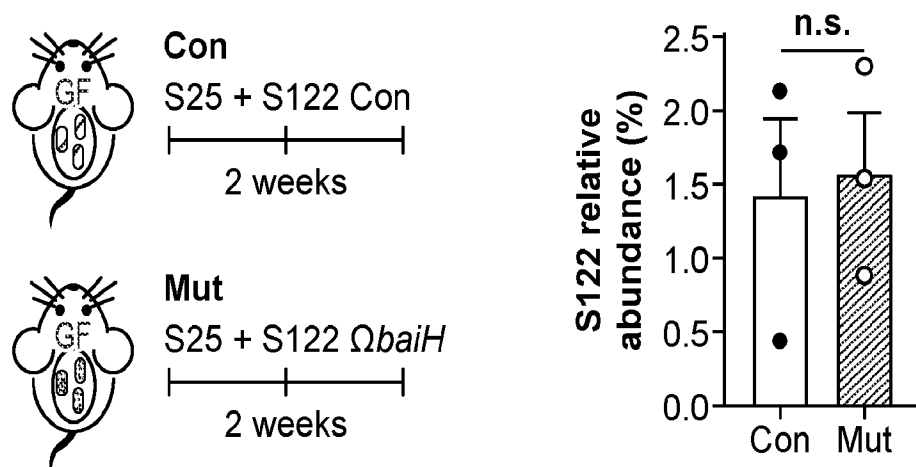


FIG. 4C

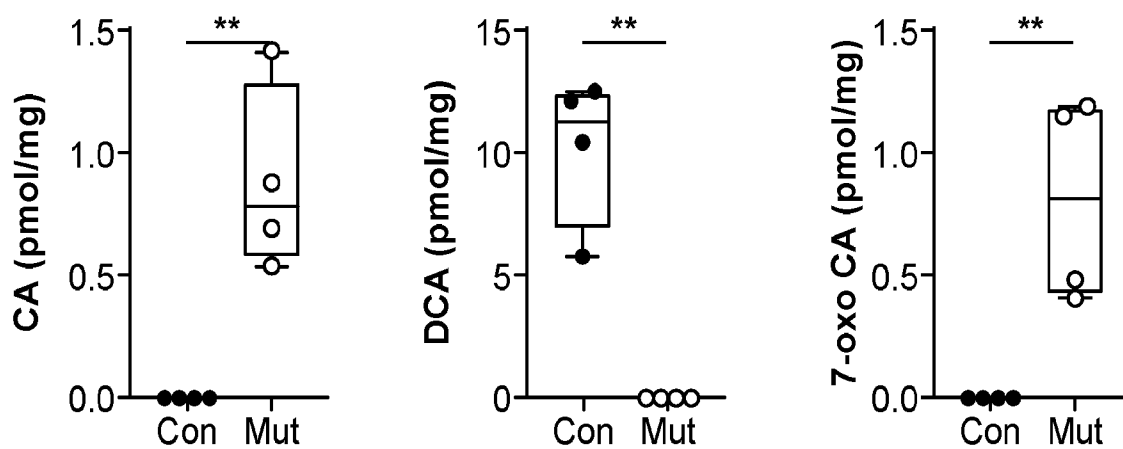


FIG. 4D

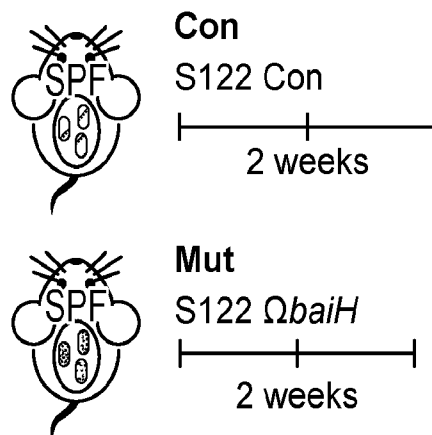


FIG. 5A

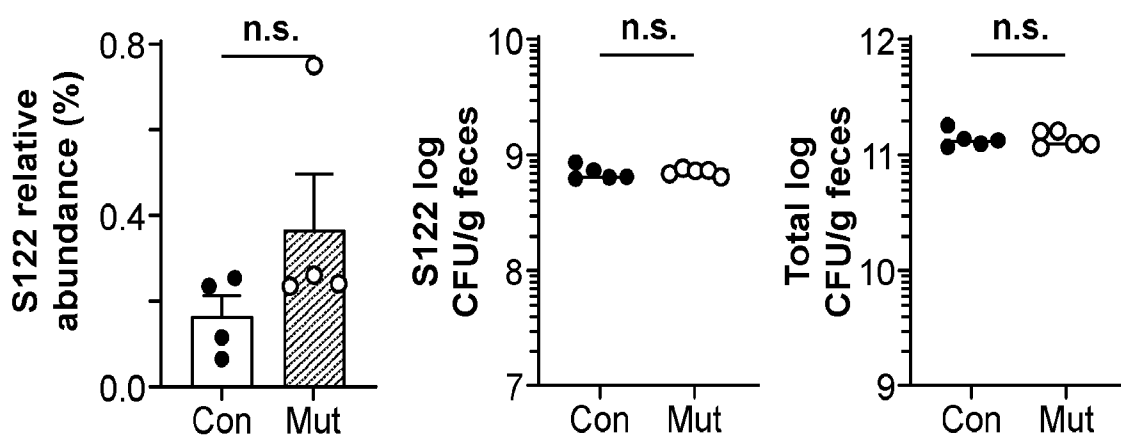


FIG. 5B

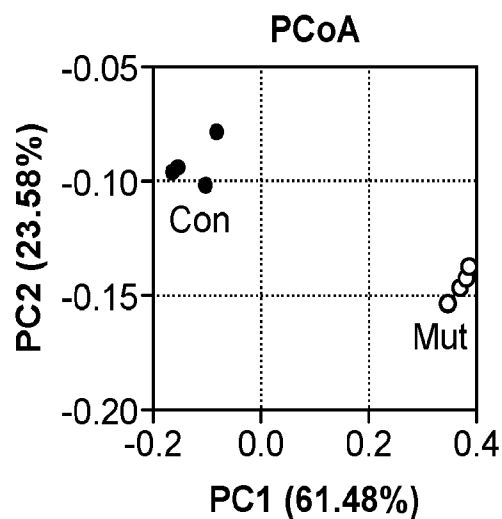


FIG. 5C

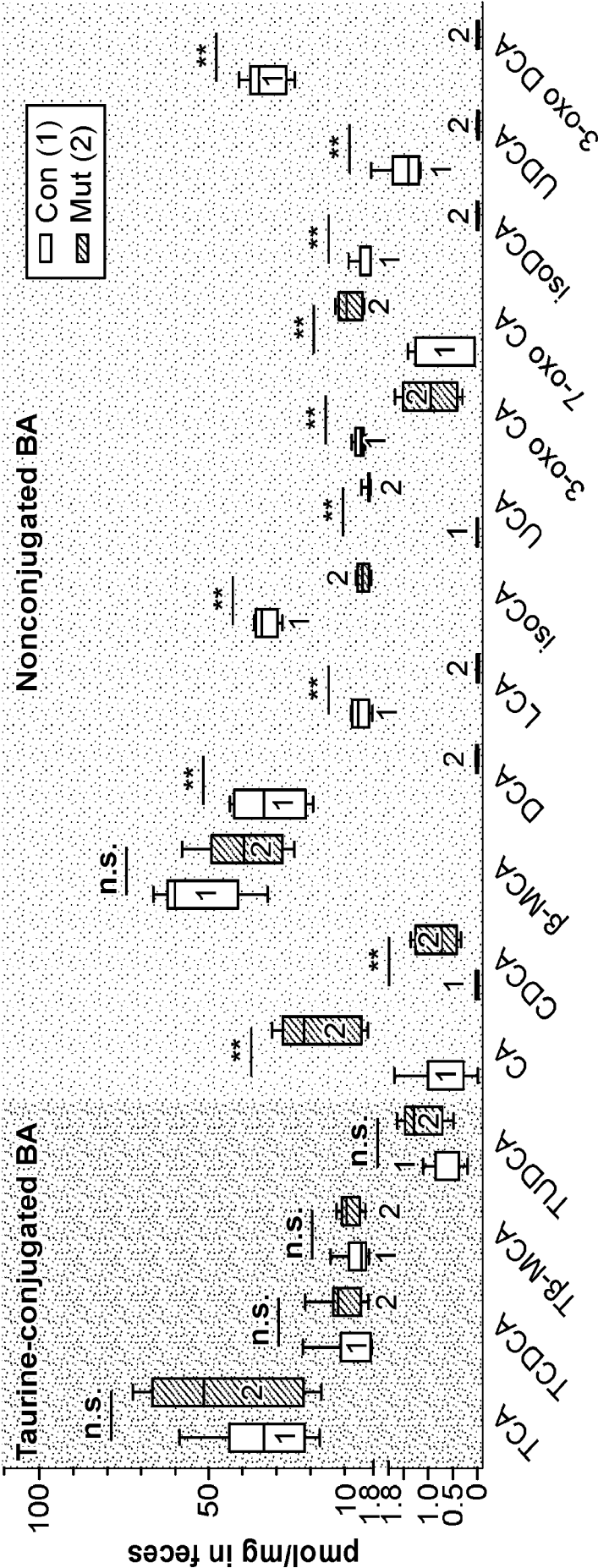


FIG. 5D

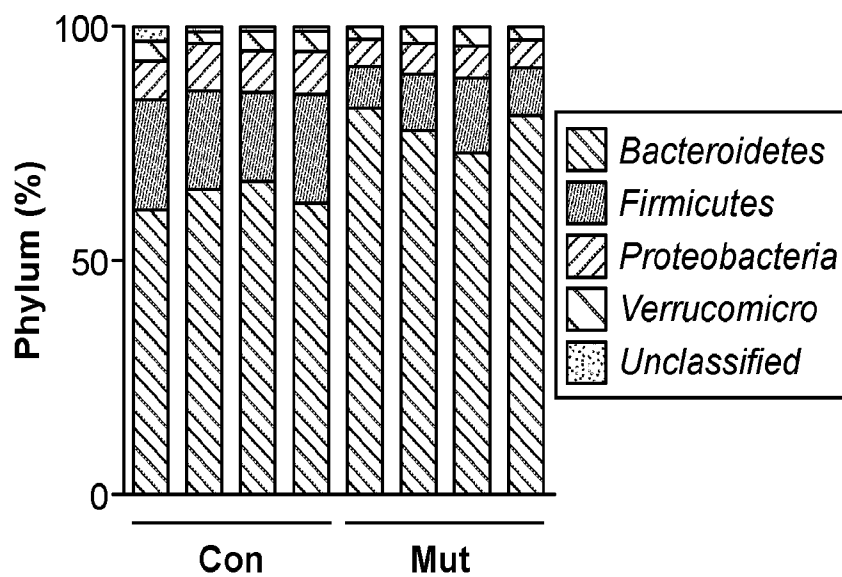


FIG. 5E

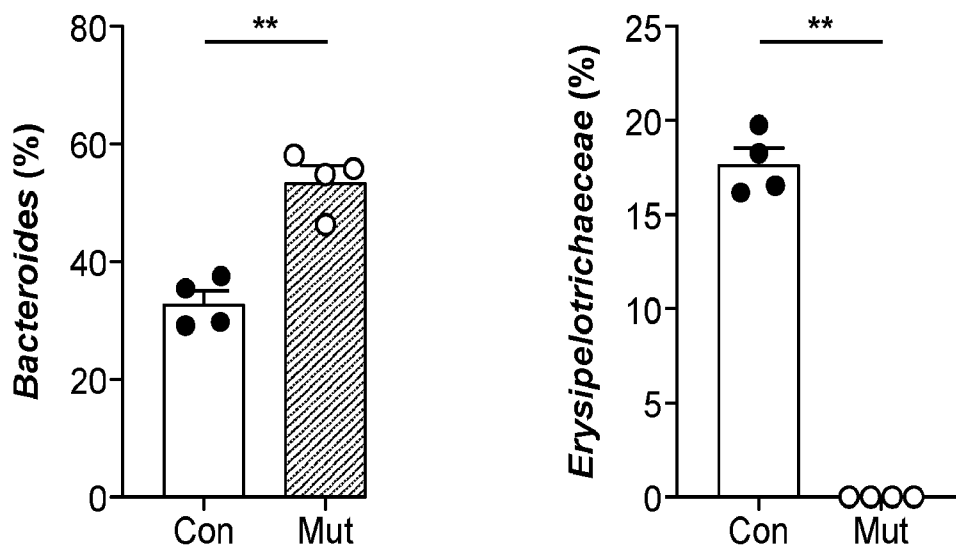


FIG. 5F

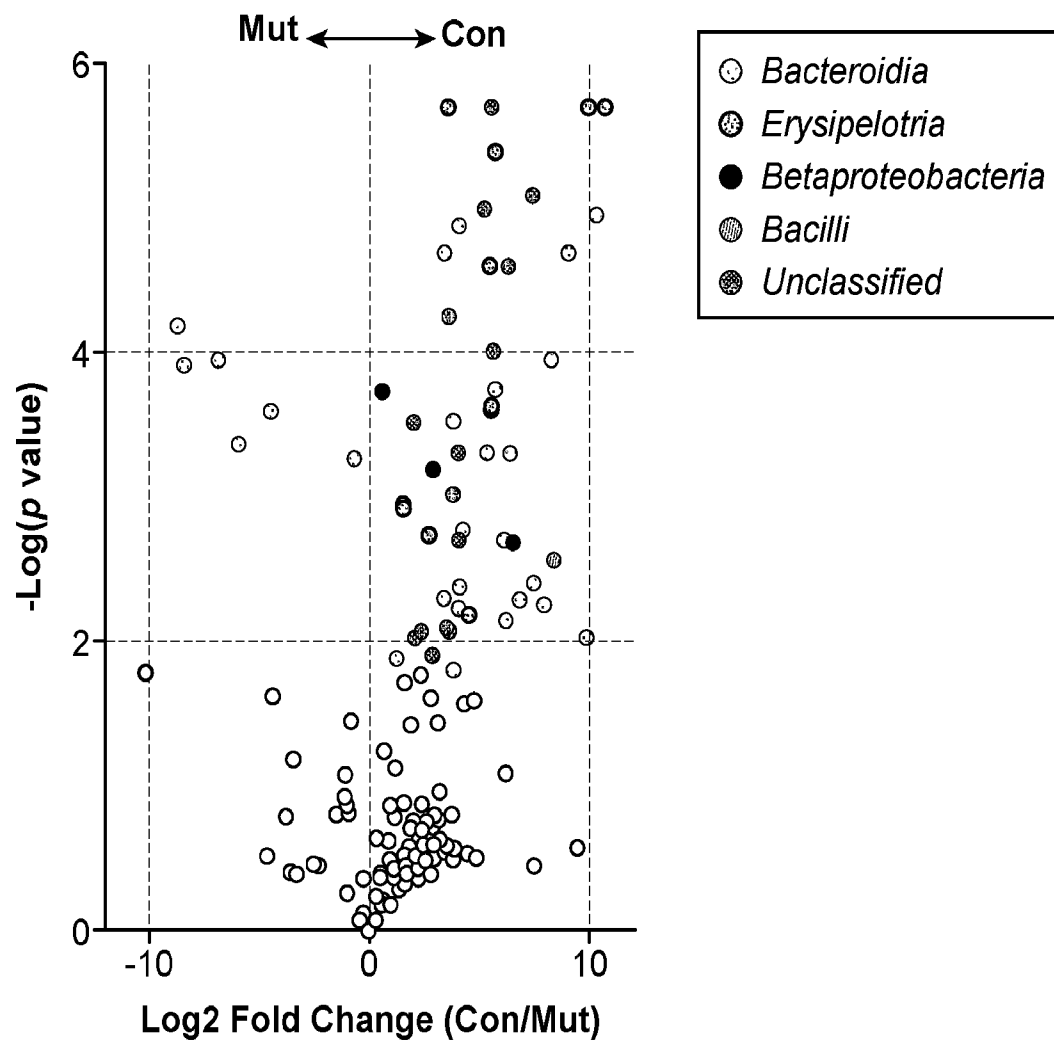


FIG. 5G

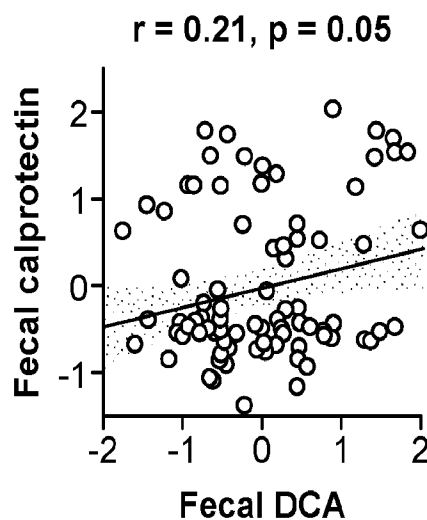


FIG. 5H

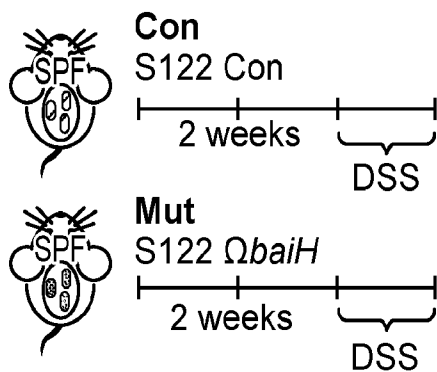


FIG. 6A

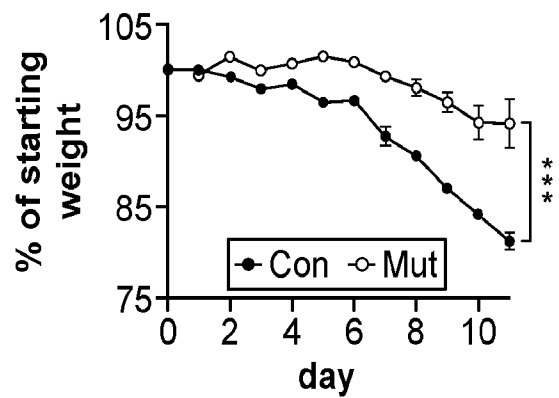


FIG. 6B

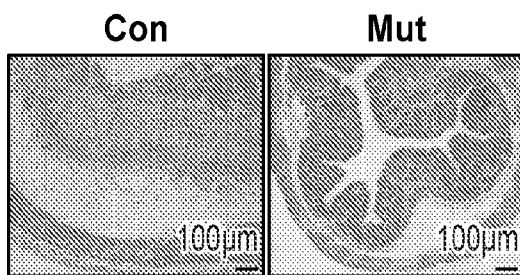


FIG. 6C

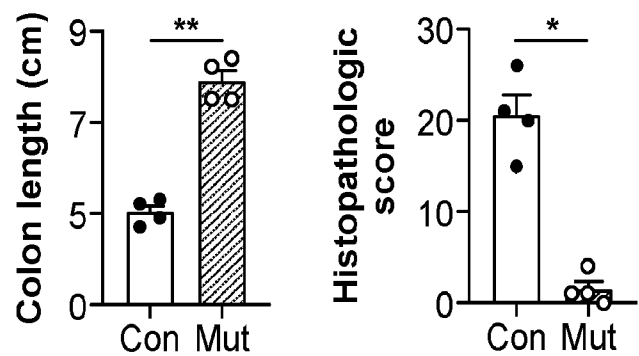


FIG. 6D

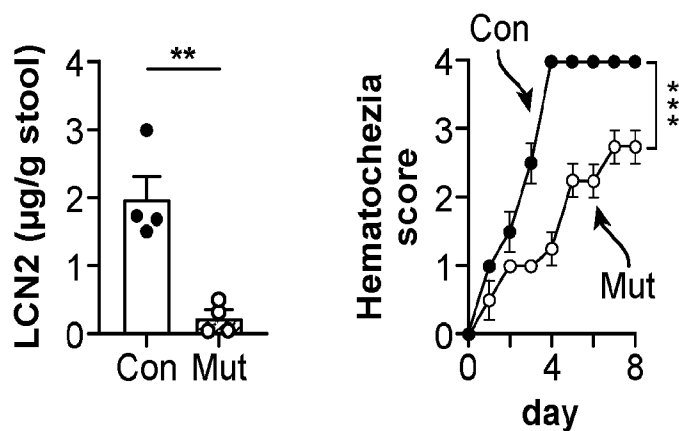


FIG. 6E

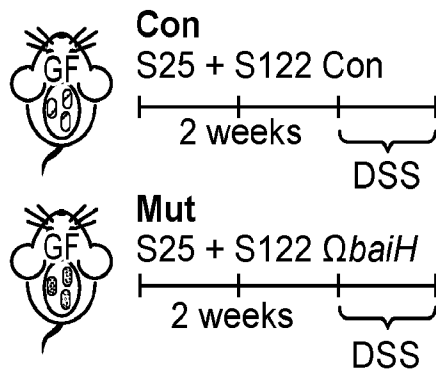


FIG. 6F

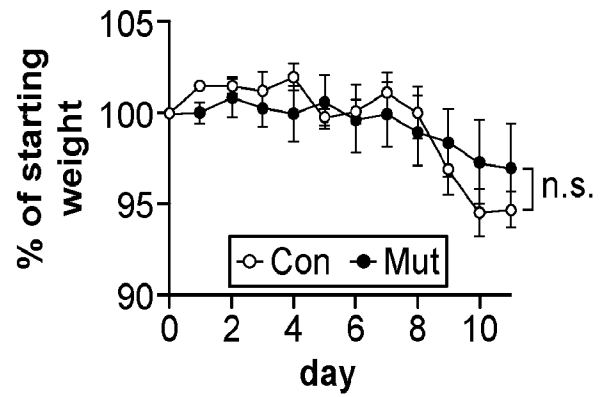


FIG. 6G

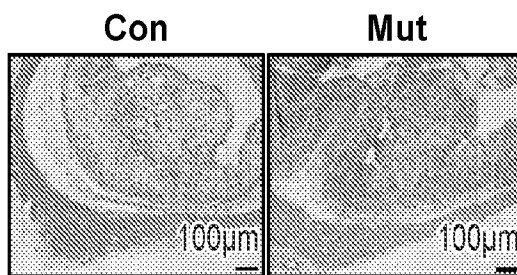


FIG. 6H

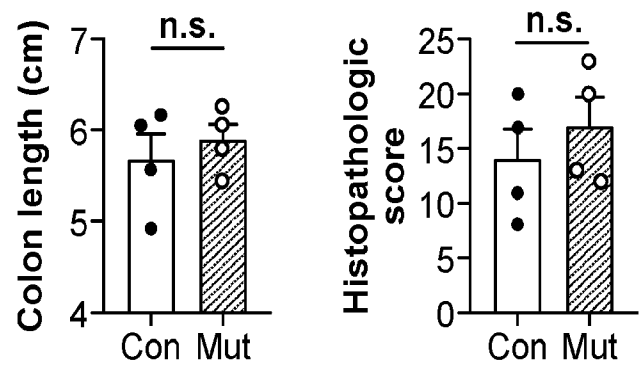


FIG. 6I

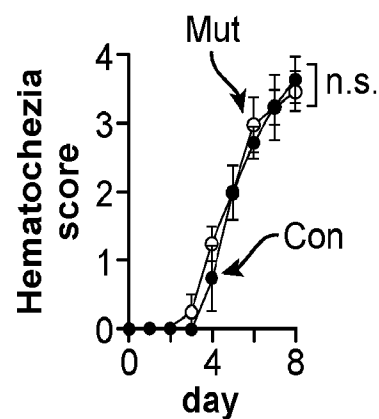
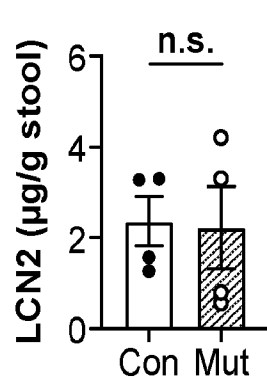


FIG. 6J

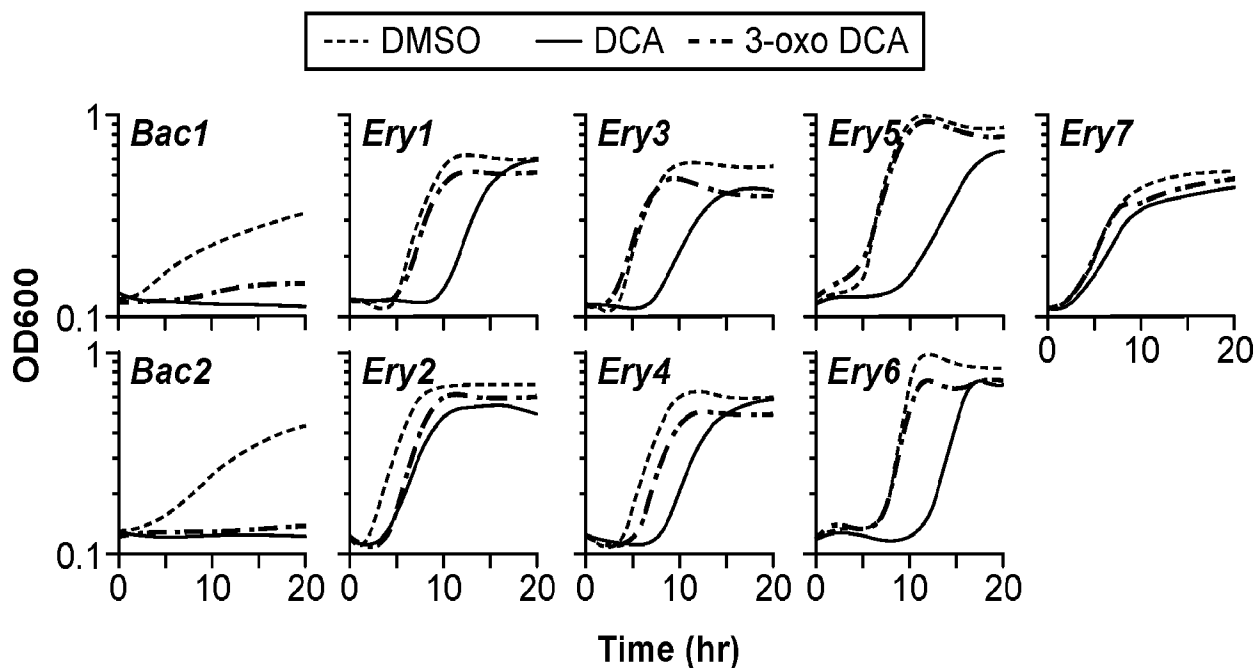


FIG. 7A

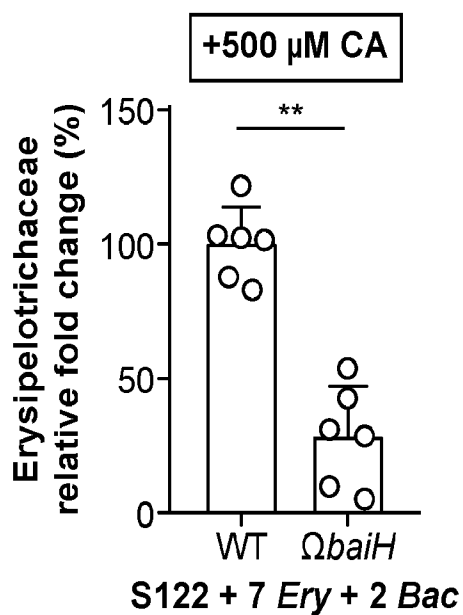


FIG. 7B

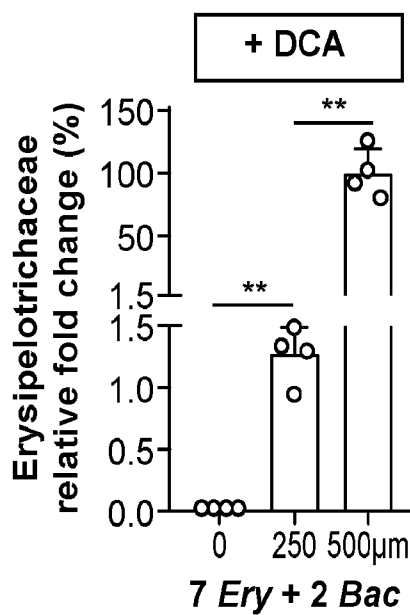


FIG. 7C

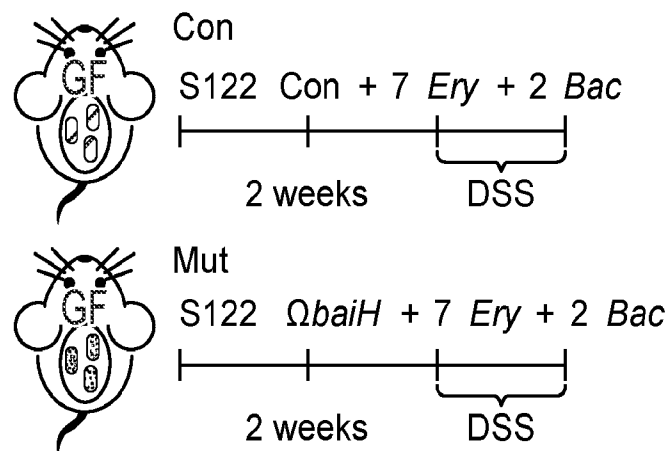


FIG. 7D

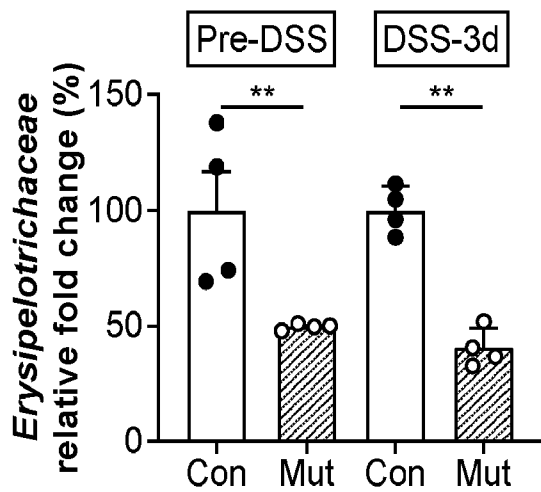


FIG. 7E

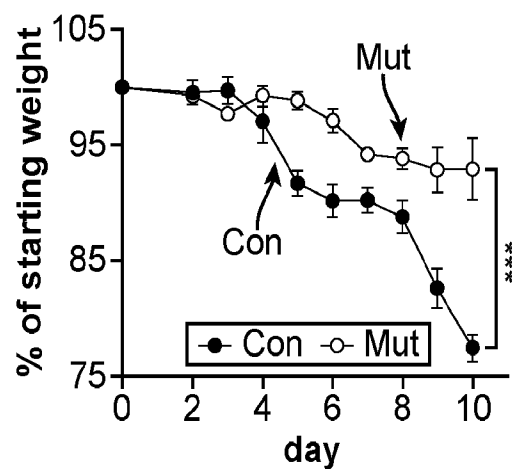


FIG. 7F

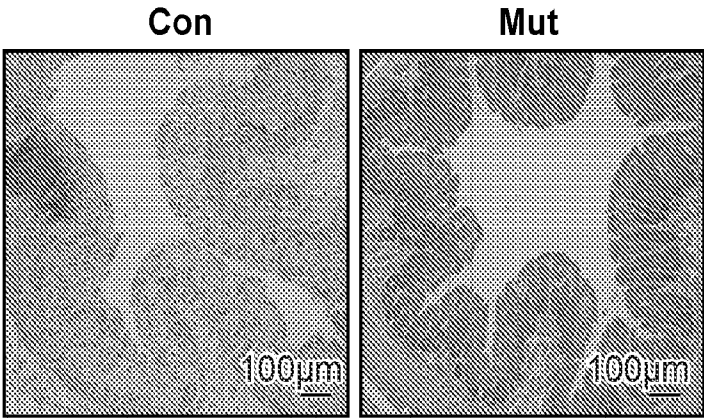


FIG. 7G

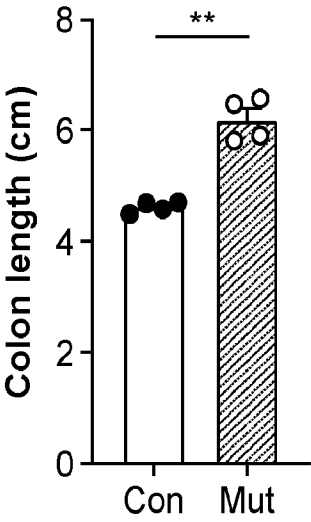


FIG. 7H

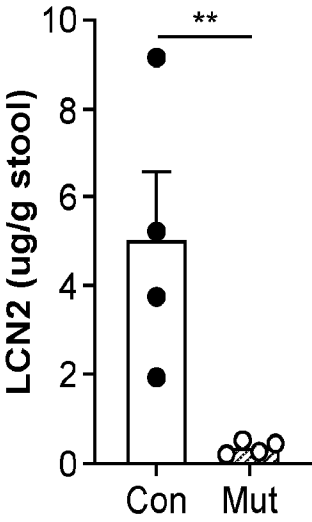


FIG. 7I

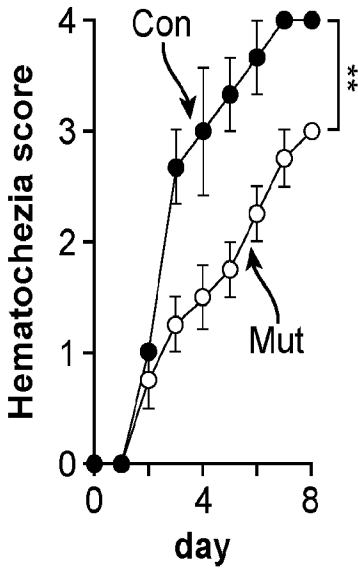


FIG. 7J

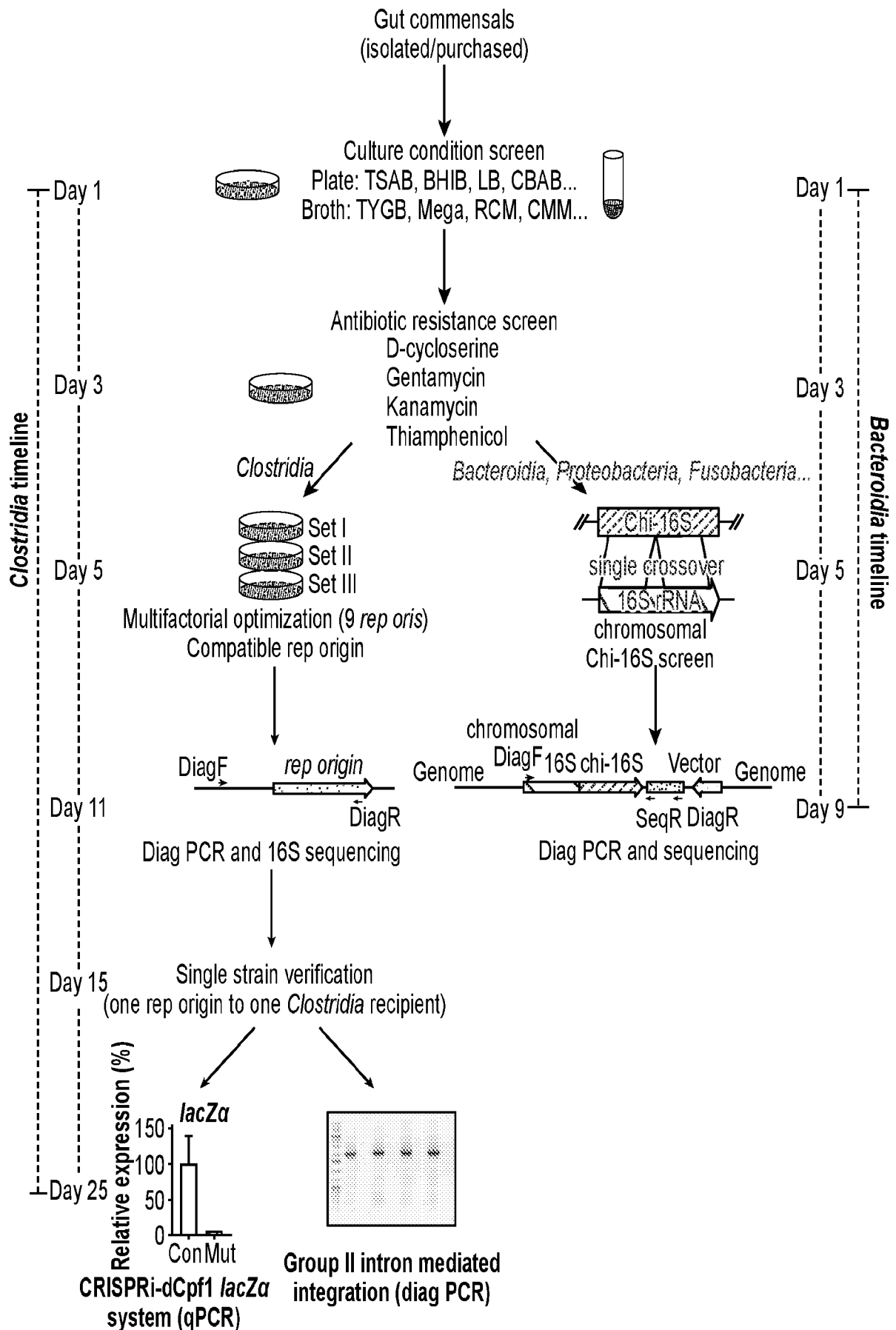


FIG. 8

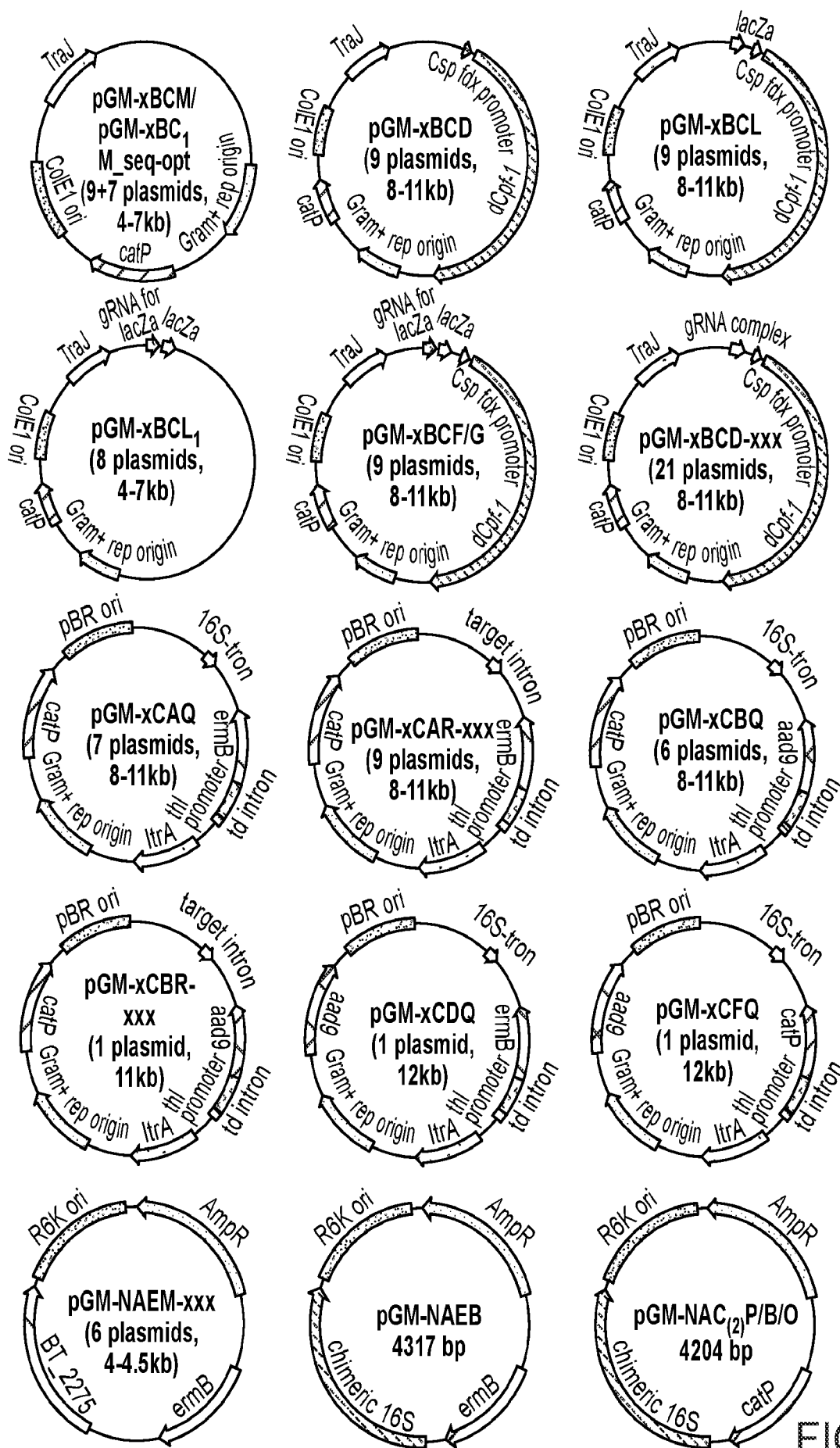


FIG. 9

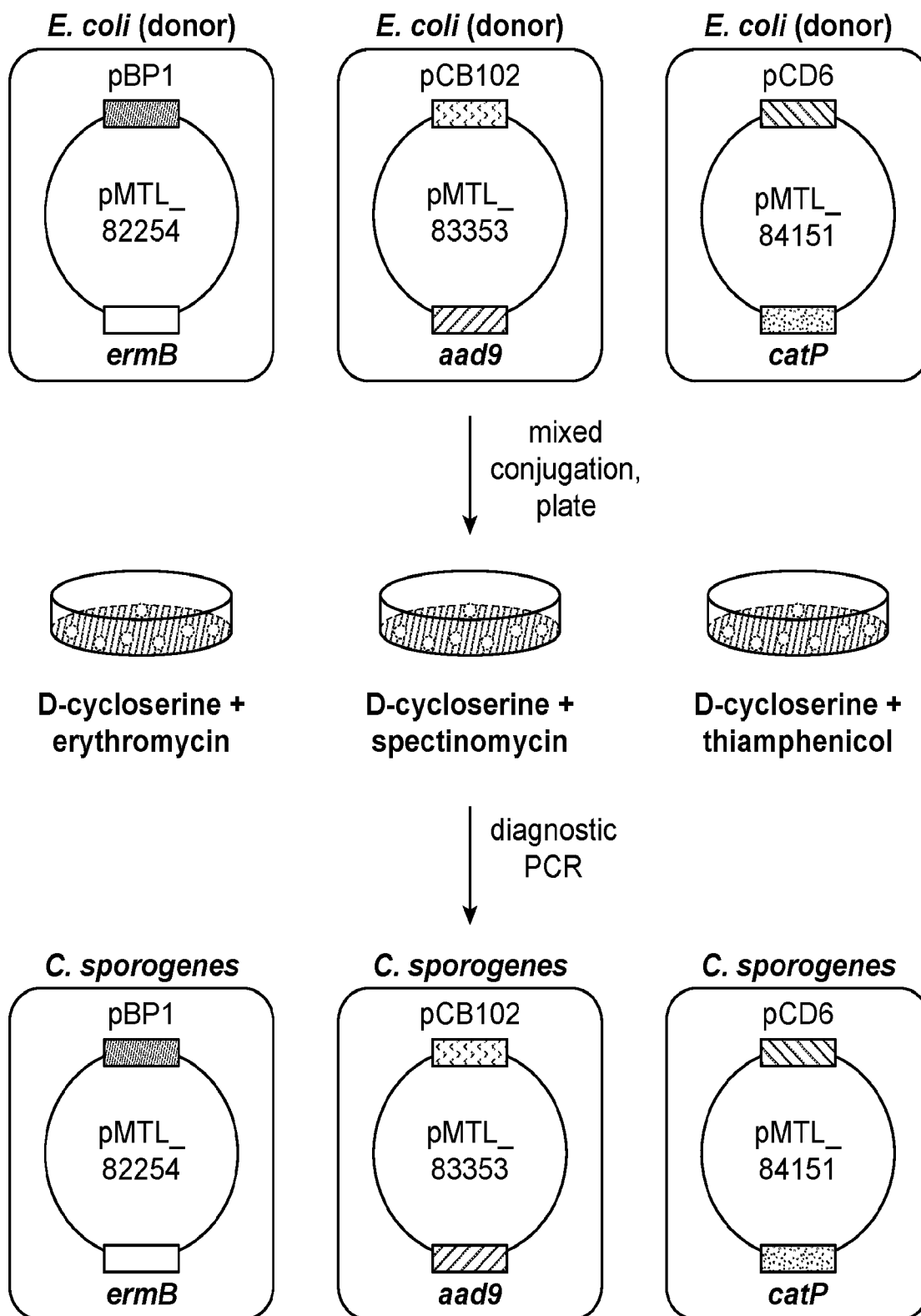


FIG. 10A

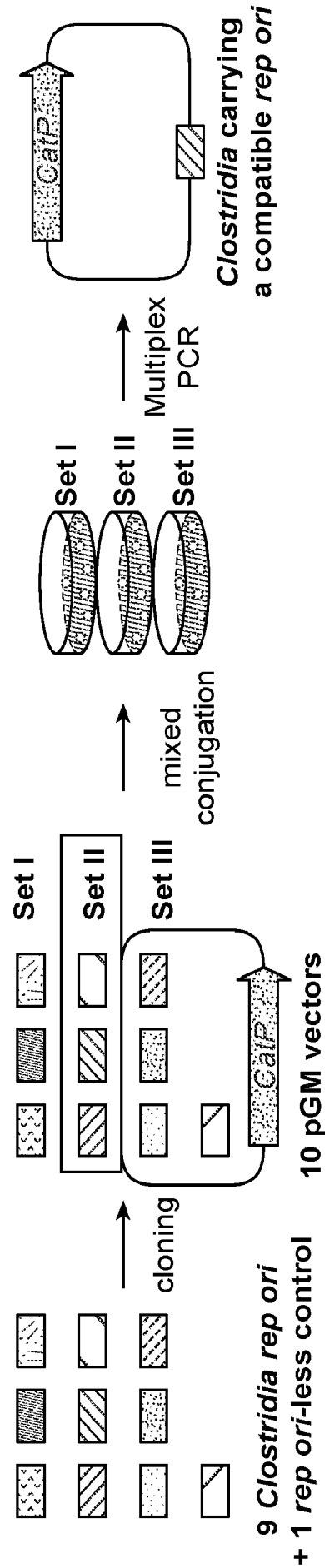


FIG. 10B

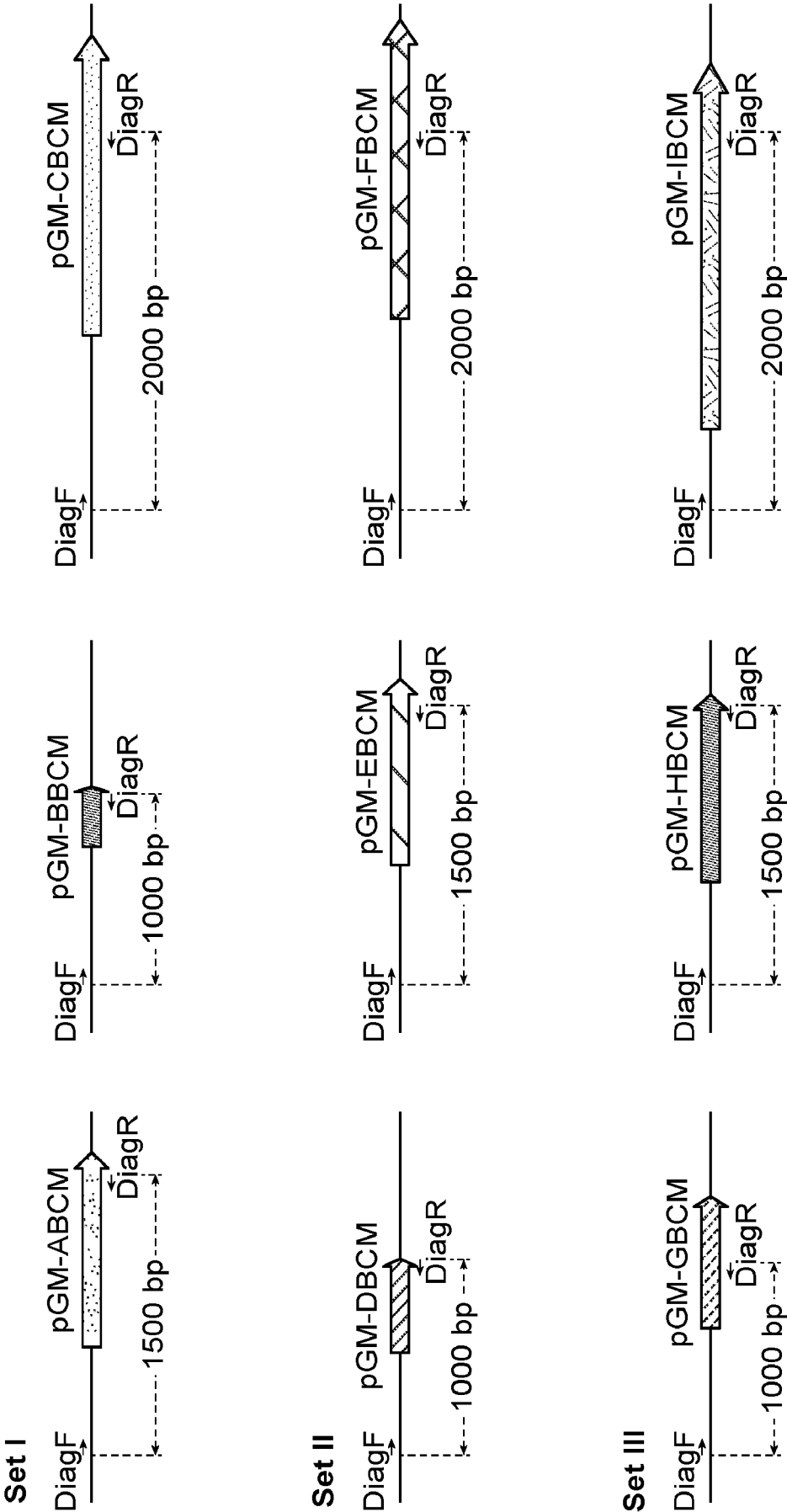


FIG. 11A

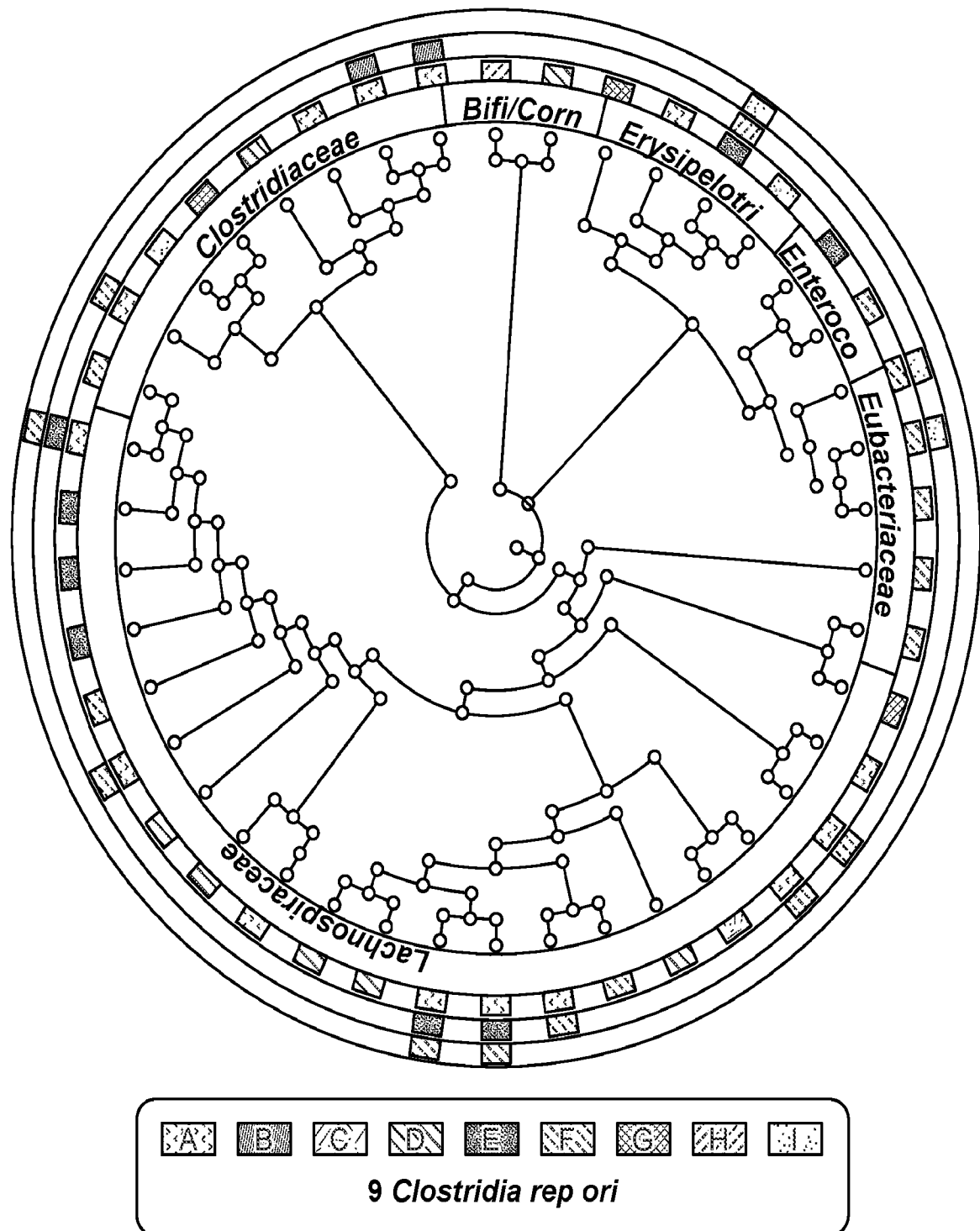


FIG. 11B

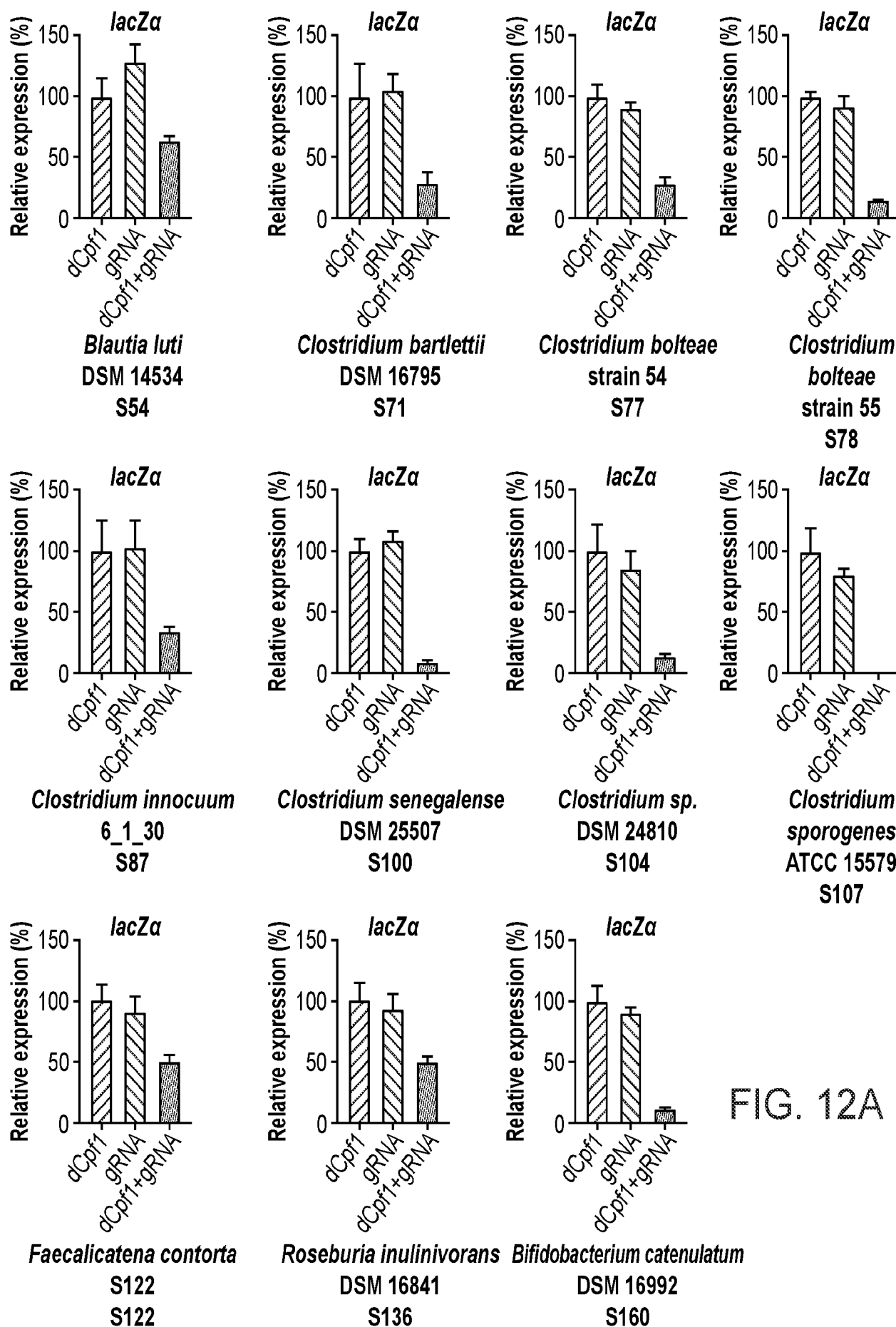


FIG. 12A

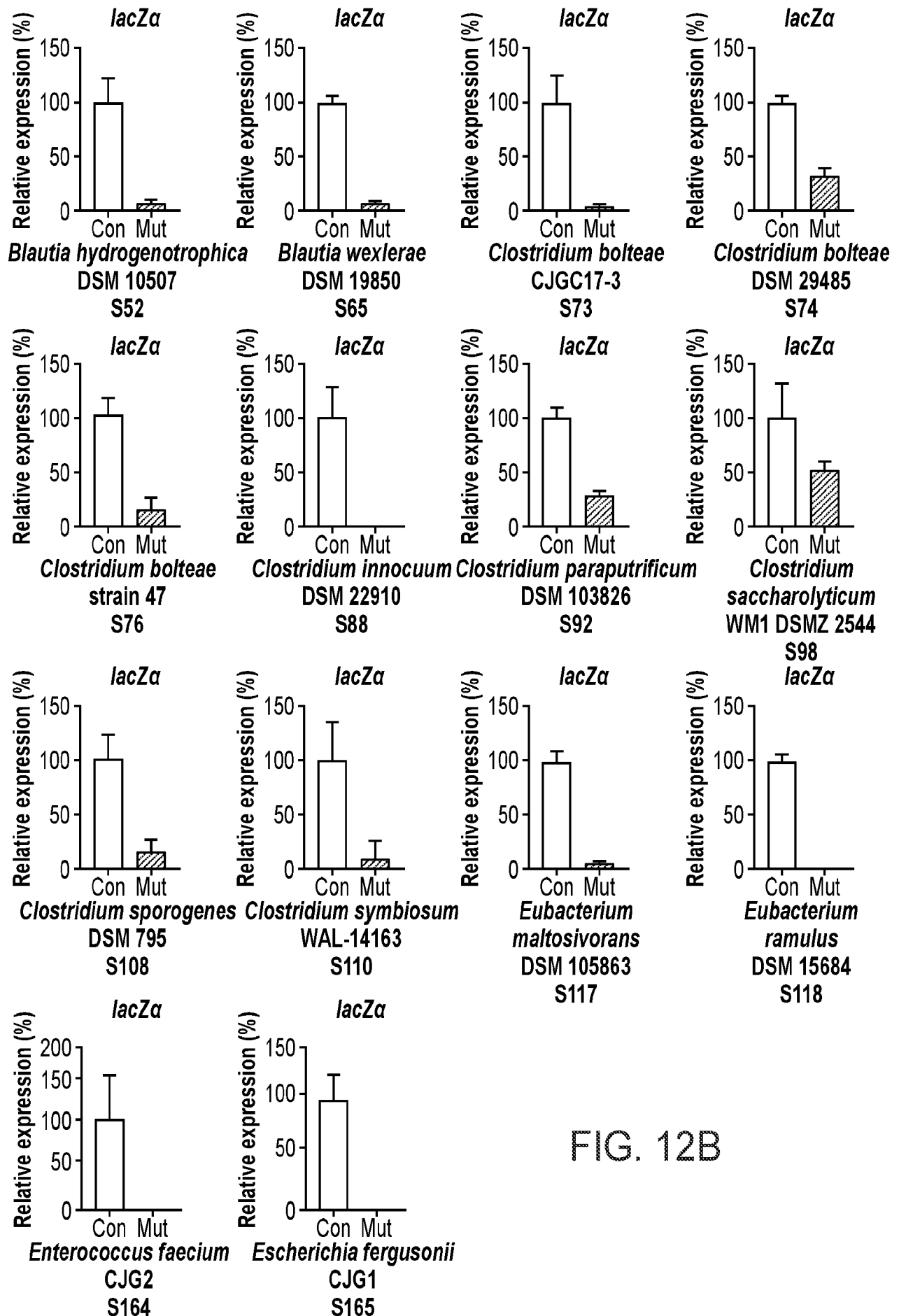


FIG. 12B

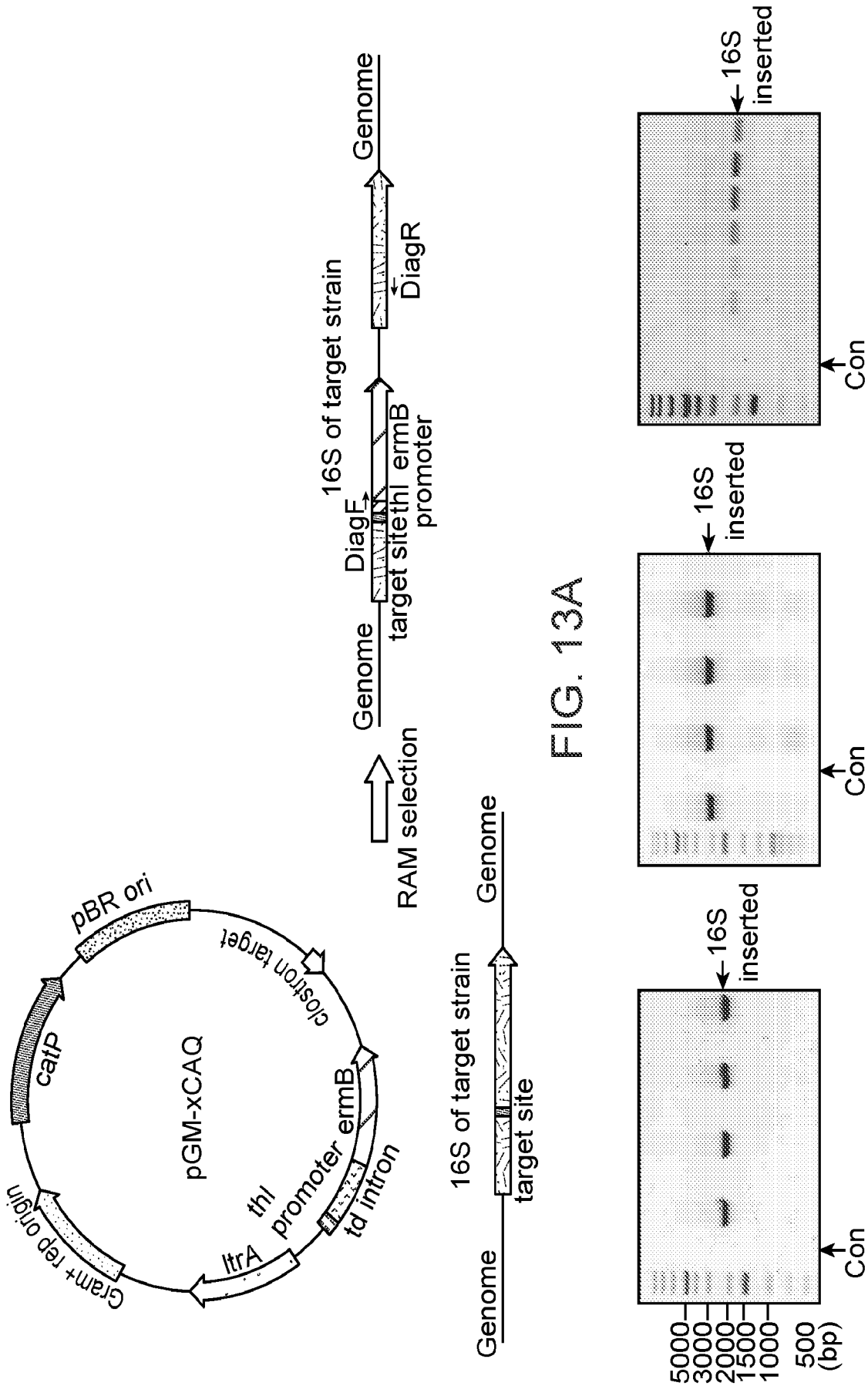


FIG. 13A

FIG. 13B

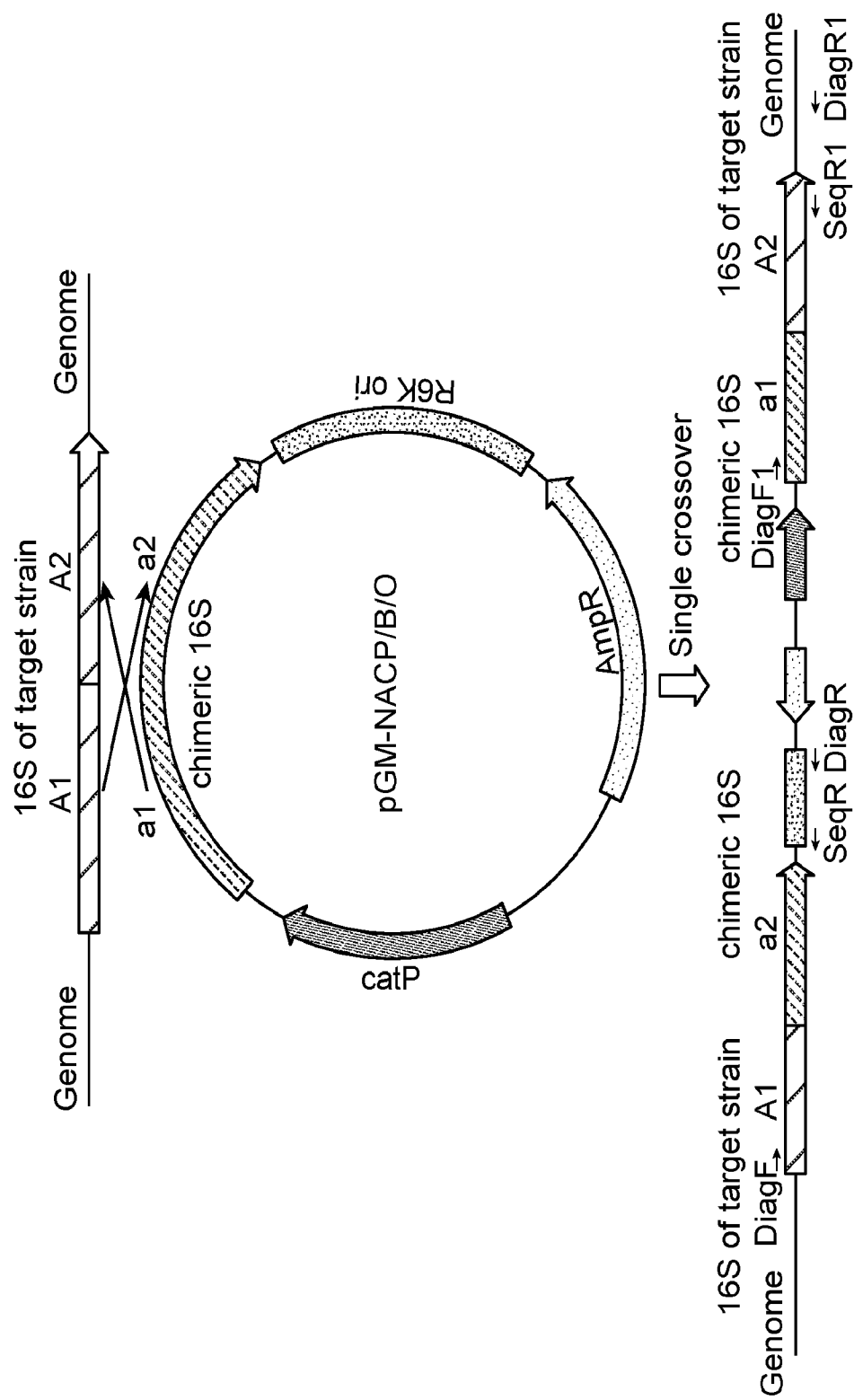


FIG. 13C

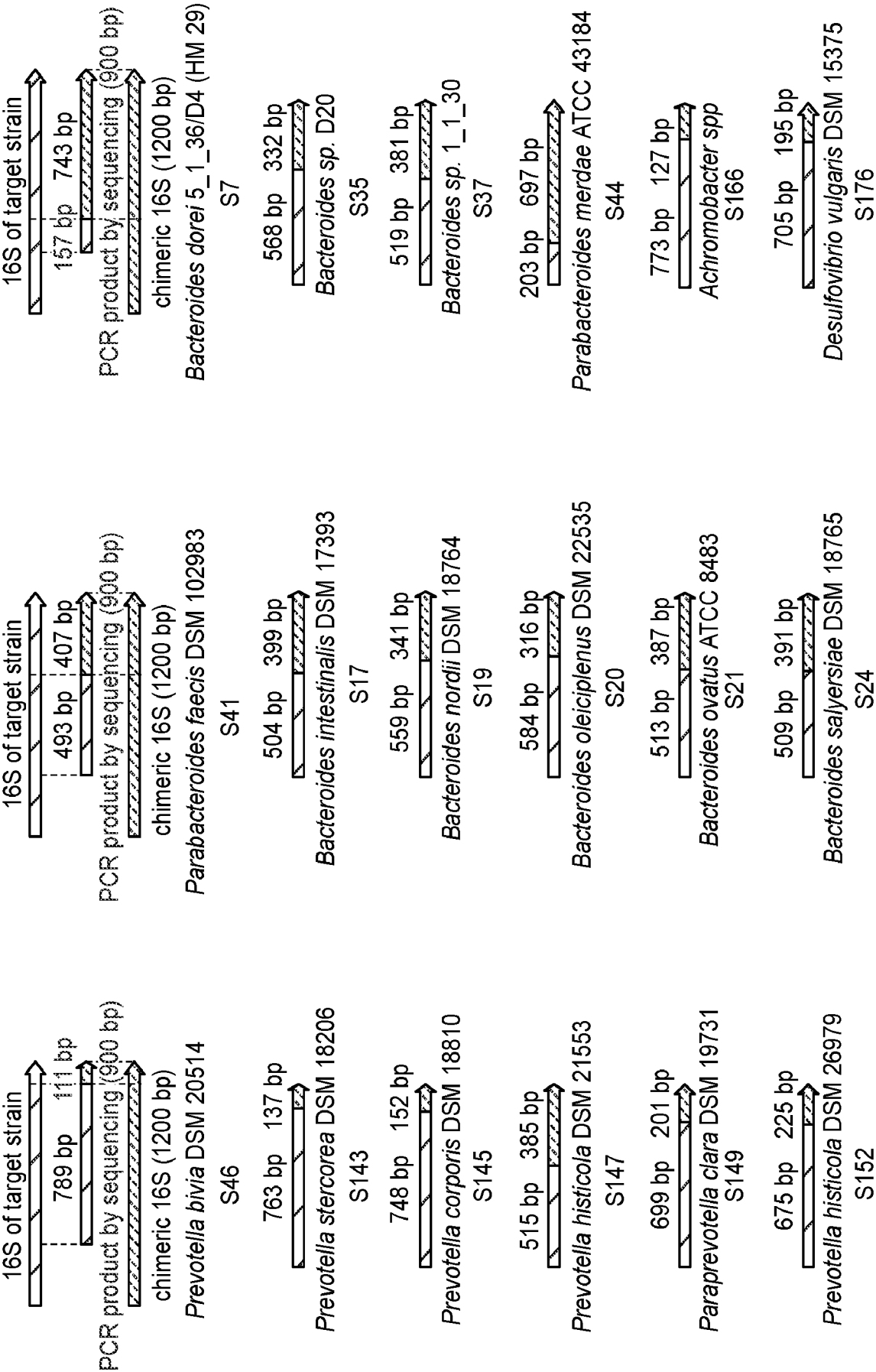


FIG. 13D

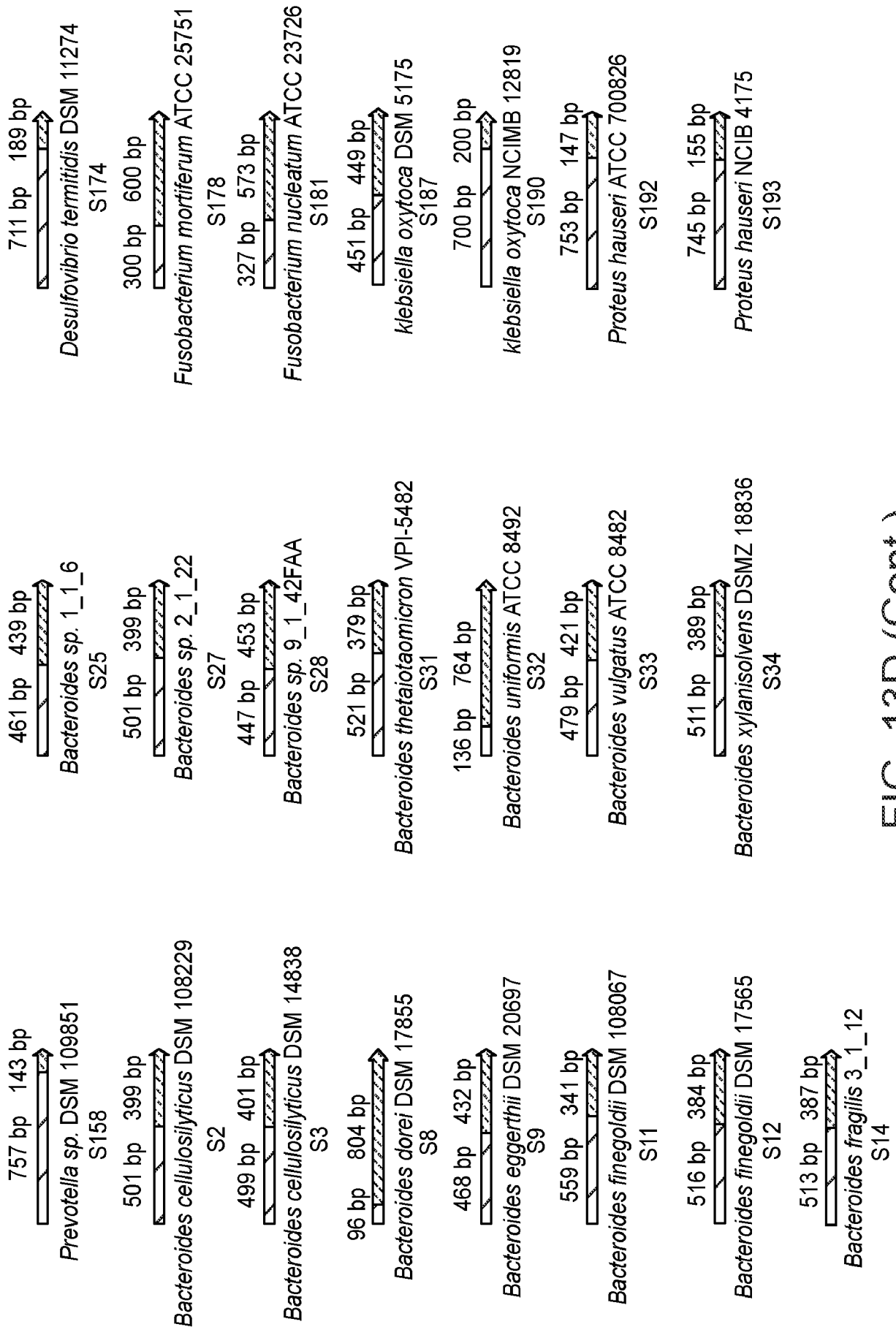


FIG. 13D (Cont.)

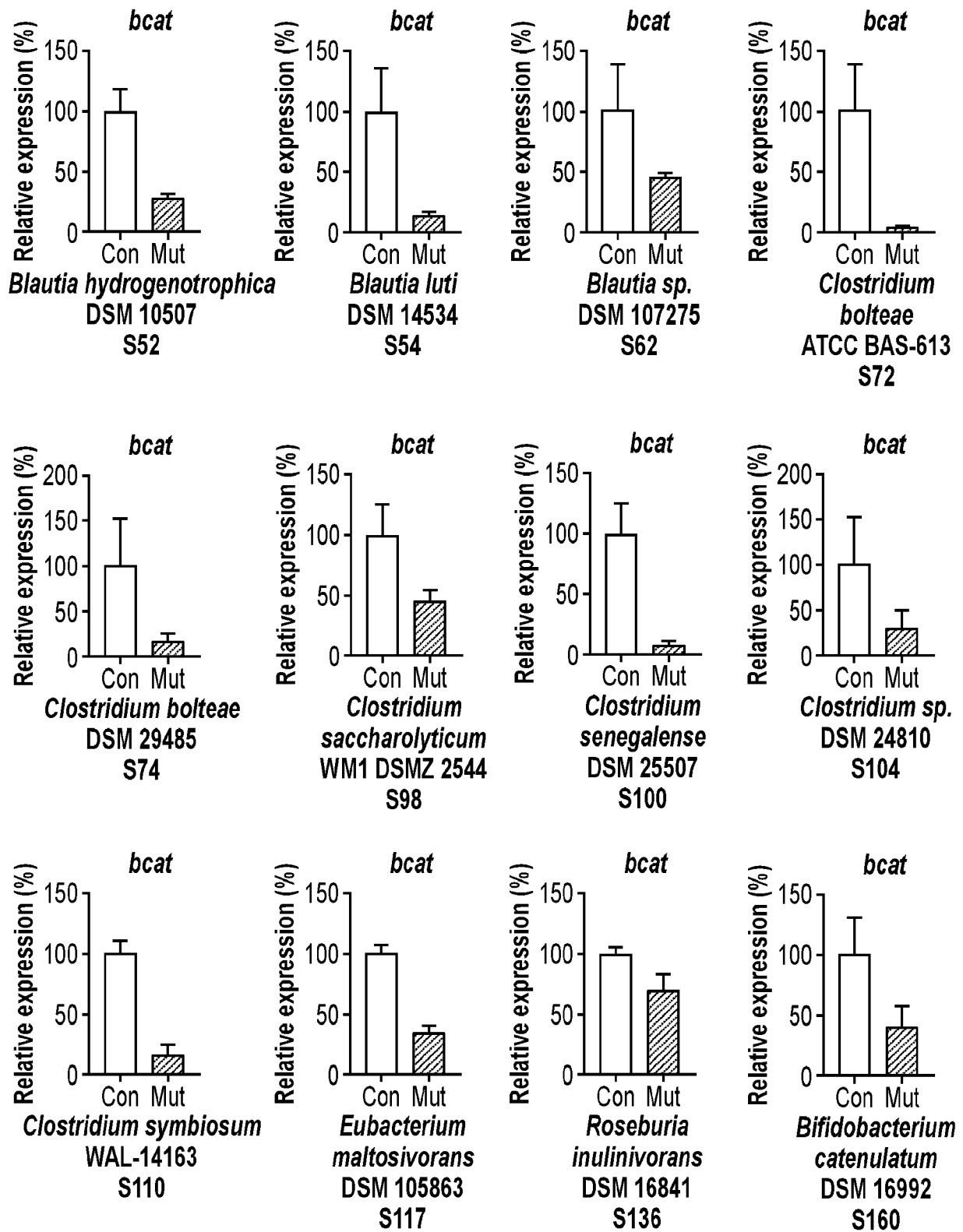


FIG. 14

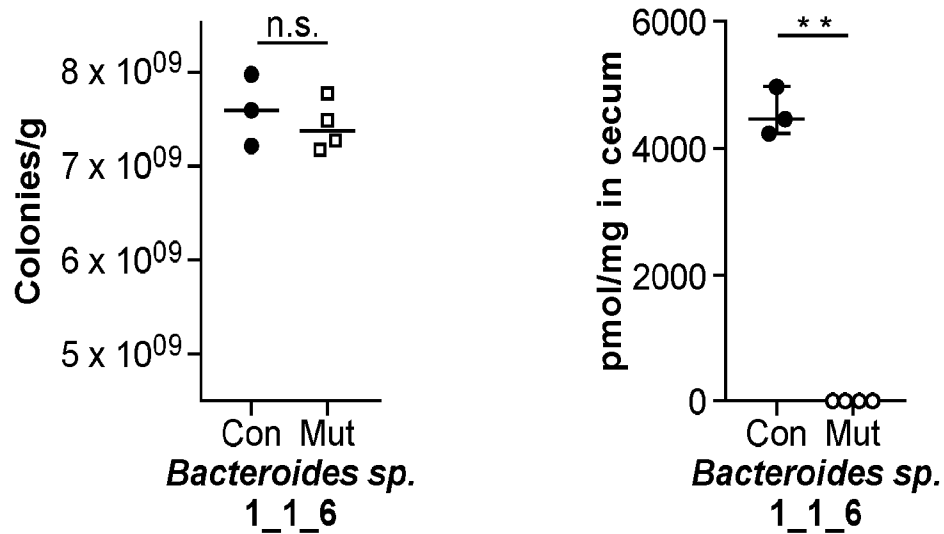


FIG. 15A

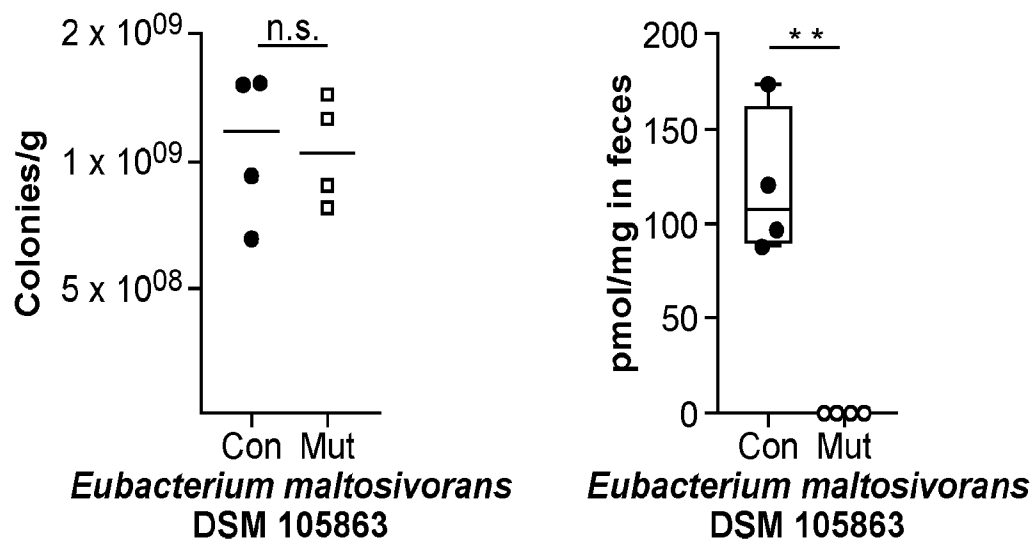


FIG. 15B

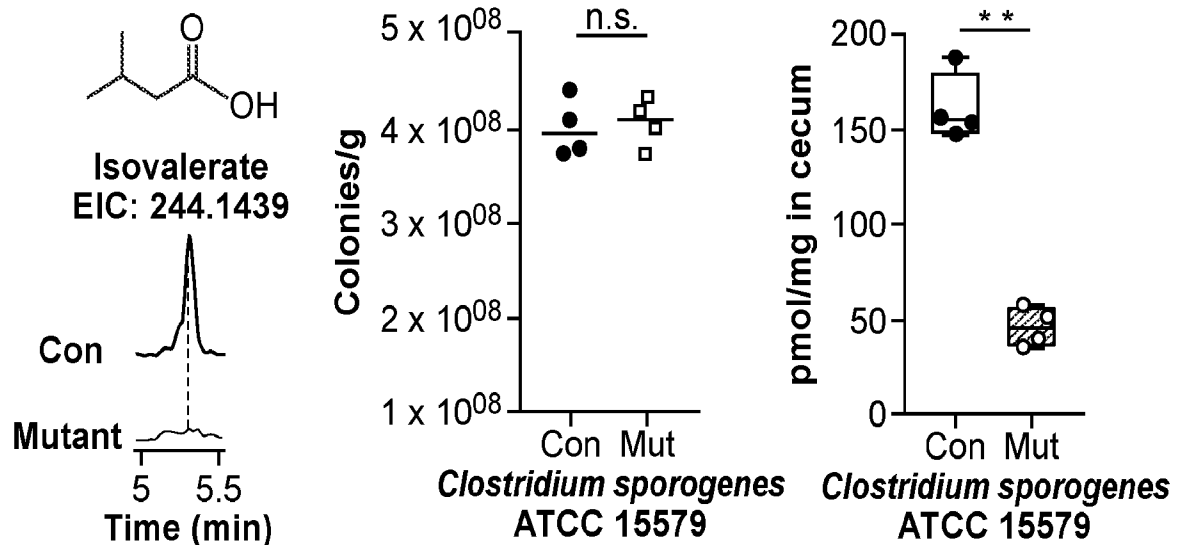


FIG. 15C

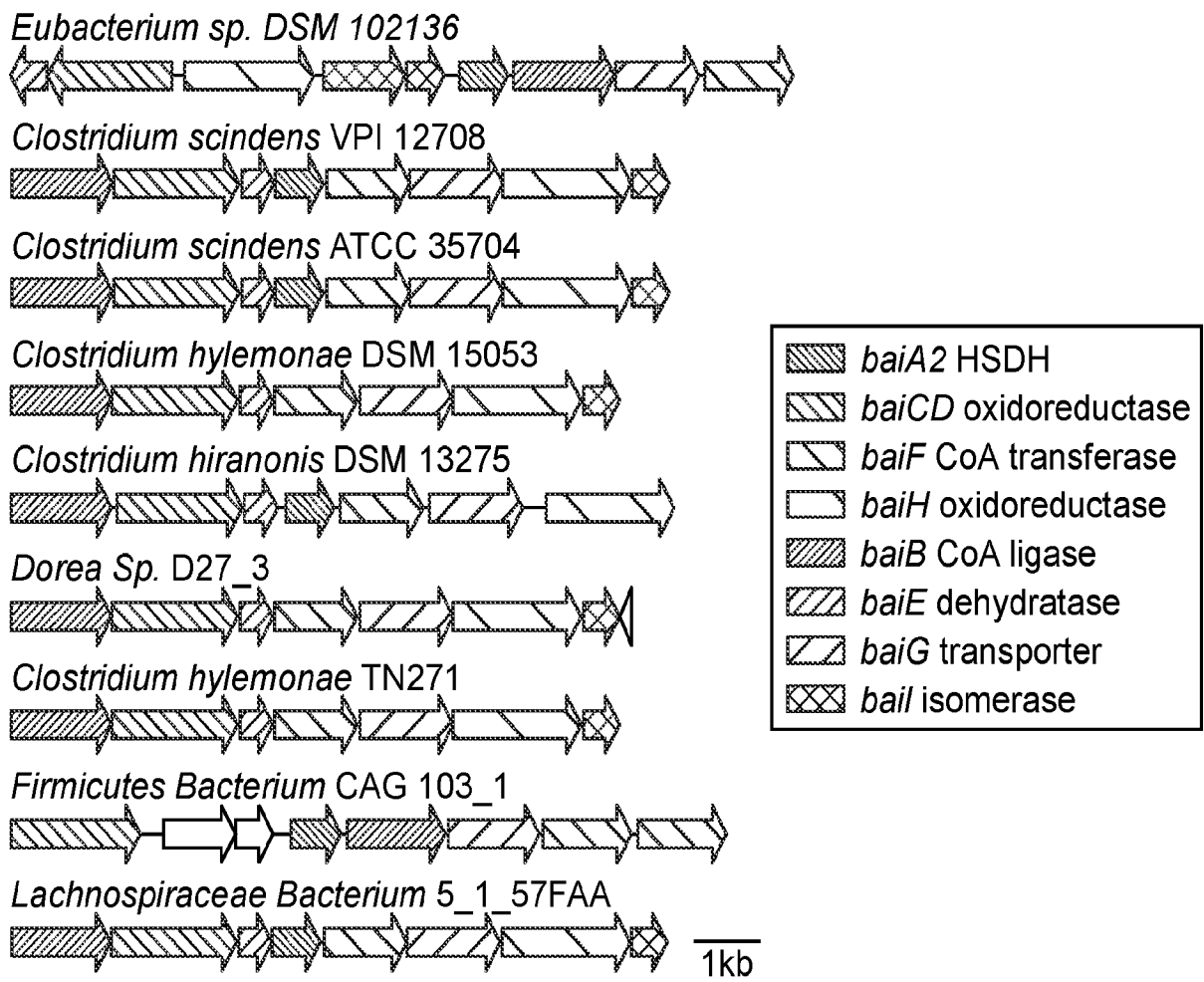


FIG. 16A

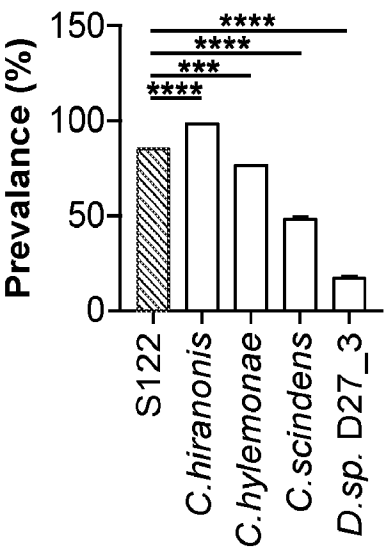


FIG. 16B

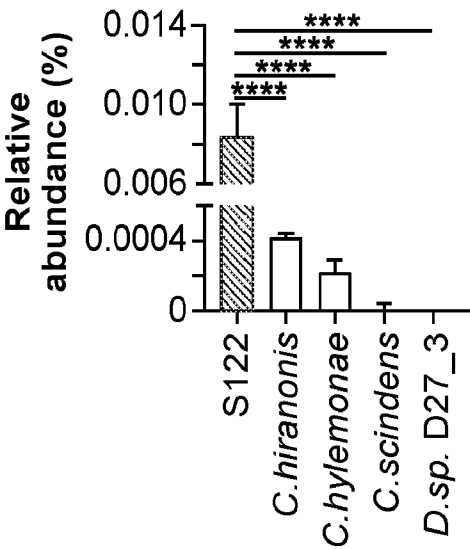


FIG. 16C

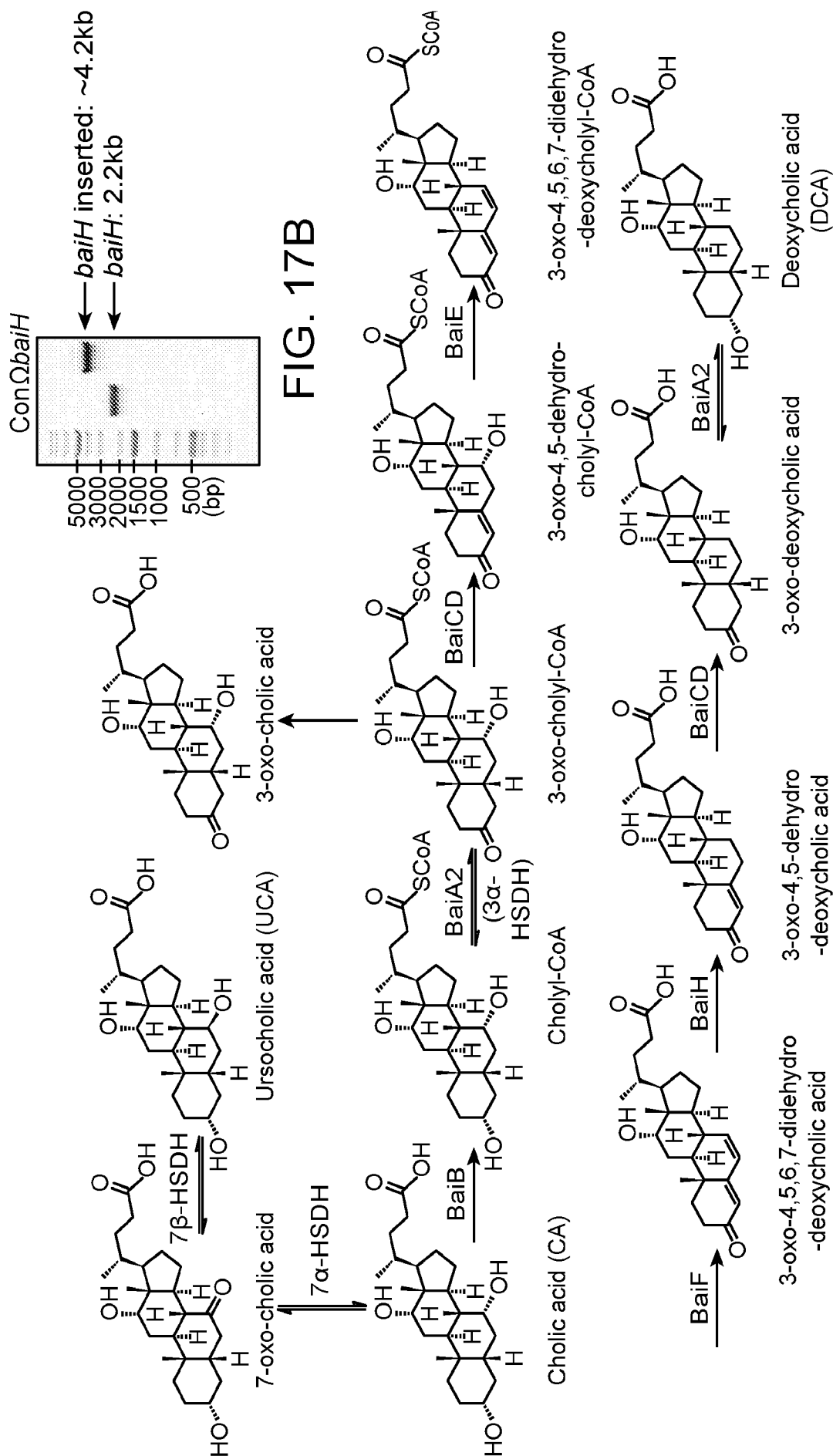


FIG. 17A

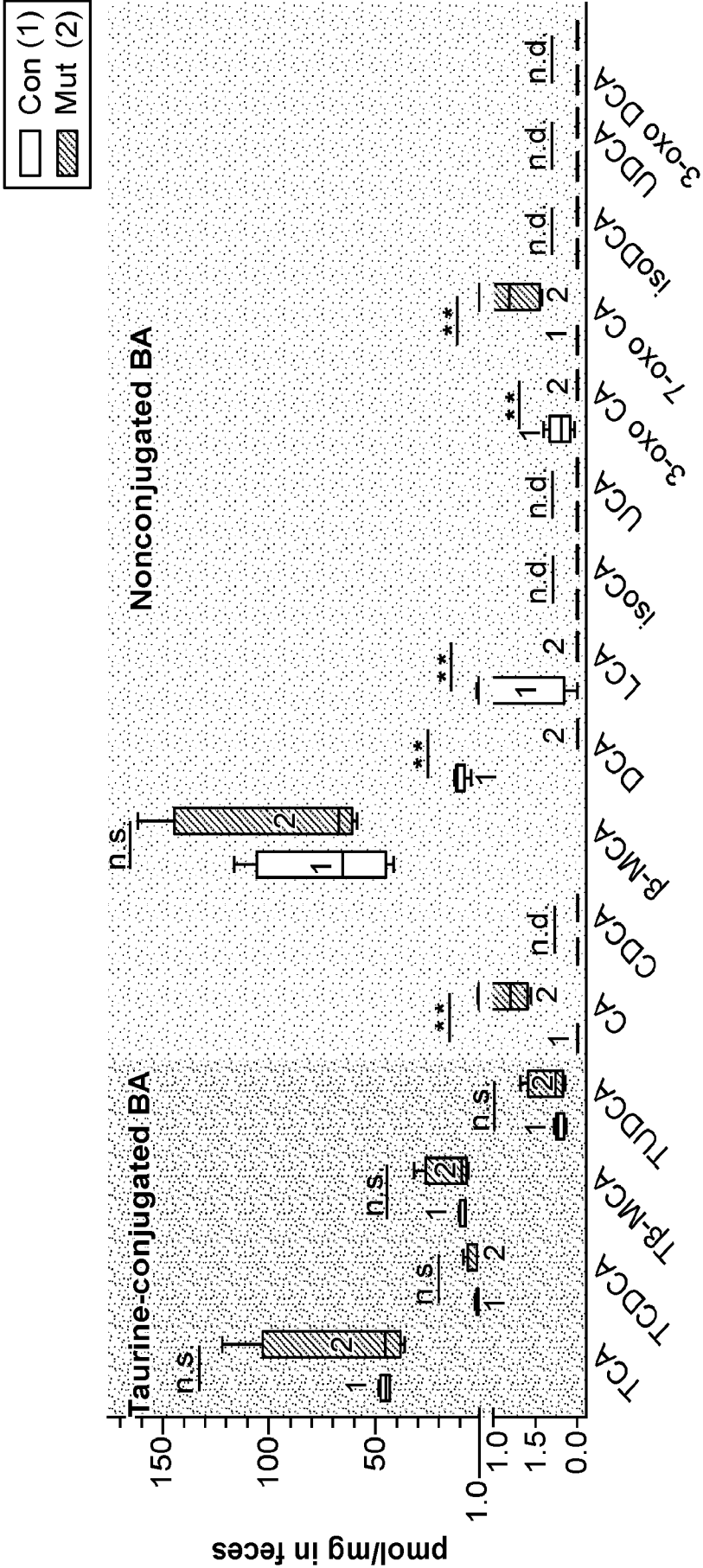


FIG. 17C

	Family	Strains
	Clostridiaceae	5
	Lachnospiraceae	14
	Eubacteriaceae	5
	Erysipelotrichaceae	3
	Enterococcaceae	2
	Bifidobacteriaceae	1
	Enterobacteriaceae	1
	Prevotellaceae	1
	Tannerellaceae	2
	Bacteroidaceae	22

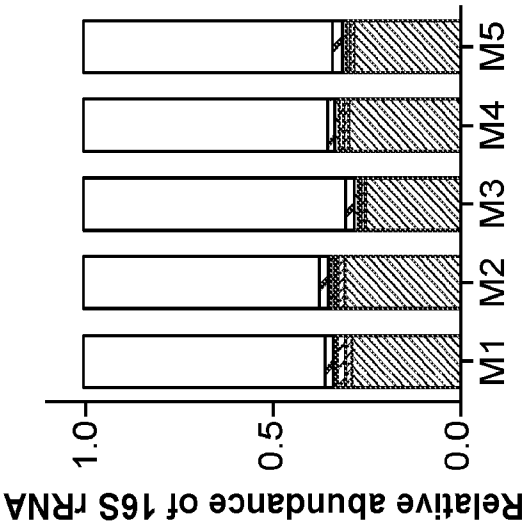


FIG. 17D

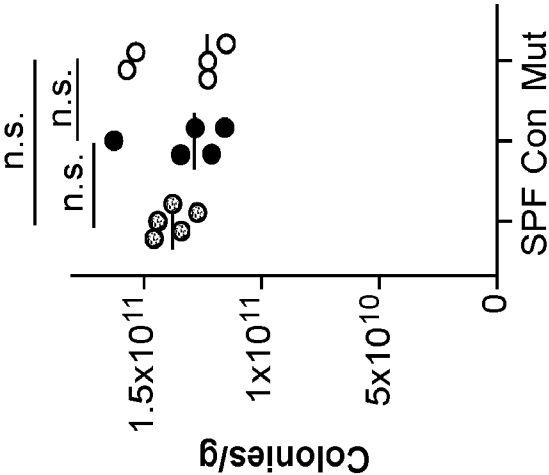
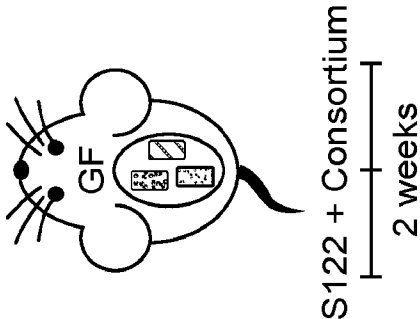
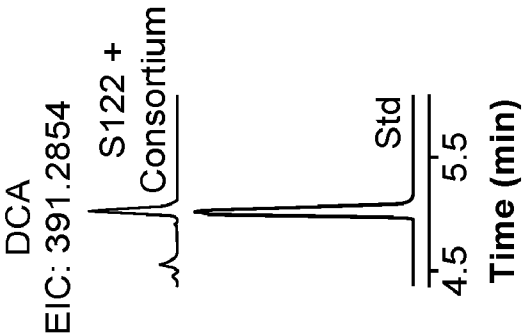


FIG. 17E

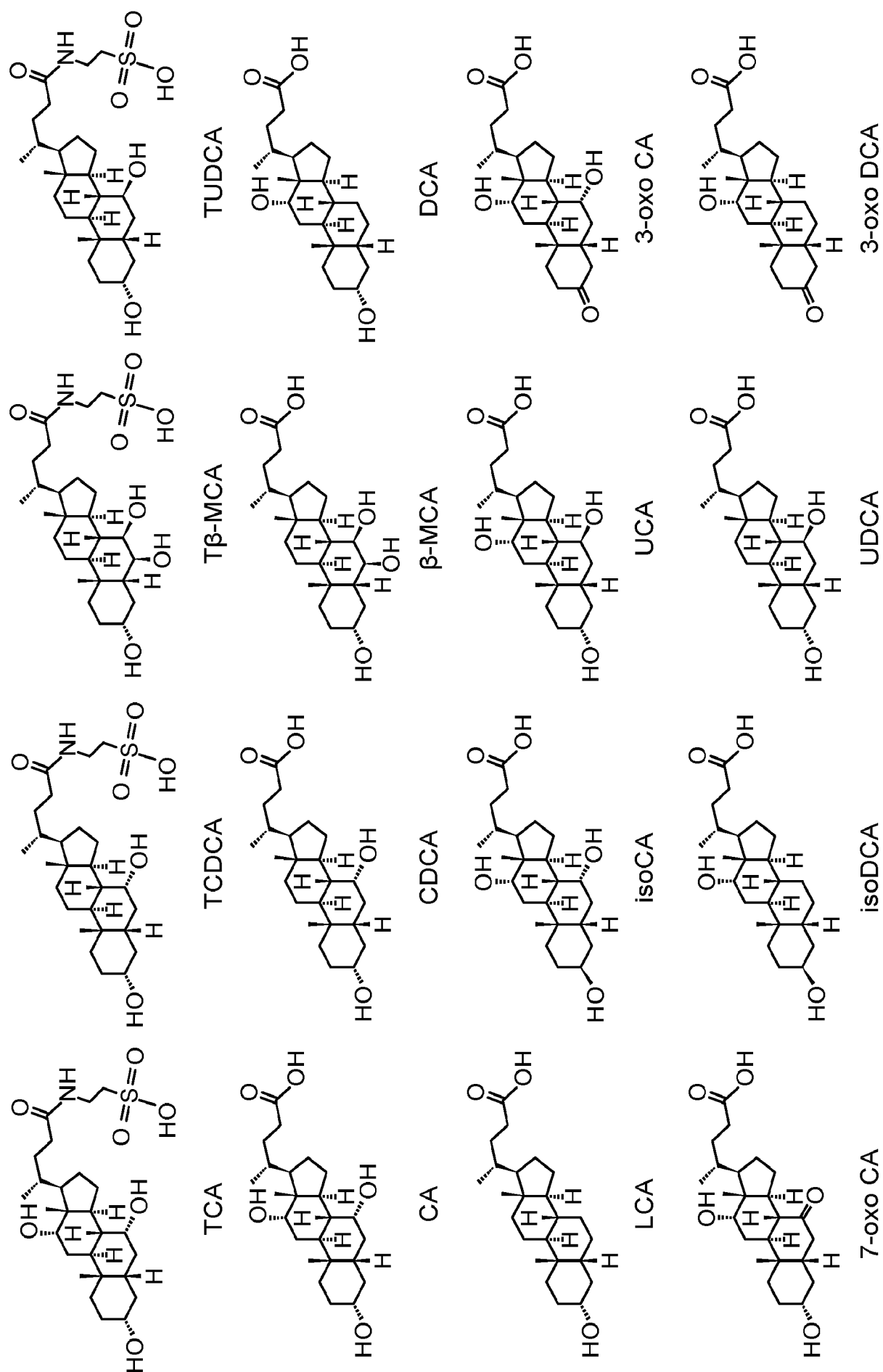
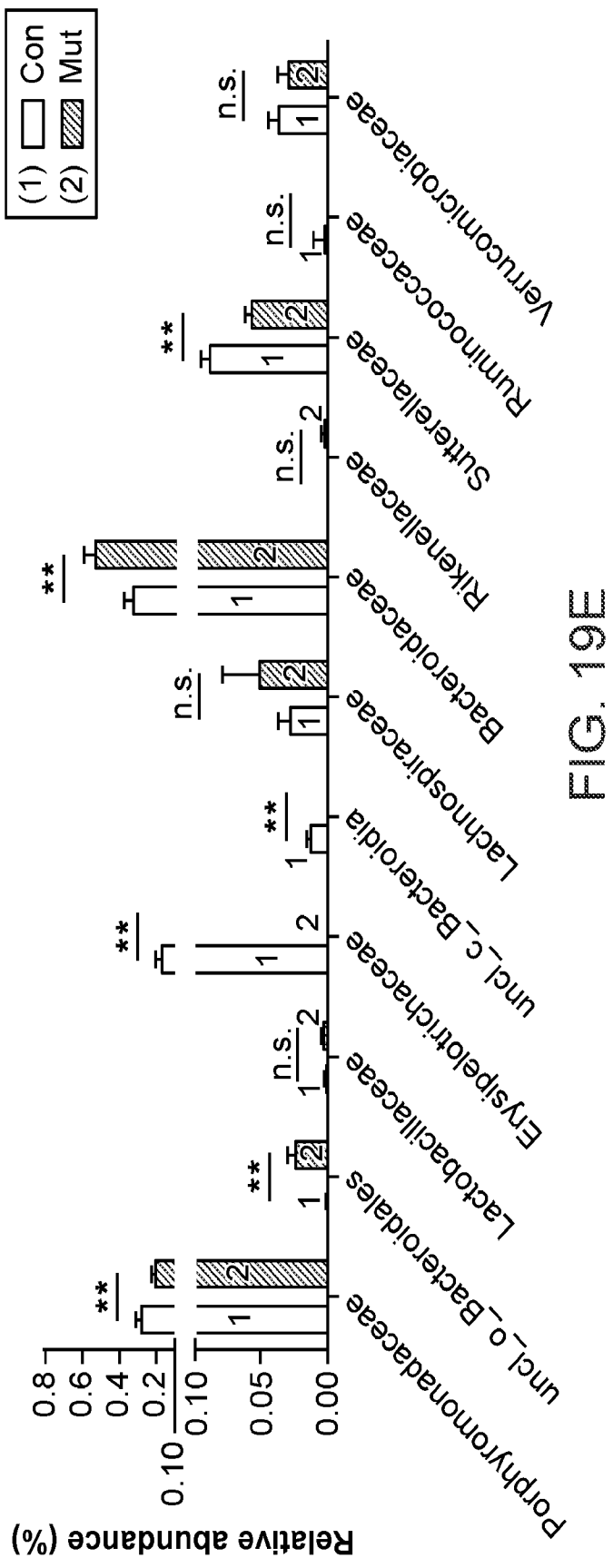
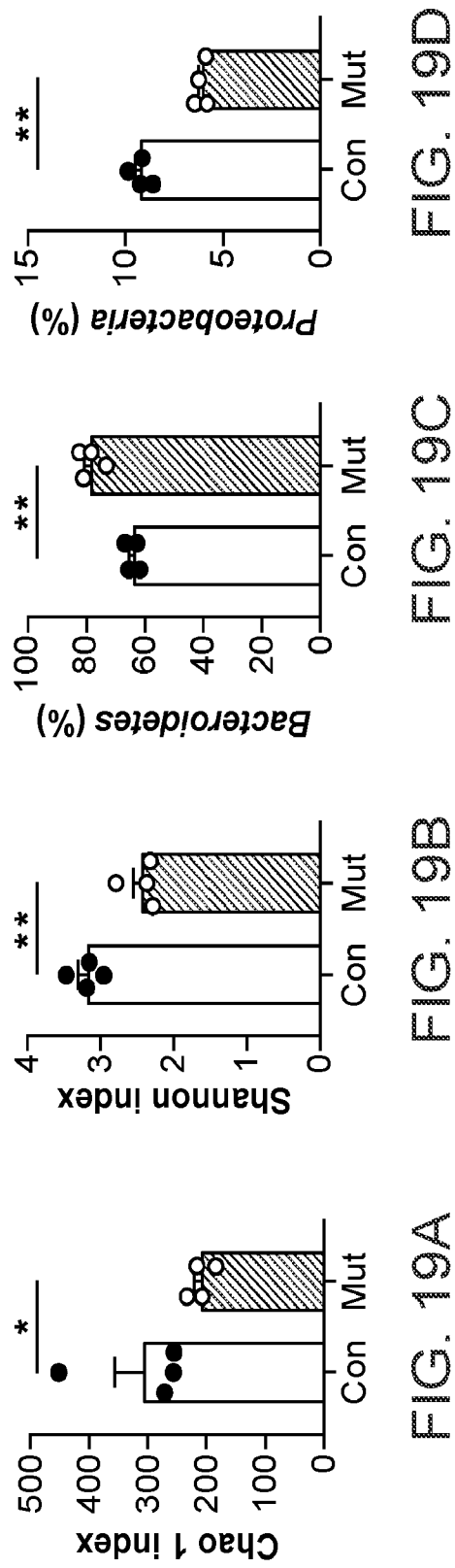


FIG. 18



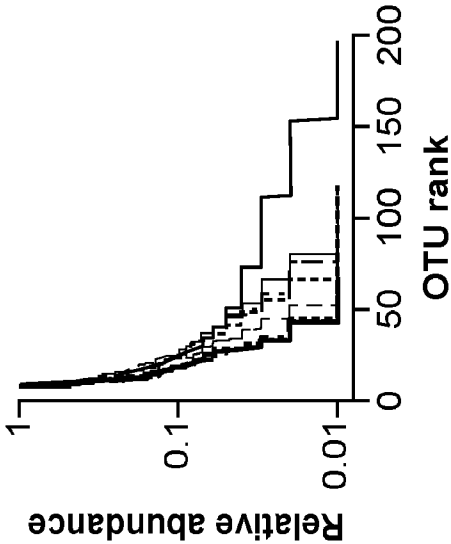
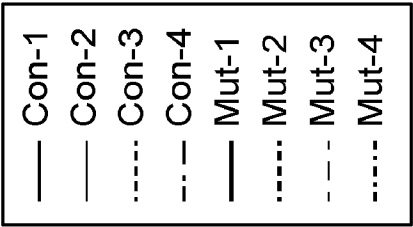


FIG. 19G

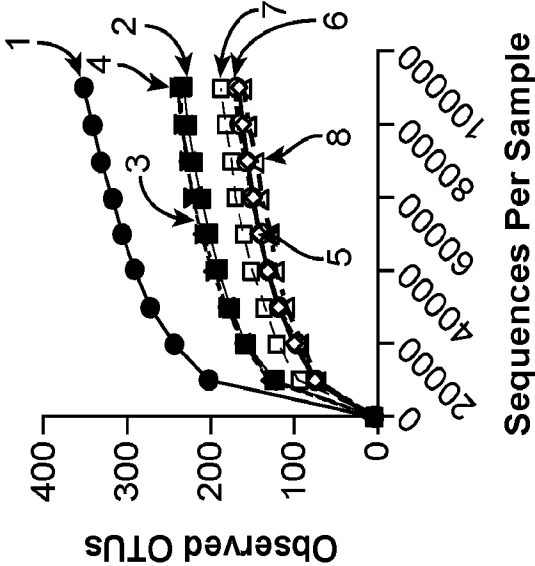
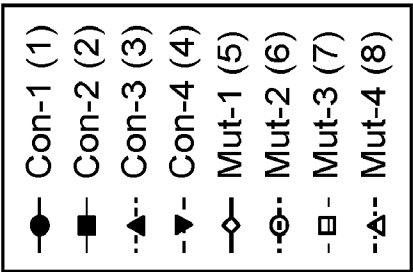


FIG. 19F

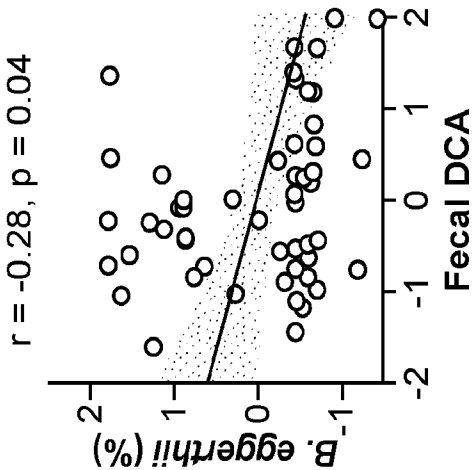


FIG. 19I

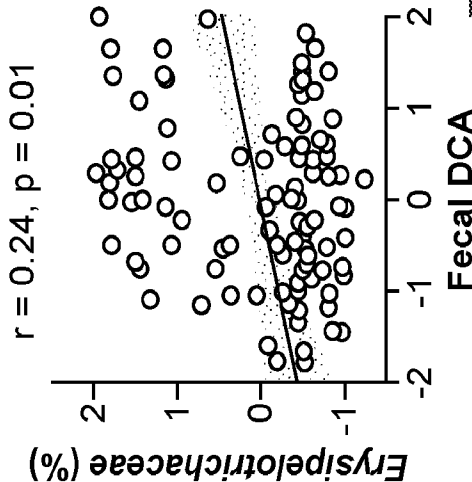


FIG. 19H



FIG. 20A



FIG. 20B

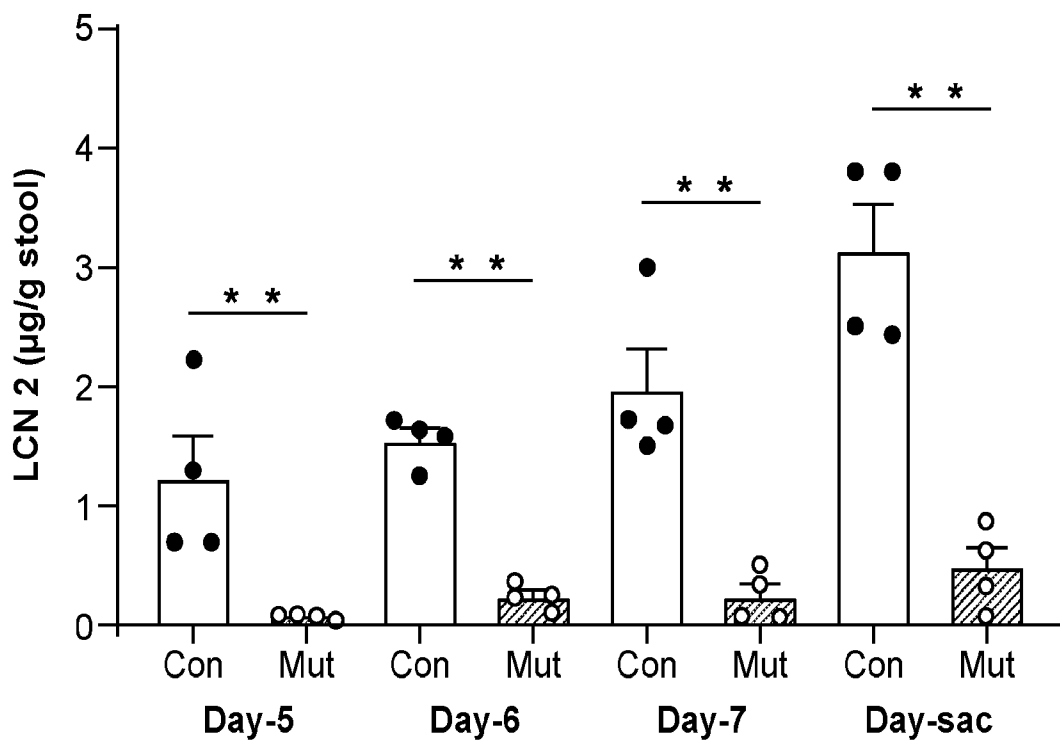


FIG. 20C

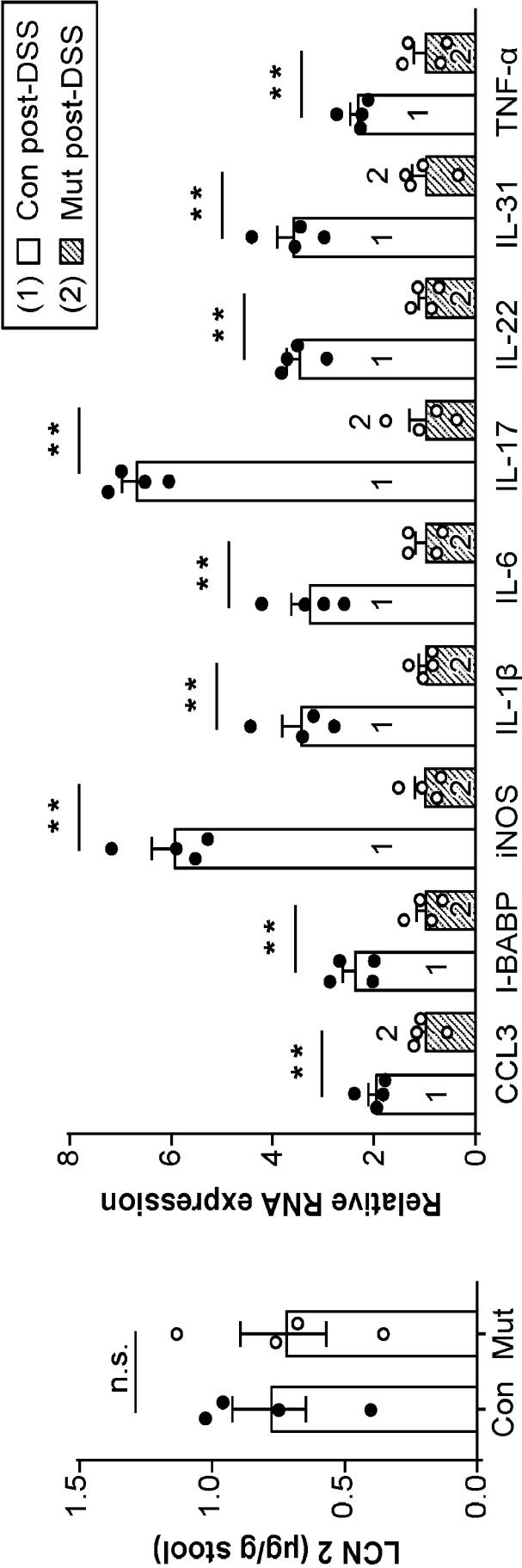


FIG. 20D

FIG. 20E

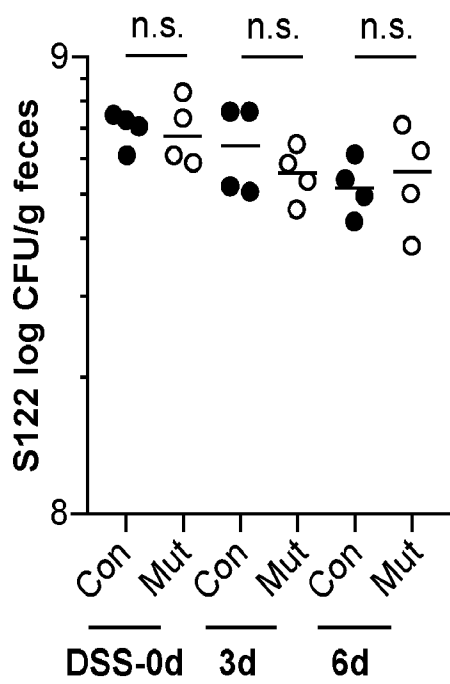


FIG. 20F

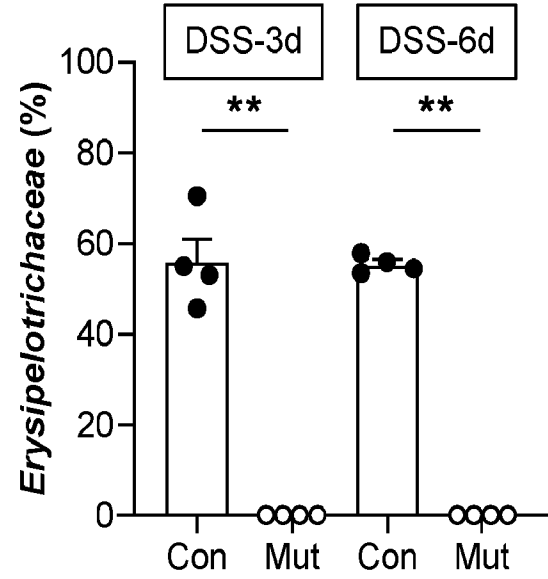


FIG. 20G

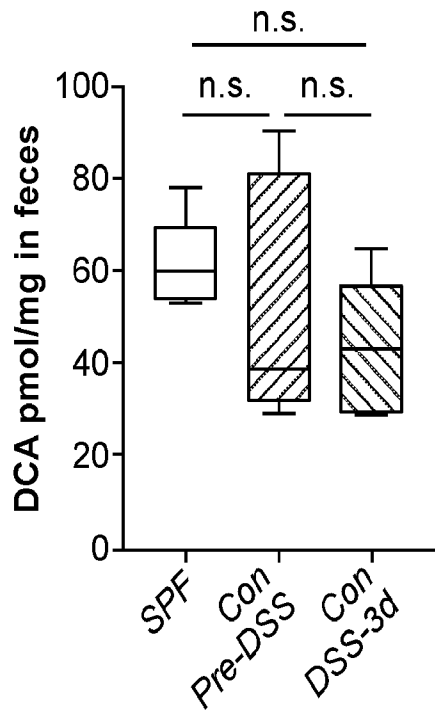


FIG. 20H

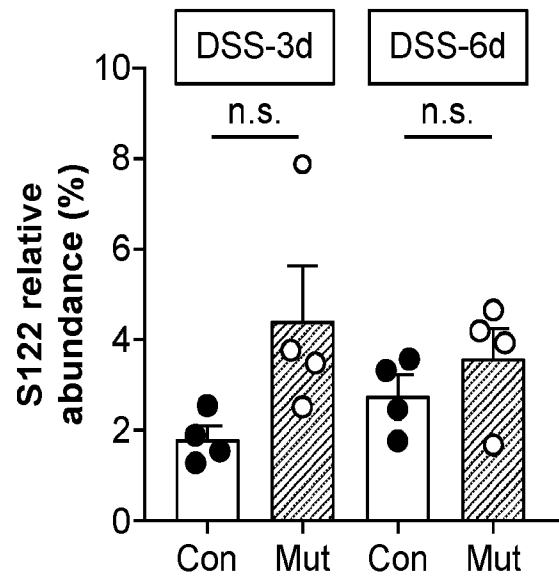


FIG. 20I

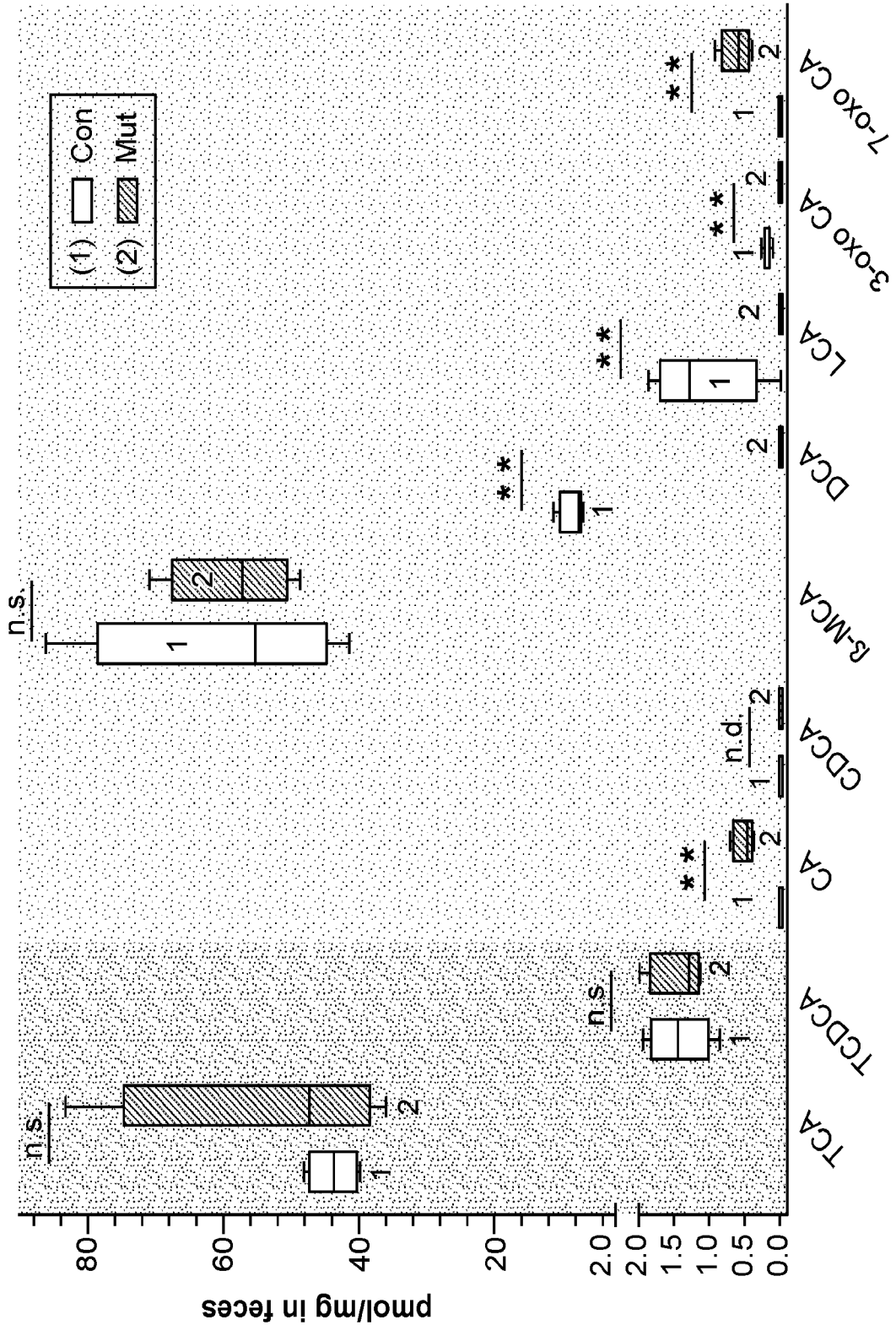


FIG. 20J

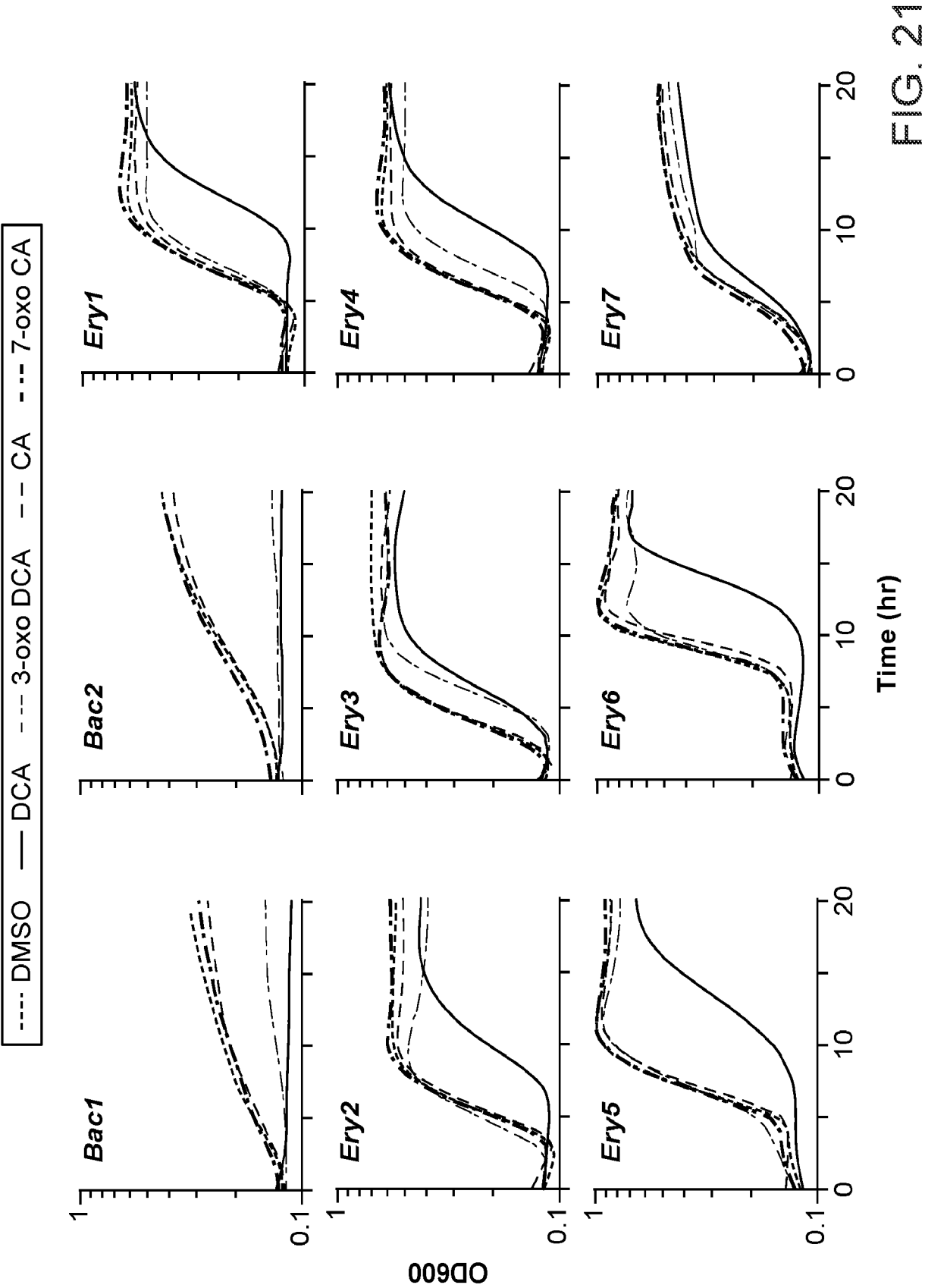


FIG. 21

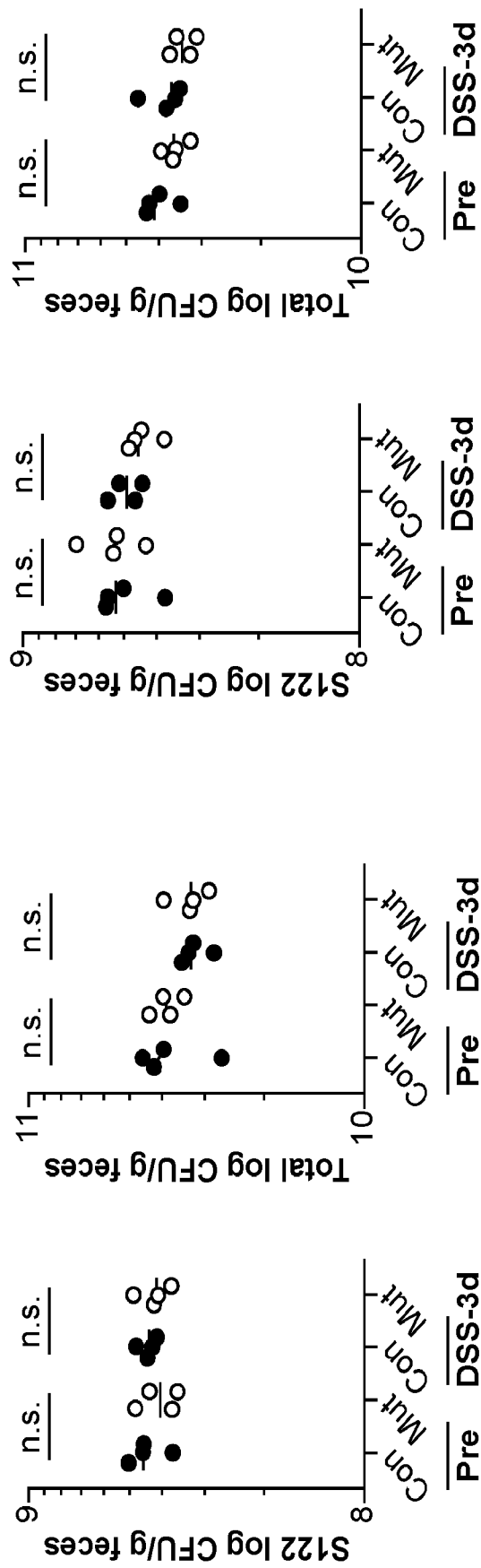
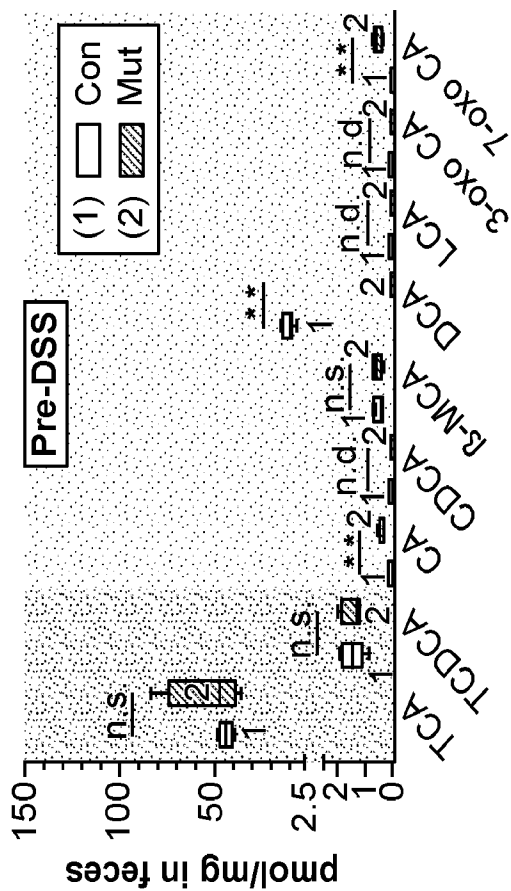
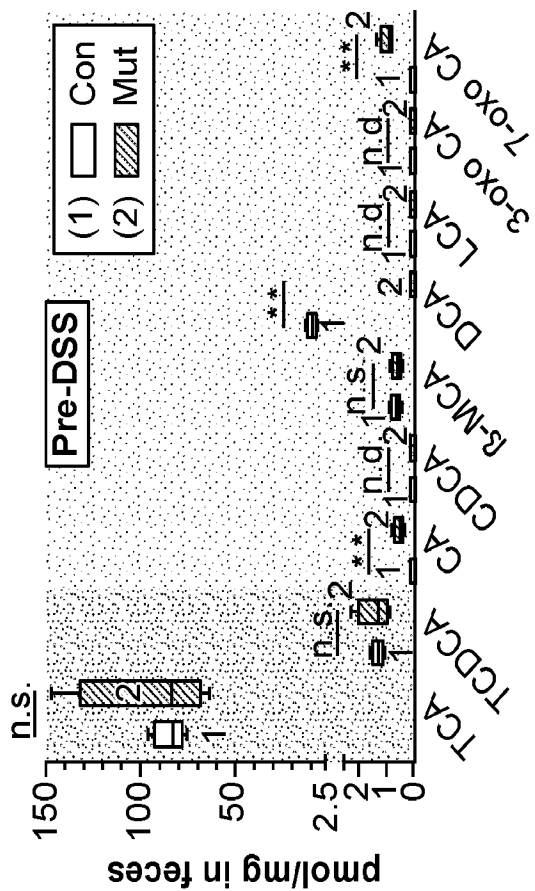
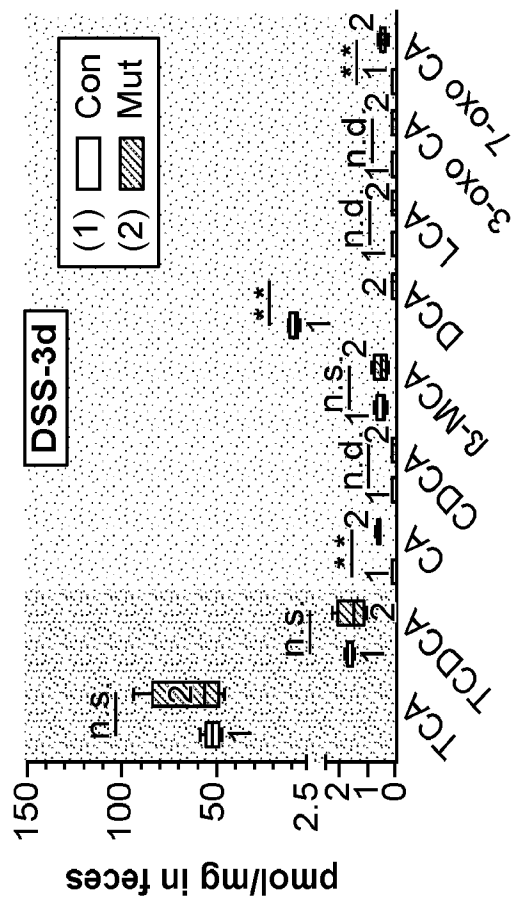
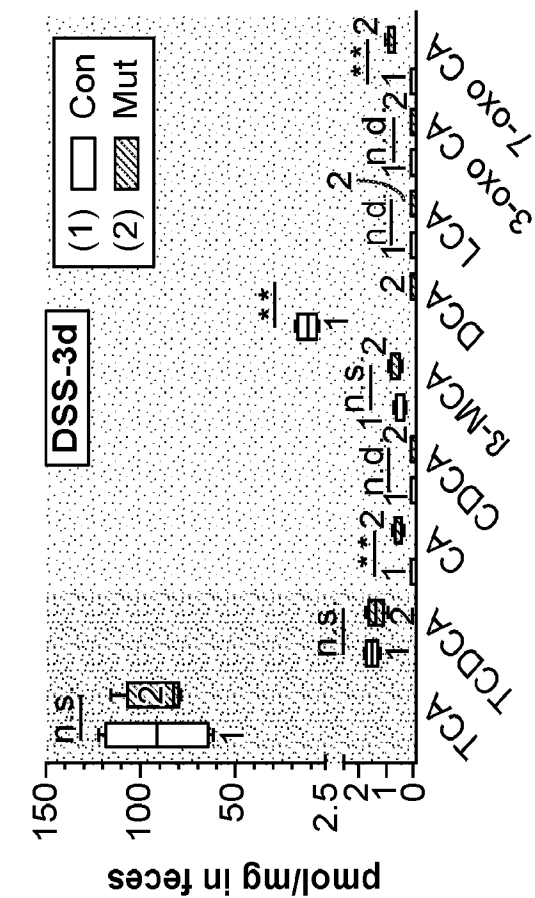


FIG. 22A

FIG. 22B



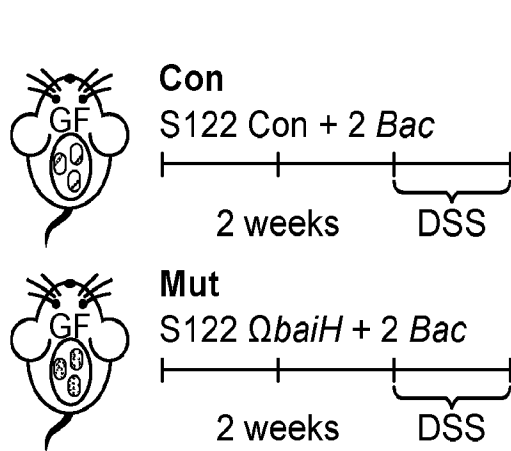


FIG. 22G

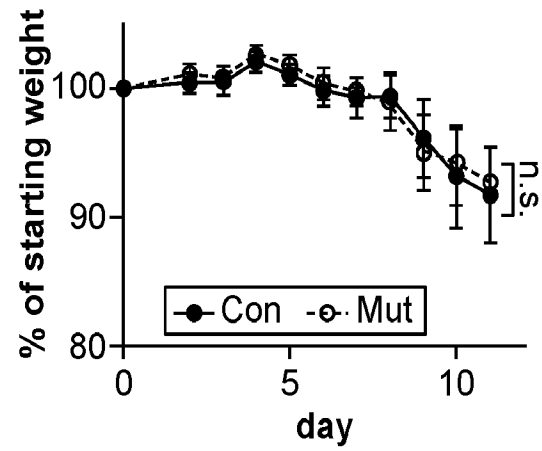


FIG. 22H

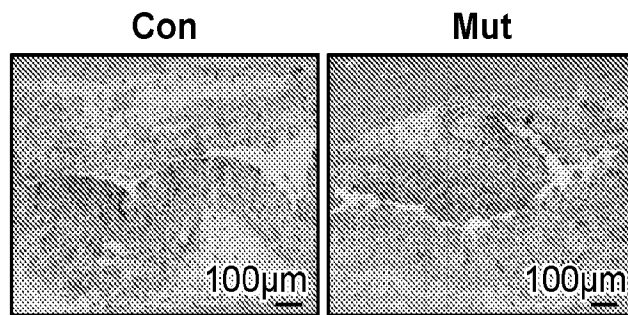


FIG. 22I

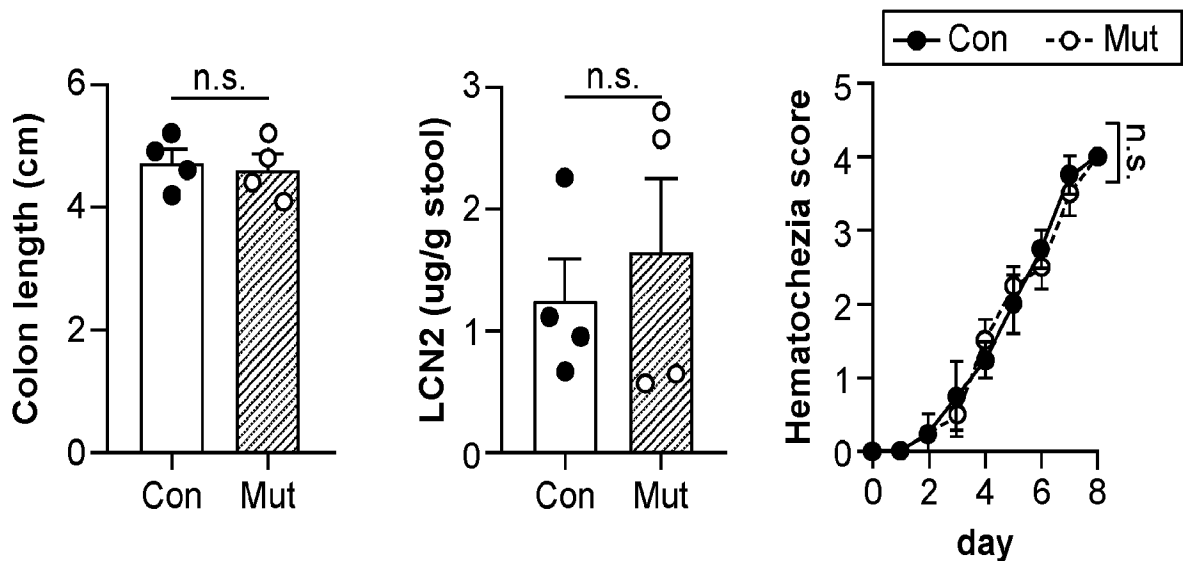


FIG. 22J

FIG. 22K

FIG. 22L

FIG. 23

Strain No.	Strain Name	Taxonomy	Culture agar	Culture broth
	Gram-Bacteroidia			
1	<i>Bacteroides caccae</i> ATCC 43185	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_caccae	TSAB	TYGB
2	<i>Bacteroides cellulosilyticus</i> DSM 108229	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_cellulosilyticus	TSAB	TYGB
3	<i>Bacteroides cellulosilyticus</i> DSM 14838	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_cellulosilyticus	TSAB	TYGB
4	<i>Bacteroides clarus</i> DSM 22519	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_clarus	TSAB	TYGB
5	<i>Bacteroides coprocola</i> DSM 17136	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_coprocola	TSAB	TYGB
6	<i>Bacteroides coprophilus</i> DSM 18228	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_coprophilus	TSAB	TYGB
7	<i>Bacteroides dorei</i> 5_1_36/D4 (HM 29)	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidales incertae sedis;g_Phocaeicola;s_dorei	TSAB	TYGB
8	<i>Bacteroides dorei</i> DSM 17855	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidales incertae sedis;g_Phocaeicola;s_dorei	TSAB	TYGB
9	<i>Bacteroides eggerthii</i> DSM 20697	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_eggerthii	TSAB	TYGB
10	<i>Bacteroides faecis</i> DSM 24798	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_faecis	TSAB	TYGB
11	<i>Bacteroides finegoldii</i> DSM 108067	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_finegoldii	TSAB	TYGB
12	<i>Bacteroides finegoldii</i> DSM 17565	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_finegoldii	TSAB	TYGB
13	<i>Bacteroides fluxus</i> DSM 22534	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_fluxus	TSAB	TYGB
14	<i>Bacteroides fragilis</i> 3_1_12	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_fragilis	TSAB	TYGB

FIG. 23 (CONT.)

15	<i>Bacteroides galacturonicus</i> DSM 3978	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_galacturonicus	TSAB	TYGB
16	<i>Bacteroides intestinalis</i> DSM 108646	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_intestinalis	TSAB	TYGB
17	<i>Bacteroides intestinalis</i> DSM 17393	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_intestinalis	TSAB	TYGB
18	<i>Bacteroides neonati</i> DSM 26805	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_neonati	TSAB	TYGB
19	<i>Bacteroides nordii</i> DSM 18764	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_nordii	TSAB	TYGB
20	<i>Bacteroides oleiciplenus</i> DSM 22535	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_oleiciplenus	TSAB	TYGB
21	<i>Bacteroides ovatus</i> ATCC 8483	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_ovatus	TSAB	TYGB
22	<i>Bacteroides plebeius</i> DSM 17135	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_plebeius	TSAB	TYGB
23	<i>Bacteroides salyersiae</i> DSM 18765	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_salyersiae	TSAB	TYGB
24	<i>Bacteroides salyersiae</i> DSM 18765	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_salyersiae	TSAB	TYGB
25	<i>Bacteroides</i> sp. 1_1_6	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;unclassified Bacteroides	TSAB	TYGB
26	<i>Bacteroides</i> sp. 2_1_16	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;unclassified Bacteroides	TSAB	TYGB
27	<i>Bacteroides</i> sp. 2_1_22	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;unclassified Bacteroides	TSAB	TYGB
28	<i>Bacteroides</i> sp. 9_1_42FAA	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;unclassified Bacteroides	TSAB	TYGB
29	<i>Bacteroides</i> sp. D2	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;unclassified Bacteroides	TSAB	TYGB

FIG. 23 (CONT.)

30	<i>Bacteroides stercoris</i> ATCC 43183 (DSMZ 19555)	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_stercoris	TSAB	TYGB
31	<i>Bacteroides thetaiotaomicron</i> VPI-5482	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_thetaiotaomicron	TSAB	TYGB
32	<i>Bacteroides uniformis</i> ATCC 8492	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_uniformis	TSAB	TYGB
33	<i>Bacteroides vulgatus</i> ATCC 8482	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_vulgatus	TSAB	TYGB
34	<i>Bacteroides xylanisolvens</i> DSMZ 18836	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_xylanisolvens	TSAB	TYGB
35	<i>Bacteroides</i> sp. D20	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;unclassified Bacteroides	TSAB	TYGB
36	<i>Bacteroides eggerthii</i> 1_2_48FAA	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_eggerthii	TSAB	TYGB
37	<i>Bacteroides</i> sp. 1_1_30	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;unclassified Bacteroides	TSAB	TYGB
38	<i>Bacteroides</i> sp. 2_1_16 (HM 58)	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;unclassified Bacteroides	TSAB	TYGB
39	<i>Bacteroides salyersiae</i> CL02T12C01	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_salyersiae	TSAB	TYGB
40	<i>Parabacteroides distasonis</i> ATCC 8503	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Tannerellaceae;g_Parabacteroides;s_distasonis	TSAB	TYGB
41	<i>Parabacteroides faecis</i> DSM 102983	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Tannerellaceae;g_Parabacteroides;s_faecis	TSAB	TYGB
42	<i>Parabacteroides goldsteinii</i> DSM 19448	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Tannerellaceae;g_Parabacteroides;s_goldsteinii	TSAB	TYGB
43	<i>Parabacteroides johnsonii</i> DSM 18315	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Tannerellaceae;g_Parabacteroides;s_johnsonii	TSAB	TYGB

FIG. 23 (CONT.)

44	<i>Parabacteroides merdae</i> ATCC 43184 (DSMZ 19495)	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Tannerellaceae;g_Parabacteroides;s_merdae	TSAB	TYGB
45	<i>Parabacteroides</i> sp. D13	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Tannerellaceae;g_Parabacteroides;unclassified Parabacteroides	TSAB	TYGB
46	<i>Prevotella bivia</i> DSM 20514	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_bivia	TSAB	Mega
47	<i>Prevotella buccae</i> DSM 19025	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_buccae	TSAB	Mega
48	<i>Prevotella buccalis</i> ATCC 35310 (DSMZ 20616)	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_buccalis	TSAB	Mega
49	<i>Prevotella copri</i> DSM 18205	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_copri	TSAB	Mega
	Gram+ Clostridia			
50	<i>Blautia faecis</i> DSM 27629	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_faecis	TSAB	Mega
51	<i>Blautia hansenii</i> DSM 20583	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_hansenii	TSAB	Mega
52	<i>Blautia hydrogenotrophica</i> DSM 10507	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_hydrogenotrophica	TSAB	CMM
53	<i>Blautia hydrogenotrophica</i> DSM 108220	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_hydrogenotrophica	TSAB	Mega
54	<i>Blautia luti</i> DSM 14534	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_luti	TSAB	Mega
55	<i>Blautia producta</i> CJG-5	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_producta	TSAB	Mega
56	<i>Blautia producta</i> DSM 14466	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_producta	TSAB	Mega
57	<i>Blautia</i> sp. DSM 101187	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;unclassified Blautia	TSAB	Mega

FIG. 23 (CONT.)

58	<i>Blautia</i> sp. DSM 102163	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;unclassified Blautia	TSAB	Mega
59	<i>Blautia</i> sp. DSM 102164	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;unclassified Blautia	TSAB	Mega
60	<i>Blautia</i> sp. DSM 102165	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;unclassified Blautia	TSAB	Mega
61	<i>Blautia</i> sp. DSM 102174	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;unclassified Blautia	TSAB	Mega
62	<i>Blautia</i> sp. DSM 107275	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;unclassified Blautia	TSAB	Mega
63	<i>Blautia</i> sp. DSM 107519	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;unclassified Blautia	TSAB	Mega
64	<i>Blautia</i> sp. DSM 108066	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;unclassified Blautia	TSAB	Mega
65	<i>Blautia</i> <i>wexlerae</i> DSM 19850	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_wexlerae	TSAB	CMM
66	<i>Blautia</i> <i>wexlerae</i> DSM 25403	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Blautia;s_wexlerae	TSAB	Mega
67	<i>Clostridiales</i> <i>bacterium</i> 3_1_39B/D5	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_unclassified Clostridiales;g_unclassified Clostridiales	TSAB	Mega
68	<i>Clostridiales</i> <i>bacterium</i> 4_1_37FAA	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_unclassified Clostridiales;g_unclassified Clostridiales	TSAB	Mega
69	<i>Clostridium</i> <i>aldense</i> WAL-18727	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_aldense	TSAB	Mega
70	<i>Clostridium</i> <i>amazonense</i> DSM 23598	k_Bacteria;p_Firmicutes;c_Clostridia;o_Eubacteriales;f_Clostridiaceae;g_Clostridium;s_amazonense	TSAB	Mega
71	<i>Clostridium</i> <i>bartlettii</i> DSM 16795 (ATCC BAA-827)	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptostreptococcaceae;g_Intestinibacter;s_bartlettii	TSAB	Mega
72	<i>Clostridium</i> <i>bolteae</i> ATCC BAA-613	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_bolteae	TSAB	Mega

FIG. 23 (CONT.)

73	<i>Clostridium bolteae</i> CJG-3	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_bolteae	TSAB	Mega
74	<i>Clostridium bolteae</i> DSM 29485	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_bolteae	TSAB	Mega
75	<i>Clostridium bolteae</i> HM-318 aka WAL-14578	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_bolteae	TSAB	Mega
76	<i>Clostridium bolteae</i> strain 47	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_bolteae	TSAB	Mega
77	<i>Clostridium bolteae</i> strain 54	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_bolteae	TSAB	Mega
78	<i>Clostridium bolteae</i> strain 55	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_bolteae	TSAB	Mega
79	<i>Clostridium bolteae</i> strain 56	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_bolteae	TSAB	Mega
80	<i>Clostridium celatum</i> DSM 1785	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;s_celatum	TSAB	Mega
81	<i>Clostridium clostridioforme</i> 2_1_49FAA	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_clostridioforme	TSAB	Mega
82	<i>Clostridium clostridioforme</i> WAL-7855	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_clostridioforme	TSAB	Mega
83	<i>Clostridium hathewayi</i> CJG-2	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Hungatella;s_hathewayi	TSAB	Mega
84	<i>Clostridium hathewayi</i> DSM 13479	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Hungatella;s_hathewayi	BHIB	Mega
85	<i>Clostridium hathewayi</i> WAL-18680	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Hungatella;s_hathewayi	TSAB	Mega
86	<i>Clostridium hiranonis</i> DSM 13275	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Peptostreptococcaceae;g_unclassified_Peptostreptococcaceae;s_hiranonis	TSAB	Mega
87	<i>Clostridium innocuum</i> 6_1_30	k_Bacteria;p_Firmicutes;c_Erysipelotrichia;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Erysipelatoclostridium;s_innocuum	TSAB	Mega

FIG. 23 (CONT.)

88	<i>Clostridium innocuum</i> DSM 22910	k_Bacteria;p_Firmicutes;c_Erysipelotrichia;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Erysipelatoclostridium;s_innocuum	TSAB	Mega
89	<i>Clostridium jeddahense</i> DSM 27834	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;s_jeddahense	TSAB	Mega
90	<i>Clostridium lavalense</i> DSM 19851	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Enterocloster;s_lavalense	TSAB	Mega
91	<i>Clostridium methylpentosum</i> DSM 5476 (ATCC 43829)	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Clostridium;s_methylpentosum	BHIB	Mega
92	<i>Clostridium paraputrificum</i> DSM 103826	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;s_paraputrificum	TSAB	Mega
93	<i>Clostridium polynesiense</i> DSM 27072	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;s_polynesiense	TSAB	Mega
94	<i>Clostridium ramosum</i> ATCC	k_Bacteria;p_Firmicutes;c_Erysipelotrichia;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Erysipelatoclostridium;s_amosum	TSAB	Mega
95	<i>Clostridium ramosum</i> CJG-1	k_Bacteria;p_Firmicutes;c_Erysipelotrichia;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Erysipelatoclostridium;s_amosum	TSAB	Mega
96	<i>Clostridium ramosum</i> DSM 1402	k_Bacteria;p_Firmicutes;c_Erysipelotrichia;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Erysipelatoclostridium;s_amosum	TSAB	Mega
97	<i>Clostridium ramosum</i> DSM 24812	k_Bacteria;p_Firmicutes;c_Erysipelotrichia;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Erysipelatoclostridium;s_amosum	BHIB	Mega
98	<i>Clostridium saccharolyticum</i> WM1 DSMZ 2544 (ATCC 35040)	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Lacrimispora;s_saccharolyticum	BHIB	Mega
99	<i>Clostridium scindens</i> ATCC 35704	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Lachnoclostridium;s_scindens	TSAB	Mega
100	<i>Clostridium senegalense</i> DSM 25507	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;s_senegalense	TSAB	Mega
101	<i>Clostridium</i> sp. 7_2_43FAA	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;unclassified Clostridium	TSAB	Mega

FIG. 23 (CONT.)

102	<i>Clostridium</i> sp. DSM 100503	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;unclassified Clostridium	TSAB	RCM
103	<i>Clostridium</i> sp. DSM 13576	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;unclassified Clostridium	RCM	Mega
104	<i>Clostridium</i> sp. DSM 24810	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;unclassified Clostridium	TSAB	RCM
105	<i>Clostridium</i> sp. DSM 24813	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;unclassified Clostridium	TSAB	Mega
106	<i>Clostridium</i> sp. DSM 46282	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;unclassified Clostridium	TSAB	RCM
107	<i>Clostridium sporogenes</i> ATCC 15579	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;s_sporogenes	TSAB	Mega
108	<i>Clostridium sporogenes</i> DSM 795	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Clostridiaceae;g_Clostridium;s_sporogenes	TSAB	Mega
109	<i>Clostridium symbiosum</i> DSM 29356	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Lachnoclostridium;s_symbiosum	TSAB	Mega
110	<i>Clostridium symbiosum</i> WAL-14163	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Lachnoclostridium;s_symbiosum	TSAB	Mega
111	<i>Clostridium symbiosum</i> WAL-14673	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Lachnoclostridium;s_symbiosum	TSAB	Mega
112	<i>Eubacterium dolichum</i> DSM 3991	k_Bacteria;p_Firmicutes;c_Erysipelotrichia;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Absiella;s_dolichum	TSAB	CMM
113	<i>Eubacterium eligens</i> ATCC 27750 (DSMZ 3376)	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Lachnospira;s_eligens	TSAB	Mega
114	<i>Eubacterium eligens</i> DSM 108064	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Eubacteriaceae;g_Eubacterium;s_eligens	TSAB	Mega
115	<i>Eubacterium limosum</i> ATCC 8486	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Eubacteriaceae;g_Eubacterium;s_limosum	TSAB	Mega
116	<i>Eubacterium limosum</i> DSM 107592	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Eubacteriaceae;g_Eubacterium;s_limosum	TSAB	RCM
117	<i>Eubacterium maltosivora</i>	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Eubacteriaceae;g_Eubacterium;s_maltosivorans	TSAB	RCM

FIG. 23 (CONT.)

	ns DSM 105863			
118	<i>Eubacterium ramulus</i> DSM 15684	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Eubacteriaceae;g_Eubacterium;s_ramulus	TSAB	Mega
119	<i>Eubacterium ramulus</i> DSM 16296	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Eubacteriaceae;g_Eubacterium;s_ramulus	BHIB	Mega
120	<i>Eubacterium rectale</i> ATCC 33656 (DSM 17629)	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Lachnospiraceae;g_Eubacterium;s_rectale	TSAB	Mega
121	<i>Eubacterium rectale</i> DSM 17629	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Lachnospiraceae;g_unclassified Lachnospiraceae;s_rectale	TSAB	CMM
122	<i>Faecalicaten a contorta</i> S122	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Eubacteriaceae;g_Eubacterium;unclassified Eubacterium	TSAB	Mega
123	<i>Eubacterium sp.</i> S123	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Eubacteriaceae;g_Eubacterium;unclassified Eubacterium	TSAB	Mega
124	<i>Eubacterium ventriosum</i> ATCC 27560 (DSM 3988)	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Eubacteriaceae;g_Eubacterium;s_ventriosum	TSAB	Mega
125	<i>Lachnoclostr idium sp.</i> DSM 107473	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Lachnospiraceae;g_Lachnoclostridium;s_unclassified Lachnoclostridium	BHIB	RCM
126	<i>Lachnospira ceae</i> 5_1_57FAA	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Lachnospiraceae;g_unclassified Lachnospiraceae	TSAB	Mega
127	<i>Lachnospira ceae</i> HM- 150	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Lachnospiraceae;g_unclassified Lachnospiraceae	TSAB	CMM
128	<i>Lachnospira ceae</i> HM-7	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Lachnospiraceae;g_unclassified Lachnospiraceae	TSAB	Mega
129	<i>Lachnospira ceae</i> sp. Strain 1_4_56FAA	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Lachnospiraceae;g_unclassified Lachnospiraceae	TSAB	Mega
130	<i>Lachnospira ceae</i> sp. Strain 2_1_58FAA	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Lachnospiraceae;g_unclassified Lachnospiraceae	TSAB	Mega
131	<i>Roseburia faecis</i> DSM 16840	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f _Lachnospiraceae;g_Roseburia;s_faecis	TSAB	RCM

FIG. 23 (CONT.)

132	<i>Roseburia hominis</i> DSM 16839	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia;s_hominis	TSAB	Mega
133	<i>Roseburia intestinalis</i> DSM 14610	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia;s_intestinalis	TSAB	CMM
134	<i>Roseburia intestinalis</i> L1-82	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia;s_intestinalis	BHIB	Mega
135	<i>Roseburia inulinivorans</i> DSM 108070	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia;s_inulinivorans	TSAB	RCM
136	<i>Roseburia inulinivorans</i> DSM 16841	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia;s_inulinivorans	TSAB	CMM
137	<i>Roseburia</i> sp. DSM 102148	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia;unclassified Roseburia	TSAB	RCM
138	<i>Roseburia</i> sp. DSM 102150	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Roseburia;unclassified Roseburia	TSAB	Mega
139	<i>Ruminococcus gnavus</i> DSM 108212	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Mediterraneibacter;s_gnavus	TSAB	Mega
140	<i>Ruminococcus lactaris</i> ATCC 29176	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Mediterraneibacter;s_lactaris	TSAB	Mega
141	<i>Ruminococcus torques</i> ATCC 27756	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Mediterraneibacter;s_torques	TSAB	Mega
	Gram- other Prevotella			
142	<i>Prevotella marshallii</i> DSM 16973	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_marshallii	TSAB	CMM
143	<i>Prevotella stercoraria</i> DSM 18206	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_stercoraria	TSAB	CMM
144	<i>Prevotella pallens</i> DSM 18710	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_pallens	TSAB	CMM
145	<i>Prevotella corporis</i> DSM 18810	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_corporis	TSAB	CMM
146	<i>Prevotella nanceiensis</i> DSM 19126	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_nanceiensis	TSAB	Mega

FIG. 23 (CONT.)

147	<i>Prevotella histicola</i> DSM 21553	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_histicola	TSAB	Mega
148	<i>Prevotella xylaniphila</i> DSM 19681	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Paraprevotella;s_xylaniphila	TSAB	Mega
149	<i>Paraprevotella clara</i> DSM 19731	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Paraprevotella;s_clara	TSAB	Mega
150	<i>Prevotella saccharolytica</i> DSM 22473	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_saccharolytica	TSAB	CMM
151	<i>Prevotella melaninogenica</i> DSM 26980	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_melaninogenica	TSAB	Mega
152	<i>Prevotella histicola</i> DSM 26979	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_histicola	TSAB	Mega
153	<i>Prevotella jejuni</i> DSM 26988	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_jejuni	TSAB	CMM
154	<i>Prevotella phocaeensis</i> DSM 103364	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_phocaeensis	TSAB	Mega
155	<i>Prevotella rara</i> DSM 105141	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_rara	TSAB	CMM
156	<i>Prevotella copri</i> DSM 108494	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_copri	TSAB	CMM
157	<i>Prevotella</i> sp. DSM 29996	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_unclassified Prevotella	TSAB	Mega
158	<i>Prevotella</i> sp. DSM 109851	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Prevotellaceae;g_Prevotella;s_unclassified Prevotella	TSAB	CMM
	Other Phyla			
159	<i>Bifidobacterium adolescentis</i> L2-32	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae;g_Bifidobacterium;s_adolescentis	TSAB	Mega
160	<i>Bifidobacterium catenulatum</i> DSM 16992	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae;g_Bifidobacterium;s_catenulatum	BHIB	Mega

FIG. 23 (CONT.)

	(ATCC 27539)			
161	<i>Bifidobacterium longum infantis</i> ATCC 55813	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae;g_Bifidobacterium;s_longum	TSAB	Mega
162	<i>Bifidobacterium pseudocatenulatum</i> DSM 20438	k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae;g_Bifidobacterium;s_pseudocatenulatum	BHIB	Mega
163	<i>Enterococcus faecalis</i> CJG3	k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Enterococcaceae;g_Enterococcus;s_faecalis	LB	LB
164	<i>Enterococcus faecium</i> CJG2	k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Enterococcaceae;g_Enterococcus;s_faecium	LB	LB
165	<i>Escherichia fergusonii</i> CJG1	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacteriales;f_Enterobacteriaceae;g_Escherichia;s_fergusonii	LB	LB
166	<i>Achromobacter</i> spp (1)	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Achromobacter	LB	LB
167	<i>Bordetella</i> spp (1)	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Bordetella	LB	LB
168	<i>Pseudomonas</i> spp (1)	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomonadales;f_Pseudomonadaceae;g_Pseudomonas	LB	LB
169	<i>Desulfovibrio diazotrophica</i> NCIMB 15228	k_Bacteria;p_Proteobacteria;c_Delta/epsilonproteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Desulfovibrio;s_diazotrophica	Desulfovibrio (postgate) medium+1.5% agar	Desulfovibrio (postgate) medium
170	<i>Desulfovibrio intestinalis</i> DSM 11275	k_Bacteria;p_Proteobacteria;c_Delta/epsilonproteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Desulfovibrio;s_intestinalis	Desulfovibrio (postgate) medium+1.5% agar	Desulfovibrio (postgate) medium
171	<i>Desulfovibrio piger</i> DSM 106036	k_Bacteria;p_Proteobacteria;c_Delta/epsilonproteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Desulfovibrio;s_piger	Desulfovibrio (postgate) medium+1.5% agar	Desulfovibrio (postgate) medium

FIG. 23 (CONT.)

172	<i>Desulfovibrio porci</i> JCM 34370	k_Bacteria;p_Proteobacteria;c_Delta/epsilon proteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Desulfovibrio;s_porci	Desulfovibrio (postgate) medium+1.5% agar	Desulfovibrio (postgate) medium
173	<i>Desulfovibrio piger</i> VPI 11112	k_Bacteria;p_Proteobacteria;c_Delta/epsilon proteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Desulfovibrio;s_piger	Desulfovibrio (postgate) medium+1.5% agar	Desulfovibrio (postgate) medium
174	<i>Desulfovibrio termitidis</i> DSM 11274	k_Bacteria;p_Proteobacteria;c_Delta/epsilon proteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Cupidesulfovibrio;s_termitidis	Desulfovibrio (postgate) medium+1.5% agar	Desulfovibrio (postgate) medium
175	<i>Desulfovibrio termitidis</i> DSM 5308	k_Bacteria;p_Proteobacteria;c_Delta/epsilon proteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Cupidesulfovibrio;s_termitidis	Desulfovibrio (postgate) medium+1.5% agar	Desulfovibrio (postgate) medium
176	<i>Desulfovibrio vulgaris</i> DSM 15375	k_Bacteria;p_Proteobacteria;c_Delta/epsilon proteobacteria;o_Desulfovibrionales;f_Desulfovibrionaceae;g_Desulfovibrio;s_vulgaris	Desulfovibrio (postgate) medium+1.5% agar	Desulfovibrio (postgate) medium
177	<i>Fusobacterium gastrovuis</i> CDW1_DSM 101753	k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Fusobacteriaceae;g_Fusobacterium;s_gastrovuis	CBAB	CB
178	<i>Fusobacterium mortiferum</i> ATCC 25751	k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Fusobacteriaceae;g_Fusobacterium;s_mortiferum	CBAB	CB
179	<i>Fusobacterium mortiferum</i> FSA-380-WT-	k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Fusobacteriaceae;g_Fusobacterium;s_mortiferum	CBAB	CB

FIG. 23 (CONT.)

	2B_DSM 108838			
180	<i>Fusobacterium nucleatum</i> ATCC 10953	k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Fusobacteriaceae;g_Fusobacterium;s_nucleatum	CBAB	PYG
181	<i>Fusobacterium nucleatum</i> ATCC 23726	k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Fusobacteriaceae;g_Fusobacterium;s_nucleatum	CBAB	CB
182	<i>Fusobacterium nucleatum</i> subsp. <i>Animalis</i> JCM 11025	k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Fusobacteriaceae;g_Fusobacterium;s_nucleatum	CBAB	CMM
183	<i>Fusobacterium perfoetens</i> FSA-380-WT-3A_DSM 105865	k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Fusobacteriaceae;g_Fusobacterium;s_perfoetens	CBAB	CB
184	<i>Fusobacterium</i> sp. DSM 13120	k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Fusobacteriaceae;g_Fusobacterium;	CBAB	CB
185	<i>Klebsiella oxytoca</i> BAS-10_DSM 29614	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacterales;f_Enterobacteriaceae;g_Klebsiella;s_oxytoca	LB	LB
186	<i>Klebsiella oxytoca</i> DSM 24121	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacterales;f_Enterobacteriaceae;g_Klebsiella;s_oxytoca	LB	LB
187	<i>Klebsiella oxytoca</i> DSM 5175	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacterales;f_Enterobacteriaceae;g_Klebsiella;s_oxytoca	LB	LB
188	<i>Klebsiella oxytoca</i> G6 PT29	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacterales;f_Enterobacteriaceae;g_Klebsiella;s_oxytoca	LB	LB
189	<i>Klebsiella oxytoca</i> M5a1_DSM 7342	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacterales;f_Enterobacteriaceae;g_Klebsiella;s_oxytoca	LB	LB
190	<i>Klebsiella oxytoca</i> NCIMB 12819	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Enterobacterales;f_Enterobacteriaceae;g_Klebsiella;s_oxytoca	LB	LB

FIG. 23 (CONT.)

191	<i>Proteus faecis</i> TJ1636_DS M 106180	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;s_ faecis	LB	LB
192	<i>Proteus hauseri</i> ATCC 700826	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;s_ hauseri	LB	LB
193	<i>Proteus hauseri</i> NCIB 4175	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;s_ hauseri	LB	LB
194	<i>Proteus mirabilis</i> WS 10039_DSM 30116	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;s_ mirabilis	LB	LB
195	<i>Proteus mirabilis</i> ATCC 35659	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;s_ mirabilis	LB	LB
196	<i>Proteus mirabilis</i> CCUG 34293_DSM 18254	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;s_ mirabilis	LB	LB
197	<i>Proteus mirabilis</i> CD1- T1542_DSM 28629	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;s_ mirabilis	LB	LB
198	<i>Proteus sp.</i> WS 10284_DSM 30119	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;	LB	LB
199	<i>Proteus vulgaris</i> DSM 103498	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;s_ vulgaris	LB	LB
200	<i>Proteus vulgaris</i> DSM 3265	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteri a;o_Enterobacterales;f_Morganellaceae;g_Proteus;s_ vulgaris	LB	LB
201	<i>Corynebacte rium spp.</i>	k_Bacteria;p_Actinobacteria;c_Actinomycetia;o_Cory nebacterales;f_Corynebacteriaceae;g_Corynebacteri um;	TSAB	Mega

FIG. 24

Strain No.	Strain Name	Method	Conjugation Donor	Conjugative plasmid	Aerobic/Anaerobic conjugation/transformation	Conjugation time	Antibiotics for inhibiting <i>E. coli</i>	Antibiotics for marker selection
	<i>Bacteroides</i>							
2	<i>Bacteroides cellulosilyticus</i> DSM 108229	Conjugation (filter mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
3	<i>Bacteroides cellulosilyticus</i> DSM 14838	Conjugation (filter mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
7	<i>Bacteroides dorei</i> 5_1_36/D4 (HM 29)	Conjugation (filter mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
8	<i>Bacteroides dorei</i> DSM 17855	Conjugation (filter mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
9	<i>Bacteroides eggerthii</i> DSM 20697	Conjugation (filter mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
11	<i>Bacteroides finegoldii</i> DSM 108067	Conjugation (filter mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
12	<i>Bacteroides finegoldii</i> DSM 17565	Conjugation (filter mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)

FIG. 24 (CONT.)

14	<i>Bacteroides fragilis</i> 3_1_12	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
17	<i>Bacteroides intestinalis</i> DSM 17393	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
19	<i>Bacteroides nordii</i> DSM 18764	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
20	<i>Bacteroides oleiciplenus</i> DSM 22535	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
21	<i>Bacteroides ovatus</i> ATCC 8483	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
24	<i>Bacteroides salyersiae</i> DSM 18765	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
25	<i>Bacteroides</i> sp. 1_1_6	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
27	<i>Bacteroides</i> sp. 2_1_22	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
28	<i>Bacteroides</i> sp. 9_1_42FAA	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
31	<i>Bacteroides thetaiotaomicron</i> VPI-5482	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
32	<i>Bacteroides uniformis</i> ATCC 8492	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
33	<i>Bacteroides vulgatus</i> ATCC 8482	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)

FIG. 24 (CONT.)

34	<i>Bacteroides xylanisolvens</i> DSMZ 18836	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
35	<i>Bacteroides</i> sp. D20	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
37	<i>Bacteroides</i> sp. 1_1_30	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
41	<i>Parabacteroides faecis</i> DSM 102983	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
44	<i>Parabacteroides merdae</i> ATCC 43184 (DSMZ 19495)	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Aerobic (plating anaerobically)	24 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
46	<i>Prevotella bivia</i> DSM 20514	Conjugation (filter mating)	<i>E.coli</i> S17	R6K/R P4	Anaerobic	48 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
143	<i>Prevotella stercorea</i> DSM 18206	Conjugation (dot mating)	<i>E.coli</i> S17	R6K/R P4	Anaerobic	48 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
145	<i>Prevotella corporis</i> DSM 18810	Conjugation (dot mating)	<i>E.coli</i> S17	R6K/R P4	Anaerobic	48 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
147	<i>Prevotella histicola</i> DSM 21553	Conjugation (dot mating)	<i>E.coli</i> S17	R6K/R P4	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (9 µg/mL)
149	<i>Paraprevotella clara</i> DSM 19731	Conjugation (dot mating)	<i>E.coli</i> S17	R6K/R P4	Anaerobic	48 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
152	<i>Prevotella histicola</i> DSM 26979	Conjugation (dot mating)	<i>E.coli</i> S17	R6K/R P4	Anaerobic	48 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
158	<i>Prevotella</i> sp. DSM 109851	Conjugation (dot mating)	<i>E.coli</i> S17	R6K/R P4	Anaerobic	48 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (15 µg/mL)
	Firmicutes/Clostridia							

FIG. 24 (CONT.)

50	<i>Blautia faecis</i> DSM 27629	Conjugatio n (dot mating)	<i>E. coli</i> Express_ R702	R702	Anaerobi c	72 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (1 µg/mL)
52	<i>Blautia hydrogenotrop hica</i> DSM 10507	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	48 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (15 µg/mL)
53	<i>Blautia hydrogenotrop hica</i> DSM 108220	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	72 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (3 µg/mL)
54	<i>Blautia luti</i> DSM 14534	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	48 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (15 µg/mL)
60	<i>Blautia sp.</i> DSM 102165	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	72 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (3 µg/mL)
62	<i>Blautia sp.</i> DSM 107275	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	48 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (15 µg/mL)
64	<i>Blautia sp.</i> DSM 108066	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	72 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (1 µg/mL)
65	<i>Blautia wexlerae</i> DSM 19850	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	48 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (15 µg/mL)
71	<i>Clostridium bartlettii</i> DSM 16795 (ATCC BAA-827)	Conjugatio n (dot mating)	<i>E. coli</i> Express_ R702	R702	Anaerobi c	48 hrs	Gentam icin (200 µg/mL)	Thiampheni col (1.5 µg/mL)
72	<i>Clostridium bolteae</i> ATCC BAA-613	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	48 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (15 µg/mL)
73	<i>Clostridium bolteae</i> CJG-3	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	48 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (15 µg/mL)
74	<i>Clostridium bolteae</i> DSM 29485	Conjugatio n (dot mating)	<i>E. coli</i> CA434	R702	Anaerobi c	48 hrs	D- cycloser ine (250 µg/mL)	Thiampheni col (15 µg/mL)
75	<i>Clostridium bolteae</i> HM-	Conjugatio n (dot mating)	<i>E. coli</i> HB101/p RK24	pRK2 4	Anaerobi c	48 hrs	Kanam ycin (200 µg/mL)	Thiampheni col (45 µg/mL)

FIG. 24 (CONT.)

	318 aka WAL-14578							
76	<i>Clostridium bolteae</i> strain 47	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
77	<i>Clostridium bolteae</i> strain 54	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
78	<i>Clostridium bolteae</i> strain 55	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
83	<i>Clostridium hathewayi</i> CJG-2	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
87	<i>Clostridium innocuum</i> 6_1_30	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
88	<i>Clostridium innocuum</i> DSM 22910	Conjugation (dot mating)	<i>E. coli</i> HB101/pRK24	pRK24	Anaerobic	48 hrs	Kanamycin (200 µg/mL)	Thiamphenicol (15 µg/mL)
92	<i>Clostridium paraputrificum</i> DSM 103826	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
93	<i>Clostridium polynesiense</i> DSM 27072	Electroporation (2.0 kV, 400 Ω, 0.625% Glycine + 0.4 M Sucrose, pre-methylate treatment of optimized plasmid)			Anaerobic			Thiamphenicol (3 µg/mL)
97	<i>Clostridium ramosum</i> DSM 24812	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	72 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (3 µg/mL)
98	<i>Clostridium saccharolyticum</i> WM1 DSMZ 2544 (ATCC 35040)	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)

FIG. 24 (CONT.)

99	<i>Clostridium scindens</i> ATCC 35704	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	72 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (3 µg/mL)
100	<i>Clostridium senegalense</i> DSM 25507	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
101	<i>Clostridium sp.</i> 7_2_43FAA	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
102	<i>Clostridium sp.</i> DSM 100503	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	72 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (3 µg/mL)
104	<i>Clostridium sp.</i> DSM 24810	Conjugation (dot mating)	<i>E. coli</i> HB101/pRK24	pRK24	Anaerobic	48 hrs	Kanamycin (200 µg/mL)	Thiamphenicol (15 µg/mL)
107	<i>Clostridium sporogenes</i> ATCC 15579	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
108	<i>Clostridium sporogenes</i> DSM 795	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
110	<i>Clostridium symbiosum</i> WAL-14163	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
115	<i>Eubacterium limosum</i> ATCC 8486	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	72 hrs	Gentamicin (200 µg/mL)	Thiamphenicol (9 µg/mL)
116	<i>Eubacterium limosum</i> DSM 107592	Conjugation (dot mating)	<i>E. coli</i> HB101/pRK24	pRK24	Anaerobic	48 hrs	Kanamycin (200 µg/mL)	Thiamphenicol (15 µg/mL)
117	<i>Eubacterium maltosivorans</i> DSM 105863	Conjugation (dot mating)	<i>E. coli</i> HB101/pRK24	pRK24	Anaerobic	48 hrs	Kanamycin (200 µg/mL)	Thiamphenicol (15 µg/mL)
118	<i>Eubacterium ramulus</i> DSM 15684	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
122	<i>Faecalicatena contorta</i> S122	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)

FIG. 24 (CONT.)

123	<i>Eubacterium</i> <i>sp.</i> S123	Conjugation (dot mating)	<i>E. coli</i> Express_R702	R702	Anaerobic	72 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (3 µg/mL)
136	<i>Roseburia inulinivorans</i> DSM 16841	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
	Other phyla							
165	<i>Escherichia fergusonii</i> CJG1	Chemical transformation			Aerobic			Chloramphenicol (25 µg/mL)
160	<i>Bifidobacterium catenulatum</i> DSM 16992 (ATCC 27539)	Conjugation (dot mating)	<i>E. coli</i> CA434	R702	Anaerobic	48 hrs	D-cycloserine (250 µg/mL)	Thiamphenicol (15 µg/mL)
163	<i>Enterococcus faecalis</i> CJG3	Conjugation (dot mating)	<i>E. coli</i> HB101/pRK24	pRK24	Aerobic	48 hrs	Kanamycin (200 µg/mL)	Thiamphenicol (15 µg/mL)
164	<i>Enterococcus faecium</i> CJG2	Conjugation (dot mating)	<i>E. coli</i> HB101/pRK24	pRK24	Aerobic	48 hrs	Kanamycin (200 µg/mL)	Thiamphenicol (15 µg/mL)
166	<i>Achromobacter</i> spp (1)	Conjugation (dot mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic	48 hrs	D-cycloserine (250 µg/mL)	Chloramphenicol (25 µg/mL)
167	<i>Bordetella</i> spp (1)	Conjugation (dot mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic	48 hrs	D-cycloserine (250 µg/mL)	Chloramphenicol (35 µg/mL)
168	<i>Pseudomonas</i> spp (1)	Conjugation (dot mating)	<i>E. coli</i> S17	R6K/RP4	Aerobic	48 hrs	D-cycloserine (250 µg/mL)	Carbenicillin (350 µg/mL)
169	<i>Desulfovibrio diazotrophica</i> NCIMB 15228	Conjugation (dot mating)	<i>E. coli</i> S17	R6K/RP4	Anaerobic	48 hrs	Kanamycin (30 µg/mL)	Chloramphenicol (6 µg/mL)
174	<i>Desulfovibrio termitidis</i> DSM 11274	Conjugation (dot mating)	<i>E. coli</i> S17	R6K/RP4	Anaerobic	48 hrs	Kanamycin (30 µg/mL)	Chloramphenicol (6 µg/mL)
176	<i>Desulfovibrio vulgaris</i> DSM 15375	Conjugation (dot mating)	<i>E. coli</i> S17	R6K/RP4	Anaerobic	48 hrs	Kanamycin (30 µg/mL)	Chloramphenicol (6 µg/mL)
177	<i>Fusobacterium gastrosuis</i> CDW1_DSM 101753	Electroporation (2.5 kV, 200 Ω)			Anaerobic			Thiamphenicol (3 µg/mL)
178	<i>Fusobacterium mortiferum</i> ATCC 25751	Electroporation (2.5 kV, 200 Ω)			Anaerobic			Thiamphenicol (3 µg/mL)

FIG. 24 (CONT.)

180	<i>Fusobacterium nucleatum</i> ATCC 10953	Eletroporation (2.5 kV, 200 Ω)			Anaerobic			Thiamphenicol (3 μ g/mL)
181	<i>Fusobacterium nucleatum</i> ATCC 23726	Eletroporation (2.5 kV, 200 Ω)			Anaerobic			Thiamphenicol (3 μ g/mL)
185	<i>klebsiella oxytoca</i> BAS-10_DSM 29614	Eletroporation (2.5 kV, 200 Ω)			Aerobic			Kanamycin (30 μ g/mL)
187	<i>klebsiella oxytoca</i> DSM 5175	Eletroporation (2.5 kV, 200 Ω)			Aerobic			Kanamycin (30 μ g/mL)
189	<i>klebsiella oxytoca</i> M5a1_DSM 7342	Eletroporation (2.5 kV, 200 Ω)			Aerobic			Kanamycin (30 μ g/mL)
190	<i>klebsiella oxytoca</i> NCIMB 12819	Eletroporation (2.5 kV, 200 Ω)			Aerobic			Kanamycin (30 μ g/mL)
192	<i>Proteus hauseri</i> ATCC 700826	Conjugation (dot mating)	<i>E.coli</i> S17	R6K/RP4	Aerobic	48 hrs	D-cycloserine (250 μ g/mL)	Chloramphenicol (20 μ g/mL)
193	<i>Proteus hauseri</i> NCIB 4175	Conjugation (dot mating)	<i>E.coli</i> S17	R6K/RP4	Aerobic	48 hrs	D-cycloserine (250 μ g/mL)	Chloramphenicol (20 μ g/mL)
195	<i>Proteus mirabilis</i> ATCC 35659	Eletroporation (2.5 kV, 200 Ω)			Aerobic			Chloramphenicol (15 μ g/mL)
201	<i>Corynebacterium</i> spp.	Eletroporation (2.0 kV, 400 Ω)			Anaerobic			Thiamphenicol (6 μ g/mL)

FIG. 25

Strain No.	Strain Name	Plasmids introduced to identify gene transfer method
	<i>Bacteroidetes</i>	
2	<i>Bacteroides cellulosilyticus</i> DSM 108229	pGM-NAC ₂ B
3	<i>Bacteroides cellulosilyticus</i> DSM 14838	pGM-NAC ₂ B
7	<i>Bacteroides dorei</i> 5_1_36/D4 (HM 29)	pGM-NAC ₂ B
8	<i>Bacteroides dorei</i> DSM 17855	pGM-NAC ₂ B
9	<i>Bacteroides eggerthii</i> DSM 20697	pGM-NAC ₂ B
11	<i>Bacteroides finegoldii</i> DSM 108067	pGM-NAC ₂ B
12	<i>Bacteroides finegoldii</i> DSM 17565	pGM-NAC ₂ B
14	<i>Bacteroides fragilis</i> 3_1_12	pGM-NAC ₂ B
17	<i>Bacteroides intestinalis</i> DSM 17393	pGM-NAC ₂ B
19	<i>Bacteroides nordii</i> DSM 18764	pGM-NAC ₂ B
20	<i>Bacteroides oleiciplenus</i> DSM 22535	pGM-NAC ₂ B
21	<i>Bacteroides ovatus</i> ATCC 8483	pGM-NAC ₂ B
24	<i>Bacteroides salyersiae</i> DSM 18765	pGM-NAC ₂ B
25	<i>Bacteroides</i> sp. 1_1_6	pGM-NAC ₂ B
27	<i>Bacteroides</i> sp. 2_1_22	pGM-NAC ₂ B
28	<i>Bacteroides</i> sp. 9_1_42FAA	pGM-NAC ₂ B
31	<i>Bacteroides thetaiotaomicron</i> VPI-5482	pGM-NAC ₂ B
32	<i>Bacteroides uniformis</i> ATCC 8492	pGM-NAC ₂ B
33	<i>Bacteroides vulgatus</i> ATCC 8482	pGM-NAC ₂ B
34	<i>Bacteroides xylanisolvens</i> DSMZ 18836	pGM-NAC ₂ B
35	<i>Bacteroides</i> sp. D20	pGM-NAC ₂ B
37	<i>Bacteroides</i> sp. 1_1_30	pGM-NAC ₂ B
41	<i>Parabacteroides faecis</i> DSM 102983	pGM-NAC ₂ B
44	<i>Parabacteroides merdae</i> ATCC 43184 (DSMZ 19495)	pGM-NAC ₂ B
46	<i>Prevotella bivia</i> DSM 20514	pGM-NAC ₂ P
143	<i>Prevotella stercorea</i> DSM 18206	pGM-NAC ₂ P
145	<i>Prevotella corporis</i> DSM 18810	pGM-NAC ₂ P
147	<i>Prevotella histicola</i> DSM 21553	pGM-NAC ₂ P
149	<i>Paraprevotella clara</i> DSM 19731	pGM-NAC ₂ P
152	<i>Prevotella histicola</i> DSM 26979	pGM-NAC ₂ P
158	<i>Prevotella</i> sp. DSM 109851	pGM-NAC ₂ P

FIG. 25 (CONT.)

	<i>Firmicutes/Clostridia</i>	
50	<i>Blautia faecis</i> DSM 27629	pGM-DBCM
52	<i>Blautia hydrogenotrophica</i> DSM 10507	pGM-ABCM, pGM-FBCM
53	<i>Blautia hydrogenotrophica</i> DSM 108220	pGM-FBCM
54	<i>Blautia luti</i> DSM 14534	pGM-ABCM, pGM-EBCM, pGM-FBCM
60	<i>Blautia sp.</i> DSM 102165	pGM-DBCM
62	<i>Blautia sp.</i> DSM 107275	pGM-ABCM, pGM-FBCM
64	<i>Blautia sp.</i> DSM 108066	pGM-DBCM
65	<i>Blautia wexlerae</i> DSM 19850	pGM-ABCM, pGM-EBCM, pGM-FBCM
71	<i>Clostridium bartlettii</i> DSM 16795 (ATCC BAA-827)	pGM-FBCM
72	<i>Clostridium bolteae</i> ATCC BAA-613	pGM-FBCM
73	<i>Clostridium bolteae</i> CJG-3	pGM-ABCM, pGM-EBCM, pGM-FBCM
74	<i>Clostridium bolteae</i> DSM 29485	pGM-ABCM, pGM-FBCM
75	<i>Clostridium bolteae</i> HM-318 aka WAL-14578	pGM-EBCM
76	<i>Clostridium bolteae</i> strain 47	pGM-FBCM
77	<i>Clostridium bolteae</i> strain 54	pGM-EBCM
78	<i>Clostridium bolteae</i> strain 55	pGM-EBCM
83	<i>Clostridium hathewayi</i> CJG-2	pGM-ABCM
87	<i>Clostridium innocuum</i> 6_1_30	pGM-ABCM
88	<i>Clostridium innocuum</i> DSM 22910	pGM-IBCM
92	<i>Clostridium paraputrificum</i> DSM 103826	pGM-ABCM, pGM-FBCM
93	<i>Clostridium polynesiense</i> DSM 27072	pGM-DBCM_seq-opt
97	<i>Clostridium ramosum</i> DSM 24812	pGM-GBCM
98	<i>Clostridium saccharolyticum</i> WM1 DSMZ 2544 (ATCC 35040)	pGM-DBCM
99	<i>Clostridium scindens</i> ATCC 35704	pGM-GBCM
100	<i>Clostridium senegalense</i> DSM 25507	pGM-ABCM
101	<i>Clostridium sp.</i> 7_2_43FAA	pGM-GBCM
102	<i>Clostridium sp.</i> DSM 100503	pGM-IBCM
104	<i>Clostridium sp.</i> DSM 24810	pGM-EBCM, pGM-HBCM, pGM-IBCM
107	<i>Clostridium sporogenes</i> ATCC 15579	pGM-ABCM, pGM-BBCM
108	<i>Clostridium sporogenes</i> DSM 795	pGM-ABCM, pGM-BBCM
110	<i>Clostridium symbiosum</i> WAL-14163	pGM-ABCM
115	<i>Eubacterium limosum</i> ATCC 8486	pGM-FBCM
116	<i>Eubacterium limosum</i> DSM 107592	pGM-FBCM, pGM-IBCM
117	<i>Eubacterium maltosivorans</i> DSM 105863	pGM-FBCM, pGM-IBCM
118	<i>Eubacterium ramulus</i> DSM 15684	pGM-FBCM
122	<i>Faecalicatena contorta</i> S122	pGM-ABCM, pGM-FBCM
123	<i>Eubacterium sp.</i> S123	pGM-ABCM

FIG. 25 (CONT.)

136	<i>Roseburia inulinivorans</i> DSM 16841	pGM-CBCM
	Other phyla	
165	<i>Escherichia fergusonii</i> CJG1	pGM-ABCM
160	<i>Bifidobacterium catenulatum</i> DSM 16992 (ATCC 27539)	pGM-HBCM
163	<i>Enterococcus faecalis</i> CJG3	pGM-EBCM
164	<i>Enterococcus faecium</i> CJG2	pGM-FBCM
166	<i>Achromobacter</i> spp (1)	pGM-NACO ₅
167	<i>Bordetella</i> spp (1)	pGM-NACO ₆
168	<i>Pseudomonas</i> spp (1)	pGM-NACO ₇
169	<i>Desulfovibrio diazotrophica</i> NCIMB 15228	pGM-NACO ₁
174	<i>Desulfovibrio termitidis</i> DSM 11274	pGM-NACO ₁
176	<i>Desulfovibrio vulgaris</i> DSM 15375	pGM-NACO ₁
177	<i>Fusobacterium gastrosuis</i> CDW1_DSM 101753	pGM-NACO ₂
178	<i>Fusobacterium mortiferum</i> ATCC 25751	pGM-NACO ₂
180	<i>Fusobacterium nucleatum</i> ATCC 10953	pGM-NACO ₂
181	<i>Fusobacterium nucleatum</i> ATCC 23726	pGM-NACO ₂
185	<i>klebsiella oxytoca</i> BAS-10_DSM 29614	pGM-NAKO ₃
187	<i>klebsiella oxytoca</i> DSM 5175	pGM-NAKO ₃
189	<i>klebsiella oxytoca</i> M5a1_DSM 7342	pGM-NAKO ₃
190	<i>klebsiella oxytoca</i> NCIMB 12819	pGM-NAKO ₃
192	<i>Proteus hauseri</i> ATCC 700826	pGM-NACO ₄
193	<i>Proteus hauseri</i> NCIB 4175	pGM-NACO ₄
195	<i>Proteus mirabilis</i> ATCC 35659	pGM-NACO ₄
201	<i>Corynebacterium</i> spp.	pGM-GBC ₁ M_seq-opt

FIG. 26

Strain No.	Strain Name	Target genes			
		on vector	on genome- <i>bcat</i>	on genome-16s targeting Group II intron (%)	on genome- <i>porA/croA/baiH</i>
50	<i>Blautia faecis</i> DSM 27629				
52	<i>Blautia hydrogenotrophica</i> DSM 10507	<i>lacZα</i>	<i>RUMHYD_RS05985</i>	100%	
53	<i>Blautia hydrogenotrophica</i> DSM 108220			63.63%	
54	<i>Blautia luti</i> DSM 14534	<i>lacZα</i>	<i>GKZ57_13035</i>	100%	
60	<i>Blautia</i> sp. DSM 102165				
62	<i>Blautia</i> sp. DSM 107275	<i>lacZα</i>	<i>C6W64_004620</i>	37.50%	
64	<i>Blautia</i> sp. DSM 108066				
65	<i>Blautia wexlerae</i> DSM 19850	<i>lacZα</i>		37.50%	
71	<i>Clostridium bartlettii</i> DSM 16795 (ATCC BAA-827)	<i>lacZα</i>			
72	<i>Clostridium bolteae</i> ATCC BAA-613	<i>lacZα</i>	<i>CGC65_03110</i>	100%	
73	<i>Clostridium bolteae</i> CJG-3	<i>lacZα</i>			
74	<i>Clostridium bolteae</i> DSM 29485	<i>lacZα</i>	<i>BCAA aminotransferase</i>		
75	<i>Clostridium bolteae</i> HM-318 aka WAL-14578				
76	<i>Clostridium bolteae</i> strain 47	<i>lacZα</i>		12.50%	
77	<i>Clostridium bolteae</i> strain 54	<i>lacZα</i>		12.50%	
78	<i>Clostridium bolteae</i> strain 55	<i>lacZα</i>			
83	<i>Clostridium hathewayi</i> CJG-2				

FIG. 26 (CONT.)

87	<i>Clostridium innocuum</i> 6_1_30	<i>lacZα</i>		37.50%	
88	<i>Clostridium innocuum</i> DSM 22910	<i>lacZα</i>		25.00%	
92	<i>Clostridium paraputrificum</i> DSM 103826	<i>lacZα</i>		37.50%	
93	<i>Clostridium polynesiense</i> DSM 27072				
97	<i>Clostridium ramosum</i> DSM 24812				
98	<i>Clostridium saccharolyticum</i> WM1 DSMZ 2544 (ATCC 35040)	<i>lacZα</i>	CLOSA_RS11625		
99	<i>Clostridium scindens</i> ATCC 35704				
100	<i>Clostridium senegalense</i> DSM 25507	<i>lacZα</i>	BCAA aminotransferase	100%	
101	<i>Clostridium</i> sp. 7_2_43FAA				
102	<i>Clostridium</i> sp. DSM 100503				
104	<i>Clostridium</i> sp. DSM 24810	<i>lacZα</i>	HMPREF0981_04422		
107	<i>Clostridium sporogenes</i> ATCC 15579	<i>lacZα</i>		100%	<i>clospo_00147 (porA)</i>
108	<i>Clostridium sporogenes</i> DSM 795	<i>lacZα</i>		75%	
110	<i>Clostridium symbiosum</i> WAL-14163	<i>lacZα</i>	HMPREF9474_RS22760		HMPREF9474_RS17470 (<i>croA</i>)
115	<i>Eubacterium limosum</i> ATCC 8486			45.80%	B2M23_RS01185 (<i>croA</i>)
116	<i>Eubacterium limosum</i> DSM 107592	<i>lacZα</i>			
117	<i>Eubacterium maltosivorans</i> DSM 105863	<i>lacZα</i>	CPZ25_012335		CPZ25_015200 (<i>croA</i>)

FIG. 26 (CONT.)

118	<i>Eubacterium ramulus</i> DSM 15684	<i>lacZα</i>			
122	<i>Faecalicatena contorta</i> S122	<i>lacZα</i>		100%	<i>baiH</i>
123	<i>Eubacterium sp.</i> S123				
136	<i>Roseburia inulinivorans</i> DSM 16841	<i>lacZα</i>	ROSEINA2194_RS1469 0		
160	<i>Bifidobacterium catenulatum</i> DSM 16992 (ATCC 27539)	<i>lacZα</i>	BIFCAT_00922		
164	<i>Enterococcus faecium</i> CJG2	<i>lacZα</i>			
165	<i>Escherichia fergusonii</i> CJG1	<i>lacZα</i>			

FIG. 27

Strain No.	Strain Name	Target gene	
		on genome-16s (%)	on genome-mmdA
2	<i>Bacteroides cellulosilyticus</i> DSM 108229	75%	
3	<i>Bacteroides cellulosilyticus</i> DSM 14838	100%	
7	<i>Bacteroides dorei</i> 5_1_36/D4 (HM 29)	62.50%	
8	<i>Bacteroides dorei</i> DSM 17855	50%	
9	<i>Bacteroides eggerthii</i> DSM 20697	100%	
11	<i>Bacteroides finegoldii</i> DSM 108067	75%	
12	<i>Bacteroides finegoldii</i> DSM 17565	50%	
14	<i>Bacteroides fragilis</i> 3_1_12	75%	
17	<i>Bacteroides intestinalis</i> DSM 17393	100%	
19	<i>Bacteroides nordii</i> DSM 18764	100%	
20	<i>Bacteroides oleiciplenus</i> DSM 22535	50%	
21	<i>Bacteroides ovatus</i> ATCC 8483	50%	
24	<i>Bacteroides salyersiae</i> DSM 18765	66.60%	
25	<i>Bacteroides</i> sp. 1_1_6	100%	BSIG_3264 (mmdA)
27	<i>Bacteroides</i> sp. 2_1_22	87.50%	HMPREF0102_01948 (mmdA)
28	<i>Bacteroides</i> sp. 9_1_42FAA	100%	
31	<i>Bacteroides thetaiotaomicron</i> VPI-5482	100%	BT_2090 (mmdA)
32	<i>Bacteroides uniformis</i> ATCC 8492	87.50%	
33	<i>Bacteroides vulgatus</i> ATCC 8482	100%	
34	<i>Bacteroides xylanisolvens</i> DSMZ 18836	100%	
35	<i>Bacteroides</i> sp. D20	37.50%	
37	<i>Bacteroides</i> sp. 1_1_30	50%	
41	<i>Parabacteroides faecis</i> DSM 102983	100%	
44	<i>Parabacteroides merdae</i> ATCC 43184 (DSMZ 19495)	87.50%	
46	<i>Prevotella bivia</i> DSM 20514	75%	
143	<i>Prevotella stercorea</i> DSM 18206	100%	
145	<i>Prevotella corporis</i> DSM 18810	87.50%	
147	<i>Prevotella histicola</i> DSM 21553	25%	
149	<i>Paraprevotella clara</i> DSM 19731	37.50%	
152	<i>Prevotella histicola</i> DSM 26979	25%	
158	<i>Prevotella</i> sp. DSM 109851	12.50%	
166	<i>Achromobacter</i> spp (1)	75%	
169	<i>Desulfovibrio diazotrophica</i> NCIMB 15228	87.50%	
174	<i>Desulfovibrio termitidis</i> DSM 11274	75%	
176	<i>Desulfovibrio vulgaris</i> DSM 15375	100%	
178	<i>Fusobacterium mortiferum</i> ATCC 25751	75%	
181	<i>Fusobacterium nucleatum</i> ATCC 23726	50%	

FIG. 27 (CONT.)

185	<i>klebsiella oxytoca</i> BAS-10_DSM 29614	75%	
187	<i>klebsiella oxytoca</i> DSM 5175	87.50%	
189	<i>klebsiella oxytoca</i> M5a1_DSM 7342	75%	
190	<i>klebsiella oxytoca</i> NCIMB 12819	100%	
192	<i>Proteus hauseri</i> ATCC 700826	75%	
193	<i>Proteus hauseri</i> NCIB 4175	87.50%	
195	<i>Proteus mirabilis</i> ATCC 35659	100%	

FIG. 28

Vector	Genotype
pGM-NAEM-001	targeting <i>16s</i> in <i>Bo</i> (S21)
pGM-NAEM-002	targeting <i>16s</i> in <i>Bt</i> (S31)
pGM-NAC ₂ M-003	targeting <i>BSIG_3264 (mmdA)</i> in S25
pGM-NAC ₂ M-004	targeting <i>HMPREF0102_01948 (mmdA)</i> in S27
pGM-NAC ₂ M-005	targeting <i>BT_2090 (mmdA)</i> in S31
pGM-ABCD-006	targeting <i>clospo_00147 (porA)</i> in S107
pGM-FBCD-007	targeting <i>RUMHYD_RS05985 (bcat)</i> in S52
pGM-ABCD-008	targeting <i>GKZ57_13035 (bcat)</i> in S54
pGM-ABCD-009	targeting <i>C6W64_004620 (bcat)</i> in S62
pGM-FBCD-010	targeting <i>CGC65_03110 (bcat)</i> in S72
pGM-FBCD-011	targeting <i>BCAA aminotransferase (bcat)</i> in S74
pGM-DBCD-012	targeting <i>CLOSA_RS11625 (bcat)</i> in S98
pGM-ABCD-013	targeting <i>BCAA aminotransferase (bcat)</i> in S100
pGM-EBCD-014	targeting <i>HMPREF0981_04422 (bcat)</i> in S104
pGM-ABCD-015	targeting <i>HMPREF9474_RS22760 (bcat)</i> in S110
pGM-FBCD-016	targeting <i>CPZ25_012335 (bcat)</i> in S117
pGM-CBCD-017	targeting <i>ROSEINA2194_RS14690 (bcat)</i> in S136
pGM-HBCD-018	targeting <i>BIFCAT_00922 (bcat)</i> in S143
pGM-ABCD-019	targeting <i>HMPREF9474_RS17470 (croA)</i> in S110
pGM-FBCD-020	targeting <i>CPZ25_015200 (croA)</i> in S117
pGM-FCAR-002	targeting <i>baiH</i> in S122
pGM-FCAR-003	targeting <i>croA</i> in S115

FIG. 29

Bacterial strains	Gene mutation/regulation	Genotype
<i>E. coli</i>		
S17 pGM-NAEM-001	None	Carrying pGM-NAEM-001
S17 pGM-NAEM-002	None	Carrying pGM-NAEM-002
S17 pGM-NAC ₂ B	None	Carrying pGM-NAC ₂ B
S17 pGM-NAC ₂ P	None	Carrying pGM-NAC ₂ P
S17 pGM-NACO ₁	None	Carrying pGM-NACO ₁
S17 pGM-NACO ₂	None	Carrying pGM-NACO ₂
S17 pGM-NAKO ₃	None	Carrying pGM-NAKO ₃
S17 pGM-NACO ₄	None	Carrying pGM-NACO ₄
S17 pGM-NACO ₅	None	Carrying pGM-NACO ₅
S17 pGM-NACO ₆	None	Carrying pGM-NACO ₆
S17 pGM-NACO ₇	None	Carrying pGM-NACO ₇
S17 pGM-NAC ₂ M-003	None	Carrying pGM-NAC ₂ M-003
S17 pGM-NAC ₂ M-004	None	Carrying pGM-NAC ₂ M-004
S17 pGM-NAC ₂ M-005	None	Carrying pGM-NAC ₂ M-005
CA434 pGM-ABCM	None	Carrying pGM-ABCM
CA434 pGM-BBCM	None	Carrying pGM-BBCM
CA434 pGM-CBCM	None	Carrying pGM-CBCM
CA434 pGM-DBCM	None	Carrying pGM-DBCM
CA434 pGM-EBCM	None	Carrying pGM-EBCM
CA434 pGM-FBCM	None	Carrying pGM-FBCM
CA434 pGM-GBCM	None	Carrying pGM-GBCM
CA434 pGM-HBCM	None	Carrying pGM-HBCM
CA434 pGM-IBCM	None	Carrying pGM-IBCM
CA434 pGM-NBCC	None	Carrying pGM-NBCC
101/24 pGM-ABCM	None	Carrying pGM-ABCM
101/24 pGM-BBCM	None	Carrying pGM-BBCM
101/24 pGM-CBCM	None	Carrying pGM-CBCM
101/24 pGM-DBCM	None	Carrying pGM-DBCM
101/24 pGM-EBCM	None	Carrying pGM-EBCM
101/24 pGM-FBCM	None	Carrying pGM-FBCM
101/24 pGM-GBCM	None	Carrying pGM-GBCM
101/24 pGM-HBCM	None	Carrying pGM-HBCM
101/24 pGM-IBCM	None	Carrying pGM-IBCM
101/24 pGM-NBCC	None	Carrying pGM-NBCC
CA434 pGM-ABCD	None	Carrying pGM-ABCD
CA434 pGM-BBCD	None	Carrying pGM-BBCD
CA434 pGM-CBCD	None	Carrying pGM-CBCD
CA434 pGM-DBCD	None	Carrying pGM-DBCD
CA434 pGM-EBCD	None	Carrying pGM-EBCD

FIG. 29 (CONT.)

CA434 pGM-FBCD	None	Carrying pGM-FBCD
CA434 pGM-GBCD	None	Carrying pGM-GBCD
CA434 pGM-HBCD	None	Carrying pGM-HBCD
CA434 pGM-IBCD	None	Carrying pGM-IBCD
101/24 pGM-ABCD	None	Carrying pGM-ABCD
101/24 pGM-BBCD	None	Carrying pGM-BBCD
101/24 pGM-CBCD	None	Carrying pGM-CBCD
101/24 pGM-DBCD	None	Carrying pGM-DBCD
101/24 pGM-EBCD	None	Carrying pGM-EBCD
101/24 pGM-FBCD	None	Carrying pGM-FBCD
101/24 pGM-GBCD	None	Carrying pGM-GBCD
101/24 pGM-HBCD	None	Carrying pGM-HBCD
101/24 pGM-IBCD	None	Carrying pGM-IBCD
CA434 pGM-ABCL	None	Carrying pGM-ABCL
CA434 pGM-BBCL	None	Carrying pGM-BBCL
CA434 pGM-CBCL	None	Carrying pGM-CBCL
CA434 pGM-DBCL	None	Carrying pGM-DBCL
CA434 pGM-EBCL	None	Carrying pGM-EBCL
CA434 pGM-FBCL	None	Carrying pGM-FBCL
CA434 pGM-GBCL	None	Carrying pGM-GBCL
CA434 pGM-HBCL	None	Carrying pGM-HBCL
CA434 pGM-IBCL	None	Carrying pGM-IBCL
101/24 pGM-ABCL	None	Carrying pGM-ABCL
101/24 pGM-BBCL	None	Carrying pGM-BBCL
101/24 pGM-CBCL	None	Carrying pGM-CBCL
101/24 pGM-DBCL	None	Carrying pGM-DBCL
101/24 pGM-EBCL	None	Carrying pGM-EBCL
101/24 pGM-FBCL	None	Carrying pGM-FBCL
101/24 pGM-GBCL	None	Carrying pGM-GBCL
101/24 pGM-HBCL	None	Carrying pGM-HBCL
101/24 pGM-IBCL	None	Carrying pGM-IBCL
CA434 pGM-ABCL ₁	None	Carrying pGM-ABCL ₁
CA434 pGM-CBCL ₁	None	Carrying pGM-CBCL ₁
CA434 pGM-EBCL ₁	None	Carrying pGM-EBCL ₁
CA434 pGM-FBCL ₁	None	Carrying pGM-FBCL ₁
CA434 pGM-HBCL ₁	None	Carrying pGM-HBCL ₁
101/24 pGM-ABCL ₁	None	Carrying pGM-ABCL ₁
101/24 pGM-CBCL ₁	None	Carrying pGM-CBCL ₁
101/24 pGM-EBCL ₁	None	Carrying pGM-EBCL ₁
101/24 pGM-FBCL ₁	None	Carrying pGM-FBCL ₁

FIG. 29 (CONT.)

101/24 pGM-HBCL ₁	None	Carrying pGM-HBCL ₁
CA434 pGM-ABC ₁ M_seq-opt	None	Carrying pGM-ABC ₁ M_seq-opt
CA434 pGM-CBC ₁ M_seq-opt	None	Carrying pGM-CBC ₁ M_seq-opt
CA434 pGM-DBC ₁ M_seq-opt	None	Carrying pGM-DBC ₁ M_seq-opt
CA434 pGM-EBC ₁ M_seq-opt	None	Carrying pGM-EBC ₁ M_seq-opt
CA434 pGM-FBC ₁ M_seq-opt	None	Carrying pGM-FBC ₁ M_seq-opt
CA434 pGM-GBC ₁ M_seq-opt	None	Carrying pGM-GBC ₁ M_seq-opt
CA434 pGM-HBC ₁ M_seq-opt	None	Carrying pGM-HBC ₁ M_seq-opt
101/24 pGM-ABC ₁ M_seq-opt	None	Carrying pGM-ABC ₁ M_seq-opt
101/24 pGM-CBC ₁ M_seq-opt	None	Carrying pGM-CBC ₁ M_seq-opt
101/24 pGM-DBC ₁ M_seq-opt	None	Carrying pGM-DBC ₁ M_seq-opt
101/24 pGM-EBC ₁ M_seq-opt	None	Carrying pGM-EBC ₁ M_seq-opt
101/24 pGM-FBC ₁ M_seq-opt	None	Carrying pGM-FBC ₁ M_seq-opt
101/24 pGM-GBC ₁ M_seq-opt	None	Carrying pGM-GBC ₁ M_seq-opt
101/24 pGM-HBC ₁ M_seq-opt	None	Carrying pGM-HBC ₁ M_seq-opt
CA434 pGM-ABCF	None	Carrying pGM-ABCF
CA434 pGM-ABCG	None	Carrying pGM-ABCG
CA434 pGM-BBCF	None	Carrying pGM-BBCF
CA434 pGM-CBCF	None	Carrying pGM-CBCF
CA434 pGM-DBCF	None	Carrying pGM-DBCF
CA434 pGM-EBCF	None	Carrying pGM-EBCF
CA434 pGM-FBCF	None	Carrying pGM-FBCF
CA434 pGM-GBCF	None	Carrying pGM-GBCF
CA434 pGM-HBCF	None	Carrying pGM-HBCF
CA434 pGM-IBCF	None	Carrying pGM-IBCF
101/24 pGM-EBCF	None	Carrying pGM-EBCF
101/24 pGM-FBCF	None	Carrying pGM-FBCF
101/24 pGM-GBCF	None	Carrying pGM-GBCF
101/24 pGM-HBCF	None	Carrying pGM-HBCF
101/24 pGM-IBCF	None	Carrying pGM-IBCF
CA434 pGM-ABCD-006	None	Carrying pGM-BBCD-006
CA434 pGM-FBCD-007	None	Carrying pGM-FBCD-007

FIG. 29 (CONT.)

CA434 pGM-ABCD-008	None	Carrying pGM-ABCD-008
CA434 pGM-ABCD-009	None	Carrying pGM-ABCD-009
CA434 pGM-FBCD-010	None	Carrying pGM-FBCD-010
CA434 pGM-FBCD-011	None	Carrying pGM-FBCD-011
CA434 pGM-DBCD-012	None	Carrying pGM-DBCD-012
CA434 pGM-ABCD-013	None	Carrying pGM-ABCD-013
101/24 pGM-EBCD-014	None	Carrying pGM-EBCD-014
CA434 pGM-ABCD-015	None	Carrying pGM-ABCD-015
101/24 pGM-FBCD-016	None	Carrying pGM-FBCD-016
CA434 pGM-CBCD-017	None	Carrying pGM-CBCD-017
CA434 pGM-HBCD-018	None	Carrying pGM-HBCD-018
CA434 pGM-ABCD-019	None	Carrying pGM-ABCD-019
101/24 pGM-FBCD-020	None	Carrying pGM-FBCD-020
CA434 pGM-FCAR-002	None	Carrying pGM-FCAR-002
CA434 pGM-FCAR-003	None	Carrying pGM-FCAR-003
<i>Bacteroidia/Clostridia</i>		
S21+pGM-NAEM-001	single crossover of <i>16s</i>	single crossover of <i>16s</i>
S31+pGM-NAEM-002	single crossover of <i>16s</i>	single crossover of <i>16s</i>
S25+pGM-NACM-003	Δ BSIG_3264	Δ BSIG_3264
S27+pGM-NACM-004	Δ HMPREF0102_01948	Δ HMPREF0102_01948
S31+pGM-NACM-005	Δ BT_2090	Δ BT_2090
S52+pGM-FBCL	None	Carrying pGM-FBCL
S52+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S54+pGM-EBCL	None	Carrying pGM-EBCL
S54+pGM-EBCL ₁	None	Carrying pGM-EBCL ₁
S54+pGM-EBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S62+pGM-FBCL	None	Carrying pGM-FBCL
S62+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S65+pGM-FBCL	None	Carrying pGM-FBCL
S65+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S72+pGM-FBCL	None	Carrying pGM-FBCL
S72+pGM-FBCL ₁	None	Carrying pGM-FBCL ₁
S72+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S73+pGM-FBCL	None	Carrying pGM-FBCL
S73+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S74+pGM-FBCL	None	Carrying pGM-FBCL
S74+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S75+pGM-GBCL	None	Carrying pGM-GBCL
S75+pGM-GBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S76+pGM-FBCL	None	Carrying pGM-FBCL
S76+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S77+pGM-EBCL	None	Carrying pGM-EBCL

FIG. 29 (CONT.)

S77+pGM-EBCL ₁	None	Carrying pGM-EBCL ₁
S77+pGM-EBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S78+pGM-EBCL	None	Carrying pGM-EBCL
S78+pGM-EBCL ₁	None	Carrying pGM-EBCL ₁
S78+pGM-EBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S87+pGM-ABCL	None	Carrying pGM-ABCL
S87+pGM-ABCL ₁	None	Carrying pGM-ABCL ₁
S87+pGM-ABCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S88+pGM-IBCL	None	Carrying pGM-IBCL
S88+pGM-IBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S92+pGM-FBCL	None	Carrying pGM-FBCL
S92+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S98+pGM-DBCL	None	Carrying pGM-DBCL
S98+pGM-DBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S100+pGM-ABCL	None	Carrying pGM-ABCL
S100+pGM-ABCL ₁	None	Carrying pGM-ABCL ₁
S100+pGM-ABCG	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S104+pGM-EBCL	None	Carrying pGM-EBCL
S104+pGM-EBCL ₁	None	Carrying pGM-EBCL ₁
S104+pGM-EBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S107+pGM-ABCL	None	Carrying pGM-ABCL
S107+pGM-ABCL ₁	None	Carrying pGM-ABCL ₁
S107+pGM-ABCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S108+pGM-ABCL	None	Carrying pGM-ABCL
S108+pGM-ABCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S110+pGM-ABCL	None	Carrying pGM-ABCL
S110+pGM-ABCG	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S116+pGM-FBCL	None	Carrying pGM-FBCL
S116+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S117+pGM-FBCL	None	Carrying pGM-FBCL
S117+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S118+pGM-FBCL	None	Carrying pGM-FBCL
S118+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S122+pGM-ABCL	None	Carrying pGM-ABCL
S122+pGM-ABCL ₁	None	Carrying pGM-ABCL ₁
S122+pGM-ABCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S136+pGM-CBCL	None	Carrying pGM-CBCL
S136+pGM-CBCL ₁	None	Carrying pGM-CBCL ₁
S136+pGM-CBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S160+pGM-HBCL	None	Carrying pGM-HBCL
S160+pGM-HBCL ₁	None	Carrying pGM-HBCL ₁
S160+pGM-HBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>

FIG. 29 (CONT.)

S164+pGM-FBCL	None	Carrying pGM-FBCL
S164+pGM-FBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S165+pGM-EBCL	None	Carrying pGM-EBCL
S165+pGM-EBCF	Down regulation of <i>lacZα</i>	Down regulation of <i>lacZα</i>
S107+pGM-ABCD	None	Carrying pGM-ABCD
S107+pGM-ABCD-006	Down regulation of <i>clospo_00147</i>	Down regulation of <i>clospo_00147</i>
S52+pGM-FBCD	None	Carrying pGM-FBCD
S52+pGM-FBCD-007	Down regulation of <i>RUMHYD_RS05985</i>	Down regulation of <i>RUMHYD_RS05985</i>
S54+pGM-ABCD	None	Carrying pGM-ABCD
S54+pGM-ABCD-008	Down regulation of <i>GKZ57_13035</i>	Down regulation of <i>GKZ57_13035</i>
S62+pGM-ABCD	None	Carrying pGM-ABCD
S62+pGM-ABCD-009	Down regulation of <i>C6W64_004620</i>	Down regulation of <i>C6W64_004620</i>
S72+pGM-FBCD	None	Carrying pGM-FBCD
S72+pGM-FBCD-010	Down regulation of <i>CGC65_03110</i>	Down regulation of <i>CGC65_03110</i>
S74+pGM-FBCD	None	Carrying pGM-FBCD
S74+pGM-FBCD-011	Down regulation of <i>BCAA aminotransferase</i>	Down regulation of <i>BCAA aminotransferase</i>
S98+pGM-DBCD	None	Carrying pGM-DBCD
S98+pGM-DBCD-012	Down regulation of <i>CLOSA_RS11625</i>	Down regulation of <i>CLOSA_RS11625</i>
S100+pGM-ABCD	None	Carrying pGM-ABCD
S100+pGM-ABCD-013	Down regulation of <i>BCAA aminotransferase</i>	Down regulation of <i>BCAA aminotransferase</i>
S104+pGM-EBCD	None	Carrying pGM-EBCD
S104+pGM-EBCD-014	Down regulation of <i>HMPREF0981_04422</i>	Down regulation of <i>HMPREF0981_04422</i>
S110+pGM-ABCD	None	Carrying pGM-ABCD
S110+pGM-ABCD-015	Down regulation of <i>HMPREF9474_RS22760</i>	Down regulation of <i>HMPREF9474_RS22760</i>
S117+pGM-FBCD	None	Carrying pGM-FBCD
S117+pGM-FBCD-016	Down regulation of <i>CPZ25_012335</i>	Down regulation of <i>CPZ25_012335</i>
S136+pGM-CBCD	None	Carrying pGM-CBCD
S136+pGM-CBCD-017	Down regulation of <i>ROSEINA2194_RS14690</i>	Down regulation of <i>ROSEINA2194_RS14690</i>
S160+pGM-HBCD	None	Carrying pGM-HBCD
S160+pGM-HBCD-018	Down regulation of <i>BIFCAT_00922</i>	Down regulation of <i>BIFCAT_00922</i>
S110+pGM-ABCD-019	Down regulation of <i>HMPREF9474_RS17470</i>	Down regulation of <i>HMPREF9474_RS17470</i>
S117+pGM-FBCD-020	Down regulation of <i>CPZ25_015200</i>	Down regulation of <i>CPZ25_015200</i>
S122+pGM-FCAR-002	Disruption of <i>baiH</i>	Disruption of <i>baiH</i>

FIG. 29 (CONT.)

S122+pGM-FCAQ	Insertion of <i>16s</i>	Insertion of <i>16s</i>
S122+pGM-FCAR-002 +pGM-FCFQ	Disruption of <i>baiH</i>	Disruption of <i>baiH</i>
S122+pGM-FCAQ +pGM-FCFQ	Insertion of <i>16s</i>	Insertion of <i>16s</i>
S115+pGM-FCAQ	Insertion of <i>16s</i>	Insertion of <i>16s</i>
S115+pGM-FCAR-003	Disruption of <i>croA</i>	Disruption of <i>croA</i>

FIG. 30

Primer	Sequence (5' → 3')
Primers used in the cloning of plasmids for screening of Gram-negative <i>Bacteroidia</i> strains	
BO_16S_F1	CGAATTCCTGCAGCCCGGG TATGAATAAGGATCGGCTAAC
BO_16S_R2	CAATTTAAGCCCGGGTAAGG
BO_16S_F3N	CCTTACCCGGGCTTAAATTG TGTGAAGGTGCTGCATGGTT
BO_16S_R4	GCTCTAGAACTAGTGGATCC GCTCCAGAAAGGAGGTGTTC
CJG_syn16s_F	CACAGGTTTCATCTTCTCCCG CGAATTCCTGCAGCCCGGG
CJG_syn16s_R	CGATACCgtcgacTCG GCTCTAGAACTAGTGGATCC
R6K_F	CGAgtcgacGGTATCG
Erm_R	CGGGAGAAGATGAACCTGTG
pMTL_cat_F	gcaataaattaaggagaggttata atggtatttgaaaaattg
pMTL_cat_R	cagttgcattgtggcctattg gttactatttatcaattcc
pEx_Erm_change_F	caataggccacatgcaactg
pEx_Erm_change_R	tataacctctccttaatttattgc
Pre16s_F	cacaggttcattcttctccc
R6K_F_RC	cgataccgtcgactcg
16s_27F	AGAGTTTGATCGTGGCTCAG
R6K_R	ctgtcagaccaagtttactcat
Primers used in the cloning of plasmids for the genetic manipulation of <i>Bacteroidia</i> strains to deplete propionate production	
S25_BSIG_3264_mmdA_pEX_F	cacaggttcattcttctccc TTACCACATGCAGGAAGCAG
S25_BSIG_3264_mmdA_pEX_R	cgataccgtcgactcg TTCTCAGCCAGTTCACATGC
S25_BSIG_3264_mmdA_diagF	CTGAAAGAGTTCATGGTGCG
S27_HMPREF0102_01948_mmdA_pEX_F	cacaggttcattcttctccc TCCGGTCAGAAAGGTTTGTC
S27_HMPREF0102_01948_mmdA_pEX_R	cgataccgtcgactcg TCGATACGAGCCTGTGTACG
S27_HMPREF0102_01948_mmdA_diagF	ATGGAACATCTCGGCTATGC
S31_BT_2090_mmdA_pEX_F	cacaggttcattcttctccc TCCTCCTTATCTGCGTGGTC
S31_BT_2090_mmdA_pEX_R	cgataccgtcgactcg TTCCTCGATACGTTCCCAAG
S31_BT_2090_mmdA_diagF	AACGGTATTTCCGCTGATTG
Primers used in the cloning of plasmids for screening of Gram-positive <i>Clostridia</i> microbes	
pmtl+RP4 oriT_F	atgagtaaacttggtctgacag gcggccgctgtatccatag
pmtl+RP4 oriT_R	CGATACCgtcgacTCG ggcatagctgtttcctg
pMTL_laz_diag_F	gcactggccgtcgttttaca
pGM-ABCM_rep_R_1500bp	catcttcgtaagtccctt
pGM-BBCM_rep_R_1000bp	gtttcatatacgcccttc

FIG. 30 (CONT.)

pGM-CBCM_rep_R_2000bp	cctttgagtgcttatgtggc
pGM-DBCM_rep_R_1000bp	cccaaattcgagtaagaggt
pGM-EBCM_rep-MU102_R_1500	ccttcctccaaatcagacaa
pGM-FBCM_rep-AMβ1_R_2000	gttatggaacgtatgcagacg
pGM-GBCM_rep-pWV01_R_1000	cgtgtaaaggactgacagcc
pGM-HBCM_rep-pMB1_R_1500	cagtcgaacagtggaaatctc
pEx_Diag_F	CGAATTCCTGCAGCCCGGG
16s_1391R	GACGGGCGGTGTGTACA
Primers used in the cloning of plasmids for testing CRISPRi-dCpf1 system in <i>C. sporogenes</i> ATCC15579	
83153_AsCpf-1_XbaI_F	gtgttc <u>tctaga</u> atgacacagttgaaggcttc
83153_AsCpf-1_XhoI_R	gtgtta <u>ctcgag</u> ttagtttctcagttcttgaatg
dAsCpf-1_D908A_F	ctcg <u>tgc</u> tat gccaataattggggtctccg
dAsCpf-1_D908A_R	ccccaattattggcata <u>gca</u> cgaggagaacgcaatctc
pMTL_rep_origin_F	gaatggcgaatggcgctagc
pMTL_rep_origin_R	caactgcccactggccg
pMTL_dCpf1_backbone_F	cggccagtgggcaagttg
pMTL_dCpf1_backbone_R	gctagcgccattcgccattc
gRNA_clo02083_z2	tcgtctcctagcTAATTTCTACTCTTGTAGATatggatatgtaactcagggagTAATTTCTACTCTTGTAGAT AGGAGAATAGAAAGAAGAA
Primers used in the cloning of plasmids for testing CRISPRi-dCpf1 lacZα system for Gram-positive Clostridia GMS	
lacZα_dCpf1_F	actgCCTGCAGGccaacacatcaagcttgaca
lacZα_dCpf1_R	actgGCGGCCGcctacgtaggagacgtcagga
dCpf1-lacZα_gRNA_F_V6_R1	ACACAGGAAACAGCTATGACTAATTTCTACTCTTGTAGATCAACGTCGTGA CTGGGAAAACCTAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAG
gRNA_Cas9_Cpf-1_R	GCTTTCCTCTGTCACTTTGC
dCpf1-lacZα_gRNA_F_V6_R2	gtcctaggtataatgctagcTAATTTCTACTCTTGTAGATACACAGGAAACAGCTA TGACTAATTTCTACTCTTGTAGAT
dCpf1_lacZα_gRNA_Gib_R	cattaattgcgttgccgtctcGCTTTCCTCTGTCACTTTGC
dCpf1-lacZα_gRNA_F_V8_R1	TGCTTCCGGCTCGTATGTTGTAATTTCTACTCTTGTAGATACACAGGAAAC AGCTATGACTAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAG
dCpf1-lacZα_gRNA_F_V8_R2	gtcctaggtataatgctagcTAATTTCTACTCTTGTAGATTGCTTCCGGCTCGTAT GTTGTAATTTCTACTCTTGTAGAT
dCpf1-lacZα_backbone_F	gagacggcgcaacgcaattaatg
8x151 without lacZα_RN	gctagcattatacctaggac
dCpf1-lacZα_qPCR_F	CGGATTCACCTGGCCGTCG
dCpf1-lacZα_qPCR_R	CAGGCTGCGCAACTGTTG

FIG. 30 (CONT.)

<i>CLOSPO_16s_qpcr_F</i>	TACACACCGCCCGTCACACC
<i>CLOSPO_16s_qpcr_R</i>	TGATCCAGCCGCAGGTTCTCC
<i>S52_16s_qPCR_F</i>	GCGTTATCCGGATTTACTGG
<i>S52_16s_qPCR_R</i>	AGCCTCAACGTCAGTTACCG
<i>S54_16s_qPCR_F</i>	GCGTTATCCGGATTTACTGG
<i>S54_16s_qPCR_R</i>	GCATTTACCGCTACACTAGG
<i>S62_16s_qPCR_F</i>	AGCTAGTTGGCAGGGTAACG
<i>S62_16s_qPCR_R</i>	CGAAATACTTCTTCGCTCACG
<i>S65_16s_qPCR_F</i>	GTCTGATGTGAAAGGCATGG
<i>S65_16s_qPCR_R</i>	AGCCTCAACGTCAGTTACCG
<i>S72_16s_qPCR_F</i>	TGGGTGTAAAGGGAGCGTAG
<i>S72_16s_qPCR_R</i>	GCCACTGGTGTTCTCTCTAA
<i>S73_16s_qPCR_F</i>	GTAATACGTAGGGGGCAAGC
<i>S73_16s_qPCR_R</i>	AGCCTCAACGTCAGTTATCG
<i>S74_16s_qPCR_F</i>	TGGGGAATATTGCACAATGG
<i>S74_16s_qPCR_R</i>	TAAATCCGGATAACGCTTGC
<i>S75_16s_qPCR_F</i>	GCGGCTTACTGGACGATAAC
<i>S75_16s_qPCR_R</i>	TGGAATGCTTACTGCGTTTG
<i>S76_16s_qPCR_F</i>	GCGGCTTACTGGACGATAAC
<i>S76_16s_qPCR_R</i>	CCCAGGTGGAATGCTTACTG
<i>S77_16s_qPCR_F</i>	GCGGCTTACTGGACGATAAC
<i>S77_16s_qPCR_R</i>	CCCAGGTGGAATGCTTACTG
<i>S78_16s_qPCR_F</i>	GCGGCTTACTGGACGATAAC
<i>S78_16s_qPCR_R</i>	TGGAATGCTTACTGCGTTTG
<i>S87_16s_qPCR_F</i>	GAAAGCCACGGCTAACTACG
<i>S87_16s_qPCR_R</i>	TCTGCACTCCAGCATACCAG
<i>S88_16s_qPCR_F</i>	GAAAGCCACGGCTAACTACG
<i>S88_16s_qPCR_R</i>	TCTGCACTCCAGCATACCAG
<i>S92_16s_qPCR_F</i>	GTAGGCGGATTTTAAAGTGG
<i>S92_16s_qPCR_R</i>	CTCAGCGTCAGTTACAGTCC
<i>S98_16s_qPCR_F</i>	CAGCAGGGAAGAAAATGACG
<i>S98_16s_qPCR_R</i>	ACTCCAGCACCAAGTTTCC
<i>S100_16s_qPCR_F</i>	CCTAGTGTAGCGGTGAAATGC
<i>S100_16s_qPCR_R</i>	GTAACTGCGGCACAGAAGG
<i>S104_16s_qPCR_F</i>	GAAAGCCACGGCTAACTACG
<i>S104_16s_qPCR_R</i>	CACTCCAGCATACCAGTTTCC
<i>S110_16s_qPCR_F</i>	TATTCCACCTGGGGAGTACG
<i>S110_16s_qPCR_R</i>	ATGCACCACCTGTCTTCTCC
<i>S116_16s_qPCR_F</i>	TGTAGCGGTGAAATGCGTAG
<i>S116_16s_qPCR_R</i>	CACTGAGTTTCCCCAACACC
<i>S117_16s_qPCR_F</i>	GAGATAACAGCCCACCAAGG
<i>S117_16s_qPCR_R</i>	CGAAAACCTTCTTCACTCACG

FIG. 30 (CONT.)

S118_16s_qPCR_F	AGTCTTGACATCCCGGTGAC
S118_16s_qPCR_R	ATCCGAAATGCTGGCTACTG
S122_16s_qPCR_F	ACTAAGAAGCCCCGGCTAAC
S122_16s_qPCR_R	TTCGACAGTTTCCAATGCAG
S136_16s_qPCR_F	TGGAAACGGCTGCTAATACC
S136_16s_qPCR_R	CGTGTCTCAGTCCCAATGTG
S160_16s_qPCR_F	TGGGGAATATTGCACAATGG
S160_16s_qPCR_R	TAATTCCGGATAACGCTTGC
S164_16s_qPCR_F	AGCGTTGTCCGGATTTATTG
S164_16s_qPCR_R	TACGCATTTACCGCTACAC
S165_16s_qPCR_F	TGCATCTGATACTGGCAAGC
S165_16s_qPCR_R	CAAGTCGACATCGTTTACGG
Primers used in the cloning of plasmids for targeting <i>BCAA aminotransferase</i> genes on genome in Gram-positive <i>Clostridia</i> strains	
gRNA_S72_CGC65_03110_dCpf1_round1	CTATATGCCGACGGACAAGCTAATTTCTACTCTTGTAGATAGGATTAATAC GATTATAATTAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAGAA
gRNA_S72_CGC65_03110_dCpf1_round2	tcgtctcctagcTAATTTCTACTCTTGTAGATCTATATGCCGACGGACAAGC TAATTTCTACTCTTGTAGAT
gRNA_dCas9_R	CCGTCTCCCGTA gcttcctctgtcactttgc
S72_CGC65_03110_qPCR_F	AGCGCTATGTGGCGAATTAT
S72_CGC65_03110_qPCR_R	AATGCTTCCATCCTCAGTGG
8x151 without <i>lacZα</i> _FN	ctgacgtctcctacgtagg
8x151 without <i>lacZα</i> _RN	gctagcattatacctaggac
gRNA_S52_RUMHYD_RS05985_dCpf1_round1	AAGGTCTGAAAGCTTACACCTAATTTCTACTCTTGTAGATTCAAACGGCAG ATGGGAGGA TAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAGAA
gRNA_S52_RUMHYD_RS05985_dCpf1_round2	TcgtctcctagcTAATTTCTACTCTTGTAGATAAGGTCTGAAAGCTTACACCTA ATTCTACTCTTGTAGAT
S52_RUMHYD_RS05985_qPCR_F	GAATTCTGCTCTCCGACTGG
S52_RUMHYD_RS05985_qPCR_R	GGTTTTACGCCGATTACAGC
gRNA_S54_GKZ57_13035_dCpf1_round1	AAGATGGCGCATGGGATGCCTAATTTCTACTCTTGTAGATCTCCGTTTACA CATATTACATAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAGAA
gRNA_S54_GKZ57_13035_dCpf1_round2	GtcctaggtataatgctagcTAATTTCTACTCTTGTAGATAAGATGGCGCATGGG ATGCC TAATTTCTACTCTTGTAGAT
S54_GKZ57_13035_qPCR_F	GCAGGTGTTTTCCAGTATGC
S54_GKZ57_13035_qPCR_R	CATCAACGAAACGGTCTTCC
gRNA_S62_C6W64_00462_0_dcpf1_round1	GCTGATTACATGTTCACTATTAATTTCTACTCTTGTAGATTCGAATGTATTT TTACATTC TAATTTCTACTCTTGTAGAT AGGAGAATAGAAAGAAGAA
gRNA_S62_C6W64_00462_0_dcpf1_round2	TcgtctcctagcTAATTTCTACTCTTGTAGATGCTGATTACATGTTCACTATTAA TTTCTACTCTTGTAGAT
S62_C6W64_004620_qPCR_F	TGGCCAGATCCAGTTATTCC

FIG. 30 (CONT.)

S62_C6W64_004620_qPCR_R_R	AGAAGTTTCCGGTTCAGACG
gRNA_S74_BCAA aminotransferase_dCpf1_round1	CTATATGCCGACGGACAAGCTAATTTCTACTCTTGTAGATAGGATTAATAC GATTATAATTAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAGAA
gRNA_S74_BCAA aminotransferase_dCpf1_round2	GtcctaggtataatgctagcTAATTTCTACTCTTGTAGATCTATATGCCGACGGAC AAGCTAATTTCTACTCTTGTAGAT
S74_BCAA aminotransferase_qPCR_F	CTGATTCCATCCTTCCATCC
S74_BCAA aminotransferase_qPCR_R	GAGATAACGGCTGCTGTTCC
gRNA_S98_CLOSA_RS116 25_dCpf1_round1	GAACCATCTTTACGGACCACTAATTTCTACTCTTGTAGATTGTCCAGACAG CATGGGCACTAATTTCTACTCTTGTAGAT AGGAGAATAGAAAGAAGAA
gRNA_S98_CLOSA_RS116 25_dCpf1_round2	TcgtctcctagcTAATTTCTACTCTTGTAGATGAACCATCTTTACGGACCACTA ATTTCTACTCTTGTAGAT
S98_CLOSA_RS11625_qPCR_R_F	ATTCTGTCATTGCCCTGTGC
S98_CLOSA_RS11625_qPCR_R_R	TCCTCAAAGCGAAGTTCACC
gRNA_S100_BCAA aminotransferase_dCpf1_round1	TACCTCATAGCTACCCTTCACTAATTTCTACTCTTGTAGATTATAATGGTGA TATGAAAACCTAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAGAA
gRNA_S100_BCAA aminotransferase_dCpf1_round2	gtcctaggtataatgctagcTAATTTCTACTCTTGTAGATTACCTCATAGCTACCCT TCAC TAATTTCTACTCTTGTAGAT
dCpf1_gRNA_Gib_R	cctacgtaggagacgtcag GCTTTCCTCTGTCACTTTGC
S100_BCAA aminotransferase_qPCR_F	GAAAACCCTAACGCAAAGG
S100_BCAA aminotransferase_qPCR_R	CTTGTTATGCCTGGTAACACC
gRNA_S104_HMPREF0981_0442 2_dCpf1_round1	CCATGGATTGGACAGCAGAGTAATTTCTACTCTTGTAGATATAGGTTACAC CCTTTTTGCTAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAGAA
gRNA_S104_HMPREF0981_0442 2_dCpf1_round2	GtcctaggtataatgctagcTAATTTCTACTCTTGTAGATCCATGGATTGGACAG CAGAGTAATTTCTACTCTTGTAGAT
S104_HMPREF0981_0442 2_qPCR_F	CAACCGTCGTCTGTGTATGC
S104_HMPREF0981_0442 2_qPCR_R	AGGCATTGCTTGATGTACC
gRNA_S110_HMPREF9474 _RS22760_dcpf1_round1	GTTATATGCAGACCGATTACTAATTTCTACTCTTGTAGATTAGTTCTGCTGT TAACGTGT TAATTTCTACTCTTGTAGAT AGGAGAATAGAAAGAAGAA
gRNA_S110_HMPREF9474 _RS22760_dcpf1_round2	tcgtctcctagcTAATTTCTACTCTTGTAGATGTTATATGCAGACCGATTACTAA TTTCTACTCTTGTAGAT
S110_HMPREF9474_RS22760_qPCR_F	AACCGAAGACGGACATATCG

FIG. 30 (CONT.)

S110_HMPREF9474_RS22760_qPCR_R	CCGTAAGGAGGCACATAAGC
gRNA_S117_CPZ25_01233_5_dCpf1_round1	TCATGGATTACACAGAAGGTTAATTTCTACTCTTGTAGATCATAATTCTTCC ACTCCAAT TAATTTCTACTCTTGTAGAT AGGAGAATAGAAAGAAGAA
gRNA_S117_CPZ25_01233_5_dCpf1_round2	gtcctaggtataatgctagcTAATTTCTACTCTTGTAGATTCATGGATTACACAGA AGGT TAATTTCTACTCTTGTAGAT
S117_CPZ25_012335_qPCR_F	GTACAGGCGAGGCTAAATGC
S117_CPZ25_012335_qPCR_R	GGAAGGATAGTTCCGTGAAGC
gRNA_S136_ROSEINA2194_RS14690_dCpf1_round1	GTAAGTATATGACAGATTACTAATTTCTACTCTTGTAGATCCAAAAGTGTA ATGGGAAAAGTAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAGAA
gRNA_S136_ROSEINA2194_RS14690_dCpf1_round2	TcgtctcctagcTAATTTCTACTCTTGTAGATGTAAGTATATGACAGATTACTA ATTTCTACTCTTGTAGAT
S136_ROSEINA2194_RS14690_qPCR_F	AGCTGACTGGGTTCCATCTG
S136_ROSEINA2194_RS14690_qPCR_R	CGGATTAAGACCTGCTGCAT
gRNA_S160_BIFCAT_0092_2_dCpf1_round1	ATCGGCTCCGTGGCCGCTCTTAATTTCTACTCTTGTAGATGGGCGGCTGCA ACAAGAAGATAATTTCTACTCTTGTAGATAGGAGAATAGAAAGAAGAA
gRNA_S160_BIFCAT_0092_2_dCpf1_round2	TcgtctcctagcTAATTTCTACTCTTGTAGATATCGGCTCCGTGGCCGCTCTTA ATTTCTACTCTTGTAGAT
gRNA_S160_BIFCAT_0092_2_dCpf1_round2N	GtcctaggtataatgctagcTAATTTCTACTCTTGTAGATATCGGCTCCGTGGCC GCTCT
S160_BIFCAT_00922_qPCR_F	AGTGGACTACTGCGTGATCG
S160_BIFCAT_00922_qPCR_R	CAGAGATGCGGCATAGTTGC
Primers used in the cloning of plasmids for targeting butyrate-producing genes on genome in Gram-positive <i>Clostridia</i> strains	
gRNA_S110_HMPREF9474_RS17470_dcpf1_round1	TTGATTACGAGGAGGATGGATAATTTCTACTCTTGTAGATTGGTTTACAAC ATATGGATG TAATTTCTACTCTTGTAGAT AGGAGAATAGAAAGAAGAA
gRNA_S110_HMPREF9474_RS17470_dcpf1_round2	gtcctaggtataatgctagcTAATTTCTACTCTTGTAGATTTGATTACGAGGAGGA TGGA TAATTTCTACTCTTGTAGAT
gRNA_S117_CPZ25_01520_0_dcpf1_round1	AACGCTTTAAACAGTGAAGTTAATTTCTACTCTTGTAGATTCTTAGCAGAA AAAGGAAAT TAATTTCTACTCTTGTAGAT AGGAGAATAGAAAGAAGAA
gRNA_S117_CPZ25_01520_0_dcpf1_round2	gtcctaggtataatgctagcTAATTTCTACTCTTGTAGATAACGCTTTAAACAGTG AAGT TAATTTCTACTCTTGTAGAT
S115_cro_123_IBSN	TTTTAAGCTTTATAATTATCCTTAATCGACGGTGTTGTGCGCCAGATAGG GTG
S115_cro_123_EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGGTGTTGATAACTTAC CTTTCTTTGT
S115_cro_123_EBS2	TGAACGCAAGTTTCTAATTTTCGGTTTCGATCCGATAGAGGAAAGTGTCT
S115_cro_DiagF	GAACCAAGGTGCAGTTGC
S115_cro_DiagR	GAGCGTATTTGTCAGAGCCAG
Primers used in the cloning of plasmids for testing Group II intron system for <i>Clostridia</i>	
EBS universal primer	CGAAATTAGAACTTGC GTTCAGTAAAC

FIG. 30 (CONT.)

WBJ_16s_tgt_685_IBSN	TTTTAAGCTTTATAATTATCCTTA GCCTTCGCCACTGTGCGCCAGATAGGGTG
WBJ_16s_tgt_685_EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCGCCACTGGTAACTTAC CTTCTTTGT
WBJ_16s_tgt_685_EBS2	TGAACGCAAGTTTCTAATTTGATTAAGGCTCGATAGAGGAAAGTGTCT
pMTL007C-E2_F	CTGTTATCACCACATTTGTACAATCTG
pMTL007C-E2_R	GGATAATTATAAAGCTTAAAAATTACAC
clostron_rep_origin_back bone_R	gctagcgccattcgccattc ggcgcgcccgatctgtggg
clostron_Spec_F	GATGCTGGTGCTTCTGGTGCTGGT atgaatacatacgaacaaattaataaag
clostron_Spec_R	GGAAGTTCGCGACTCATAGAA ccaattagaatgaatatttccc
Csp-316s_marker_F	TTCTATGAGTCGCGAAGTTCCTATTC
Csp-316s_marker_R	ACCAGCACCAGAAGCACCAGCATC
16s_tron_diagR_v4	GGTGCTTCTGGTGCTGGT
16s_1391R_3to5	CTGCCCCGCACACATGT
Primers used in the cloning of plasmids for the genetic manipulation of <i>baiH</i> gene in the <i>bai</i> operon of <i>Faecalicatena contorta</i> S122 (S122)	
EBS universal primer	CGAAATTAGAACTTTCGTTTCAGTAAAC
WBJ_BaiH_tgt_645_IBSN	TTTTAAGCTTTATAATTATCCTTA GAAGTCTTCGCTGTGCGCCAGATAGGGTG
WBJ_BaiH_tgt_645_EBS1d	CAGATTGTACAAATGTGGTGATAACAGATAAGTCTTCGTTATAACTTACC TTTCTTTGT
WBJ_BaiH_tgt_645_EBS2	TGAACGCAAGTTTCTAATTTGATTAAGGCTCGATAGAGGAAAGTGTCT
pMTL007C-E2_F	CTGTTATCACCACATTTGTACAATCTG
pMTL007C-E2_R	GGATAATTATAAAGCTTAAAAATTACAC
aad9_carb_007C2_thiam_F	ctgattttctgggcccggcc caatgaataggtttac
aad9_carb_007C2_thiam_R	cttagggtaacaaaaaacaccg acgcgttatgagtaaacctg
pmtl_007C2_thiam_mark er_F	cgggtgtttttgttacctaag
pMTL007C_Clostron_87_87_Erm_ltrA R	ggccggcccaggaaaatcag
clostron_Thiam_F	GATGCTGGTGCTTCTGGTGCTGGT atggtatttgaaaaaattgata
clostron_Thiam_R	GGAAGTTCGCGACTCATAGAA cttagggtaacaaaaaacaccg
Csp-316s_marker_F	TTCTATGAGTCGCGAAGTTCCTATTC
Csp-316s_marker_R	ACCAGCACCAGAAGCACCAGCATC
WBJ_BaiH_tgt_DiagF	ATCGAAGTCAACGGAGAATG
WBJ_BaiH_tgt_DiagR	CCGTCTTCGAGATTCGTTTA
Primers used for the qPCR of inflammatory genes	
CCL3_qPCR_F	GAGCTGACACCCCGACTGCC
CCL3_qPCR_R	CCAGGTCTCTTTGGAGTCAG
I-BABP_qPCR_F	gatcatcacagaggtccagca
I-BABP_qPCR_R	cttagcgctcataggtcacatc
iNOS_qPCR_F	acgagacggataggcagaga

FIG. 30 (CONT.)

iNOS_qPCR_R	cacatgcaaggaagggaact
IL-1 β _qPCR_F	tcacagcagcacatcaacaa
IL-1 β _qPCR_R	tgctctatcctcgaaggtc
IL-6_qPCR_F	ccggagaggagacttcacag
IL-6_qPCR_R	tccacgatttcccagagaac
IL17_qPCR_F	CCCTGGACTCTCCACGCAA
IL17_qPCR_R	TCCCTCCGATTGACACAGC
IL-22_qPCR_F	ATGGACAGAGCCAATGGGAA
IL-22_qPCR_R	GTGTCCTGGTCCAGGCAGAG
IL-31_qPCR_F	CTGACATCACATATTGGCAT
IL-31_qPCR_R	TCTGAACCGAAGCATGCATAC
TNF- α _qPCR_F	acggcatggatctcaaagac
TNF- α _qPCR_R	gtgggtgagcacgtagt
Bac_Erysi_16s_qPCR_F-2	TGACGGGGGCCCGCACAAGC
Bac_Erysi_16s_qPCR_R-2	CACGAGCTGACGACAACCAT
Erysi_16s_qPCR_F	GTGCTGCAGTTAACGCAA
Erysi_16s_qPCR_R	GCGGGACTTAACCCAACAT
Primers used for generating gRNA-only control for dCpf1 system	
pGM_lacza_dcpf-1_gRNA_No_dcpf-1_F	ctgaatggcgaatggcgctag
pGM_lacza_dcpf-1_gRNA_No_dcpf-1_R	gcggccgcctacgtaggaga
Primers used for generating sequence-optimized <i>Clostridia</i> shuttle vectors	
Pfdx_synCatP_F	GTGTAGTAGCCTGTGAAATAAG
Pfdx_synCatP_R	CTATTTTATCAATTTTTTCAAATACCATTAAACACACCTCCTTAAAAATTACA C
Pfdx+synCatP_Fus_R	CTCACGTTAAGGGATTTTGGTC
Pfdx+synCatP_Fus_R_com p_F	GACCAAAATCCCTTAACGTGAG
pmtl_ColE_R	GAATCAGGGGATAACGCAGG
pmtl_TraJ_opt_F	CCTGCGTTATCCCCTGATTG
pmtl_TraJ_opt_R	GGGCCCATCGGTCTTGC
pmtl_ori_opt_F	GCAAGACCGATCGGGCCC
pmtl_ori_opt_R	CTTATTTACAGGCTACTACAC
A_ori_syn_F1_R	CCGTAAGTTCCCTTATACGAG
A_ori_syn_F2_F	CTCGTATAAGGGAATTACGG
A_ori_syn_F2_R	CTTATTTACAGGCTACTACAC CGAATTTGTAAATAAACCCAC
C_ori_adaptor_F	GCAAGACCGATCGGGCCC CTTAAGTCTAAAAATTAGGGG
C_ori_adaptor_R	CTTATTTACAGGCTACTACAC GAAATGCAAGTTTCTAACTAACATAC
D_ori_adaptor_F	GCAAGACCGATCGGGCCC GCATTCACTTCTTTTCTATATAAAT
D_ori_adaptor_R	CTTATTTACAGGCTACTACAC CTTAATCGCATTTTCATAGATTG
E_ori_adaptor_F	GCAAGACCGATCGGGCCC cctgattctgtggataaccg
E_ori_adaptor_R	CTTATTTACAGGCTACTACAC gttcatttgatatgcctcc

FIG. 30 (CONT.)

F_ori_adaptor_F	GCAAGACCGATCGGGCCC ctcacgttaagggattttgg
F_ori_adaptor_R	CTTATTTACAGGCTACTACAC caggaaaatcagatccaacc
G_ori_adaptor_F	GCAAGACCGATCGGGCCC gcagcgaagatgttgctg
G_ori_adaptor_R	CTTATTTACAGGCTACTACAC cctgtattgctattaatcgc
H_ori_adaptor_F	GCAAGACCGATCGGGCCC gaatggcgaatggcgctagc
H_ori_adaptor_R	CTTATTTACAGGCTACTACAC caacttgcccactggccg
Primers used for generating 16s rRNA targeting primers for microbes of other phyla	
pEX_R6K+Amp_R	gagaaactcttcgggatgc
pMTL_pEX_CatP_F	gcatcccgaagagtttctc gccagtgggcaagttgaa
pMTL_pEX_CatP_R	cgataccgtcgactcg ggagttaaactagggtaac
pMTL_pEX_Kan_F	gcatcccgaagagtttctc CTGCGCTAGCATGCCTATTTG
pMTL_pEX_Kan_R	cgataccgtcgactcg GAACCGACGACAAGCTGACG
16s_Fuso_pEX_F	CGAATTCCTGCAGCCCGGG GATCCTGGCTCAGGATGAACG
16s_Fuso_pEX_R	GCTCTAGAACTAGTGGATCC CTTACGGTTAGGCCTGCTAC
16s_Koxy_pEX_F	CGAATTCCTGCAGCCCGGG CTAACACATGCAAGTCGAAC
16s_Koxy_pEX_R	GCTCTAGAACTAGTGGATCC CCACGATTACTAGCGATTCC
16s_pEX_amplify_F	CGAATTCCTGCAGCCCGGG AGAGTTTGATCGTGGCTCAG
16s_pEX_amplify_R	GCTCTAGAACTAGTGGATCC GACGGGCGGTGTGTACA
pEx_Diag_F	CGAATTCCTGCAGCCCGGG
pEx_Diag_R	GCTCTAGAACTAGTGGATCC

FIG. 31

1. Formula of TSAB (Tryptic Soy Agar + blood) plates.				
Component	Amount (in 1 L)	Final conc.	Vendor (cat. #) [if any]	
Tryptic Soy Agar	40 g	4% (w/v)	BD (236950)	
Horse blood	50 mL	5% (v/v)	Fisher Scientific (50863761)	
L-Cysteine hydrochloride 50 mg/mL, pH 2.0, in H ₂ O	10 mL	1% (v/v)	Sigma (C1276)	
Vitamin K3 1 mg/mL, in EtOH	1 mL	0.1% (v/v)		
Firstly, 40 g of Tryptic Soy Agar is added into 1 L water and autoclaved at 121 °C for 15 min. After cooling the mixture down till the temperature is under 60 °C, 50 mL of horse blood, 10 mL of 50 mg/mL L-Cysteine hydrochloride solution and 1 mL of 1 mg/mL Vitamin K3 are added (antibiotics are added at this stage if needed), then the medium is poured into sterile 100 mm x 15 mm Petri Dishes and stock at 4 °C cold room.				

FIG. 31 (CONT.)

2. Formula of BHIB (Brain Heart Infusion Agar + blood) plates.				
Component	Amount (in 1 L)	Final conc.	Vendor (cat. #) [if any]	
Brain Heart Infusion Agar	52 g	5.2% (w/v)	BD (241830)	
Horse blood	50 mL	5% (v/v)	Fisher Scientific (50863761)	
L-Cysteine hydrochloride 50 mg/mL, pH 2.0, in H ₂ O	10 mL	1% (v/v)	Sigma (C1276)	
Vitamin K3 1 mg/mL, in EtOH	1 mL	0.1% (v/v)		
Firstly, 52 g of Brain Heart Infusion Agar is added into 1 L water and autoclaved at 121 °C for 15 min. After cooling the mixture down till the temperature is under 60 °C, 50 mL of horse blood, 10 mL of 50 mg/mL L-Cysteine hydrochloride solution and 1 mL of 1 mg/mL Vitamin K3 are added (antibiotics are added at this stage if needed), then the medium is poured into sterile 100 mm x 15 mm Petri Dishes and stock at 4 °C cold room.				
3. Formula of TGYB liquid medium for culturing <i>Bacteroidia</i>.				
Component	Amount (in 1 L)	Final conc.	Mfr.	Vendor (cat. #) [if any]
Bacto Tryptone	10 g	1% (w/v)	BBL	BD (211921)
Yeast Extract	5 g	0.5% (w/v)	Bacto	BD (212750)
D-(+)-Glucose	2 g	0.2% (w/v)	Sigma	Sigma (G8270)
L-Cysteine hydrochloride	0.5 g	0.05% (w/v)	Sigma	Sigma (C1276)
1 M Potassium phosphate buffer, pH 7.4	100 mL	10% (v/v)		
TYG salts solution	40 mL	4% (v/v)		
Vitamin K solution	1 mL of 1 mg/mL	0.000001% (w/v)	Sigma	Sigma (M5625)
0.8% (w/v) CaCl ₂	1 mL			
FeSO ₄ ·7H ₂ O	1 mL of 0.4 mg/mL			
Then add: 1mL hematin-histidine solution and 4mL Resazurin anaerobic indicator				

FIG. 31 (CONT.)

Histidine-Hematin 1. Prepare 0.2 M histidine, pH 8.0 • Mix 4.2 g Histidine-HCl monohydrate (Sigma H7875) in 80 mL Milli-Q water. • Adjust the pH from 4 to 8 with 10 N NaOH (the histidine will go into solution as the pH rises). • Bring the final volume to 100 mL with Milli-Q water. 2. Mix 12 mg hematin (Sigma H3281) with 10 mL of 0.2 M histidine, pH 8.0. Dissolve by end-over-end rotation or vigorous shaking for several hours. Filter-sterilize using 0.2 µm filter. Resazurin anaerobic indicator (0.25 mg/mL) • Dissolve 25 mg resazurin (Sigma R2127) in 100 mL distilled H₂O. Store protected from light at 4 °C.				
4. Formula of Mega liquid medium for culturing Gram-positive strains.				
Component	Amount (in 500 mL)	Final conc.	Mfr.	Vendor (cat. #) [if any]
Trypticase™ Peptone	5 g	1% (w/v)	BBL	BD (211921)
Yeast Extract	2.5 g	0.5% (w/v)	Bacto	BD (212750)
D-(+)-Glucose	1 g	0.2% (w/v)	Sigma	Sigma (G8270)
L-Cysteine hydrochloride	0.25 g	0.05% (w/v)	Sigma	Sigma (C1276)
1 M Potassium phosphate buffer, pH 7.4 [†]	50 mL	10% (v/v)		
TYG salts solution [†]	20 mL	4% (v/v)		
Vitamin K solution [†]	500 µL of 1 mg/mL	0.000001% (w/v)	Sigma	Sigma (M5625)
0.8% (w/v) CaCl ₂ [†]	500 µL			
FeSO ₄ ·7H ₂ O [†]	500 µL of 0.4 mg/mL			
Resazurin	2 mL of 0.25 mg/mL	0.000001% (w/v)	Sigma	Sigma (R2127)
Histidine-Hematin [†]	500 µL			
Milli-Q water (dH ₂ O)	150 mL			
† = See "Stock solution recipes" below for details on preparing this component				

FIG. 31 (CONT.)

Next, carbohydrate supplements are added:				
Component	Amount (in 500 mL)	Final conc.	Mfr.	Vendor (cat. #) [if any]
D-(+)-Cellobiose	0.5 g	0.1% (w/v)	Sigma	Sigma (C7252)
D-(+)-Maltose monohydrate	0.5 g	0.1% (w/v)	Sigma	Sigma (M5885)
D-(-)-Fructose	0.5 g	0.1% (w/v)	Sigma	Sigma (F0127)
Soluble starch	12.5 mL of 2% (w/v)	0.05% (w/v)		
Then add:				
Component	Amount (in 500 mL)	Final conc.	Mfr.	Vendor (cat. #) [if any]
Tween 80	1 mL of 25% (v/v)	0.05% (v/v)		
Meat extract	2.5 g	0.5% (w/v)	Sigma	Sigma (70164)
Trace Mineral Supplement	5 mL	1% (v/v)	ATCC	ATCC (MD-TMS)
Vitamin Supplement	5 mL	1% (v/v)	ATCC	ATCC (MD-VS)
SCFA supplement [†]	1.4 mL	0.28% (v/v)		
† = See "Stock solution recipes" below for details on preparing this component				
Adjust pH to pH 7 by and filter-sterilize. Store protected from light at 4 °C (wrap in foil).				
<u>STOCK SOLUTION RECIPES</u>				
1 M potassium phosphate buffer, pH 7.4				
1. Prepare 1 M KH ₂ PO ₄ (monobasic).				
• 68.045 g KH ₂ PO ₄ (anhydrous, f.w. = 136.09) in Milli-Q water to 500 mL				
2. Prepare 1 M K ₂ HPO ₄ (dibasic).				
• 174.18 g K ₂ HPO ₄ (anhydrous, f.w. = 174.18) in Milli-Q water to 1 L				
3. Add monobasic to dibasic to achieve pH 7.4.				
(You typically need ~430 mL monobasic added to 1 L dibasic)				
Vitamin K solution				
• Dissolve 40 mg menadione (Vitamin K3, Sigma M5625) in 40 mL 100% EtOH.				
TYG salts solution				
• MgSO ₄ ·7H ₂ O (Sigma 230391) 0.5 g				
• NaHCO ₃ (Sigma S5761) 10.0 g				
• NaCl (Sigma S7653) 2.0 g				
• Milli-Q water to 1 L				

FIG. 31 (CONT.)**FeSO₄·7H₂O (0.4 mg/mL)**

- Dissolve 40 mg FeSO₄·7H₂O (Sigma F8633) in 100 mL Milli-Q water.

0.8% (w/v) CaCl₂

- Dissolve 0.4 g CaCl₂·2H₂O (Sigma C7902) in 50 mL Milli-Q water.

Resazurin anaerobic indicator (0.25 mg/mL)

1. Dissolve 25 mg resazurin (Sigma R2127) in 100 mL distilled H₂O.
2. Store protected from light at 4 °C.

Histidine-Hematin

1. Prepare 0.2 M histidine, pH 8.0
 - Mix 4.2 g Histidine-HCl monohydrate (Sigma H7875) in 80 mL Milli-Q water.
 - Adjust the pH from 4 to 8 with 10 N NaOH (the histidine will go into solution as the pH rises).
 - Bring the final volume to 100 mL with Milli-Q water.
2. Mix 12 mg hematin (Sigma H3281) with 10 mL of 0.2 M histidine, pH 8.0. Dissolve by end-over-end rotation or vigorous shaking for several hours. Filter-sterilize using 0.2 µm filter.

SCFA supplement

Acetic acid, glacial (Sigma A6283) 17 mL
Propionic acid (Sigma P5561) 6 mL
Butyric acid (Sigma B103500) 4 mL
Isovaleric acid (Sigma 129542) 1 mL

FIG. 31 (CONT.)

5. Formula of Chopped Meat Medium (CMM) (1 L liquid) for culturing Gram-positive strains.				
Ground beef (fat-free) 500.0 g Distilled water 1.0 L 1N NaOH 25.0 mL Mix meat, water and NaOH and bring to a boil with stirring. Cool to room temperature, skim fat from surface, and filter, retaining both meat particles and filtrate. To filtrate, add sufficient distilled water to restore volume to 1.0 L. To this filtrate, add: Trypticase Peptone (BD 211921) 30.0 g Yeast extract 5.0 g K₂HPO₄ 5.0 g 0.025% Resazurin 4.0 mL Boil and cool medium and add: L-Cysteine hydrochloride 0.5 g Hemin Solution (see below) 10.0 mL Vitamin K1 Solution (see below) 0.2 mL Adjust medium for final pH 7.0. Dispense liquid into bottles containing meat particles. Vitamin K1 Solution: Vitamin K1 0.15 mL 95% Ethanol 30.0 mL Store solution in brown bottle under refrigeration. Discard after one month. Hemin Solution: Hemin 50.0 mg 1N NaOH 1.0 mL Distilled water to 100.0 mL Autoclave at 121 °C for 15 minutes. w/ 15% Sterilized Rumen Fluid from Bar Diamond Ranch (or clarified by centrifugation if prefer) add 1:100 of 100X carbohydrate stock Cellobiose 0.5 g Maltose 0.5 g				

FIG. 31 (CONT.)¹

Fructose 0.5 g Dissolve in 5 mL water, sterile filter				
6. Formula of CBAB (Columbia agar + blood) for culturing <i>Fusobacterium</i>.				
Component	Amount (in 1 L)	Final conc.	Vendor (cat. #) [if any]	
Columbia agar	40 g	4% (w/v)	BD (279240)	
Horse blood	50 mL	5% (v/v)	Fisher Scientific (50863761)	
Firstly, 40 g of Columbia agar is added into 1 L water and autoclaved at 121 °C for 15 min. After cooling the mixture down till the temperature is under 60 °C, 50 mL of horse blood is added (antibiotics are added at this stage if needed), then the medium is poured into sterile 100 mm x 15 mm Petri Dishes and stock at 4 °C cold room.				
7. Formula of <i>desulfovibrio</i> (postgate) medium for culturing <i>Desulfovibrio</i>.				

FIG. 31 (CONT.)

Solution A				
Component	Amount (in 1 L)	Final conc.	Vendor (cat. #) [if any]	
K ₂ HPO ₄	0.5 g	0.05% (w/v)	ACROS ORGANICS (424195000)	
NH ₄ Cl	1.0 g	0.1% (v/v)	Fisher Chemical (12125-02-9)	
Na ₂ SO ₄	1.0 g	0.1% (v/v)	Fisher Chemical (S421-1)	
CaCl ₂ x 2 H ₂ O	0.1 g	0.01% (w/v)	ACROS ORGANICS (349615000)	
MgSO ₄ x 7 H ₂ O	2.0 g	0.2% (v/v)	Fisher Chemical (M65-500)	
Na-DL-lactate	2.0 g	0.2% (v/v)	Sigma (71720)	
Yeast extract	1.0 g	0.1% (v/v)	Fisher Bioreagents (BP1422-2)	
Distilled water	980 mL			
Solution B				
Component	Amount (in 1 L)	Final conc.	Vendor (cat. #) [if any]	
FeSO ₄ x 7 H ₂ O	0.5 g	0.05% (w/v)	ACROS ORGANICS (423731000)	
Distilled water	10 mL			
Solution C				
Component	Amount (in 1 L)	Final conc.	Vendor (cat. #) [if any]	
Na-thioglycolate	0.1 g	0.01% (w/v)	TCI (M0053)	
Ascorbic acid	0.1 g	0.01% (w/v)	Fisher Chemical (BP351500)	
Distilled water	10 mL			

FIG. 32

Taxon	Con-1	Con-2	Con-3	Con-4	Mut-1	Mut-2	Mut-3	Mut-4
Phylum								
Bacteroidetes	61.74	65.83	67.41	62.80	82.81	78.63	73.62	81.49
Firmicutes	22.60	20.65	18.43	22.81	8.69	11.75	15.45	9.82
Proteobacteria	8.65	9.90	9.30	9.15	5.81	6.29	6.50	5.92
Verrucomicrobia	4.03	2.89	3.96	4.30	2.42	3.08	4.17	2.59
Unclassified	2.97	0.73	0.90	0.94	0.27	0.25	0.25	0.18
Tenericutes	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00
Class								
Bacteroidia	61.72	65.80	67.38	62.77	82.81	78.62	73.61	81.48
Bacilli	0.33	0.14	0.22	0.25	4.80	2.60	10.97	6.67
Erysipelotrichi	18.23	16.53	16.17	19.74	0.02	0.01	0.00	0.01
Betaproteobacteria	8.28	9.85	9.24	9.12	5.75	6.20	6.40	5.80
Verrucomicrobiae	4.03	2.89	3.96	4.30	2.42	3.08	4.17	2.59
Clostridia	4.04	3.98	2.04	2.82	3.87	9.14	4.47	3.15
Unclassified	3.23	0.77	0.96	0.99	0.32	0.30	0.36	0.25
Alphaproteobacteria	0.10	0.03	0.01	0.01	0.01	0.03	0.00	0.01
Gammaproteobacteria	0.04	0.01	0.01	0.00	0.00	0.00	0.00	0.04
Mollicutes	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00
Sphingobacteriia	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Deltaproteobacteria	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Order								
Bacteroidales	61.72	65.80	67.38	62.77	82.81	78.62	73.61	81.48
Lactobacillales	0.33	0.14	0.22	0.25	4.80	2.60	10.97	6.67
Erysipelotrichales	18.23	16.53	16.17	19.74	0.02	0.01	0.00	0.01
Burkholderiales	8.21	9.76	9.13	9.00	5.74	6.20	6.38	5.80
Verrucomicrobiales	4.03	2.89	3.96	4.30	2.42	3.08	4.17	2.59
Clostridiales	4.04	3.98	2.04	2.82	3.87	9.14	4.47	3.15
Unclassified	3.33	0.86	1.07	1.11	0.33	0.31	0.38	0.25
Rickettsiales	0.07	0.03	0.01	0.01	0.01	0.03	0.00	0.01
Aeromonadales	0.04	0.01	0.01	0.00	0.00	0.00	0.00	0.04
Mycoplasmatales	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00
Sphingobacteriales	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Desulfovibrionales	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Family								
Porphyromonadaceae	31.41	27.51	27.60	30.36	19.98	19.96	23.09	21.83
uncl_o_Bacteroidales	0.04	0.04	0.04	0.03	3.20	2.21	2.08	2.32
Lactobacillaceae	0.35	0.16	0.23	0.27	0.24	0.18	0.24	0.50
Erysipelotrichaceae	18.95	16.76	16.24	20.11	0.02	0.02	0.00	0.01
uncl_c_Bacteroidia	1.11	1.48	1.36	1.54	0.00	0.00	0.00	0.00

1

[illegible]

FIG. 33

OT U_ ID	C o n- 1	C o n- 2	C o n- 3	C o n- 4	M ut -1	M ut -2	M ut -3	M ut -4	Taxonomy
OT U1	28 80 7	35 14 3	37 14 2	29 38 4	57 70 2	54 63 7	45 77 2	55 44 5	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_ovatus
OT U7	19 31 0	17 71 3	20 36 6	19 39 6	15 56	17 88	21 97	16 50	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U1 0	25 1	70	89	17 9	24 2	17 0	22 4	47 7	k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Lactobacillaceae;g_Lactobacillus;s_
OT U1 7	17 61 3	15 89 7	15 51 5	19 04 2	20	11	1	10	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U3	80 77	96 21	89 69	88 80	55 81	60 49	61 80	56 56	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;s_
OT U3 7	68	9	11	13	17 30 5	16 98 1	17 48 8	18 79 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U4 1	42 69	22 14	98 5	26 68	3	0	5	3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U1 3	40 30	28 80	39 48	42 90	24 10	30 77	41 57	25 79	k_Bacteria;p_Verrucomicrobia;c_Verrucomicrobiae;o_Verrucomicrobiales;f_Verrucomicrobiaceae;g_Akkermansia;s_muciniphila
OT U2 0	26 62	20 99	19 97	27 74	38 62	89 88	44 70	31 47	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae;g_Dorea;s_
OT U8 3	56	31 0	11 6	40	0	0	3	4	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U5 5	17	0	3	1	45 67	24 36	10 76 1	61 96	k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Enterococcaceae;g_Enterococcus;s_
OT U9 6	16	1	3	2	30 62	20 28	19 49	21 94	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U8 8	47 01	48 85	39 71	57 09	0	1	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U9 7	58	73	13 2	72	0	0	1	0	k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Lactobacillaceae;g_Lactobacillus;s_reuteri
OT U1 06	10 11	14 07	13 09	14 22	0	2	1	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_

FIG. 33 (CONT.)

OT	4	4	4	4	3	5	6	3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1	5	4	9	7	5	6	0	8	s;f_S24-7;g_;s_
14	8	8	6	1					
OT	2	1	6	2	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1	3	5	6	0					s;f_S24-7;g_;s_
13	6	7		2					
OT	2	2	3	2	7	5	1	8	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1	1	9	7	5	2	9	3	5	s;f_S24-7;g_;s_
37	8	7	2	8			1		
OT	4	4	4	5	0	1	1	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotric
U1	8	7	6	4					hales;f_Erysipelotrichaceae;g_Allobaculum;s_
26	0	8	7	8					
OT	1	1	5	1	0	1	1	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1	9	0	4	4					s;f_S24-7;g_;s_
28	1	1		0					
OT	1	7	3	9	1	2	6	2	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1	0	3	3	2	9	4	6	6	s;f_S24-7;g_;s_
68	9								
OT	1	2	1	4	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1		4	5						s;f_S24-7;g_;s_
29									
OT	0	0	0	0	5	4	5	6	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1					5	4	7	2	s;f_S24-7;g_;s_
52					2	0	8	9	
OT	2	1	1	4	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1		3	4						s;f_S24-7;g_;s_
38									
OT	1	6	2	7	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1	0	3	5	9					s;f_S24-7;g_;s_
70	9								
OT	1	1	1	1	7	1	1	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U2	1	3	4	0		5	8	6	s;f_S24-7;g_;s_
14	7	7	2	6					
OT	1	0	0	0	4	3	5	4	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1					6	8	0	6	s
49					8	7	5	6	
OT	5	1	1	1	0	0	0	0	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_;g_
U1		7							;s_
53		0							
		8							
OT	0	0	0	0	9	5	1	9	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1					3		6		s;f_S24-7;g_;s_
51							4		
							4		
OT	2	0	1	1	3	3	4	4	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U3					6	8	2	7	s;f_S24-7;g_;s_
94					5	5	5	8	
OT	8	1	1	1	2	2	4	2	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale
U1	1	3	4	0	4	2	5	1	s;f_Bacteroidaceae;g_Bacteroides
20		7	5	1					

FIG. 33 (CONT.)

OT U1 42	9 3 9	1 2 5	1 3 5	1 4 5	7 4 5	1 1 0	3 0 0	6 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U1 72	1 5 9	1 4 1	1 4 6	1 9 2	0 0 0	0 0 0	0 0 0	0 0 0	k_Bacteria
OT U1 25	1 1 8 0	1 5 7	2 3	1 7	2 0 0	0 0 0	0 0 0	0 0 0	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae
OT U1 27	2 0 0	3 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U1 59	2 0 0	0 0 0	1 0 0	1 0 0	0 0 0	0 0 0	1 0 0	0 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U1 40	8 6 0	6 3 0	2 1 0	6 0 0	0 0 0	1 1 0	1 0 0	0 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U1 56	2 2 6	3 0 5	2 9 2	3 5 0	0 0 0	0 0 0	0 0 0	0 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U1 46	3 0 0	1 0 0	0 0 0	0 0 0	2 0 0	2 5 0	4 0 0	1 8 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U3 69	3 0 0	1 4 0	4 0 0	3 0 0	0 0 0	0 0 0	0 0 0	0 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U1 85	1 2 2	1 3 9	1 2 3	1 4 8	0 0 0	0 0 0	1 0 0	0 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U1 79	6 9 0	8 6 0	8 9 0	5 9 0	3 0 0	9 0 0	5 0 0	6 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U1 19	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Ruminococcus;s_
OT U1 55	1 8 1	2 7 0	2 3 8	3 2 1	0 0 0	0 0 0	0 0 0	0 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U1 30	5 0 2	1 0 2	1 0 0	2 0 0	0 0 0	0 0 0	0 0 0	0 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U2 09	4 4 0	5 7 0	8 7 0	5 2 0	8 0 0	1 7 0	1 4 0	9 0 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U1 45	4 4 0	5 4 0	7 3 0	7 1 0	3 0 0	6 0 0	1 8 0	5 0 0	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria

FIG. 33 (CONT.)

OT U1 47	1	0	1	0	2 4	1 4	2 7	2 4	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 87	6 4	4 6	3 5	5 7	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 33	1 4 4	2 2 8	1 7 0	2 3 1	0	0	0	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_
OT U1 23	2 9	4 3	4 6	2 9	6 7	6 0	8 0	8 3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U4 29	1	0	0	0	1 9 0	1 3 2	1 7 9	2 1 6	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 73	7 3	6 5	7 1	9 5	0	0	0	0	k_Bacteria
OT U2 11	8	1 7	9	7	1 0	1 1	1 5	1 3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_
OT U1 48	1	0	0	0	1 0 8	1 4 2	2 9 1	1 6 3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 43	4 5	3 5	2 4	5 0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U3 93	0	0	0	0	1 7 6	1 2 1	2 1 0	1 7 1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 79	0	0	0	0	1 4 4	8 5	1 2 2	9 9	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U2 16	7	9	2	6	3	5 3 3	4 5	1 5	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Rikenellaceae;g_;s_
OT U3 65	3 3	4 7	4 0	2 4	7 0	6 2	5 8	5 0	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;s_
OT U3 90	0	0	1	0	1 5 9	9 3	1 6 6	1 3 8	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 97	1	0	0	0	1 3 3	1 3 5	1 3 3	1 4 6	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 78	1 2	3 5	3 9	2 7	5 5	6 2	7 1	5 2	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;s_

FIG. 33 (CONT.)

OT U1 95	4 4	8 6	8 7	9 3	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 01	0	0	0	1	1 1 2	8 6	6 0	8 7	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 84	1 0 5	1 2 6	1 2 2	1 6 3	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 69	4 5	1 9	5	3 1	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 73	5 4	5 0	4 8	4 8	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U4 06	1	0	0	0	7 8	9 5	2 1 7	1 1 8	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 80	1 9	2 7	2 1	3 5	5	3	1	5	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U5 04	1 0 3	1 0 3	1 0 0	1 4 3	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U2 06	1 3	1 8	9	2 3	0	0	0	0	k_Bacteria
OT U3 68	2	2	0	2	1	0	3	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 60	4 0	3 3	4 8	4 6	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U1 81	5 1	9 1	7 6	7 6	1	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U2	0	0	6 4	2 9	5 8	6 4	0	0	Unclassified
OT U1 74	4 1	4 1	4 4	4 9	0	0	0	0	k_Bacteria
OT U3 31	2 8	7 8	7 4	7 9	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 77	0	0	0	0	8 3	4 5	4 6	6 3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 36	2 9	3 5	3 8	4 1	0	0	0	0	k_Bacteria

FIG. 33 (CONT.)

OT U1 76	0	2	2	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 74	0	1	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U2 67	2 3	2 5	3 0	2 4	0	2	3	1	k_Bacteria;p_Bacteroidetes
OT U1 8	4 2	0	0	0	0	0	3 2	5 1	k_Bacteria
OT U3 56	0	2	1	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 41	1	2	4	2	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U5 03	1 2	1 7	2 5	2 3	8	2	1 3	2	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;s_
OT U3 88	0	0	0	0	7	6	1 5	1 5	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 73	4 6	8 3	9 1	8 0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 66	2 1	1 5	2	1 5	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U2 17	1 7 3	0	0	1	0	1	0	0	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae
OT U1 62	2 6	1 6	1 0	1 8	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 82	2	3	1	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U4 00	0	1	0	0	6 6	4 4	7 2	8 7	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U1 78	4	7	6	2	1 2	5	1 8	7	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;s_
OT U1 75	0	3	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes

1
FIG. 33 (CONT.)

OT U1 91	3 5	4 9	5 7	3 5	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotric hales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U2 07	3 6	1 4	1 5	0	0	0	0	4 1	k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_ Aeromonadales;f_g;s_
OT U2 01	7	1 2	9	1 1	1 9	2 4	1 2	1 4	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Bur kholderiales;f_Alcaligenaceae;g_Sutterella;s_
OT U1 39	0	0	0	0	0	1 2 7	0	0	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lac hnospiraceae;g;s_
OT U2 15	1 6	1 2	3	1 6	0	0	0	0	k_Bacteria
OT U4 21	0	0	0	0	3 8	4 1	7 6	5 8	k_Bacteria;p_Proteobacteria
OT U3 82	1	1	0	0	3	3	2	2	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_S24-7;g;s_
OT U1 58	8	5	1	9	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_S24-7;g;s_
OT U3 98	0	0	0	0	1 2	0	2 0 4	2	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s
OT U6	5 6	4	8	4	5	2	6	1	k_Bacteria
OT U5 01	3 1	7 3	4 6	6 1	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s
OT U1 64	1 0	1 4	7	7	0	4	1 2	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_Bacteroidaceae;g_Bacteroides
OT U4 41	7 1	4 3	2 4	6 5	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_S24-7;g;s_
OT U2 94	1 5	1 7	3 3	2 6	0	0	1	0	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Bur kholderiales;f_Alcaligenaceae;g_Sutterella;s_
OT U3 95	0	0	0	0	7	0	1 9 7	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_S24-7;g;s_
OT U4 74	4 5	4 5	5 8	5 3	0	0	0	0	k_Bacteria
OT U4 76	9	2 1	1 7	1 1	1	1	7	3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s

FIG. 33 (CONT.)

OT U4 90	4 2	4 5	3 7	6 1	0	0	0	0	k_Bacteria
OT U3 84	5	9	3	8	3	2	8	5	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 09	0	0	0	0	8	2	1 8 1	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U2 18	0	2	1	0	0	0	0	0	k_Bacteria
OT U2 84	1 2	1 1	3	1 3	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 36	7 2	4 3	8	5 7	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 1	6 4	0	6	8	1	8	3	5	k_Bacteria
OT U1 18	0	0	2 7	2 4	2 5	2 2	0	0	k_Bacteria
OT U1 84	7	2	1	4	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 61	1 2	5	3	7	0	0	0	0	k_Bacteria
OT U4	0	3 4	2 8	2 0	0	0	0	0	k_Bacteria
OT U1 50	0	0	0	0	1 7	2 2	2 9	1 9	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U3 81	4	1 4	2 0	9	5	8	1 0	4	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_acidifaciens
OT U5 4	5 7	4	5	2	9	5	7	0	k_Bacteria
OT U1 66	1 1	6	5	1 2	5	2	1 2	5	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;s_
OT U1 9	5 5	4	9	4	7	4	3	0	k_Bacteria
OT U3 67	0	2	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_

FIG. 33 (CONT.)

OT U4 87	1 3	7	1 5	1 2	2	3	3	4	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U4 7	4 6	4	4	8	3	7	1	0	k_Bacteria
OT U4 10	0	0	0	0	1 6	3 4	7 3	2 3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 88	1 3	2 0	3 7	1 8	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U3 4	5 3	1	7	1 0	4	6	4	4	k_Bacteria
OT U2 05	8	5	2	4	0	0	0	0	k_Bacteria
OT U1 77	3	2	2	4	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U4 62	2 2	1 3	1 2	1 4	0	0	0	0	k_Bacteria
OT U8	5 4	2	3	2	4	1	5	4	k_Bacteria
OT U2 8	4 8	3	6	6	4	1	4	1	k_Bacteria
OT U5 2	3 8	4	9	5	2	4	2	3	k_Bacteria
OT U2 99	3 5	2	5	0	0	0	2 0	2 1	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Comamonadaceae;g_Comamonas;s_
OT U1 71	9	7	3	1 9	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U5	0 9	2	0	0	0	0	0	3 0	Unclassified
OT U4 08	0	0	0	0	3 0	2 5	4 4	3 1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 93	2 6	3 0	2 9	4 5	0	0	1	0	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria
OT U3 51	7	7	6	1	1	0	2	0	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;s_

FIG. 33 (CONT.)

OT U4 07	1	0	0	0	3 2	3 4	3 4	2 9	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U4 2	4 5	6	1 0	8	1	2	4	2	k_Bacteria
OT U1 63	1 0	1 1	1 2	1 2	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U3 18	1 0	1 4	1 1	1 8	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 3	4 2	3	7	6	0	2	7	0	k_Bacteria
OT U7 0	3 7	0	8	2	4	2	2	4	k_Bacteria
OT U1 4	4 6	1	4	6	7	1	2	2	k_Bacteria
OT U4 02	3	3	8	5	1 3	7	1 5	4	k_Bacteria;p_Verrucomicrobia;c_Verrucomicrobiae;o_Verrucomicrobiales;f_Verrucomicrobiaceae;g_Akkermansia;s_muciniphila
OT U4 83	1 0	1 0	1 1	9	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U9	4 2	2	7	3	6	1	2	1	k_Bacteria
OT U1 96	1 0	1 4	9	1 3	0	0	0	0	k_Bacteria
OT U2 88	4	5	1 2	1 0	0	0	2	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U1 44	0	0	0	0	1	0	0	1	k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Lactobacillaceae;g_Lactobacillus
OT U1 1	4 2	0	4	4	3	2	3	1	k_Bacteria
OT U4 18	1	0	0	0	3 3	1 3	3 3	3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U4 91	1 9	3 3	3 9	2 1	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U2 4	4 6	1	7	4	4	2	4	3	k_Bacteria

FIG. 33 (CONT.)

OT U3 6	2 8	0	3	3	2	5	3	4	k_Bacteria
OT U3 9	3 9	4	3	0	2	1	5	2	k_Bacteria
OT U1 83	0	9	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U2 5	3 6	5	7	5	2	1	0	1	k_Bacteria
OT U1 6	3 1	2	2	2	1	6	4	4	k_Bacteria
OT U2 1	4 0	2	3	3	4	2	6	2	k_Bacteria
OT U3 0	3 5	1	5	3	4	3	4	0	k_Bacteria
OT U3 2	3 6	1	4	4	4	2	3	1	k_Bacteria
OT U2 10	7 2	2 2	1 4	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 04	1	0	0	0	2 0	2 9	3 7	1 8	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U3 5	3 5	2	5	3	1	2	2	2	k_Bacteria
OT U3 52	1	1	0	0	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U4 85	1 7	4 0	1 7	2 8	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 92	6	2	9	3	0	0	1	0	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Alcaligenaceae;g_Sutterella;s_
OT U1 67	2 0	1 8	1 0	2 0	0	0	0	0	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae
OT U2 9	2 6	3	6	8	3	2	6	1	k_Bacteria

FIG. 33 (CONT.)

OT U3 3	3 8	0	6	3	3	1	0	1	k_Bacteria;p_Proteobacteria
OT U4 56	8	3	0	5	0	0	3	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 5	3 0	3	2	3	2	3	2	1	k_Bacteria
OT U6 1	2 9	1	2	3	2	4	5	3	k_Bacteria
OT U2 7	3 0	1	3	3	4	1	3	2	k_Bacteria
OT U6 7	2 9	1	5	4	4	2	3	1	k_Bacteria
OT U1 57	2 4	2 8	1 5	2 5	1	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 4	3 1	3	2	1	5	2	3	1	k_Bacteria
OT U6 8	2 7	2	2	3	1	4	3	1	k_Bacteria
OT U6 0	3 0	3	3	5	2	2	2	0	k_Bacteria
OT U1 2	1 7	0	0	0	0	0	2 2	0	Unclassified
OT U1 92	7	4	1	3	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 8	3 7	3	3	1	2	6	3	1	k_Bacteria
OT U2 02	7	7	1 1	1 9	0	0	0	0	k_Bacteria
OT U5 7	2 9	1	2	3	4	0	3	1	k_Bacteria
OT U4 6	3 2	3	1	0	2	1	2	1	k_Bacteria

FIG. 33 (CONT.)

OT U8 1	2 7	4	4	3	2	1	1	1	k_Bacteria
OT U1 94	7	1 5	1 3	1 7	0	0	0	0	k_Bacteria
OT U4 80	7	2 5	1 5	1 4	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 0	2 9	1	4	1	4	2	2	1	k_Bacteria
OT U5 02	1 2	2 3	1 2	3 4	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U5 8	8	2 8	0	0	0	0	0	3	k_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rickettsiales;f_mitochondria
OT U6 5	2 6	0	4	1	1	3	6	2	k_Bacteria;p_Proteobacteria
OT U1 97	0	0	1	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U2 2	2 8	1	3	2	3	1	1	0	k_Bacteria;p_Proteobacteria
OT U2 6	3 0	3	5	1	1	2	2	1	k_Bacteria
OT U1 07	2 3	2	5	3	2	0	3	2	k_Bacteria
OT U6 3	2 7	1	5	3	2	1	1	1	k_Bacteria
OT U8 9	3 3	2	3	2	0	1	1	3	k_Bacteria
OT U1 01	2 7	2	4	4	1	2	0	2	k_Bacteria
OT U9 2	2 4	0	7	4	3	4	1	1	k_Bacteria
OT U7 8	3 2	2	1	3	2	4	2	1	k_Bacteria

FIG. 33 (CONT.)

OT U3 91	8	0	6	1	8	2 3	8	5	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lac hnospiraceae
OT U1 93	0	0	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s
OT U1 65	0	5	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_S24-7;g_;s_
OT U9 5	2 8	1	3	2	0	1	2	0	k_Bacteria
OT U7 1	2 6	0	4	4	2	3	3	2	k_Bacteria;p_Proteobacteria
OT U2 3	1 5	1	3	3	2	1	1	3	k_Bacteria
OT U5 9	2 2	2	3	1	0	0	3	1	k_Bacteria
OT U5 0	2 2	1	2	3	1	2	0	3	k_Bacteria
OT U9 8	2 5	0	3	2	3	1	1	1	k_Bacteria
OT U8 0	2 0	2	5	2	2	2	3	0	k_Bacteria
OT U7 5	0	0	0	5	5	2 4	0	0	k_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Ri ckettsiales;f_mitochondria
OT U7 2	2 6	1	2	2	1	1	2	0	k_Bacteria
OT U3 76	3	1 0	1	2	0	5	6	0	k_Bacteria;p_Tenericutes;c_Mollicutes;o_Mycoplasmatal es;f_Mycoplasmataceae;g_;s_
OT U4 34	2 4	2	3	2	1	1	4	1	k_Bacteria
OT U1 16	2 0	0	3	3	4	1	1	0	k_Bacteria
OT U3 45	0	3 6	0	0	0	0	0	0	k_Bacteria

FIG. 33 (CONT.)

OT U1 15	2 7	3	4	2	2	0	1	1	k_Bacteria
OT U9 4	2 0	2	1	2	3	0	2	0	k_Bacteria;p_Proteobacteria;c_Alphaproteobacteria
OT U6 9	2 5	1	4	1	2	3	5	1	k_Bacteria
OT U3 87	4	3	6	2	3	1	5	3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U5 6	2 4	0	2	0	3	1	1	0	k_Bacteria
OT U8 5	2 7	0	1	3	3	0	3	1	k_Bacteria;p_Proteobacteria
OT U4 75	0	3	1	4	0	2	2	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U5 3	2 0	1	4	2	3	1	3	0	k_Bacteria
OT U4 9	2 5	0	0	3	0	2	0	0	k_Bacteria
OT U4 05	1	2	0	2	7	1 7	2 0	1 0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U5 1	1 6	1	3	4	2	2	3	0	k_Bacteria
OT U6 6	2 5	0	1	1	2	1	3	0	k_Bacteria
OT U7 6	2 7	1	6	1	2	1	1	0	k_Bacteria
OT U1 99	1	0	1	0	0	0	0	0	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria
OT U1 88	8	3	6	8	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotrichales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U1 80	1	4	0	4	0	1	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_

FIG. 33 (CONT.)

OT U3 86	0	0	0	0	0	0	0	1	k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Lactobacillaceae;g_Lactobacillus
OT U1 08	1 6	1	5	2	3	3	2	1	k_Bacteria;p_Proteobacteria
OT U3 8	2 2	1	3	1	1	3	1	0	k_Bacteria
OT U1 11	2 4	2	1	2	1	1	0	0	k_Bacteria
OT U8 6	2 1	0	7	2	0	2	0	2	k_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rickettsiales;f_mitochondria
OT U6 4	1 7	1	0	3	3	1	2	2	k_Bacteria;p_Proteobacteria
OT U7 4	1 8	0	3	1	0	0	2	2	k_Bacteria;p_Proteobacteria
OT U1 17	2 1	2	0	1	3	1	1	1	k_Bacteria
OT U3 48	0	2	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U3 49	2	3	0	2	0	0	0	0	k_Bacteria
OT U4 5	0	2 6	0	0	0	0	0	0	k_Bacteria
OT U7 9	1 7	1	1	0	0	1	1	2	k_Bacteria
OT U3 96	0	0	0	0	1 1	7	1 3	1 3	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_s_
OT U9 9	2 0	1	0	2	0	0	1	0	k_Bacteria;p_Proteobacteria
OT U3 71	0	0	1 3	8	0	0	0	0	k_Bacteria
OT U1 09	2 0	1	0	2	2	0	1	1	k_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rickettsiales;f_mitochondria

FIG. 33 (CONT.)

OT U9 3	0	0	0	2 0	0	0	0	0	Unclassified
OT U7 7	1 9	1	0	0	0	2	0	1	k_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rickettsiales;f_mitochondria
OT U8 2	1 2	1	2	3	0	1	1	0	k_Bacteria;p_Proteobacteria
OT U8 4	1 6	1	3	2	1	2	0	1	k_Bacteria
OT U7 3	1 6	0	1	1	0	0	4	2	k_Bacteria
OT U9 1	1 7	0	4	2	1	1	0	1	k_Bacteria
OT U4 25	0	0	0	0	1 3	7	1 0	1 1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 28	0	1	7	2	0	1	0	7	k_Bacteria;p_Verrucomicrobia;c_Verrucomicrobiae;o_Verrucomicrobiales;f_Verrucomicrobiaceae;g_Akkermansia;s_muciniphila
OT U1 12	1 4	0	4	2	0	1	1	1	k_Bacteria
OT U8 7	1 8	1	0	0	2	1	0	1	k_Bacteria
OT U3 54	5	3	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 79	0	5	2	4	1	1	2	2	k_Bacteria;p_Verrucomicrobia;c_Verrucomicrobiae;o_Verrucomicrobiales;f_Verrucomicrobiaceae;g_Akkermansia;s_muciniphila
OT U3 46	4	3	5	3	0	0	0	0	k_Bacteria
OT U1 89	1	5	8	5	0	0	0	0	k_Bacteria;p_Proteobacteria;c_Betaproteobacteria
OT U4 77	9	1 2	1	1 2	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 86	3	7	4	2	0	0	0	0	k_Bacteria

FIG. 33 (CONT.)

OT U1 10	1 4	0	1	2	0	0	0	2	k_Bacteria
OT U4 68	4	4	2	2	0	0	0	1	k_Bacteria;p_Verrucomicrobia;c_Verrucomicrobiae;o_Ve rrucomicrobiales;f_Verrucomicrobiaceae;g_Akkermansia; s_muciniphila
OT U1 54	1 1	3	2	1	1	1	1	0	k_Bacteria
OT U4 17	0	0	0	0	1	6	1 3	1 1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_Bacteroidaceae;g_Bacteroides
OT U3 57	2	3	2	4	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotric hales;f_Erysipelotrichaceae
OT U4 19	0	1	0	0	4	3	1 0	4	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_Bacteroidaceae;g_Bacteroides
OT U6 2	1	0	0	0	7	7	0	0	Unclassified
OT U1 90	0	0	0	0	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotric hales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U3 83	1	1	2	3	1	1	2	1	k_Bacteria;p_Verrucomicrobia;c_Verrucomicrobiae;o_Ve rrucomicrobiales;f_Verrucomicrobiaceae;g_Akkermansia; s_muciniphila
OT U3 53	0	0	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_S24-7;g_;s_
OT U3 55	0	4	0	2	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_S24-7;g_;s_
OT U3 42	0	2	0	1	1	0	1	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s
OT U4 27	0	0	0	0	2	3	1 0	1 1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_S24-7;g_;s_
OT U9 0	8	0	2	1	1	2	1	1	k_Bacteria
OT U2 24	3	3	2	3	0	0	0	0	k_Bacteria;p_Firmicutes;c_Erysipelotrichi;o_Erysipelotric hales;f_Erysipelotrichaceae;g_Allobaculum;s_
OT U4 14	0	0	0	0	4	2	7	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidale s;f_S24-7;g_;s_

FIG. 33 (CONT.)

OT U4 26	0	0	0	0	1	4	1	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 13	0	0	0	0	7	1	1	6	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 20	1	0	1	0	0	0	2	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 96	0	0	2	5	0	2	0	0	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae
OT U2 97	0	0	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 21	3	0	7	1	0	0	0	0	Unclassified
OT U2 03	0	0	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 27	8	0	0	0	0	0	3	1	k_Bacteria
OT U4 98	5	5	4	9	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 04	8	0	2	0	0	2	1	0	k_Bacteria;p_Proteobacteria
OT U3 58	0	2	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U4 22	0	0	0	0	5	6	1	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U4 12	0	0	0	0	6	1	1	5	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U3 39	0	0	3	1	1	2	1	1	k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Lachnospiraceae
OT U3 61	0	1	0	1	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 70	0	0	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_

FIG. 33 (CONT.)

OT U2 12	0	0	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 75	0	0	2	4	2	2	0	0	k_Bacteria
OT U4 78	2	1 3	2	2	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U1 05	0	0	0	0	0	0	7	0	Unclassified
OT U1 24	9	0	1	0	2	1	1	0	k_Bacteria;p_Proteobacteria;c_Alphaproteobacteria
OT U3 99	0	1	0	0	6	1	2	2	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U5 00	1	6	5	5	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides
OT U1 00	0	0	0	0	0	0	6	0	Unclassified
OT U1 22	0	0	0	9	0	0	0	0	Unclassified
OT U3 63	0	1	1	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U3 59	1	3	4	2	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U1 02	0	0	8	0	0	0	0	0	k_Bacteria
OT U4 60	5	5	1	3	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_
OT U2 08	1	0	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales
OT U4 15	0	0	2	0	0	0	8	2	k_Bacteria;p_Bacteroidetes;c_Sphingobacteriia;o_Sphingobacteriales;f_Sphingobacteriaceae;g_Sphingobacterium;s_multivorum
OT U3 62	0	2	0	1	0	0	0	1	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_S24-7;g_;s_

FIG. 33 (CONT.)

OT U4 64	5	0	0	0	0	0	0	0	k_Bacteria
OT U3 02	2	0	0	0	0	0	0	0	Unclassified
OT U4 61	4	0	0	0	0	0	0	0	k_Bacteria
OT U3 72	1	0	0	0	0	0	0	0	Unclassified
OT U2 95	1	0	0	0	0	0	0	0	k_Bacteria
OT U3 89	0	1	0	0	0	0	0	1	Unclassified
OT U3 01	0	0	0	0	0	0	0	0	k_Bacteria
OT U1 32	1	0	0	0	0	0	0	0	k_Bacteria
OT U4 65	4	0	0	0	0	0	0	0	k_Bacteria
OT U1 35	0	0	0	0	0	0	0	0	k_Bacteria
OT U4 16	0	0	0	0	0	0	3	1	k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Enterococcaceae
OT U4 82	1	3	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes
OT U4 48	4	0	0	0	0	0	0	0	k_Bacteria
OT U2 13	0	0	0	0	0	0	0	0	k_Bacteria
OT U1 33	1	0	0	0	0	0	0	0	k_Bacteria
OT U3 85	0	0	0	0	0	0	0	0	k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Bacteroidaceae;g_Bacteroides;s_acidifaciens

FIG. 33 (CONT.)

[illegible]

FIG. 33 (CONT.)

[illegible]

FIG. 33 (CONT.)

[illegible]

FIG. 33 (CONT.)

[illegible]

FIG. 34

Index	Con-1	Con-2	Con-3	Con-4	Mut-1	Mut-2	Mut-3	Mut-4
#PE_reads	118974	118976	119190	117713	121639	119605	119682	118457
#Nochimer a	105582	106154	102847	105161	106943	105434	101888	103505
AvgLen(bp)	452.57	455.24	455.42	455.58	455.41	454.07	455.31	455.38
GC(%)	52.7	52.69	52.58	53.17	50.91	51.02	51.58	51.07
Effective(%)	88.74	89.22	86.29	89.34	87.92	88.15	85.13	87.38

FIG. 35

Gram+ rep origin (Source)	<i>E. coli</i> rep origin	Antibiotic marker	Application-specific
A. pBP1 (<i>C. botulinum</i>)	A. R6K ori	C. P _{pmtl-catp} ⁻ <i>catP</i>	M. MCS
B. pCB102 (<i>C. butyricum</i>)	B. ColE1 ori	C ₁ . P _{fdx} ⁻ <i>catP</i>	C. No <i>rep ori</i> control
C. pCD6 (<i>C. difficile</i>)	C. pBR ori	C ₂ . P _{pex} ⁻ <i>catP</i>	D. P _{fdx} + dCpf-1
D. pIM13 (<i>B. subtilis</i>)		E. <i>ermB</i>	L. <i>lacZα</i> + P _{fdx} + dCpf-1 ; L ₁ . gRNA-1 for <i>lacZα</i> + <i>lacZα</i>
E. <i>Cthem</i> -based rep origin (<i>C. thermocellum</i>)		A. <i>catP+ermB</i> (RAM)	F. gRNA-1 for <i>lacZα</i> + <i>lacZα</i> + P _{fdx} + dCpf-1
F. pAB1-based (<i>E. faecalis</i>)		B. <i>catP+aad9</i> (RAM)	G. gRNA-2 for <i>lacZα</i> + <i>lacZα</i> + P _{fdx} + dCpf-1
G. pWV01-based (<i>L. lactis</i>)		D. <i>aad9+ermB</i> (RAM)	Q. 16s-targeting Group II intron
H. pMB1 (<i>B. longum</i>)		F. <i>aad9+catP</i> (RAM)	R. other specific sites-targeting Group II intron
I. pIP404-based (<i>C. perfringens</i>)		K. P _{pmtl-catp} ⁻ <i>kanR</i>	P. <i>Prevotella</i> chimeric 16s
N. no gram+ rep origin			B. <i>Bacteroidales</i> chimeric 16s
			O. chimeric 16s of other Phyla (O ₁ : <i>Desulfovibrio</i> , O ₂ : <i>Fusobacterium</i> , O ₃ : <i>Klebsiella</i> , O ₄ : <i>Proteus</i> , O ₅ : <i>Achromobacter spp</i> , O ₆ : <i>Bordetella spp</i> , O ₇ : <i>Pseudomonas spp</i>)

FIG. 36

Put ativ e RM stie s	pG M- AB C M	pGM- ABC ₁ M_se q-opt	p G M - CB C M	pGM - CBC ₁ M_se q-opt	pG M- DB C M	pGM- DBC ₁ M_se q-opt	p G M - EB C M	pGM - EBC ₁ M_se q-opt	p G M - FB C M	pGM - FBC ₁ M_se q-opt	pG M- GB C M	pGM- GBC ₁ M_se q-opt	pG M- HB C M	pGM- HBC ₁ M_se q-opt
AA GC TT	3	0	3	0	3	0	3	0	4	1	3	0	1	0
AC AC AG	2	1	4	3	2	1	3	1	2	1	2	1	1	1
AC TG GG	1	0	1	0	1	0	1	0	2	1	1	0	1	0
AG GC CT	1	0	1	0	1	0	1	0	1	0	1	0	1	0
CA GC TG	1	0	1	0	1	0	1	0	2	1	1	0	2	0
CCC GG G	2	0	2	0	2	0	2	0	2	0	2	0	1	0
CCC GT	6	4	5	2	5	2	6	3	7	4	6	3	8	2
CC GG	20	8	18	6	18	6	18	6	20	8	18	6	18	6
CC NG G	22	9	20	7	20	7	21	7	24	11	20	7	21	8
CCS SG G	6	1	6	1	6	1	6	1	6	1	6	1	5	2
CC WG G	14	6	12	4	12	4	13	4	16	8	12	4	12	5
CT AG	10	4	32	28	7	3	6	2	14	10	7	3	3	1
CT GC AG	3	0	3	0	3	0	3	0	3	0	3	0	2	0
GA ATT C	3	0	3	0	3	0	3	0	3	0	3	0	1	0
GA CG C	8	3	8	3	10	5	9	4	8	3	9	4	17	6

FIG. 36 (CONT.)

GASTC	4	2	4	2	4	2	5	2	5	3	6	4	3	2
GATC	16	5	15	5	15	5	18	7	23	13	15	5	15	5
GCGC	16	10	15	9	15	9	19	10	17	11	19	13	36	13
GCNGC	25	13	25	13	27	14	29	14	28	16	29	16	53	30
GGCC	28	10	26	9	26	9	27	9	28	11	26	9	34	15
GGNCC	11	6	10	5	11	5	10	5	14	8	10	5	13	6
GTCGA C	2	0	2	0	2	0	2	0	2	0	2	0	1	0
SUM	20 4	82	21 6	97	19 4	73	20 6	75	23 1	111	20 1	81	24 9	102