



- (51) International Patent Classification:
H04W 88/06 (2009.01)
- (21) International Application Number:
PCT/US2012/028600
- (22) International Filing Date:
9 March 2012 (09.03.2012)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
61/450,741 9 March 2011 (09.03.2011) US
61/485,887 13 May 2011 (13.05.2011) US
61/497,663 16 June 2011 (16.06.2011) US
- (71) Applicant (for all designated States except US): **WASHINGTON UNIVERSITY** [US/US]; One Brookings Drive, St. Louis, MO 63130 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **GORDON, Jeffrey, I** [US/US]; One Brookings Drive, St. Louis, MO 63130 (US). **FAITH, Jeremiah, J.** [US/US]; One Brookings Drive, St. Louis, MO 63130 (US). **MCNULTY, Nathan, P.** [US/US]; One Brookings Drive, St. Louis, MO 63130 (US). **REY, Federico, E.** [US/US]; One Brookings Drive, St. Louis, MO 63130 (US). **GOODMAN, Andrew** [US/US]; One Brookings Drive, St. Louis, MO 63130 (US). **KALLSTROM, George** [US/US]; One Brookings Drive, St. Louis, MO 63130 (US). **RIDAURA, Venessa** [US/US]; One Brookings Drive, St. Louis, MO 63130 (US).
- (74) Agents: **RILEY-VARGAS, Rebecca** et al.; Polsinelli Shughart PC, Mark Twain Plaza III, 105 West Vandalia, Suite 400, Edwardsville, IL 62025 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: CULTURED COLLECTION OF GUT MICROBIAL COMMUNITY

(57) Abstract: The present invention encompasses cultured collections of a gut microbial community, models comprising such cultures, and methods of use thereof.



CULTURED COLLECTION OF GUT MICROBIAL COMMUNITY

GOVERNMENTAL RIGHTS

[0001] This invention was made with government support under DK70977 and DK30292 awarded by the National Institutes of Health. The government has certain rights in the invention.

INCORPORATION BY REFERENCE

[0002] This application claims priority to US provisional applications 61/450,741, filed March 9, 2011, 61/485,887, filed May 13, 2011, and 61/497,663, filed June 16, 2011, each of which is hereby incorporated by reference in their entirety.

BACKGROUND OF THE INVENTION

[0003] The largest microbial community in the human body resides in the gut and comprises somewhere between 300 and 1000 different microbial species. The human body, consisting of about 100 trillion cells, carries about ten times as many microorganisms in the intestines. The gut microbiome contains at least two orders of magnitude more genes than are found in the *Homo sapiens* genome. However, efforts to dissect the functional interactions between microbial communities and their environmental or animal habitats are complicated by the long-standing observation that, for many of these communities, the great majority of organisms have not been cultured in the laboratory, and some may not have been previously identified. Furthermore, experiments to determine the effect of a perturbation on a gut microbial community are hampered because teasing out the effects of a particular perturbation on each of the hundreds or thousands of different species in a gut microbial community using current techniques is difficult at best and may well be impossible. A need exists, therefore, for methods of culturing and dissecting the gut microbial community populations both in vitro and in vivo.

REFERENCE TO COLOR FIGURES

[0004] The application file contains at least one photograph executed in color. Copies of this patent application publication with color photographs will be provided by the Office upon request and payment of the necessary fee.

BRIEF DESCRIPTION OF THE FIGURES

[0005] **FIG. 1** Comparison of the taxonomic representation of bacterial species and gene content in complete versus cultured human fecal microbial communities before and after their introduction into gnotobiotic mice. **(A)** 16S rRNA sequences from complete microbiota were compared with those identified from microbial communities cultured from the same human donors. At each taxonomic level, the proportion of reads in the complete community belonging to a taxonomic group observed in the cultured sample is shown in blue; the proportion of reads belonging to a taxonomic group not observed in the cultured sample (or lacking taxonomic assignment) is shown in black. (Data shown are the average of two unrelated human donors.) In vitro samples refer to comparisons between human fecal samples and plated material. In vivo samples refer to comparisons between gnotobiotic mice colonized with a complete human fecal microbiota and mice colonized with the readily cultured microbes from the same human fecal sample. **(B)** Annotated functions identified in the microbiomes of complete and cultured human gut communities. Each point represents a KO designation plotted by relative abundance (average across two donors, per 100,000 sequencing reads). Black points represent KO comparisons between the in vitro samples; orange points represent comparisons between in vivo samples. **(C)** The distribution of taxa and their relative abundance along the length of the intestine are similar in the two groups of animals. Relative abundances of class-level taxa at six locations are shown; data represent the average of mice colonized from two unrelated donors. SI, small intestine divided into 16 equal-size segments and sampled at SI-2 (proximal), SI-5 (middle), and SI-13 (distal). PCoA suggests that gut biogeography, rather than donor or culturing, explains the majority (58%) of variance between samples (**FIG. 4 A–C**).

[0006] **FIG. 2** Abundance of readily cultured taxa in fecal samples from two unrelated human donors, as determined by SILVA-VOTE, Ribosomal Database Project

(RDP)-based 16S rRNA annotation, and annotation-independent (OTU %ID cutoff) methods. (A–F) Analyses were performed as described in FIG. 1. Unsupervised hierarchical clustering of 16S rRNA datasets generated from either complete uncultured (G) or readily cultured (H) human gut microbial communities separates all samples from Donor 1 (red) from Donor 2 (blue). Unweighted pair group method with arithmetic mean (UPGMA) clustering of unweighted UniFrac distances between samples (rarefied to 1,000 de-noised, chimera-checked sequences each) is shown; nonphylogenetic distance metrics (Jaccard, Hellinger, Bray-Curtis) produce similar results (data not shown). In both G and H, bootstrap support separating Donor 1 samples from Donor 2 samples is 100% [100 iterations; 500 sequences subsampled from each complete (nonrarefied) dataset].

[0007] **FIG. 3** Relative abundance of functional annotations in the uncultured (complete) and readily cultured fecal communities of two unrelated donors. From each donor, complete and cultured fecal samples also were introduced into germfree mice. After a 4-wk acclimatization period on a standard LF/PP diet, fecal microbiomes were characterized by shotgun pyrosequencing. Reads were mapped to KO (A and B), EC (C and D), and level 2 KEGG pathways (E and F). In these graphs, each point represents a functional annotation, and the axes represent the relative abundance (per 100,000 shotgun pyrosequencer reads) of these predicted functions in comparisons of complete versus cultured microbiomes (black points) and in comparisons of complete versus cultured microbiomes after each had been introduced into germfree mice (orange points). In each comparison, the goodness of fit (R^2) values increase in communities that share the same environment (mouse gut) regardless of the donor. (G) Annotation-independent comparison of functions encoded in the microbiomes of uncultured complete or cultured fecal communities: capture of antibiotic resistance genes. Shades represent number of *E. coli* clones (per GB of subcloned DNA from complete or readily cultured microbial communities from each of two human donors), resulting from each of 15 antibiotic selections. (H) Proof-of-principle study connecting captured genes with their associated bacterial sources. Fecal DNA fragments cloned into *E. coli* from Donor 1, but not from Donor 2, conferred resistance to the aminoglycoside amikacin. Replating these fecal samples directly on high levels of

amikacin (4,100 µg/mL) reveals that this functional difference is mirrored in the source communities. Mean values $\pm$ SEM of triplicate samples (separate frozen aliquots of the original fecal material) are plotted. $*P < 0.005$ based on student's t test (unpaired, two-tailed, assuming equal variance; $P < 0.02$ by heteroscedastic test).

[0008] **FIG. 4** Biogeography of complete and readily cultured human gut microbial communities in gnotobiotic mice and the impact of colonization on host adiposity. (A–C) Principal coordinate analysis (PCoA) of weighted UniFrac distances between samples collected along the length of the gut indicates that mice colonized with readily cultured microbial communities have microbiota similar to those colonized from an uncultured source. (A) Principal coordinate 1 (PC1) separates samples by location. (B) Principal coordinate 2 (PC2) separates samples by donor. (C) No other coordinate explains $\geq 5\%$ of the total variance between samples. (D) Epididymal fat pad:body weight ratios in germfree mice and those colonized with complete or readily cultured microbial communities from the two human donors. Ratios represent the average $\pm$ SEM from $n = 5$ mice per group (except the Donor 2 cultured community; $n = 3$). Asterisks indicate statistically significant differences based on an unpaired, two-tailed student's t test. $*P < 0.005$; $**P < 0.001$; N/S, not significant.

[0009] **FIG. 5** Human gut microbial communities composed only of cultured members exhibit in vivo dynamics similar to those in their complete counterparts. (A) PCoA of UniFrac distances between 16S rRNA datasets generated from fecal samples from gnotobiotic mice, colonized with complete or cultured human fecal microbial communities from two unrelated donors and sampled over time. From day 33–46, mice were switched from their standard LF/PP chow to a high-fat, high-sugar Western diet. Time series analysis of community structure as viewed along the first two principal coordinates from A shows that interpersonal (donor) differences separated communities on PC1 (B), and host diet separated communities on PC2 (C). Principal coordinate 3 (PC3) separated samples from mice colonized with complete communities from those colonized with cultured populations (FIG. 6A). Nonphylogenetic distance metrics produced similar results (FIG. 6 D–H). (D) Evidence that the community response to diet is driven by readily cultured bacteria and that members of the same taxonomic group manifest distinct responses to diet perturbations. Species-level taxa significantly

influenced by diet (student's t test $P \leq 0.01$ after Bonferroni correction; $n = 97$ taxa tested) in either the complete communities (blue names), the cultured communities (green names), or both (red names) are plotted over time (arrows). Each column represents the average relative abundance in fecal samples harvested from three to five individually caged mice that were sampled at various times: (i) during the initial LF/PP diet phase; (ii) during the subsequent shift to the Western diet; and (iii) upon return to LF/PP chow. Members of family-level groups with at least one diet-responsive species are shown (excluding rare species with average abundance $<0.1\%$ across each time point). The names of all taxa are shown in **FIG. 7. (E)** The functional gene repertoire in the fecal microbiomes of humanized gnotobiotic mice. Each point represents a KEGG level 2 pathway; the number of hits to each pathway per 100,000 shotgun pyrosequencing reads is plotted for mice consuming LF/PP (x axis) or Western (y axis) diets. Data represent the averages of mice colonized with microbial communities from two unrelated donors. The results show that the fecal microbiome associated with the Western diet is enriched for genes in pathways associated with PTS (red arrows) both in mice colonized with complete uncultured human gut communities (black points) and mice colonized with communities of readily cultured members (orange points). Donor-specific data and results from alternate annotation schemes are shown in **FIG. 8.**

[0010] **FIG. 6** Diet shapes complete and readily cultured human gut microbial communities in a similar manner. **(A)** PCoA of unweighted UniFrac distances between fecal samples obtained from mice colonized with complete or cultured microbial communities from two unrelated human donors. On day 33 after gavage, mice were switched from an LF/PP chow to a high-fat, high-sugar Western diet (gray rectangle). On day 47 they were returned to the original LF/PP diet. Variance along principal coordinate 3 (PC3) is plotted against time. **(B)** Scree plot from PCoA analysis. Only PC1-PC3 (red) account for $\geq 5\%$ of the variance between samples. **(C)** In gnotobiotic mice, communities composed of readily cultured human gut microbes and communities containing a complete human gut microbiota undergo similar diet-dependent changes in abundance of Bacteroidia and Erysipelotrichi upon changes in host diet. Each mouse in each treatment group was caged individually, and each group that received a given community was housed in a separate gnotobiotic isolator. Mean values $\pm$ SEM and P

values (* $P < 0.05$; ** $P < 0.01$ based on a paired, two-tailed student's t test) are indicated when mice were consuming a LF/PP diet (black bars) and then switched to the Western diet (white bars). **(D–H)** Diet-dependent community-wide shifts in bacterial species representation as measured by PCoA analysis based on a nonphylogenetic (binary Jaccard) distance measurement. **(D)** PCoA plot of binary Jaccard distances between all samples. **(E–G)** Separate PCoA values plotted against time. **(H)** Scree plot of variance explained by PCoA axes.

[0011] **FIG. 7** Relative abundances of species-level taxa in fecal samples obtained from gnotobiotic mice over time. **(A–C)** All identified taxa present at an abundance of $\geq 0.1\%$ in at least a single time point are shown. Species significantly influenced by diet in either the complete community (blue names), the readily cultured community (green names), or both (red names) are plotted over time (arrows) during the initial LF/PP, subsequent Western, and final LF/PP phases of the diet oscillation experiment. Significance ($P \leq 0.01$ after Bonferroni correction) was determined by unpaired, two-tailed student's t test, assuming equal variances; $n = 97$ taxa tested. The assumption of equal variances was tested by F test ($P < 0.02$).

[0012] **FIG. 8** KEGG level 2 pathway-based analysis of fecal microbiomes obtained from LF/PP- and Western diet-fed mice colonized with a complete or cultured human gut microbiota from two human donors. **(A–D)** Phosphotransferase system (PTS) pathways are marked in red and highlighted with arrows. **(E and F)** Multiple predicted PTS pathway components are enriched in the fecal microbiomes of mice colonized with complete **(E)** or cultured **(F)** human gut microbial communities and maintained on a high-fat, high-sugar Western diet. KO level-predicted functional annotations are colored by average fold-difference in their representation in microbiomes obtained from mice on the different diets (Western versus LF/PP). Data represent averages from mice colonized with the complete or cultured fecal communities from two unrelated human donors.

[0013] **FIG. 9** The community composition of microbes cultured from humanized gnotobiotic mice can be reshaped by altering host diet. **(A)** Culture collections were generated from fecal samples obtained from gnotobiotic mice colonized with complete or cultured human gut microbial communities and maintained on LF/PP

or Western diets. **(B)** PCoA of nonphylogenetic (binary Jaccard) distances between cultured samples indicates that manipulation of host diet can be used to shape the composition of communities recovered in culture from these animals. Analysis of phylogenetic (UniFrac) distances between samples produced similar clustering by donor and host diet (**Fig. 10**).

[0014] **FIG. 10** Plated communities of human gut microbes can be reshaped through diet selection in gnotobiotic mice. **(A)** PCoA analysis of unweighted UniFrac distances between communities collected from mice before and after a diet switch and plated on GMM. Scree plots display variance explained by PCoA analysis of binary Jaccard **(B)** or unweighted UniFrac **(C)** distances between samples.

[0015] **FIG. 11** Experimental parameters for en masse culturing, taxonomic assignment, inoculation of germfree mice, and arrayed strain collections. **(A)** Most readily cultured OTU in a human fecal sample are observed in six GMM plates. Rarefaction analysis describes the number of new OTU observed with each additional agar plate added to the dataset (10 plates were prepared independently). Points reflect the mean value after 100 iterations (plates sampled in random order without replacement); error bars represent one SD. The mean number of new OTU identified per plate drops below 20 with six or more plates (red). **(B)** Distributions of percent identities between pairwise comparisons of V2 16S rRNA gene sequences from representatives of bacterial taxa with varying degrees of shared phylogeny. We selected 4,041 16S rRNA sequences from the SILVA database (v102) that contain complete V2 regions and full (species-level) SILVAVOTE annotations. Sequences were aligned using PyNast. After removal of gap-only columns, the %ID between V2 regions was calculated for each pairwise comparison. The resulting %ID distributions are plotted for members of two different species within the same genus (interspecies), two different genera of the same family (intergenus), and so forth. **(C)** Comparison of three methods for assigning taxonomy to V2 16S rRNA sequences. **(D and E)** Sequences identified in gnotobiotic mice colonized with a readily cultured human gut microbiota do not reflect nongrowing or dormant cells. **(D)** Alpha-diversity analysis of fecal microbial communities of mice that had been inoculated with the control sample described in SI Materials and Methods. Diversity is similar at the 7-d and 14-d time points. **(E)** Time-course beta-

diversity analysis of gnotobiotic mice inoculated with the control sample. UniFrac distances were calculated between all samples and represented spatially by PCoA. In this figure, variance along the major coordinate of variance (PC1) is plotted against time; colors represent individual mice. **(F)** Communities of cultured human gut microbes can be clonally archived in 384-well format by limiting dilution. At a dilution of a human fecal sample that produces 70% empty wells, a Poisson distribution predicts that 25% of wells will contain a clonal population of cells and that 5% of wells will be nonclonal (arrow). **(G)** Optimization of dilutions for arrayed strain collections. A 1.28×10^8 -fold dilution of the stored fecal aliquot and subsequent inoculation into 384-well trays (0.17 mL per well) produces 70% empty wells. **(H)** Rarefaction analysis indicates that ten 384-well trays are sufficient to capture nearly all the genus-level diversity that can be retrieved from a single fecal sample using this arrayed culturing method.

[0016] **FIG. 12** Personal culture collections archived in a clonally arrayed, taxonomically defined format. **(A)** After limiting dilution of the sample into 384-well trays to the point at which most turbid wells are clonal, a two-step, barcoded pyrosequencing scheme allows each culture well to be associated with its corresponding bacterial 16S rRNA sequence. In the first round of PCR, one of the V2-directed 16S rRNA primers incorporates 1 of 96 error correcting barcodes (BC1, highlighted in red) that designates the location (row and column) within a quadrant of the 384-well tray where the sample resides. The primer also contains a 12-bp linker (blue). All amplicons generated from all wells in a given quadrant from a single plate then are pooled and subjected to a second round of PCR in which one of primers, which targets the linker sequence, incorporates another error-correcting barcode (BC2; green) that designates the quadrant and plate from which the samples were derived, plus an oligonucleotide (gray) used for 454 pyrosequencing. Amplicons generated from the second round of PCR then are pooled from multiple trays and subjected to multiplex pyrosequencing. This approach allows unambiguous assignment of 16S rRNA reads to well and plate locations using a minimum number of barcodes and primers; e.g., 96 BC1 primers and 96 BC2 primers allow 96^2 (9,216) wells to be analyzed. **(B)** Representation of the original (complete) microbial community in the arrayed strain collection.

[0017] **FIG. 13** Study design for refined diets. Two sets of gnotobiotic mice harboring a synthetic microbiota composed of ten sequenced human gut bacterial species were presented a total of 17 diets differing in their concentrations of casein (protein), corn oil (fat), sucrose (simple sugar), and cornstarch (polysaccharide). **(A)** The model community used for all experiments consisted of sequenced bacterial species from the four most abundant phyla in the adult human gut microbiota: Bacteroidetes (blue), Firmicutes (green), Actinobacteria (yellow), and Proteobacteria (red). **(B,C)** The first screen consisted of 11 refined diets: 9 of these diets (A-I in panel B) represent all possible combinations of high, medium, and low protein and fat; the two additional diets (J and K in panel C) contained high sucrose/low starch and high starch/low sucrose, respectively.

[0018] **FIG. 14** Total community abundance (biomass) and the abundance of each community member can best be explained by changes in casein. **(A)** The total DNA yield per fecal pellet increased as the amount of casein in the host diet increased (shown are mean $\pm$ S.E.M. for each tested concentration of casein). **(B)** Changes in species abundance as a function of changes in the concentration of casein in the host diet were also apparent for all 10 species; 7 species were positively correlated with casein concentration (e.g., *B. caccae*) while the remaining three species were negatively correlated with casein concentration (e.g. *E. rectale*). Data points from the first and second set of mice given the refined diets (see **Table 9** for explanation) are shown in purple and green, respectively, while the mean and standard error for all diets at a given concentration of casein are shown in red and tan, respectively.

[0019] **FIG. 15** Mean community member abundance for each diet. The height of each bar indicates the total DNA yield/biomass for a given diet. Casein concentrations (g/kg) for each diet are displayed in gray above each bar. See **FIG. 13** and **Table 6** for a description of diets A-Q.

[0020] **FIG. 16** Total community DNA yield as a function of protein concentration. Nine different 15-week-old gnotobiotic male C57Bl/6J mice harboring the 10-member community were each given a randomly selected diet containing varying amounts of three different refined protein sources (soy, egg white solids, and lactalbumin), and two different refined fat sources (olive oil and lard) (see **Table 9** for

diet schema). Each mouse was sampled on days 5, 6, and 7 of the diet period. DNA was extracted from each fecal pellet and the three samples from each mouse were averaged to produce the final DNA yield per fecal pellet (see **Table 17** for results of DNA measurements).

[0021] **FIG. 17** Changes in species abundance as a function of changes in the concentration of casein in the host diet. Changes are apparent for all species in the model microbiota (note that the responses of *E. rectale* and *B. caccae* are shown in **Fig. 14B** of the main text). Data obtained from the first and second set of mice are shown in blue and green, respectively, while mean values $\pm$ S.E.M are shown in red and tan, respectively.

[0022] **FIG. 18** Simulation of competition for limiting resources. (A) Using equations 4 and 5 for species1 and species2, both species were initialized to a population size of 2 and a diet switch was initiated every ten days, increasing casein abundance at each switch (red numbers indicate % casein for each diet period). (B) The steady-state values of the simulation in panel A mirror the findings in our mouse datasets where the increase in a bacterial species (species1) that is casein limited leads to a decrease in species2 with increasing dietary casein.

[0023] **FIG. 19** Example of community member responses to complex human foods. Changes in species abundance as a function of diet ingredients were apparent for all 10 species (**Table 16**). *B. ovatus* increased in absolute abundance with increased concentration of oats in the diet (A), while most of the ten bacterial species (including *E. rectale* and *C. aerofaciens*) responded to multiple ingredients (B and C). The mean and standard error for all diets are plotted (no error bars are shown when replicate points are not available). The colored z-axis mesh grid on the 3D plots is a triangle-based linear interpolation of the data with color changes corresponding to the values in the color bar on the right.

[0024] **FIG. 20** Estimation of steady state. Nine adult male gnotobiotic mice harboring the ten-member model human gut community were fed a low-fat, low-protein diet (diet A in **FIG. 13B**) for 7-days and then switched to a high-fat, high-protein (diet I in **FIG. 13B**) at time-point zero for 13 days. The relative abundance of each of the taxa was subsequently defined using shotgun sequencing of fecal DNA to determine their

Informative Genome Fraction (IGF). (A) The relative abundance of each bacterium changes rapidly within hours of a diet switch, reaching steady state levels by the third day (shown are the two species with the greatest increase and decrease respectively in relative abundance). Mean values $\pm$ SEM are plotted at each time point. (B) Changes in total fecal DNA yield also increased rapidly in the first 24 h after the diet switch, reaching steady state levels on the fourth day.

[0025] **FIG. 21** Reliable replication of human donor microbiota in gnotobiotic mice. (A) Assembly of bacterial communities in mice that had received microbiota transplants from the obese and lean co-twins in DZ pair 1. PCoA plot based on unweighted Unifrac distance matrix and 97%ID OTUs in sampled fecal communities. Spheres correspond to a single fecal sample obtained at a given time point from a given mouse and are colored according to the co-twin microbiota donor, and the experiment (n=3 independent experiments). Note that assembly is reproducible within members of a group of mice that have received a given microbiota, and between experiments. (B) Transplantation of fecal microbiota from human donors to recipient mice is not only reproducible within members of a recipient group but also captures interpersonal differences. Mean values $\pm$ SEM for pairwise UniFrac distance measurements are plotted. 'Self-Self' comparison, same mouse sampled at different time points within a given experiment; 'Mouse-Mouse (same donor)', mice colonized with the same human donor's fecal microbiota sample (3-8 mice/donor; 1-4 independent experiments/donor sample); 'Mouse-Human donor', comparison of fecal bacterial communities in recipient group of mice versus their human donor's microbiota; 'Mouse-Unrelated Humans', comparison of fecal microbiota from recipients of given donor's microbiota compared to the fecal microbiota of all other unrelated individuals (across twin pair comparison). *, p-value <0.05, **, p-value <0.001; Monte Carlo simulation, 100 iterations.

[0026] **FIG. 22** Unweighted UniFrac analysis of samples collected along the length of the gut. Mean values $\pm$ SEM for pairwise UniFrac distance measurements are plotted. 'Self-Self', comparison of community structures from different regions of the gut (small intestinal segments 1, 2, 5, 9, 13, 15 each analyzed separately with pair wise comparisons of segments). 'Mice colonized with the same donor', mice colonized with the same human donor's fecal microbiota sample (3-8 mice/donor; 1-4 independent

experiments/donor sample); 'Mouse colonized with sibling donor', where the sibling represents the discordant co-twin; 'Unrelated human donors', comparison of fecal microbiota from recipients of given donor's microbiota versus fecal microbiota of all other unrelated individuals (across twin pair comparison). *p-value <0.05, ** p-value <0.001; Monte Carlo simulation, 100 iterations.

[0027] **FIG. 23** Correlation between the representation of genes with assigned KEGG EC annotations in human donor 1 **(A)**, human donor 2 **(B)** and human donor 3 **(C)** microbiome and their representation in the cecal microbiomes of the corresponding gnotobiotic mouse transplant recipients. Each sphere represents an EC. Mean values $\pm$ SEM are plotted for each EC in a given group of mice (n=4 recipient mice/human donor). The Spearman correlation r-value is indicated and is significant in all cases (p<0.0001).

[0028] **FIG. 24** Transmission of the increased adiposity phenotypes of obese co-twins in discordant pairs by transplantation of their fecal microbiota into gnotobiotic mice. **(A)** Body composition, defined by quantitative magnetic resonance, performed one day and two weeks post-colonization, of each mouse in each recipient group. Mean values $\pm$ SEM are plotted of the increase in % body fat over 14 d in all recipient mice for each of the 4 obese co-twins' or lean co-twins' fecal microbiota (n=3-12 animals/donor microbiota; 41 mice/BMI bin; total of 82 mice). ***, p<0.001, as judged by a two-tail Mann-Whitney U-test. **(B)** More detailed and prolonged time course study for recipients of fecal microbiota from the co-twins in discordant pair 1 (mean values $\pm$ SEM plotted; n=4 mice/donor microbiota). A linear mixed model revealed that the difference between the gain in adiposity between the two recipient groups of mice is statistically significant (p<0.001).

[0029] **FIG. 25** Pathway maps representing ECs enriched in the fecal meta-transcriptome of mice colonized with an obese compared to lean co-twin's fecal microbiome.

[0030] **FIG. 26** Metabolites with significant differences in their levels in the ceca of gnotobiotic recipients of obese compared to lean co-twin fecal microbiome transplants. **(A)** cellobiose and lactose levels as defined by non-targeted GC/MS. **(B)**

Targeted GC/MS analysis of cecal SCFA. *, $p < 0.05$; **, $p < 0.01$ (two-tailed unpaired Student's t-test).

[0031] **FIG. 27** Transplantation of a bacterial culture collection from an obese co-twin into germ-free mice produces an increased adiposity phenotype that is ameliorated by exposure to co-housed mice harboring a culture collection from her lean co-twin. (A) Design of follow-up co-housing experiment. 8 week-old, male germ-free C57Bl/6J mice received culture collections from the lean (Ln) co-twin or the obese (Ob) co-twin in DZ twin pair 1. Five days post gavage (5dpc) mice were dually co-housed in one of three configurations: control groups consisted of Ob-Ob, or Ln-Ln cagemates; the experimental group consisted of Obch-Lnch cagemates. The number of types of cages in each gnotobiotic isolator is shown. All mice were fed a low-fat, plant polysaccharide-rich diet. Adiposity phenotypes were measured by quantitative magnetic resonance. (B) Adiposity change from dpc1 measured by quantitative MR at dpc15 after 10 d of co-housing. (C-F). PCoA plot of UniFrac distances between transplanted bacterial communities showing the effects of co-housing over time. Also shown are the results obtained when two GF mice were co-housed (GFch) with a Lnch and a Obch mouse ($n=3$ cages). (G) Family-level taxa whose representation are significantly different between Ob mice and Ln, Obch, Lnch, or GFch mice ($p < 0.05$ after Bonferroni correction) and discriminatory between the Ob and other groups (Random Forests, feature importance score > 0.07). (H) GC/MS analysis of levels of short chain fatty acids in the ceca of the indicated groups of mice. Concentrations of acetate, propionate and butyrate were significantly higher in the ceca of control Ln, and Obch, Lnch, and GFch mice compared to Ob controls (*, $p < 0.05$; two tailed Student's t-test). (I) Non-targeted GC/MS reveals that in contrast to Ob controls, cecal levels of cellobiose and lactose are undetectable in Ln mice and in all three groups of co-housed animals.

[0032] **FIG. 28** Co-housing experiments designed to test the effects of a bacterial consortium assembled from the clonally arrayed culture collection from the lean co-twin in DZ pair 1. (A, B) Experimental design. Effects of co-housing Obch and Ln37ch mice on (C) adiposity (note that dashes connect animals that were co-housed, arrows highlight the Obch mice whose adiposity decreased during the period of co-housing, *, $p < 0.05$, **, $p < 0.01$ compared to Ob controls as defined by Mann-Whitney U

test), (D) fecal bacterial community structure (PCoA of unweighted UniFrac distance matrix based on community 97%ID OTU composition defined from V2-16S rRNA gene sequencing).

DETAILED DESCRIPTION OF THE INVENTION

[0033] The present invention discloses in vitro and in vivo cultures of a gut microbial community, models comprising such cultures, and methods of use thereof. Such a culture, while not a complete reproduction of a gut microbial community, maintains a phylotypic composition such that the culture reflects the original gut microbial community it was derived from. The original gut microbial community may be a complete gut microbial community, or a prior culture of a complete gut microbial community. As used herein, a "complete" gut microbial community refers to the natural in vivo composition of the gut microbial community of a given individual. Advantageously, a culture of the invention allows the analysis of the effect of perturbations on a complete gut microbial community by analyzing the effect of the perturbation on a culture derived from the gut microbial community.

[0034] The present invention comprises several different in vitro cultures, including a cultured collection of a gut microbial community and a clonally arrayed culture collection of a gut microbial community. Additionally, the invention comprises in vivo cultures, wherein an animal comprises a cultured collection of a gut microbial community or a clonally arrayed culture collection of a gut microbial community. Furthermore, the invention comprises methods of using a cultured collection of a gut microbial community, a clonally arrayed culture collection of a gut microbial community, or an animal comprising a culture of a gut microbial community.

(I) In vitro cultures of a gut microbial community

[0035] The present invention encompasses in vitro cultures of a gut microbial community. For instance, the present invention encompasses a cultured collection of a gut microbial community and a clonally arrayed cultured collection of a gut microbial community as detailed below. Generally speaking, an in vitro culture will have a phylotypic composition similar to the original gut microbial community.

(a) phylotypic composition

[0036] The term "phylotypic composition," as used herein, refers to the composition of a gut microbial community as defined by phylotypes. A phylotype is a biological type that classifies an organism by its phylogenetic, e.g. evolutionary, relationship to other organisms. The term phylotype is taxon-neutral, and therefore, may refer to the species composition, genus composition, class composition, etc. or, in alternative embodiments, may refer to organisms with a specified genetic similarity (e.g. 97% similar at a sequence level, or 97% similar at a gene function level).

[0037] In some embodiments, an in vitro culture of a gut microbial community may comprise between about 1 and 100% of the phylotypes present in the original gut microbial community. In certain embodiments, an in vitro culture of a gut microbial community may comprise at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10% of the phylotypes present in the original gut microbial community. In other embodiments, an in vitro culture of a gut microbial community may comprise at least about 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95 or 100% of the phylotypes present in the original gut microbial community. In still other embodiments, an in vitro culture of a gut microbial community may comprise at least about 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90% of the phylotypes present in the original gut microbial community. In yet other embodiments, an in vitro culture of a gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the phylotypes present in the original gut microbial community. In a preferred embodiment, an in vitro culture of a gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the phylotypes present in the original gut microbial community. In another preferred embodiment, an in vitro culture of a gut microbial community may comprise greater than 99.0% of the phylotypes present in the original gut microbial community.

[0038] In exemplary embodiments, an in vitro culture of a gut microbial community may comprise at least about 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90% of the

phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In yet other embodiments, an in vitro culture of a gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In a preferred embodiment, an in vitro culture of a gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In another preferred embodiment, an in vitro culture of a gut microbial community may comprise greater than 99.0% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community.

[0039] In certain exemplary embodiments, an in vitro culture of a gut microbial community may comprise at least about 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90% of the metagenome, transcriptome, or proteome of the original gut microbial community. In yet other embodiments, an in vitro culture of a gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the metagenome, transcriptome, or proteome of the original gut microbial community. In a preferred embodiment, an in vitro culture of a gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the metagenome, transcriptome, or proteome of the original gut microbial community. In another preferred embodiment, an in vitro culture of a gut microbial community may comprise greater than 99.0% of the metagenome, transcriptome, or proteome of the original gut microbial community.

[0040] The phylotypic composition of a cultured or complete gut microbial community may be evaluated using several different methods. Non-limiting examples of methods that may be used to evaluate the phylotypic composition of a complete or cultured gut microbial community may include the biological classification of individual isolated microbial colonies, the analysis of the biological functions represented in a sample, and the metagenomic analysis of genetic material isolated from the complete or cultured gut microbial community.

[0041] In some embodiments, the phylotypic composition of a gut microbial community may be evaluated by analyzing biological functions represented in a sample of the community. Suitable biological functions may include enzyme functions or drug resistance, such as antibiotic resistance. For instance, the capture and characterization of antibiotic resistance genes may be used to evaluate the biological functions represented in a sample. Non-limiting examples of antibiotics that may be used to capture and characterize antibiotic resistance genes may include amikacin, amoxicillin, carbenicillin, cefdinir, chloramphenicol, ciprofloxacin, cefepime, gentamicin, oxytetracycline, penicillin, piperacillin, piperacillin+Tazobactam, tetracycline, trimethoprim, and rimethoprim+sulfamethoxazole.

[0042] In other embodiments, the phylotypic composition of a gut microbial community may be evaluated using metagenomic analysis of genetic material isolated from the gut microbial community. For instance, a conserved region in the composite genomes of the gut microbial community may be sequenced, or the composite genome of a gut microbial community may be shotgun sequenced. In one embodiment, the phylotypic composition of a gut microbial community may be evaluated by sequencing a conserved 16S ribosomal RNA (rRNA) gene of the composite genomes of the gut microbial community. By way of non-limiting example, DNA from a complete or cultured collection of a gut microbial community may be extracted and the variable region 2 (V2) of bacterial 16S rRNA genes may be pyrosequenced.

[0043] In yet other embodiments, the phylotypic composition of a gut microbial community may be evaluated by shotgun sequencing of the composite genomes followed by analysis of predicted functions contained in the composite genomes of the gut microbial community. In one embodiment, the phylotypic composition of a gut microbial community may be evaluated by shotgun sequencing of the composite genomes followed by analysis of predicted functions contained in the composite genomes of the gut microbial community by querying against a known database, such as the KEGG Orthology (KO) database.

[0044] The phylotypic composition in a gut microbial community may be evaluated at various stages during sample collection, extraction, culture, and storage to produce a profile of diversity in a sample. In some embodiments, the phylotypic

composition in the gut microbial community may be evaluated after extraction from the animal host but before culture. In other embodiments, the representation of the taxa in the gut microbial community may be evaluated after culture. In preferred embodiments, the taxa in the gut microbial community may be evaluated both after extraction from the animal host and after culture.

(b) cultured collection of a gut microbial community

[0045] One aspect of the present disclosure provides a cultured collection of a gut microbial community. As used herein, a "cultured collection of a gut microbial community" refers to an in vitro collection of cultured microorganisms derived from an original gut microbial community. Cultivation of a cultured collection of a gut microbial community may alter the microbial community structure and representation of members of the original gut microbial community, resulting in a "cultured collection" with a phylotypic composition similar to the original gut microbial community. In an exemplary embodiment, the cultured collection is stable, meaning that over time the members comprising the collection do not substantially change. As used herein, "substantially change" means less than about 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1% difference between the members comprising the cultured collection when it is evaluated at two separate time points. In some embodiments, the cultured collection is stable for 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, or more than 90 days. In other embodiments, the cultured collection is stable for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more than 12 months. In still other embodiments, the cultured collection is stable for 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more than 10 years.

[0046] Culture conditions may be optimized to maximize the phylotypic composition of members of the original gut microbial community during culture. Non-limiting examples of methods that may be used to optimize culture conditions to maximize the phylotypic composition during culture may include using a low concentration of nutrients to limit growth of aggressive members of the gut microbial community, optimizing plating density to produce dense but distinct colonies, and

optimizing the incubation period to balance the growth of aggressive and slow growing members of a gut microbial community.

[0047] In some embodiments, low concentrations of a nutrient may be used to limit growth of aggressive members of the gut microbial community to maximize the phylotypic composition of members of a gut microbial community during culture. Non-limiting examples of commonly used nutrients in microbial culture that may be used at low concentrations include glucose, tryptone and yeast extract.

[0048] In other embodiments, plating density is optimized to produce dense but distinct colonies to maximize the phylotypic composition of members of a gut microbial community during culture. In a preferred embodiment, a sample of a gut microbial community may be plated at a density of about 4000 to about 6000 colonies per 150mm diameter culture plate. In another preferred embodiment, a sample of a gut microbial community may be plated at a density of about 2000 to about 7000 colonies per 150mm diameter culture plate. In an exemplary embodiment, a sample of a gut microbial community may be plated at a density of about 5000 colonies per 150mm diameter culture plate.

[0049] The number of colonies cultured from a sample of a gut microbial community can and will vary depending on the desired phylotypic composition in the resulting cultured collection of the gut microbial community. The number of colonies needed to culture a gut microbial community may be determined using any of the methods used to assess the phylotypic composition of a gut microbial community described above. In some embodiments, the number of colonies cultured from a sample of a gut microbial community is about 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 11,000, 12,000, 13,000, 14,000, 15,000, 16,000, 17,000, 18,000, 19,000, 20,000, or more than 20,000. In other embodiments, the number of colonies cultured from a sample of a gut microbial community is about 20,000 to 40,000. In yet other embodiments, the number of colonies cultured from a sample is greater than 40,000. In an exemplary embodiment, the number of colonies cultured from a sample of a gut microbial community is about 30,000 colonies.

[0050] The incubation period during the culture of a gut microbial community may be optimized to maximize the phylotypic composition. In some embodiments,

plates comprising a gut microbial community may be incubated for about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 days or more. In one embodiment, plates comprising a gut microbial community may be incubated for about 5 days.

[0051] In some embodiments, non-commercially available components that may help increase the phylotypic composition during culture may be used. Non-limiting examples of such components may include sterile rumen or human fecal extracts.

[0052] In one embodiment, a gut microbial community isolated from a subject is cultured on solid agar media.

[0053] In an exemplary embodiment, a cultured collection of a gut microbial community may comprise at least about 50, 60, 70, 80, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the phylotypes present in the original gut microbial community. In a preferred embodiment, a cultured collection of a gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the phylotypes present in the original gut microbial community. In another preferred embodiment, a cultured collection of a gut microbial community may comprise greater than 99.0% of the phylotypes present in the original gut microbial community.

[0054] In exemplary embodiments, a cultured collection of a gut microbial community may comprise at least about 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In yet other embodiments, a cultured collection of a gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In a preferred embodiment, a cultured collection of a gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In another preferred embodiment, a cultured collection of a gut microbial community may comprise greater than 99.0% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community.

[0055] In certain exemplary embodiments, a cultured collection of a gut microbial community may comprise at least about 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90% of the metagenome, transcriptome, or proteome of the original gut microbial community. In yet other embodiments, a cultured collection of a gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the metagenome, transcriptome, or proteome of the original gut microbial community. In a preferred embodiment, a cultured collection of a gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the metagenome, transcriptome, or proteome of the original gut microbial community. In another preferred embodiment, a cultured collection of a gut microbial community may comprise greater than 99.0% of the metagenome, transcriptome, or proteome of the original gut microbial community.

(c) clonally arrayed culture collection of a gut microbial community

[0056] Another aspect of the present disclosure provides a clonally arrayed culture collection of a gut microbial community. A "clonally arrayed culture collection" of a gut microbial community, as used herein, refers to a collection of cultured microbes each derived from a single microbial cell from a gut microbial community.

[0057] A clonally arrayed culture collection of a gut microbial community may be derived from a complete gut microbial community isolated from a subject, or from a previously cultured gut microbial community. A complete or previously cultured gut microbial community may be sampled as described in section I(d) below.

[0058] In some embodiments, a clonally arrayed culture collection of a gut microbial community may be generated by picking and isolating individual colonies from a gut microbial community cultured on plates. In certain embodiments, a clonally arrayed culture collection of a gut microbial community may be generated using a most probable number (MPN) technique, also known as the method of Poisson zeroes. In essence, the MPN technique allows for creating clonally arrayed species collections by inoculating culture wells with a diluted sample of a gut microbial community so that a certain percentage of the inoculated wells does not receive a microbe.

[0059] In some embodiments, a clonally arrayed culture collection of a gut microbial community is generated using a dilution point that yields about 30 to 90% empty wells. In other embodiments, a clonally arrayed culture collection of a gut microbial community is generated using a dilution point that yields about 50, 40, 60, 70, 80, or 90% empty wells. In a preferred embodiment, a clonally arrayed culture collection of a gut microbial community is generated using a dilution point that yields about 70% empty wells. At this dilution point, only about 5% of the wells will receive more than one cell in the inoculum, or non-clonal wells.

[0060] Handling, culture and storage conditions for generating a clonally arrayed culture collection of a gut microbial community are under strictly anaerobic conditions as described in section I(d) below. A clonally arrayed culture collection of a gut microbial community may be in multiwell culture plates. Non-limiting examples of multiwell culture plates that may be used for generating and storing a clonally arrayed culture collection of a gut microbial community include 6-well, 12-well, 24-well, 48-well, 96-well and 384-well plates. In an exemplary embodiment, a clonally arrayed culture collection of a gut microbial community may be in 384-well plates.

[0061] A clonally arrayed culture collection of a gut microbial community may be taxonomically defined. A two-step barcoded pyrosequencing scheme illustrated in **FIG. 12** may be used to allow each culture well to be associated with its corresponding bacterial 16S rRNA sequence. In essence, the two-step barcoded pyrosequencing scheme uses two DNA amplification reactions using barcoded primers. This, combined with pyrosequencing, allows unambiguous assignment of 16S rRNA reads to well and plate locations using a minimum number of barcodes and primers.

[0062] In some embodiments, a clonally arrayed culture collection derived from an original gut microbial community contains about 100 to about 5000 taxonomically defined isolates. In other embodiments, a clonally arrayed culture collection may contain about 500, 600, 700, 800, 900, 1000, 2000, 3000, or 4000 to about 5000 taxonomically defined isolates. In still other embodiments, a clonally arrayed culture collection may contain about 800, 900, 1000, or 2000 to about 5000 taxonomically defined isolates. In an exemplary embodiment, the clonally arrayed culture collection may contain about 1,000 taxonomically defined isolates.

[0063] In an exemplary embodiment, a clonally arrayed culture collection of a gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the phylotypes present in the original gut microbial community. In a preferred embodiment, a clonally arrayed culture collection of a gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the phylotypes present in the original gut microbial community. In another preferred embodiment, a clonally arrayed culture collection of a gut microbial community may comprise greater than 99.0% of the phylotypes present in the original gut microbial community.

[0064] In exemplary embodiments, a clonally arrayed culture collection of a gut microbial community may comprise at least about 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In yet other embodiments, a clonally arrayed culture collection of a gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In a preferred embodiment, a clonally arrayed culture collection of a gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In another preferred embodiment, a clonally arrayed culture collection of a gut microbial community may comprise greater than 99.0% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community.

[0065] In certain exemplary embodiments, a clonally arrayed culture collection of a gut microbial community may comprise at least about 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90% of the metagenome, transcriptome, or proteome of the original gut microbial community. In yet other embodiments, a clonally arrayed culture collection of a gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the metagenome, transcriptome, or proteome of the original gut microbial

community. In a preferred embodiment, a clonally arrayed culture collection of a gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the metagenome, transcriptome, or proteome of the original gut microbial community. In another preferred embodiment, a clonally arrayed culture collection of a gut microbial community may comprise greater than 99.0% of the metagenome, transcriptome, or proteome of the original gut microbial community.

(d) gut microbial community and collection

[0066] An in vitro culture of a gut microbial community may be derived from a subject that is a rodent, a human, a livestock animal, a companion animal, or a zoological animal. In one embodiment, a culture of a gut microbial community may be derived from a rodent, e.g. a mouse, a rat, a guinea pig, etc. In another embodiment, an in vitro culture of a gut microbial community may be derived from a livestock animal. Non-limiting examples of suitable livestock animals may include pigs, cows, horses, goats, sheep, llamas and alpacas. In still another embodiment, an in vitro culture of a gut microbial community may be derived from a companion animal. Non-limiting examples of companion animals may include pets such as dogs, cats, rabbits, and birds. In still yet another embodiment, an in vitro culture of a gut microbial community may be derived from a zoological animal. As used herein, a "zoological animal" refers to an animal that may be found in a zoo. Such animals may include non-human primates, large cats, wolves, and bears. In an exemplary embodiment, an in vitro culture of a gut microbial community may be derived from a human.

[0067] An in vitro culture of a gut microbial community may be derived from the same subject over a predetermined time period. For instance, in some embodiments, the microbial community may be sampled at an interval of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, or 200 days.

[0068] In some embodiments, an in vitro culture of a gut microbial community may be derived from a subject with an endemic gut microbial community. In other embodiments, an in vitro culture of a gut microbial community may be derived from a sterile subject inoculated with a gut microbial community from another subject. In yet

other embodiments, an in vitro culture of a gut microbial community may be derived from a sterile animal inoculated with a previously cultured gut microbial community (e.g. a cultured collection or a clonally arrayed cultured collection as described herein). In other embodiments, an in vitro culture of a gut microbial community may be derived from a sterile animal inoculated with a defined mixture of gut microbes. In yet other embodiments, an in vitro culture of a gut microbial community may be derived from a sterile animal inoculated with a mixture of gut microbes from a clonally arrayed culture collection of a gut microbial community.

[0069] The gut environment in a suitable subject is anaerobic. Any prolonged exposure to aerobic conditions may lead to a significant alteration in the gut microbial community structure. Therefore, to reflect the gut microbial community structure in a subject, sample collection, extraction, culture and storage conditions should be maintained under strictly anaerobic conditions upon harvesting the sample from the animal host. Methods for providing anaerobic conditions for sample collection, extraction, culture and storage are known in the art and include performing all operations in anaerobic chambers and incubators.

[0070] Anaerobic conditions must also be maintained in sample extraction buffers and growth media. Anaerobic conditions may be maintained in the sample extraction buffers and growth media by using reducing agents. Non-limiting examples of reducing agents may include cysteine.

[0071] Methods of collecting a gut microbial community are known in the art. In certain embodiments, a gut microbial community may be extracted from luminal material collected from the gastrointestinal system, such as from the proximal, central, or distal portions of the small intestine, cecum, or colon. In other embodiments, a gut microbial community may be extracted from a freshly excreted fecal sample. Generally speaking, a freshly excreted fecal sample should be transferred to an anaerobic chamber within 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 minutes of its collection. In one embodiment, a freshly excreted fecal sample is transferred to an anaerobic chamber within 5 minutes of its collection.

[0072] Generally speaking, a newly collected sample of a gut microbial community may be suspended in buffer. In a preferred embodiment, the sample is

allowed to separate in the buffer, allowing large insoluble particles to settle, thus improving downstream handling steps. In an exemplary embodiment, a gut microbial community may be suspended in pre-reduced PBS buffer.

(II) In vivo culture of a gut microbial community

[0073] Yet another aspect of the present disclosure provides an animal comprising a gut microbial community consisting of cultured microbial members. In essence, to generate an animal comprising a gut microbial community consisting of cultured microbial members, a sterile animal may be colonized with a cultured gut microbial community. Such an animal may be referred to as gnotobiotic. Methods of colonizing sterile animals with a gut microbial community are known in the art and consist of introducing an extract comprising a gut microbial community directly into the animal by oral gavage. Oral gavage is the administration of fluids directly into the lower esophagus or stomach using a feeding needle or tube introduced into the mouth and threaded down the esophagus.

[0074] In some embodiments, the animal is a laboratory animal. Non-limiting examples of a laboratory animal may include rodents, canines, felines, and non-human primates. In certain embodiments, the animal is a rodent. Non-limiting examples of rodents may include mice, rats, guinea pigs, etc. The genotype of the sterile animal can and may vary depending on the intended use of the animal. In embodiments where the animal is a mouse, the mouse may be a C57BL/6 mouse, a Balb/c mouse, a 129sv, or any other laboratory strain. In an exemplary embodiment, the mouse is a C57BL/6J mouse. In other embodiments, the animal is a livestock animal, such as swine.

[0075] Sterile animal husbandry methods are known in the art. Sterile animals are typically born under aseptic conditions, which may include removal from the mother by Caesarean section. Sterile animals are generally housed in a sterile or microbially-controlled laboratory environment in which they remain free of all microbes such as bacteria, exogenous viruses, fungi, and parasites.

[0076] In some embodiments, a sterile animal may be colonized with an in vitro culture of a gut microbial community. An in vitro culture of a gut microbial community may be derived from an animal as described in section I above. In certain

embodiments, a sterile animal may be colonized with a cultured collection of a gut microbial community. In other embodiments, a sterile animal may be colonized with a clonally arrayed culture collection of a gut microbial community.

[0077] In yet other embodiments, a sterile animal may be colonized with a subset of a clonally arrayed culture collection of a gut microbial community. For instance, a sterile animal may be colonized with one or more clonal members of a clonally arrayed culture collection of a gut microbial community. In one embodiment, a sterile animal may be colonized with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more than 10 clonal members of a clonally arrayed culture collection of a gut microbial community. In another embodiment, a sterile animal may be colonized with about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 or more than 100 clonal members of a clonally arrayed culture collection of a gut microbial community. In yet another embodiment, a sterile animal may be colonized with about 100, 200, 300, 400, 500, 600, 700, 800, 900 1000 or more than 1000 clonal members of a clonally arrayed culture collection of a gut microbial community. In still another embodiment, a sterile animal may be colonized with about 1000, 2000, 3000 or more clonal members of a clonally arrayed culture collection of a gut microbial community.

[0078] In an exemplary embodiment, an in vivo culture of a cultured gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the phylotypes present in the original gut microbial community. In a preferred embodiment, an in vivo culture of a cultured gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, or 99.5% of the phylotypes present in the original gut microbial community. In another preferred embodiment, an in vivo culture of a cultured gut microbial community may comprise greater than 99.0% of the phylotypes present in the original gut microbial community.

[0079] In exemplary embodiments, an in vivo culture of a cultured gut microbial community may comprise at least about 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In yet other embodiments, an in vivo culture of a cultured gut

microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In a preferred embodiment, an in vivo culture of a cultured gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community. In another preferred embodiment, an in vivo culture of a cultured gut microbial community may comprise greater than 99.0% of the phylum, class, order, family, genus or species phylotypes present in the original gut microbial community.

[0080] In certain exemplary embodiments, an in vivo culture of a cultured gut microbial community may comprise at least about 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90% of the metagenome, transcriptome, or proteome of the original gut microbial community. In yet other embodiments, an in vivo culture of a cultured gut microbial community may comprise at least about 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 or 100% of the metagenome, transcriptome, or proteome of the original gut microbial community. In a preferred embodiment, an in vivo culture of a cultured gut microbial community may comprise at least about 98.0, 98.1, 98.2, 98.3, 98.4, 98.5, 98.6, 98.7, 98.9, 99.0, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, or 99.9% of the metagenome, transcriptome, or proteome of the original gut microbial community. In another preferred embodiment, an in vivo culture of a cultured gut microbial community may comprise greater than 99.0% of the metagenome, transcriptome, or proteome of the original gut microbial community.

(III) Models and Methods of the invention

[0081] Yet another aspect of the present disclosure provides methods of using an in vitro or in vivo culture of the invention. Importantly, an in vitro or in vivo culture of the invention may be used as a model of a complete gut microbial community. For instance, as described in more detail below, an in vitro or in vivo culture may be used to determine the effect of a perturbation on a gut microbial community or host thereof. As used herein, "perturbation" refers to any compound or condition

administered or applied to a gut microbial community. Advantageously, the effect of the perturbation on an in vitro or in vivo culture of the invention or a host thereof may be representative of the effect of the perturbation on the complete gut microbial community that the culture was derived from.

(a) perturbations and effects

[0082] As described above, a “perturbation,” as used herein, refers to any compound or condition administered or applied to a gut microbial community. For instance, in one embodiment, a perturbation may be diet related. Non-limiting examples of diet related perturbations may include foods, specific food ingredients, specific nutrients (e.g. vitamin, mineral, protein, carbohydrate, etc.), or combinations thereof.

[0083] In another embodiment, a perturbation may be environmentally related. Non-limiting examples of environmentally related perturbations may include temperature, humidity, or other climate related variables, exposure to other gut microbial communities, exposure to other environmental microbial communities, exposure to different living conditions (e.g. different physical conditions, different psychological conditions, different sleeping conditions, different work conditions, etc.), or exposure to pathogens.

[0084] In yet another embodiment, a perturbation may be pharmaceutical. Non-limiting examples of a pharmaceutical may be a drug, a prebiotic, a probiotic, or a nutraceutical. In some embodiments where the perturbation is a drug, the drug is an approved drug. In other embodiments where the perturbation is a drug, the drug is undergoing clinical studies or regulatory testing. In certain embodiments, a drug may be a small molecule, a protein, an antibody, a nucleic acid (e.g. antisense, aptamer, miRNA, RNAi, etc.), or other pharmaceutical.

[0085] In an alternative embodiment, a perturbation may be genetic.

[0086] In an exemplary embodiment, the perturbation is a food or food ingredient. In another exemplary embodiment, the perturbation is a drug, prebiotic, or probiotic.

[0087] Non-limiting examples of the types of effects that can be determined using a model and method of the invention include effects of the perturbation on the

composition of the gut microbial community, effects of the perturbation on the metabolism of the gut microbial community, and effects of the perturbation on host biology due to changes in the gut microbial community. In one embodiment, a model and method of the invention may be used to determine the effect of a perturbation on the composition of the gut microbial community. In this regard, "composition of the gut microbial community," may include the phylotypic composition, the metagenomic composition, the transcriptome composition, or the proteome composition of the gut microbial community. In another embodiment, a model and method of the invention may be used to determine the effect of a perturbation on the metabolism of the gut microbial community. In yet another embodiment, a model and method of the invention may be used to determine the effects of the perturbation on host biology. By way of non-limiting examples, the effects may be changes in host metabolism, changes in host transcription, changes in host protein expression, changes in host immune response, and changes in host gut cellular biology.

[0088] In an exemplary embodiment, a method of the invention comprises applying the perturbation to the gut microbial community and determining the impact of the perturbation on the spatial and/or functional organization of the gut microbial community and the niches (professions) of its component members, the impact of the perturbation on the capacity of the community to respond to changes in diet, the impact of the perturbation on the ability of component members to forage adaptively on host-derived mucosal substrates, the impact of the perturbation on the physical and functional interactions that occur between the changing microbial communities and the intestinal epithelial barrier, or the impact of the perturbation on the interaction of the gut microbial community and the immune system of the host.

(b) diet related perturbations

[0089] One method of the invention encompasses a method of determining the effect of a diet related perturbation on a gut microbial community or host thereof. Such a method comprises applying the perturbation to the gut microbial community and determining the effect of the perturbation on the gut microbial community or host thereof. As detailed above, the gut microbial community may be an in vitro or in vivo

cultured gut microbial community. In exemplary embodiments, differences in the cultured gut community before and after the application of the perturbation advantageously represent the effect of the perturbation on the original gut microbial community or host thereof.

[0090] In some embodiments, a method of the invention encompasses a method of determining the effect of a food or food ingredient on a cultured gut microbial community. Such a method comprises evaluating the cultured gut microbial community before and after the perturbation, wherein the difference in the cultured collection represents the effect of the perturbation on the original gut microbial community.

[0091] In another embodiment, the invention encompasses a method of evaluating how the nutritional value of a food ingredient varies with the composition of a subject's gut microbial community. The method generally comprises administering a food ingredient (or food) to one or more subjects with varying gut microbial communities and evaluating the differences in nutritional value of the food ingredient between the subjects in conjunction with evaluating the differences in the gut microbial community of the subjects. Nutritional value of a food or food ingredient may be measured using any method known in the art. In certain embodiments, nutritional value may be determined in terms of growth of the host, metabolic activity of the microbiome, metabolic activity of the host, or microbiome biomass. Such methods may provide information on which foods (or food ingredients) provide better nutrition to a particular group of subjects. For instance, it may be determined for a particular population that the nutritional value of one food ingredient is better than a second food ingredient. Hence, such a method may be used to increase the feed efficiency of a particular diet for either agricultural animals, performance animals, or humans.

[0092] In an exemplary embodiment of the above method, the gut microbial community is a cultured gut microbial community.

[0093] In a further embodiment, for a malnourished population, such a method may be used to determine the best food or food ingredient for ameliorating the malnourishment. As used herein, "malnourishment" refers to the inadequate or excessive consumption of dietary ingredients leading to the development of disease.

[0094] In other embodiments, a method of the invention encompasses a method of predicting the variations in the abundance of a member of a gut microbiome of a host in response to a proposed diet. Generally speaking, the method comprises (a) determining the abundance of a member of a gut microbiome in a host, (b) determining the amount of the diet ingredients protein, fat, polysaccharide and simple sugar in a proposed diet, (c) determining the linear coefficient for a particular gut microbiome member in relation to protein, fat, polysaccharide, and simple sugar, (d) predicting the absolute abundance of the member of the gut microbiome if the host were to be fed the proposed diet in (b) based on the linear coefficient from (c) for a particular diet ingredient and the amount of the diet ingredient, and determining the difference between (a) and (d), wherein the difference is the predicted variation in the abundance of a gut microbiome member in response to the proposed diet.

[0095] In yet other embodiments, a method of the invention encompasses a method of predicting the abundance of a member of a gut microbiome of a host in response to a proposed diet. The method comprises (a) determining the amount of the diet ingredients protein, fat, polysaccharide and simple sugar in a proposed diet, (b) determining the linear coefficient for a particular gut microbiome member in relation to protein, fat, polysaccharide, and simple sugar, and (c) predicting the absolute abundance of the member of the gut microbiome if the host were to be fed the proposed diet in (a) based on the linear coefficient from (b) for a particular diet ingredient and the amount of the diet ingredient.

[0096] In still other embodiments, a method of the invention encompasses a method of specifically manipulating the abundance of a member of a gut microbiome of a host to a target level by changing the diet of the host. The method comprises (a) determining the linear coefficient for a particular gut microbiome member in relation to protein, fat, polysaccharide, and simple sugar, (b) determining the amount of protein, fat, polysaccharide and sugar in a diet necessary to achieve the target level of the gut microbiome member based on the linear coefficients from (a), and (c) feeding a diet to the host that contains the amount of protein, fat, polysaccharide and sugar determined in (b).

[0097] For each of the above embodiments, methods of determining the abundance of a member of a gut microbiome are known in the art. Similarly, methods of determining the amount of the diet ingredients protein, fat, polysaccharide and simple sugar in a proposed diet are known in the art. The linear coefficient for a particular gut microbiome member for a particular food ingredient may be determined using a model of a human gut microbiome community. The abundance of a member of a gut microbiome may be calculated with the equation $y_i = \beta_0 + \beta_{\text{protein}} X_{\text{protein}} + \beta_{\text{polysaccharide}} X_{\text{polysaccharide}} + \beta_{\text{sucrose}} X_{\text{sucrose}} + \beta_{\text{fat}} X_{\text{fat}}$, where y_i is the abundance of the member of the gut microbiome, β_0 is the calculated parameter for the intercept, X is the amount in g/(kg of total diet) of the diet ingredient, and β_{protein} , $\beta_{\text{polysaccharide}}$, β_{sucrose} , and β_{fat} are the linear coefficients for a particular gut microbiome member for each of the diet components. In some embodiments, β_0 for a particular gut microbiome member for a particular food ingredient may be determined using a gnotobiotic mouse model of a human gut microbiome community.

(c) environmental perturbations

[0098] One method of the invention encompasses a method of determining the effect of an environmental perturbation on a gut microbial community or host thereof. Such a method comprises applying the perturbation to the gut microbial community and determining the effect of the perturbation on the gut microbial community or host thereof. As detailed above, the gut microbial community may be an in vitro or in vivo cultured gut microbial community. In exemplary embodiments, differences in the cultured gut community before and after the application of the perturbation advantageously represent the effect of the perturbation on the original gut microbial community or host thereof.

(d) pharmaceutical perturbations

[0099] Yet another aspect of the present disclosure provides a method of evaluating the impact of a pharmaceutical on a gut microbial community. The method typically comprises evaluating a culture of a gut microbial community in the presence and absence of the pharmaceutical, and identifying the differences, if any, between the

culture exposed to the pharmaceutical, and the culture not exposed to the pharmaceutical. For instance, in one embodiment, an in vivo model (as detailed in **section II** above) of a particular subject's gut microbial community may be used to determine how a particular pharmaceutical would impact that subject's gut microbial community. Such a method may be used to determine the reaction of the subject's gut microbial community to the pharmaceutical without having to administer the drug or pharmaceutical to the subject itself. Such reactions may include any changes in the composition of the gut microbial community, changes in the metabolism of the gut microbial community, or changes in host biology stemming from a change in the gut microbial community.

(e) methods of identifying agents that impact a gut microbial community

[0100] In some instances the invention encompasses methods of identifying agents that impact a gut microbial community. In one embodiment, such a method may comprise applying a perturbation to a gut microbial community and determining the changes the perturbation evokes in the community. Specifically, in certain embodiments, changes in one or more taxa are identified. These taxa may then be applied to a cultured gut microbial community, either individually or in combinations, to determine their impact on the cultured gut microbial community. In this manner, agents that impact a gut microbial community may be identified.

[0101] In one embodiment, the invention encompasses a method of discovering a probiotic. The method generally comprises identifying a microbe that thrives after administration of a particular food or food ingredient to a subject. For instance, an in vitro culture may be created before and after administration of a food or food ingredient to a subject. Differences in the before and after gut microbial cultures may be evaluated to determine a microbe that thrives upon the administration of the particular food or food ingredient. Similarly, a particular food or food ingredient may be administered to a sterile animal comprising a known culture of a gut microbial community. Changes in the gut microbial community may be evaluated to identify a microbe that thrives upon the administration of the particular food or food ingredient. Methods of evaluating a gut microbial community, and method of identifying a microbe

that is thriving in a gut microbial community are known in the art, and may include those detailed herein.

(f) disease models

[0102] Another aspect of the present disclosure provides a method of creating a disease model. The method generally comprises 1) identifying a gut microbial community that is related to, the cause of, or the result of a particular disease or disorder, and 2) reproducing that gut microbial community in an in vitro or in vivo model as described above.

[0103] Yet another aspect of the present disclosure provides a method of treating a disease. The method typically comprises identifying a difference between a normal gut microbial community and a gut microbial community of a subject afflicted with the disease or disorder, and altering the gut microbial community of the subject afflicted with the disease or disorder to more closely resemble a normal gut microbial community.

(g) administration of a specific cultured gut microbiome

[0104] Yet another aspect of the present disclosure provides a method of altering the gut microbiome of a subject, the method comprising administering a cultured gut microbiome to the subject. For instance, it may be determined, using a method of the invention, that a particular gut microbiome culture may be advantageous to a subject. Such a microbiome may be administered via oral gavage, as described herein, or in any other manner suitable for administering the cultured collection. By way of non-limiting example, a gut microbiome may be administered to a subject early in its life to form a gut microbiome that is best suited for the growth of the subject in a particular environment. Suitable subjects may include animals (e.g. performance animals, food animals, companion animals, etc.) and humans.

DEFINITIONS

[0105] The term "metagenomics" refers to the application of modern genomic techniques to the study of the composition and operations of communities of

microbial organisms sampled directly in their natural environments, by passing the need for isolation and lab cultivation of individual species.

[0106] The term “sterile animal” refers to an animal that has no microorganisms living in or on it. In one embodiment, the sterile animal is a sterile mouse.

[0107] The term “gnotobiotic animal” refers to an animal in which only certain known strains of bacteria and other microorganisms are present.

EXAMPLES

[0108] The following examples illustrate various embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples that follow represent techniques discovered by the inventors to function well in the practice of the invention. Those of skill in the art should, however, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments that are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention, therefore all matter set forth or shown in the accompanying drawings is to be interpreted as illustrative and not in a limiting sense.

Introduction to Examples 1-9.

[0109] Efforts to dissect the functional interactions between microbial communities and their environmental or animal habitats are complicated by the long-standing observation that, for many of these communities, the great majority of organisms have not been cultured in the laboratory. Methodological differences between culture-independent and culture-based approaches have contributed to the challenge of deriving a realistic appreciation of exactly how much discrepancy exists between the culturable components of a microbial ecosystem and total community diversity. (**Table 1** gives examples of these methodological differences.)

[0110] The largest microbial community in the human body resides in the gut: Its microbiome contains at least two orders of magnitude more genes than are found in our *Homo sapiens* genome. Culture-independent metagenomic studies of the

human gut microbiota are identifying microbial taxa and genes correlated with host phenotypes, but mechanistic and experimentally demonstrated links between key community members and specific aspects of host biology are difficult to establish with these methods alone. The goals of the examples presented below are (i) to evaluate the representation of readily cultured phylotypes in the human gut microbiota; (ii) to profile the dynamics of these cultured communities in a mammalian gut ecosystem; and (iii) to determine whether a clonally arrayed, personalized strain collection could be constructed to serve as a foundation for reassembling varying elements of a human's gut microbiota in vitro or in vivo.

Example 1. Estimating the abundance of readily cultured bacterial phylotypes in the distal human gut

[0111] To estimate the abundance of readily cultured bacterial phylotypes in the distal human gut, primers were used to amplify variable region 2 (V2) of bacterial 16S ribosomal RNA (rRNA) genes present in eight freshly discarded fecal samples obtained from two healthy, unrelated anonymous donors living in the United States ($n = 1$ complete sample per donor at $t = 1, 2, 3,$ and 148 d). Amplicons were subjected to multiplex pyrosequencing, and the results were compared with those generated from DNA prepared from $\sim 30,000$ colonies cultured from each sample, under strict anaerobic conditions and harvested after 7 d at 37°C on a rich gut microbiota medium (GMM) composed of commercially available ingredients ("cultured" samples; details of the culturing technique are given in Materials and Methods, and a description of GMM is given in **Table 2**). The resulting 16S rRNA datasets were de-noised to remove sequencing errors, reads were grouped into operational taxonomic units (OTU) of $\geq 97\%$ nucleotide sequence identity (ID), and chimeric sequences were eliminated (Materials and Methods).

[0112] In total, 632 distinct 97%ID OTU were observed in the complete samples, and 316 were identified in the cultured samples. The average abundance of cultured OTU in the complete samples was 0.4%, but the average abundance of uncultured OTU (i.e., those observed in the complete but not the cultured samples) was

significantly lower (0.06%; $P < 10^{-6}$ by an unpaired, two-tailed student's *t* test, not assuming equal variances).

Example 2. Evaluating the representation of readily cultured taxa in the human gut microbiota.

[0113] To evaluate the representation of readily cultured taxa in the human gut microbiota at varying phylogenetic levels, taxonomic designations were assigned to each 97%ID OTU (Materials and Methods). Each 16S rRNA read from the complete fecal sample was scored as "cultured" if it had a taxonomic assignment that also was identified in the corresponding cultured population. If a 97%ID OTU in the complete sample could not be placed in any known taxonomic group, it was scored as "cultured" only if the same 97%ID OTU was observed in the cultured sample. This analysis indicated that 99% of the 16S rRNA reads derived from the complete fecal samples from either donor belong to phylum-, class- and order-level taxa that also are present in the corresponding cultured sample; $89 \pm 4\%$ of the reads are derived from readily cultured family-level taxa, and $70 \pm 5\%$ and $56 \pm 4\%$ belong to readily cultured genus- and species-level taxa, respectively (**FIG. 1A Upper**). Two alternate taxonomic binning methods, the Ribosomal Database Project (RDP) Bayesian classifier v2.0 and an arbitrary %ID cutoff, produced similar results (**FIG. 2 A–F**). Control experiments described in Materials and Methods indicate that at least 98% of the reads generated from 30,000 pooled colonies are not derived from nongrowing or lysed bacteria (the percentage of reads from the original fecal samples that is derived from dead cells is unknown).

[0114] Unsupervised hierarchical clustering of the complete and cultured microbial communities, across the two donors and four time points, revealed that cultured samples cluster separately from those that had not been cultured. Both phylogenetic and nonphylogenetic metrics segregate cultured samples by donor, suggesting that the distinctiveness of each donor's microbiota is preserved in their collections of readily cultured representatives (**FIG. 2 G and H**).

[0115] Shotgun DNA pyrosequencing was performed to determine the degree to which predicted functions contained in the composite genomes of the

complete human fecal microbial communities were represented in the corresponding collection of cultured microbes [$n = 4$ samples (one complete and one cultured from each of two donors); $119,842 \pm 43,086$ high-quality shotgun reads per microbiome; average read length, 366 nt]. On average, 90% of the 2,302 distinct KEGG Orthology (KO) annotations identified in the two uncultured samples also were observed in the cultured communities (**FIG. 1B, FIG. 3 A and B**). This high percentage of functional representation also was observed when the microbiomes were subjected to alternate annotation schemes: On average, 94% of 929 enzyme commission (EC) assignments and 95% of 216 level 2 KEGG pathways associated with the complete fecal samples also were detected in the cultured communities (**FIG. 3 C–F**).

Example 3. Comparing functions represented in the complete and cultured microbiota.

[0116] To compare further the functions represented in the complete and cultured microbiota independent of annotation, antibiotic-resistance genes were captured from their microbiomes in expression vectors in *Escherichia coli*. Each *E. coli* library contained ~1 GB of 1.5- to 4-kB fragments of microbiome DNA subcloned into an expression vector and was screened against a panel of 15 antibiotics and clinically relevant antibiotic combinations (**Table 3**). Genes encoding resistance to the same 14 antibiotics were captured in libraries prepared from complete and cultured fecal samples (**FIG. 3G and Table 4**). In one example, a screen for DNA fragments that confer resistance to the aminoglycoside amikacin produced candidate genes from the microbiomes of both complete and cultured microbial communities from Donor 1 but not from Donor 2. Two genes conferring amikacin resistance (either the 16S rRNA methylase *rmtD* or the aminoglycoside phosphotransferase *aphA-3*) were identified in 70% of the DNA fragments captured in selections for this phenotype. Direct culturing of the original fecal communities in the presence of amikacin confirmed that this resistance function is significantly enriched in the readily cultured microbiota of Donor 1 compared with Donor 2 ($P < 0.005$ based on triplicate samples; unpaired, two-tailed student's *t* test assuming equal variances) (**FIG. 3H**); and PCR analysis showed that many of the amikacin resistant fecal strains harbor *rmtD* or *aphA-3*. Sequencing the 16S rRNA

genes of a subset of these isolates indicated that *rmtD* is present in strains of *Bacteroides uniformis*, *B. caccae*, and *B. thetaiotaomicron* in this donor (although, notably, not in the sequenced type strains of these species) and that *aphA-3* is contained in the genome of a member of the genus *Desulfotomaculum* (order Clostridiales).

Example 4. Determining whether an individual's readily cultured community exhibits behavior in vivo mirroring that of the individual's complete microbial community.

[0117] To determine whether a community composed of an individual's readily cultured bacteria exhibits behavior in vivo mirroring that of the individual's complete microbial community, 9-wk-old C57Bl6/J germfree mice were colonized with a complete or cultured microbiota from each of the two human donors ($n = 5$ recipient mice per sample type). A fecal sample from each donor was divided after collection, and one aliquot was gavaged directly into one group of recipient mice; the other aliquot was cultured on GMM plates for 7 d, as above, harvested, and introduced into a second group of recipient animals. Mice were maintained on a standard autoclaved low-fat, plant polysaccharide-rich (LF/PP) chow diet before and 4 wk after gavage. 16S rRNA analysis of fecal samples collected from these mice at the end of the 4-wk period indicated that the complete and the cultured communities were influenced similarly by host selection: $91 \pm 3\%$ of the 16S rRNA reads identified from mice colonized with a human donor's complete fecal microbiota were derived from genus-level taxa that also were identified in the mice colonized with the cultured microbial community from the same donor (**FIG. 1A Lower**). Importantly, control experiments demonstrated that the harvested, actively growing colonies gavaged into each germfree mouse are able to exclude nongrowing species that might be present on GMM plates from establishing themselves in recipient animals (details are given in Materials and Methods).

[0118] Luminal material was collected from the proximal, central, and distal portions of the small intestine, cecum, and colon of mice colonized with either the complete or cultured communities from each of the two human donors. V2-directed

bacterial 16S rRNA sequencing revealed similar geographic variations in community structures (FIG. 1C and FIG. 4 A–C).

Example 5. Determining whether the similarities in community composition in vivo extend to similarities in community gene content.

[0119] To determine whether the similarities in community composition in vivo extend to similarities in community gene content, the same fecal DNA samples that had been prepared from these mice after 4 wk on the LF/PP diet for 16S rRNA analyses were subjected to shotgun pyrosequencing ($n = 4$ samples; $87,357 \pm 30,710$ reads per sample). As with the 16S rRNA analysis, comparisons of the representation of KOs in the various microbiome samples revealed an even greater correlation between complete and cultured communities after they had been subjected to in vivo selection than before their introduction into mice (FIG. 1B, FIG. 3 A–F).

Example 6. Assessing whether a complex community of cultured microbes could restore epididymal fat pad weights to the levels associated with complete microbial communities.

[0120] Previous comparisons of adult germfree mice with those that harbor gut microbial communities (either conventionally raised animals or formerly germfree animals colonized from mouse or human donors) have shown that the presence of a complete gut microbiota is associated with increased adiposity. In comparison, colonization of germfree mice with a single, readily cultured, prominent human gut symbiont (*Bacteroides thetaiotaomicron*) or with a defined community of 12 bacterial species prominently represented in the distal human gut is insufficient to restore epididymal fat pad stores to levels observed in conventionally raised animals (data not shown). To assess whether a complex community of cultured microbes could restore epididymal fat pad weights to the levels associated with complete microbial communities, we evaluated mice colonized with the complete or the cultured fecal communities from the two human donors. All animals displayed significantly greater fat pad to body weight ratios than germfree controls, and no significant difference was

observed in adiposity between mice colonized with the donors' complete or cultured microbiota (**FIG. 4D**).

Example 7. Testing whether a microbial community consisting only of cultured members recapitulates known diet-induced changes in microbial community structure in vivo.

[0121] Mice colonized with a complete human microbiota undergo drastic changes in microbial community structure (even after a single day) when shifted from LF/PP chow to a high-fat, high-sugar Western diet. To test whether a microbial community consisting only of cultured members recapitulates this behavior in vivo, the four groups of gnotobiotic mice colonized with the complete or cultured microbes from two unrelated human donors were monitored by fecal sampling before, during, and after a 2-wk period when they were placed on the Western diet (samples were collected at days 4, 7, and 14 of the first LF/PP phase, then 1 d before and 3, 7, and 14 d after initiation of the Western diet phase, and finally 1, 3, 8, and 15 d after the return to the LF/PP diet). 16S rRNA-based comparisons of fecal communities were performed using both phylogenetic and nonphylogenetic distance metrics. With either metric, principal coordinates analysis (PCoA) revealed that mice colonized with the complete or cultured samples maintain communities that cluster first by donor (principal coordinate 1; PC1) and that the complete and cultured communities from both donors respond to the diet shift in a similar manner [principal coordinate 2 (PC2); **FIG. 5A–C and FIG. 6A and B**]. Like the transplanted complete microbiota examined here and in previous reports, the cultured microbiota responded to this Western diet by increasing the relative proportion of representatives of one class of Firmicutes (the Erysipilotrichi) and decreasing the relative proportion of the Bacteroidia class (**FIG. 6C**). Notably, of the 18 species-level phylotypes significantly affected by diet shift in the mice containing the complete microbiota of both human donors, 14 were detected and demonstrated the same statistically significant response in mice colonized with readily cultured taxa (**FIG. 5D and FIG. 7**).

[0122] The fecal microbiomes of LF/PP-fed mice harboring complete or cultured communities from each of the two unrelated donors were compared with

microbiomes sampled after these mice consumed the Western diet for 14 d ($101,222 \pm 24,271$ reads per sample). The representation of level 2 KEGG pathway functions was highly concordant on both diets, with one exception: Genes encoding phosphotransferase system (PTS) pathways for carbohydrate transport were significantly overrepresented in Western diet-fed mice harboring complete or cultured communities from either donor (**FIG. 5E and FIG. 8A–D**). Higher-resolution KO level annotations confirmed that the diet-based PTS pathway enrichment reflected increased representation of multiple carbohydrate transporters (**FIG. 8E and F**). These results emphasize that the similar taxonomic restructuring of complete and cultured communities in response to diet is accompanied by similar changes in community gene content.

Example 8. Using gnotobiotic mice as biological filters to recover collections of readily cultured microbes.

[0123] Because human gut communities composed of readily cultured members exhibit responses to host diet that mirror those characteristic of a complete microbiota, the possibility that gnotobiotic mice can be used as biological filters to recover collections of readily cultured microbes, obtained from selected human hosts, that are enriched for certain properties [e.g., the ability to prosper (bloom) when exposed to specific foods or food ingredients] was explored. To this end, fecal samples from mice colonized with the complete or corresponding cultured human gut microbial communities from the two unrelated donors and fed the LF/PP or Western diets were collected directly into anaerobic medium and then plated on prereduced GMM plates (**FIG. 9A**). After 7-d incubation, V2-directed 16S rRNA profiling of these plated microbial collections confirmed that these populations of cultured microbes can be reshaped deliberately in vivo and then recovered in vitro (**FIG. 9B and FIG. 10**). On either diet, cultured populations showed significantly greater resemblance to the in vivo communities from mice consuming the same diet than to the in vivo communities from the same mice consuming the alternate diet ($P < 10^{-11}$, unpaired two-tailed student's *t* test of within donor distances shown in **FIG. 9B**, assuming equal variances).

Example 9. Creating arrayed species collections representing the bacterial diversity of the gut microbiota.

[0124] The strict anaerobic techniques used here, compounded with highly diverse colony morphologies across taxa, complicate efforts to pick and isolate individual colonies at a scale sufficient to capture the bacterial diversity represented on the culture plates. Therefore, a most probable number (MPN) technique was used for creating arrayed species collections in a multiwell format without colony picking. First, the dilution point for a fecal sample that yields 70% empty wells (no detectable growth) after inoculation into 384-well trays and 7-d anaerobic incubation was empirically determined. Assuming that the distribution of cells into the wells follows a Poisson distribution, a dilution that leaves 70% of wells empty should yield nonclonal wells (that is, wells that received more than one cell in the inoculum) only 5% of the time; the remainder should be clonal (**FIG. 11F and G**). At this dilution, ten 384-well trays should yield ~1,000 clonal wells. The two-step barcoded pyrosequencing scheme outlined in **FIG. 12A** was developed to assign a 16S rRNA sequence to the isolate(s) present in each turbid well.

[0125] This approach was used to create an archived, personalized culture collection of ten 384-well trays from one of the human donors. 16S rRNA sequences could be assigned to more than 99% of growth-positive wells. One advantage of clonally arrayed collections is that the effects of 16S rRNA primer bias encountered using DNA templates prepared from complex microbial populations are minimized when wells contain a single taxon. This point is illustrated by the known bias of most commonly used primers against *Bifidobacteria* spp.: Members of this genus were better represented among the set of 16S rRNA genes produced from individual wells than among those observed in complex communities harvested from GMM plates.

[0126] After the archived trays had been frozen under anaerobic conditions and stored at -80°C for 7 mo, recovery of organisms from wells exceeded 60%. Full-length 16S rRNA sequences generated from these recovered strains matched the assignments from the barcoded pyrosequencing data in every case, suggesting that the dilutions did follow a Poisson distribution as predicted. Like 16S rRNA-based community profiling, such collections may miss rare, but important, members of the

microbiota; seeding additional 384-well trays with the diluted sample will capture additional phylotypes (**FIG. 11H**). In total, this individual's culture collection contained 1,172 taxonomically defined isolates from four different phyla, seven classes, eight orders, 15 families, 23 genera, and 48 named bacterial species. Novel isolates were encountered at the family-, genus-, and species- levels, and 69% of the complete community had a genus-level representative in the arrayed collection (**FIG. 12B**). As a frame of reference, we identified a total of 159 human fecal or gut bacterial species from humans worldwide (including pathogens) in the German Resource Centre for Biological Material (DSMZ) culture collection (Materials and Methods). As such, personalized microbiota collections can complement those of international repositories by capturing strains that coexist in a shared habitat where community structure and host parameters can be measured.

[0127] The ability to capture this level of diversity after MPN dilution in these arrayed collections indicates that it is unlikely that interspecies syntrophic relationships by themselves are sufficient to explain the diversity observed on the GMM agar plates. On the other hand, these personalized arrayed culture collections should help identify obligate syntrophic relationships (e.g., by analyzing the patterns of co-occurrence of taxa in wells harboring more than one phylotype or by comparing arrayed collections in which one set of trays contains a candidate syntroph deliberately added to all wells).

Discussion for Examples 1-9.

[0128] The Examples presented above show that it is possible to capture a remarkable proportion of a person's fecal microbiota using straightforward anaerobic culturing conditions and easily obtained reagents. Variations in culturing conditions, including components that are not commercially available (e.g., sterile rumen or human fecal extracts) and other approaches for more closely approximating a native gut habitat, undoubtedly will allow additional members of the human gut microbiota to be cultured in vitro. These personal culture collections can be generated from humans representing diverse cultural traditions and various physiologic or pathophysiologic states. A key opportunity is provided when anaerobic culture initiatives are combined

with gnotobiotic mouse models, thereby allowing culture collections to be characterized and manipulated in mice with defined (including engineered) genotypes who are fed diets comparable to those of the human donor, or diets with systematically manipulated ingredients. Temporal and spatial studies of these communities can be used to identify readily cultured microbes that thrive in certain physiological and nutritional contexts, creating a discovery pipeline for new probiotics and for preclinical evaluation of the nutritional value of food ingredients. Based on their in vivo responses, clonally archived cultured representatives of a person's microbiota can be selected for complete genome sequencing (including multiple strains of a given species-level phylotype) to identify potential functional variations that exist or evolve within a species occupying a given host's body habitat. Coinciding with the introduction of yet another generation of massively parallel DNA sequencers, this approach also should allow vast scaling of current sequencing efforts directed at characterizing human (gut) microbial genome diversity, evolution, and function. Recovered organisms also could be used as source material for functional metagenomic screens (bio-prospecting). Guided by the results of metagenomic studies of human microbiota donors, components of a personalized collection that have coevolved in a single host can be reunited in varying combinations in gnotobiotic mice, potentially after genome-wide transposon mutagenesis of selected taxa of interest, for further mechanistic studies of their interactions and impact on host phenotypes.

Materials and Methods for Examples 1-9.

[0129] **Culturing of Fecal Microbiota.** Freshly discarded fecal samples from two anonymous unrelated human donors were transferred into an anaerobic chamber (Coy Laboratory Products) within 5 min of their collection. Samples were placed in prereduced PBS with 0.1% cysteine (PBSC) $15 \text{ mL}^{-1} \text{ g}^{-1}$ feces. The fecal material was suspended by vortexing for 5 min, and the suspension was allowed to stand at room temperature for 5 min to permit large insoluble particles to settle to the bottom of the tube. 16S ribosomal RNA (rRNA) sequencing of the starting material and of the supernatant obtained after the settling step showed no significant differences in community composition. This settling step dramatically increased the reproducibility of

subsequent dilutions. Diluted (10^{-4}) samples were plated on plates 150 mm in diameter containing prereduced, nonselective Gut Microbiota Medium (GMM) (**Table 2**) so that colonies were dense but distinct (~5,000 colonies per plate) after a 7-d incubation at 37°C under an atmosphere of 75% N₂, 20% CO₂, and 5% H₂. GMM is modified from tryptone-yeast-glucose agar with 6% NaCl (TYGS) medium and contains only commercially available components; to discourage colony overgrowth, the concentration of glucose, tryptone, and yeast extract is reduced fivefold compared with TYGS. Colonies were harvested en masse from each of six plates by scraping with a cell scraper (BD Falcon) into 10 mL of prereduced PBSC. Stocks were generated by adding prereduced glycerol containing 0.1% cysteine to the fecal or cultured samples (final concentration of glycerol, 20%). Stocks were stored in anaerobic glass vials in a standard -80°C freezer.

[0130] To determine the optimal number of plates to be surveyed for each fecal sample, a freshly discarded sample from one of the anonymous human donors was processed as above, and the 10^{-4} dilution was plated on 10 prereduced GMM plates. De-noised, chimera-checked variable region 2 (V2)-directed 16S rRNA reads generated from the pooled colonies obtained from each plate separately after a 7-d incubation were assigned to operational taxonomic units (OTU) at 97% nucleotide sequence identity (ID) using QIIME v1.1. Each additional plate contributed new OTU, although, as shown by the rarefaction curves plotted in **FIG. 11A**, the contribution of each added plate fell to <20 new OTU after approximately six plates. For this reason, each subsequent cultured sample reflects pooled scraped material from six plates.

[0131] **Gnotobiotic Mouse Husbandry.** Germfree adult male C57BL/6J mice were maintained in plastic gnotobiotic isolators. Mice were housed under a strict 12-h light/dark cycle and fed a standard, autoclaved low-fat/plant polysaccharide-rich (LF/PP) chow diet (B&K Universal.) ad libitum. Mice were colonized by gavage (0.2 mL of the resuspended fecal material or pooled cultured organisms recovered from GMM after 7-d incubation as above, per germfree recipient). Animals receiving different microbial inoculations were placed in separate gnotobiotic isolators before gavage; once gavaged, they were caged individually.

[0132] After 4 wk of acclimatization on the LF/PP diet, mice were transitioned to Western diet (Harlan-Teklad TD96132) ad libitum for 2 wk and then were returned to the LF/PP diet for 2 wk. 16S rRNA analysis of fecal samples collected 1, 4, 7, and 14 d after gavage indicated that both complete and cultured communities reached a steady state well before the diet transition. During the initial LF/PP diet phase, fecal samples were collected at postgavage days 4, 7, 14, and either on day 32 when the input community was a complete microbiota or on day 25 in the case of an input cultured community. Mice were sampled on days 1, 3, 7, and 14 after the shift to the Western diet and on days 1, 3, 8, and 15 upon return to LF/PP chow. The animals then were fasted for 24 h and returned to the LF/PP diet for 1 wk before they were killed.

[0133] Fecal samples for subsequent culture were collected from each mouse directly into BBL thioglycollate medium (BD), transported to an anaerobic chamber within 30 min, then diluted and plated on GMM as above. Fecal samples from fasted mice were not cultured because 16S rRNA analysis did not show significant changes in community composition at either the 12-h or 24-h fasting time points. After mice were killed, the intestine was subdivided into 16 segments of equivalent length numbered from 1 (proximal) to 16 (distal). Contents from small intestine segments 2, 5, and 13, plus cecum and colon contents, were snap frozen in liquid nitrogen and stored at -80°C .

[0134] **DNA Extraction and Purification.** Fecal samples (0.1 g) were resuspended into 710 μL of 200 mM NaCl, 200 mM Tris, 20 mM EDTA (Buffer A) plus 6% SDS). After the addition of 0.5 mL of 0.1-mm zirconia/silica beads (BioSpec Products) and 0.5 mL of phenol/chloroform/isoamyl alcohol, pH 7.9 (Ambion), cells were lysed by mechanical disruption with a bead-beater (BioSpec Products) for 3 min. Samples were centrifuged for 3 min at $6,800 \times g$, and the aqueous phase was collected and subjected to a second phenol-chloroform-isoamyl alcohol extraction using Phase Lock Gel tubes (5 Prime). DNA in the aqueous phase was precipitated by the addition of an equal volume of isopropanol and 0.1 volumes of 3 M sodium acetate, pH 5.5 (Ambion). After overnight incubation at -20°C , samples were centrifuged for 20 min at 4°C at $18,000 \times g$, and the supernatant was removed. Pelleted DNA was washed once

with 0.5 mL of 100% ethanol, and dried using a vacuum evaporator. DNA pellets then were resuspended in 0.2 mL of Tris-EDTA containing 4 µg RNaseA (Qiagen). Crude DNA extracts were column-purified using the Rapid PCR Purification Kit (Marligen Biosciences). DNA concentrations were adjusted to 50 ng/µL for subsequent 16S rRNA or shotgun pyrosequencing.

[0135] **16S rRNA Sequencing.** The V2 region of bacterial 16S rRNA genes was subjected to PCR amplification. PCR reactions were carried out in triplicate using 2.5× Master Mix (5 Prime), forward primer FLX-8F; 5'-GCCTTGCCAGCCCGCTCAGTCAGAGTTTGATCCTGGCTCAG-3' (SEQ ID NO 1); (the 54 FLX Amplicon primer B sequence is underlined, and the 16S rRNA primer sequence 8F is italicized), barcoded reverse primer FLX-BC-338R; 5'-GCCTCCCTCGCGCCATCAGNNNNNNNNNNNNNCATGCTGCCTCCCGTAGGAGT-3'; (SEQ ID NO 2) (the 454 FLX Amplicon primer A sequence is underlined, "Ns" indicate the barcode sequence, and the 16S rRNA primer sequence 338R is shown in italics), and 50 ng input DNA purified as described above. Reactions were incubated for 2 min at 95°C, followed by 30 cycles of 20 s at 95°C, 20 s at 52°C, and 1 min at 65°C. Triplicate reactions were pooled, inspected by gel electrophoresis, and purified on AMPure beads (Agencourt Biosciences). For each barcoded primer, negative control reactions lacking input DNA were conducted in parallel. Purified samples were combined at equimolar concentrations and sequenced with FLX chemistry on a 454 pyrosequencer (Roche).

[0136] **16S rRNA Sequence Analysis.** Metadata for all 500 samples, including barcodes, are provided in **Table 5**. For 16S rRNA sequence analysis, sequences were preprocessed to remove reads with low-quality scores (sliding window set to 50 bp), ambiguous characters, and incorrect lengths (<200 or >300 bp). Reads passing these criteria were assigned to specific samples based on their error-corrected barcode sequence, de-noised using default parameters, grouped into OTU at 97%ID, and a representative sequence was selected from each OTU using default parameters in QIIME v.1.1. These representative sequences then were filtered for possible chimeric sequences using ChimeraSlayer (microbiomeutil.sourceforge.net) with default parameters (sequences designated "unknown" were not discarded). Filtered datasets

were subsampled to 1,000 sequences per sample (the only exceptions were the datasets from the human-derived complete and cultured samples collected at the day 148 time point shown in **FIG. 1A**, which were subsampled to 5,000 sequences). Beta-diversity calculations were conducted using Jaccard (nonphylogenetic) and UniFrac (phylogenetic) metrics in QIIME v1.1.

[0137] **Weighted Taxonomic Analysis.** To quantify the representation of cultured and uncultured lineages in microbial communities, the presence or absence of each phylum-, class-, order-, family-, genus-, or species-level phylotype assigned to sequences in the complete sample(s) was determined in the cultured sample(s). Analysis of full-length and V2- region delimited sequences from the 16S rRNA genes of taxonomically defined bacteria indicated that discrete %ID cutoffs do not correspond closely to established taxonomic levels: Histograms of the distributions of %ID values of 16S rRNA sequences between representatives of two species in the same genus, two genera in the same family, and so forth, overlap to a large extent. (**FIG. 11B** shows comparisons of 16S V2 regions, which are used commonly in multiplex pyrosequencing studies, between 4,041 bacterial species selected from the SILVA database v102.) For this reason, a taxonomic assignment method was primarily used to compare uncultured and cultured communities across taxonomic levels rather than approximating taxonomic groups by selecting arbitrary %ID cutoffs to represent each taxonomic level. (For example, **FIG. 11B** illustrates that there is no clear % ID cutoff that distinguishes species-level from genus-level groups or family-level from order-level groups.) As an alternate, taxonomy- independent approach, de-noised 16S rRNA sequences were grouped into OTU at a range of %ID cutoffs (80%, 90%, 95%, 97%) using uclust in QIIME v1.1. At each %ID cutoff, OUT were filtered for chimeras as above. The proportion of reads in the uncultured sample that belonged to OTU also identified in the corresponding cultured sample were determined as in **FIG. 1A** (see **FIG. 2A–F** for a comparison of results obtained from different assignment methods).

[0138] **SILVA-VOTE: A Computational Pipeline for Improved Accuracy in Taxonomic Assignments of V2 16S rRNA Sequences.** Commonly used tools for taxonomy assignment often failed to assign correctly V2 16S rRNA sequences derived from known human gut microbes. To generate a nonredundant, curated 16S rRNA

database for taxonomy assignment, the v102 SILVA database was downloaded prefiltered for redundancy at a 99%ID (SSURef_102_SILVA_NR_99.fasta; <http://www.arb-silva.de>). This database is composed of 262,092 full-length sequences from the small subunit rRNAs of Eukaryotes, Bacteria, and Archaea. A total of 297 sequences whose accession numbers had been removed from or modified by GenBank or were not associated with a complete National Center for Biotechnology Information (NCBI) taxonomy (i.e., phylum, class, order, family, genus, and species designations) were excluded. The remaining sequences were aligned using PyNast as implemented in QIIME v1.1: 224,899 sequences were aligned successfully and contained more than 90% of the V2 region. These V2 sequences were filtered for redundancy by clustering and selection of a representative sequence from each cluster, using uclust at a 99% identity. To assign consensus taxonomies to the representative sequences, we applied a 75% majority voting scheme: For each taxonomic level, the representative sequence was assigned a taxonomic designation if more than 75% of the sequences within the cluster shared the same assignment; otherwise, the cluster was labeled “unknown” at that taxonomic level. Taxonomic designations of sequences within a cluster that included the nonunique identifiers “unknown,” “uncultured,” “candidatus,” or “bacterium” were not considered in the 75% majority vote for taxonomy assignment of the representative sequence. Species-level annotations with numbers or decimal points (which in almost all cases refer to strains rather than species) also were excluded. After removal of sequence clusters with little or no consensus taxonomy (i.e., with 50% or more of the taxonomic levels labeled “unknown” after the voting analysis), 34,181 nonredundant, annotated, bacterial 16S rRNA V2 sequences remained and were designated as our reference database.

[0139] To assign taxonomy to the 16S rRNA V2-amplicon pyrosequencing reads, significant matches to the 34,181-sequence reference database were identified by BLAST (the top 100 hits with an e-value cutoff of $\leq 10^{-30}$ were retained). All the BLAST hits with a score within 10% of the score of the best BLAST hit were considered for the taxonomy assignment. Taxonomy was assigned for each phylogenetic level independently by using a majority voting scheme: A read was assigned a taxonomic designation if 50% or more of the selected reference sequences (whose BLAST scores

were within 10% of the top score for that query sequence) shared the same taxonomic assignment. As above, sequences designated “unknown” were not taken into account for the voting. When no assignment was conserved in >50% of the selected BLAST hits, the query sequence was labeled as “nonidentified” at that taxonomic level. Data were normalized by the abundance of each taxonomic group in the original (uncultured) sample. For analysis of microbial communities from mice, taxonomic groups observed in fewer than two replicate animals were omitted.

[0140] To test this method, 16S rRNA V2 sequences were extracted from the genomes of 66 human gut microbes. Taxonomy was assigned to these test sequences in QIIMEv.1.1 using three methods: (i) the Ribosomal Database Project Bayesian classifier (v.2.0); (ii) a BLAST top hit-based query of the Greengenes core sequence set [composed of 4,938 sequences; downloaded September 15, 2009; (greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/core_set_aligned_fast a. imputed) (7)]; and (iii) with SILVA-VOTE. Comparison of these results suggests that SILVA-VOTE yields a significantly increased number of correct taxonomic assignments, particularly at the genus and species levels (FIG. 11C).

[0141] **Control Experiments to Address the Influence of Lysed or Nongrowing Cells on 16S rRNA Datasets from Colonies Collected from Agar Plates.** Although each average-sized colony among the ~30,000 colonies obtained from each cultured sample likely contributed ~10⁹ cells to the pooled population, in theory genetic material from lysed or nongrowing cells also could contribute to the sequences obtained from the plated samples. To test this possibility directly, fecal samples were diluted to the same level as above and plated onto GMM and also onto plates containing ingredients that should not support growth of bacteria and thus represent the background expected if 100% of the plated material was nongrowing or lysed [control PARC plates contained Phosphate buffer, noble Agar, Resazurin (oxygen indicator), and Cysteine (reducing agent)]. After a 7-d anaerobic incubation, no colonies were detected on the PARC plates. Twenty randomly selected single colonies from the GMM plates were picked, an aliquot was reserved for 16S rRNA gene sequencing, and the remainder was pooled with the scraped surfaces of the PARC plates. Sequencing this pool revealed that >98% of the 16S rRNA reads could be attributed to the 20 colonies

from the GMM plates; among the remainder, none belonged to OTU represented by more than two reads per 1,000. Together, these findings suggest that at least 98% of the reads generated from 30,000 pooled colonies are not derived from nongrowing or lysed bacteria.

[0142] **Testing the Possible Contribution of Lysed or Nongrowing Cells to Microbial Communities in Gnotobiotic Mice Gavaged with a Readily Cultured Human Gut Microbiota.** As noted in the main text, the initial cultured inoculum was prepared by scraping GMM plates en masse. In theory, this input could include cells that did not actually grow in these conditions but instead remained dormant, below the limit of detection by 16S rRNA sequencing, over the 7-d in vitro incubation period. To determine whether such non-growing taxa contributed to the distal gut communities of mice that received a readily cultured microbiota, a control sample containing material harvested from PARC plates and pooled with 20 visible colonies picked directly from GMM plates was introduced into five age-matched, individually caged, germfree mice fed a LF/PP diet. Fecal samples were collected at 3, 7, and 14 d postgavage and were subjected to V2-targeted 16S rRNA pyrosequencing. Community composition, as determined both by alpha-diversity and beta-diversity metrics, was stable after the 7-d time point (average UniFrac distance within 7-d or 14-d samples = 0.319; average distance between 7-d and 14-d samples = 0.321; $P > 0.89$ based on unpaired, two-tailed student's t test; **FIG. 11D–E**). When 16S rRNA sequences obtained from the 14-d fecal samples were compared with sequences obtained from the 20 picked colonies, it was discovered that the mice harbored only two OTU, both mapping to *Akkermansia muciniphila*, that could not be attributed to the 20 colonies and that were not observed on GMM plates in any other experiments. *Akkermansia muciniphila* type strain ATCC BAA-835 contains three 16S rRNA genes and grows readily on GMM. This species was a minor component in the fecal microbiota of the two donors (no or one read per 1,000 reads from eight samples collected over time); in fecal communities sampled from mice that received the readily cultured component of either donor's microbiota, abundance averaged 1.8% across all time points. These findings indicate that just 20 actively growing colonies are able to exclude

virtually all nongrowing species that may be present on GMM plates from colonizing germfree mice.

[0143] **Shotgun Pyrosequencing.** Five hundred-nanogram aliquots of DNA prepared from selected complete and cultured microbiota were sheared and ligated to the default 454 Titanium multiplex identifiers (MIDs; Roche Rapid Library Preparation Method Manual, GS FLX Titanium Series, October 2009). 16S rRNA sequencing of samples from individually caged mice colonized with the same community indicated a high degree of similarity between individual animals; for this reason, fecal DNAs from replicate mice were pooled ($n = 3\text{--}5$ mice per pool), and the 12 pooled samples, each labeled with a unique MID, were sequenced in a single 454 Titanium run.

[0144] Shotgun pyrosequencing reads were parsed by MID and filtered to remove short sequences (<60 bp), low-quality sequences (three or more N bases in the sequence or two continuous N bases), and replicate sequences ($>97\%$ ID over the length of the read, with identical sequences over the first 20 bases). Reads reflecting host DNA contamination were identified by BLAST (against the mouse genome for samples isolated from mice and against the human genome for all other samples) and were removed in silico ($\geq 75\%$ identify, $E\text{-value} \leq 10^{-5}$, $\text{bitscore} \geq 50$). Remaining sequences were queried against the KEGG Orthology (KO) database (v52) with a Blastx e-value cutoff of 10^{-5} . KO assignments were mapped further to Enzyme Classification (EC) and KEGG pathway annotations.

[0145] **Bio-Prospecting for Antibiotic-Resistance Genes in Uncultured and Readily Cultured Microbiota.** DNA fragments from complete and cultured communities were cloned into an expression vector, electroporated into *Escherichia coli*, and screened for their ability to confer resistance to 15 different antibiotics. To this end, $10\mu\text{g}$ of DNA purified from the two human donors' fecal microbiota, from the derived cultured communities, plus pooled contents of the arrayed strain collection were sheared to 1.5- to 4-kB fragments (Bioruptor XL), followed by size selection (1% agarose gel electrophoresis). Sheared DNA then was endrepaired (Epicentre EndIt Kit), column-purified (Qiagen Qia- Quick PCR Purification Kit), and concentrated in a vacuum evaporator (SpeedVac). The expression vector pZE21-MCS1 was prepared by

PCR amplification using primers flanking the HincII site [pZE21_126_146FOR, 5'-GACGGTATCGATAAGCTTGAT-3' (SEQ ID NO 3); pZE21_111_123rcREV, 5'-GACCTCGAGGGGGGGG-3' (SEQ ID NO 4)] to linearize the vector, gel-purification of the linear product, dephosphorylation (calf intestinal phosphatase), and column purification. Approximately 500ng of the DNA fragments were ligated to 100 ng of the linearized vector in an overnight ligation reaction (Epicentre FastLink Ligation Kit). The ligation reaction was desalted by dialysis in double-distilled H₂O and electroporated into *E. coli* MegaX DH10B T1R cells (Invitrogen). After 1 h of recovery, a small (1 µL) aliquot of the library was titered with serial dilutions onto LB agar plates containing 50µg/mL kanamycin (to select for pZE21 transformants) and incubated at 37°C for 16 h. The insert size distribution for each library was characterized by gel electrophoresis of amplicons obtained using primers flanking the HincII site in the multiple cloning site of pZE21 MCS1. The total size of each library was estimated by multiplying average insert size by the number of cfu in a given library. The remainder of the recovered cells was inoculated into 10mL LB containing 50µg/mL kanamycin and grown overnight, with shaking, at 27°C for ~16 h. The culture subsequently was diluted with an equal volume of LB medium containing 30% glycerol and stored at -80°C before screening.

[0146] For functional selections, 100µL of each library freezer stock (corresponding to $0.5-1 \times 10^8$ cfu) was plated on an LB agar plate containing kanamycin (50 µg/mL) plus one of 15 different antibiotics (**Table 3**). The total number of cells plated on each antibiotic represented ~10 copies of each original unique transformant. Antibiotic-resistant colonies were scored after plates had been incubated at 37°C for 16 h. Inserts contained in colonies with amikacin-, piperacillin- and piperacillin/tazobactam-resistant phenotypes were subjected to bidirectional Sanger sequencing (Beckman Coulter Genomics) using primers pZE21_81_104_57C (5'-GAATTCATTAAAGAGGAGAAA GGT-3'; SEQ ID NO 5) and pZE21_151_174rc_58C (5'-TTTCGTTTTATTTGATGCCTCTAG-3'; SEQ ID NO 6). Resulting reads were trimmed to remove low-quality and vector sequences and subjected to within-library contig assembly (≥ 200 bp of 97%ID sequence required). Contigs and unassembled reads were mapped by BLAST to the National Center for Biotechnology Information (NCBI) nonredundant database and to a custom database of 122human gutmicrobial

genomes. All sequence datasets have been deposited in the NCBI Sequence Read Archive (SRA) under accession no. SRA026271.

[0147] Amikacin-resistant strains were quantified from each donor, in triplicate, by plating diluted fecal samples on GMM with and without amikacin (4,100 µg/mL; lower concentrations produced high background). Amikacin-resistant colonies were quantified after 5-d incubation under anaerobic conditions, and colony counts were normalized to the total number of colonies obtained in the absence of the antibiotic. A total of 48 fecal isolates (12 from the GMM+amikacin selection and 12 from the nonselective plates, from each of two donors) were chosen for a PCR-based survey for the amikacin-resistance genes captured in the *E. coli* libraries described above and for 16S rRNA sequencing.

[0148] **Preparation of an Arrayed Species Collection.** A single vial of the -80°C anaerobic glycerol stock containing an aliquot of a fecal sample from Donor 2 was diluted into prereduced TYGs medium lacking resazurin in an anaerobic chamber and was dispensed into prereduced 384-well flat-bottomed trays (0.17 mL per well). To determine the dilution at which a high percentage of wells received a single viable cell in the initial inoculation (**FIG. 11F**), two- and fourfold serial dilutions were performed (from 10⁻⁶ to 10⁻¹⁰) in a trial inoculation (48 wells per dilution; 0.17 mL per well). Trays were sealed with sterile foil lids and incubated anaerobically at 37°C for 5 d; the dilution at which ~30% of wells were turbid (OD₆₃₀ > 0.2) was chosen for the subsequent large-scale culturing (**FIG. 11G**). To this end, a second vial of the frozen anaerobic glycerol stock from the same donor was added to 500 mL of prereduced TYGS medium lacking resazurin at the calculated dilution and dispensed into ten 384-well culture trays (170 µL per well). Trays were sealed and incubated as above. Cells then were resuspended in each well of each tray by pipetting, and 25 µL aliquots were transferred to each of two archive trays containing 25 µL prereduced TYGS (resazurin included) plus 40% glycerol per well. The arrayed archive trays were sealed with aluminum foil, frozen on dry ice inside the anaerobic chamber, and transported on dry ice to a conventional -80°C freezer for storage. Cultures stored in this fashion remain anaerobic, as judged colorimetrically using resazurin in the medium and by recovery of strict anaerobes (as long as they are transported frozen, on dry ice, into an anaerobic chamber for strain

recovery). Another 50- μ L aliquot from the culture trays (not from the archive trays) was measured by OD₆₃₀ and stored at -80°C for PCR amplification.

[0149] Taxonomies were assigned to each strain in the 3,840-well collection by two-step barcoded 454 FLX pyrosequencing. The V2 16S rRNA region of the DNA present in each well was amplified with an invariant V2-directed forward primer and 1 of 96 barcoded V2-directed reverse primers. A 1- μ L aliquot from each well was transferred to a new tray, and cells were lysed in 10 μ L of lysis buffer (25 mM NaOH, 0.2 mM EDTA; incubation for 30 min at 95°C) followed by the addition of 10 μ L of neutralization buffer (40 mM Tris-HCl). To reduce the amplification of background DNA present from dead or lysed cells, the neutralized lysate was diluted 1:10 into EB buffer (Qiagen). To barcode each bacterial strain uniquely before amplicon sequencing, the V2 region of their 16S rRNA gene was targeted for PCR using 2 $\times$ Master Mix (Phusion HF), 2 μ L input DNA, primers 454_16S_8F and 1 of 96 barcoded (Roche Multiplex Identifiers) reverse primers (454_16S_338R_barcode1) that include a 12-bp tail sequence in a 10- μ L reaction (384-well format). Duplicate reactions were incubated for 30 s at 98°C, followed by 30 cycles of 10 s at 98°C, 30 s at 61°C, and 30 s at 72°C.

[0150] Reactions were combined so that each pool contained one representative associated with each barcode (four pools per 384-well tray), passed over a PCR cleanup column (Qiagen), and diluted to 0.5 ng/ μ L. Pools then were subjected to a second round of PCR amplification with 0.5 ng of pool DNA in a 25- μ L reaction. Reactions were incubated for 30 s min at 98°C, followed by five cycles of 10 s at 98°C, 30 s at 54°C, and 30 s at 72°C, followed by an additional 25 cycles of 10 s at 98°C and 30 s at 70°C. In this second PCR, the reverse primers were replaced with a second barcoded linker primer (454_linker_barcode2) specific to the 12-bp tail sequence added in the first PCR. In this way, the 16S rRNA V2 regions of the bacterial genomes in each initial well were associated with a unique twobarcode pointer sequence (**FIG. 12A** and legend). After the second-round PCR, reactions were pooled and run over a PCR cleanup column (Qiagen), and DNA in the expected size range (200–300 bp) was gel purified (Qiagen). The final product was quantified and subjected to multiplex 454 FLX pyrosequencing at a depth expected to yield 250 sequences per well (25% of one sequencing run).

[0151] The resulting reads were assigned to wells in the archive trays based on their associated barcodes. Of the 3,840 wells, 1,181 (30.8%) were turbid as defined by $OD_{630} \geq 0.2$; of the turbid wells, 1,172 (99.2%) had at least a single sequence with the correct barcode combination. Barcode combinations mapping to wells with culture $OD_{630} < 0.2$ also were identified in the sequencing dataset. However, the total number of reads that mapped to these wells was much lower than to turbid wells (51,925 turbid versus 14,427 nonturbid), and the percentage of mapped wells was much lower for the nonturbid subset (62.2% vs. 99.2%).

[0152] The taxonomy of the most abundant sequence associated with each barcode combination was assigned using SILVA-VOTE. These most abundant sequences also were clustered into 97%ID OTU using uclust in QIIME 1.1. To evaluate the diversity captured at varying taxonomic levels, representative 97%ID OUT sequences were assigned taxonomy using SILVA-VOTE. Reads designated "nonidentified" by SILVA-VOTE were not considered to represent an additional taxonomic group unless they were associated with a distinct higher-order taxonomic classification (e.g., sequences annotated as "Family Clostridiaceae; Genus nonidentified" were scored as representing a different genus-level group than sequences annotated as "Family Ruminococcaceae; Genus nonidentified"). Rarefaction analysis of the number of additional taxa added with each additional 384-well culture tray is shown in **FIG. 11H**. Mapping was verified by recovery of strains from the archive trays, colony purification, and full-length Sanger sequencing of their 16S rRNA gene.

[0153] **Identification of Human Gut Isolates in the German Resource Centre for Biological Material Culture Collection.** The German Resource Centre for Biological Material (DSMZ) bacterial culture collection (www.dsmz.de/microorganisms/bacteria_catalogue.php; October 14, 2010) was searched under the terms "gut," "faeces," "feces," "fecal," and "stool." Search results were filtered to exclude strains from nonhuman sources. Strains that matched the search terms without host species information were included, as were noncommensals (i.e., pathogens).

Tables for Examples 1 to 9

Table 1. Common methodological differences between culture-independent and culture-based surveys of human gut microbial diversity.		
	Culture-Independent Metagenomic Methods	Culture-Based Methods
Survey depth	Thousands to millions of 16S rRNA gene sequences generated per microbial community sample	Tens to hundreds of isolates cultured per sample
Accuracy of (bacterial) 16S rRNA gene assignments	<u>454 FLX (FLX standard) pyrosequencing platform:</u> 10x-100x over-estimate of species-level diversity due to artifacts (sequencing errors, chimeras) <u>Illumina sequencing platform:</u> $\geq 100x$ over-estimate of species-level diversity due to artifacts	Taxonomy typically defined from full length 16S rRNA genes using the more accurate Sanger method of dideoxy-chain termination sequencing, or from high-coverage assemblies of isolate genomes
Documentation	"Culturability" determined by matching 16S rRNA reads to public databases	Many readily cultured taxa not documented in public databases (in one study, 64%)

Table 2. Gut microbiota medium (GMM)

Component	Amount/L	Concentration	Comments
Tryptone Peptone	2 g	0.2%	
Yeast Extract	1 g	0.1%	
D-glucose	0.4 g	2.2 mM	
L-cysteine	0.5 g	3.2 mM	
Cellobiose	1 g	2.9 mM	
Maltose	1 g	2.8 mM	
Fructose	1 g	2.2 mM	
Meat Extract	5 g	0.5%	
KH ₂ PO ₄	100 mL	100 mM	1M stock solution pH 7.2
MgSO ₄ -7H ₂ O	0.002 g	0.008 mM	
NaHCO ₃	0.4 g	4.8 mM	
NaCl ₂	0.08 g	1.37 mM	
CaCl ₂	1 mL	0.80%	0.8g/100mL stock
Vitamin K (menadione)	1 mL	5.8 mM	1 mg/mL stock solution
FeSO ₄	1 mL	1.44 mM	0.4 mg FeSO ₄ /mL stock solution
Histidine Hematin Solution	1 mL	0.1%	1.2 mg hematin/mL in 0.2M histidine
Tween 80	2 mL	0.05%	25% stock solution
ATCC Vitamin Mix	10 mL	1%	
ATCC Trace Mineral Mix	10 mL	1%	
Acetic acid	1.7 mL	30 mM	
Isovaleric acid	0.1 mL	1 mM	
Propionic acid	2 mL	8 mM	
Butyric acid	2 mL	4 mM	
Resazurin	4 mL	4 mM	0.25 mg/mL stock solution
Noble Agar	12 g	1.2%	

Table 3. Antibiotics used in functional metagenomic selections			
Name	Class	Abbreviation	MIC (mg/mL)
Amikacin	Aminoglycoside	AM	64
Amoxicillin	β lactam	AX	16
Carbenicillin	β lactam	CA	64
Cefdinir	Cephalosporin	CF	2
Cloramphenicol	Amphenicol	CH	8
Ciprofloxacin	Flouroquinolone	CI	4
Cefepime	Cephalosporin	CP	8
Gentamicin	Aminoglycoside	GE	16
Oxytetracycline	Tetracycline	OX	8
Penicillin	β lactam	PE	128
Piperacillin	β lactam	PI	16
Piperacillin+Tazobactam	β lactam+ β lactamase inhibitor	PI+TZ	16(PI), 4(TZ)
Tetracycline	Tetracycline	TE	8
Trimethoprim	Pyrimidine derivative	TR	8
Trimethoprim+Sulfamethoxazole	Pyrimidine derivative+Sulfonamide	TR+SX	2(TR), 38(SX)

Table 4. BLAST analysis of metagenomic sequences identified by selection in *E. coli*

<i>E. coli</i> libraries containing microbiome fragments were selected on each of 15 antibiotics and antibiotic combinations. Microbiome DNA fragments identified from three of these selections (amikacin, piperacillin, piperacillin plus tazobactam) were subjected			
Antibiotic	Gene identified	Donor	Complete/Cultured
Amikacin	rmtD; 16S rRNA methylase rmtD	1	Complete*
Amikacin	aphA-3; aminoglycoside phosphotransferase type III	1	Complete, Cultured
Amikacin	bcrA; bacitracin transport ATP-binding protein	1	Cultured
Amikacin	marR locus; multiple antibiotic resistance repressor	1	Cultured
Piperacillin	beta-lactamase; non-experimental evidence, no additional details recorded	1,2	Complete, Cultured
Piperacillin	Megamonas hypermegale beta-lactamase class D	1	Cultured
Piperacillin	Bacteroides uniformis beta-lactamase cblA	1,2	Cultured
Piperacillin	beta-lactamase/D-alanine carboxypeptidase ampC	2	Cultured
Piperacillin	Beta-lactamase precursor ampC	2	Cultured
Piperacillin	Clostridium nexile beta-lactamase class A	2	Cultured
Piperacillin	ABC-type multidrug transport system	2	Cultured
Piperacillin + Tazobactam	beta-lactamase; non-experimental evidence, no additional details recorded	1	Complete, Cultured
Piperacillin + Tazobactam	Clostridium bolteae beta-lactamase class A	1	Cultured
Piperacillin + Tazobactam	Megamonas hypermegale beta-lactamase class D	1	Cultured
Piperacillin + Tazobactam	Bacteroides uniformis Multi Antimicrobial Exclusion family	1	Cultured
Piperacillin + Tazobactam	Bacteroides caccae beta-lactamase class A	1	Cultured
Piperacillin + Tazobactam	Bacteroides uniformis beta-lactamase cblA	1	Cultured
Piperacillin + Tazobactam	beta-lactamase/D-alanine carboxypeptidase ampC	2	Cultured

*rmtD sequences were found in readily cultured amikacin-resistant fecal strains from Donor 1

Table 5. Control PARC medium				
	Component	Amount/L	Concentration	Comments
P	KH ₂ PO ₄	100 mL	100 mM	1M stock solution pH 7.2
A	Noble Agar	12 g	1.2%	
R	Resazurin	4 mL	4 mM	0.25 mg/mL stock solution
C	L-cysteine	0.5 g	3.2 mM	

Example 10. Modeling the response of a microbiota to changes in host diets.

[0154] Owing to its many roles in human health, there is great interest in deciphering the principles that govern the operations of an individual's gut microbiota. Current estimates indicate that each of us harbors several hundred bacterial species in our intestine and different diets lead to large and rapid changes in the composition of the microbiota. Given the dynamic interrelationship between diet, the configuration of the microbiota, and the partitioning of nutrients in food to the host, inferring the rules that govern the microbiota's responses to dietary ingredients represents a challenge.

[0155] Gnotobiotic mice colonized with simple, defined collections of sequenced representatives of the various phylotypes present in the human gut microbiota provide a simplified in vivo model system where metabolic niches, host-microbe, and microbe-microbe interactions can be examined using a variety of techniques. These studies have focused on small communities exposed to a few perturbations. In this example, gnotobiotic mice harboring a 10-member community of sequenced human gut bacteria were used to model the response of a microbiota to changes in host diet. The aim was to predict the absolute abundance of each species in this microbiota based on knowledge of the composition of the host diet. Another aim was to gain insights into the niche preferences of members of the microbiota, and to discover how much of the response of the community was a reflection of their phenotypic plasticity.

[0156] The ten bacterial species were introduced into germ-free mice to create a model community with representatives of the four most prominent bacterial phyla in the healthy human gut microbiota (**FIG. 13A**). Their genomes encode major metabolic functions that have been identified in anaerobic food webs, including the ability to break down complex dietary polysaccharides not accessible to the host

(*Bacteroides thetaiotaomicron*, *Bacteroides ovatus* and *Bacteroides caccae*), consume oligosaccharides and simple sugars (*Eubacterium rectale*, *Marvinbryantia formatexigens*, *Collinsella aerofaciens*, *Escherichia coli*), and ferment amino acids (*Clostridium symbiosum*, *E. coli*). Two species capable of removing the end products of fermentation were also included: a H₂-consuming, sulfate-reducing bacterium (*Desulfovibrio piger*) and a H₂-consuming acetogen (*Blautia hydrogenotrophica*).

[0157] To perturb this community, a series of refined diets were used where each ingredient represented the sole source of a given macronutrient (casein=protein, corn oil=fat, cornstarch=polysaccharide, and sucrose =simple sugar) and where the concentrations of these four ingredients were systematically varied (**FIG. 13B, C and Table 6**). Each individually caged male C57Bl/6J mouse was fed a randomly selected diet with diet switches occurring every two-weeks (n=13 animals; **Table 7** shows the variation of diet presentation between animals). 13 gnotobiotic mice harboring the 10-member community were each fed a randomly selected diet every two weeks for eight weeks (i.e., four total diets per mouse). Shotgun sequencing of total fecal DNA allowed the determination of the absolute abundance of each community member, based on assignment of reads to the various species' genomes, in samples obtained from each mouse on days 1, 2, 4, 7, and 14 of a given diet period. Analyses of the shotgun sequencing data revealed that steady state levels of community members were achieved within 24 h of a diet change. Therefore, the values from all five time points sampled within a diet period were averaged to obtain the mean absolute abundance of each community member for each of the refined diet periods.

Table 7. Variation of diet presentation between animals.

mouse:	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13
1 st diet period	E	E	E	E	E	E	E	E	E	E	E	E	E
2 nd diet period	B	G	E	K	F	J	H	D	J	A	C	I	E
3 rd diet period	B	G	C	J	F	I	K	A	K	H	E	D	E
4 th diet period	I	F	D	H	E	J	C	G	J	A	K	B	E

[0158] To predict the abundance of each species in the model human gut microbiome given only knowledge of the concentration of each of the four perturbed diet ingredients, a linear model was used,

$$y_i = \beta_0 + \beta_{\text{casein}}X_{\text{casein}} + \beta_{\text{starch}}X_{\text{starch}} + \beta_{\text{sucrose}}X_{\text{sucrose}} + \beta_{\text{oil}}X_{\text{oil}}$$

where y_i is the absolute abundance of species i , X_{casein} , X_{starch} , X_{sucrose} , and X_{oil} are the amounts (in g/kg of mouse diet) of casein, corn starch, sucrose, and corn oil respectively in a given host diet, β_0 is the estimated parameter for the intercept, and β_{casein} , β_{starch} , β_{sucrose} , and β_{oil} are the estimated parameters for each of the perturbed diet components. Since each mouse underwent a sequence of three diet permutations presented in different order, and each of the diet periods covered all of the 11 possible diets (**Table 7**), it was possible to use two of these three diet intervals to fit the model for the equation (13 mice x 2 diets per mouse = 26 samples per bacterial species) and then the ability to predict the abundance of each bacterial species for the 13 samples was measured in the remaining (third) diet. Averaging this cross-validation from all three subsets, the model explained over 61% of the variance in the abundance of the community members (abundance weighted mean $R^2 = 0.61$; see **Table 8** for species-specific R^2).

Example 11. Predicting response of microbiota to a diet

[0159] Although the cross-validation provided evidence that the response of this microbiota was predictable from knowledge of these diet ingredients, a more conclusive validation of the model would be its ability to make predictions for new diets. Therefore, six additional diets were designed with new combinations of the four refined ingredients. Using a design similar to the first experiment, eight different 10-week-old gnotobiotic male C57Bl/6J mice harboring the 10-member community were each given a randomized sequence of diets selected from the six new diets (shaded diets L-Q in **FIG. 13B**), or one of the previous diets (**Table 9**). Fitting the model parameters with the data from the first experiment, we were able to explain 61% of the variance in the abundance of the community members on the new diets, showing virtually equivalent results to the cross validation procedure (see **Table 8**).

Table 9. Variation of diet presentation between animals.								
mouse:	m1	m2	m3	m4	m5	m6	m7	m8
1 st diet period	Q	M	P	N	L	E	E	O
2 nd diet period	N	D	O	M	E	Q	P	L
3 rd diet period	P	N	M	O	F	E	L	Q

[0160] These results indicate that the linear model explains the majority of the variation in abundance of each organism using only a knowledge of the species in the community and the concentrations of casein, cornstarch, sucrose, and corn oil in the diet, without having to explicitly consider the effects of microbe-microbe or microbe-host interactions, or diet order. As described in Materials and Methods below, several other models were also tested including adding interactions between the variables, quadratic terms, and interactions with quadratic terms. After correcting for the number of parameters in the model using Akaike information criterion, the linear model was still the best performing.

Example 12. Inferring the association of ingredients with the abundance of each community member.

[0161] To further dissect the community response to these diet perturbations, we need to infer which set of diet ingredients is associated with the abundance of each community member. Feature selection algorithms assume that the response variable (in this case, the abundance of each organism) is potentially affected by only a fraction of the variables in the model, and use statistical methods to choose the subset of variables that most informatively predict the abundance of each species. Using stepwise regression as a feature selection procedure with the equation above, all species in the 10-member community had the diet variable X_{casein} significantly associated with their abundance (**Table 10**).

[0162] *E. coli* and *C. symbiosum* were the only bacteria with more than one variable significantly associated with their abundance (casein and sucrose for *E. coli* and casein and starch for *C. symbiosum*). Further exploring this finding, we found

casein highly correlated with the yield of total DNA per fecal pellet across all diets (**FIG. 14A and FIG. 15**). A component of casein, presumably amino acids and/or nitrogen, limits the biomass of the community: this resource limitation was observed even for combinations of three additional refined protein and two additional fat sources (soy, lactalbumin, egg-white solids, olive oil and lard; n=9 different diets given to another group of 9 C57Bl/6J male mice; **FIG. 16; Table 11**). However, the observed changes in species abundance are not a simple consequence of a constant relative abundance of each community member that is scaled upwards as casein is increased: three community members, *E. rectale*, *D. piger*, and *M. formatexigens*, decreased in absolute abundance by 1.4-2.4 -fold from the low casein to high casein diets even though total community biomass tripled (**FIG. 14B, 17; Table 12**). Similar changes in species abundance and total community DNA levels were observed when casein concentrations were altered in gnotobiotic mice harboring a 9- member or an 8-member subset of the original community (minus *B. hydrogenotrophica* or minus *D. piger* and *B. hydrogenotrophica*) (**Table 13**).

Example 13. Correlation of diet ingredients with mRNA expression.

[0163] Microbial RNA-Seq was used on fecal RNA samples, prepared from mice on each diet (mean=2.1±0.7 replicates per diet; **Table 14**), to determine if perturbations in diet ingredients correlated with underlying changes in mRNA expression by community members. Each of the 36 RNA-Seq datasets was composed of 36 nt-long reads (3.20±1.35 x 10⁶ mRNA reads/sample). Transcript abundances were normalized for each of the 10 species to reads per million per kilobase (RPKM). After correcting for multiple-hypotheses, no statistically significant changes in gene expression were found within a given bacterial species as a function of any of the diet perturbations. While community members do not appear to significantly alter their gene expression, they do respond by increasing or decreasing their absolute abundances (**FIG. 15**), thereby adjusting the total available transcript pool in the microbiota for processing dietary components. For example, as casein levels are increased across the diets, *B. caccae* increases its contribution to the gene pool/community transcriptome; so the number of transcripts per unit of casein remains roughly constant.

[0164] Since RNA-Seq provides accurate estimates of absolute transcript levels, transcript abundance information was used as a proxy to predict the major metabolic niche occupied by each community member. For species positively correlated with casein, it was found that high expression of mRNAs predicted to be involved in pathways using amino acids as substrates for nitrogen, as energy and/or as carbon sources. By contrast, the three species that negatively correlated with dietary casein concentration showed no clear evidence of high levels of expression of genes involved in catabolism of amino acids. The changes in abundance of the negatively correlated species (e.g., *E. rectale*) can be explained by competition with another member of the community that increases with casein (see FIG. 18).

Example 14. Use of the modeling framework with typically consumed human diets.

[0165] The power of the refined diets used lies in the capacity to precisely control individual diet variables and to aid data interpretation from more complex diets. To test if the modeling framework used here generalizes to diets containing food more typically consumed in human diets, 48 meals were created consisting of random combinations and concentrations of four ingredients selected from a set of eight pureed human baby foods (apples, peaches, peas, sweet potatoes, beef, chicken, oats, and rice; Table 15). The meals were administered for periods of 7d to the same eight gnotobiotic mice used for the follow-up refined diet experiments described above and in FIG. 13E. Each mouse received a sequence of 6 baby food diets. The order of presentation of the baby food diets was varied between animals (see Table 15). The absolute abundance of each bacterial community member was measured on days 1, 5, 6, and 7 for each diet. Using the linear modeling approach described above, over half of the variation in species abundance could be explained using only knowledge of the concentrations of the pureed foods present in each meal ($R^2=0.62$). Stepwise regression was used to identify the type of pureed food(s) present in a given mixed meal that was most significantly associated with changes in each bacterial species (Table 16; FIG. 19).

[0166] Defining the interrelationship between diet and the structure and operations of the human gut microbiome is key to advancing understanding of the nutritional value of food, for creating new guidelines for feeding humans at various stages of their lifespan, for improving global human health, and for developing new ways to manipulate the properties of the microbiota to prevent or treat various diseases. The experiments and model described above highlight the extent to which host diet can explain the configuration of the microbiota, both for refined diets where all of the perturbed diet components are digestible by the host, and for human diets whose ingredients are only partially known. These models can now be tested using larger defined gut microbial communities representing those of humans living in different cultural settings, and with more complex diets, including various combinations of food ingredients that they consume.

Materials and Methods for Examples 10 to 14.

Assembling a model human microbiota in gnotobiotic mice

[0167] *B. caccae* ATCC 43185 (GenBank genome accession number NZ_AAVM000000000), *B. ovatus* ATCC 8483 (NZ_AAXF000000000), *B. thetaiotaomicron* VPI-5482 (NC_004663), *B. hydrogenotrophica* DSM 10507 (NZ_ACBZ000000000), *M. formatexigens* DSM 14469 (NZ_ACCL000000000), *C. symbiosum* ATCC 14940, *C. aerofaciens* ATCC 25986 (NZ_AAVN000000000), *E. coli* str. K-12 substr. MG1655 (NC_000913), and *E. rectale* ATCC 33656 (NC_012781) were obtained from public strain repositories (ATCC or DSMZ). A draft genome assembly for *C. symbiosum* ATCC 25986 is available at the Washington University Genome Center public web site (genome.wustl.edu/pub/organism/Microbes/Human_Gut_Microbiome/Clostridium_symbiosum/assembly/Clostridium_symbiosum-1.0/output/). *D. piger* GOR1 was isolated from a healthy human by plating serial dilutions of freshly voided feces under strictly anaerobic conditions (80% H₂/20% CO₂ at 15 psi) onto plates containing medium with the following components (quantities expressed per liter): K₂HPO₄ (0.3g); KHPO₄ (0.3g); (NH₄)SO₄ (0.3g); NaCl (0.6g); MgSO₄·7H₂O (0.13g); CaCl₂·2H₂O (0.008g), yeast extract (0.5g); NH₄Cl (1.0g); NaHCO₃ (5.0g); dithiothreitol (0.5g); sodium formate (3.0g); Noble

agar (10g); 5 ml of a 0.2% (w/v) solution of $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$, 1 ml of a 0.2% (w/v) solution of resazurin; cysteine (1g), 10 ml of trace mineral solution (ATCC), and 10 ml of a vitamin solution (ATCC). The genome sequence of *D. piger* was determined by 454 FLX and FLX Titanium pyrosequencing. For both *C. symbiosum* and *D. piger*, genes were identified using Glimmer3.0, tRNAScan 1.23, and RNAmmer 1.2. All 10 genomes were annotated using PFAM v23; and String COG version 7.1. Annotations for all 40,669 predicted protein-coding genes in the 10 genomes can be found at gordonlab.wustl.edu/modeling_microbiota/.

[0168] Each community member was grown anaerobically in 5 ml of TYGS medium in Balch tubes. Inoculation times were staggered so that all organisms reached stationary phase within a 24 h window. Just prior to gavage, equal volumes (1 ml) of each culture were pooled and mixed regardless of the final stationary phase density reached by each mono-culture (OD_{600} values ranged from 0.4 to >2.0). Each germ-free mouse was subsequently gavaged with 300 μl of the pooled cultures.

Refined diet composition, experimental design, and data processing

[0169] A set of eleven diets was initially designed (FIG. 13B,C and Table 6), each differing in their concentrations of casein (protein), corn oil (fat), cornstarch (polysaccharide), and sucrose (simple sugar). Nine of the diets consisted of all possible combinations of high, medium, and low casein and corn oil, with a fixed amount of cornstarch and the remainder as sucrose (FIG. 13B; diets A-I). Using sucrose as the 'remainder' for these initial nine diets generated a negative correlation between sucrose concentration and casein/fat concentration. Therefore, two diets, one with high starch and low sucrose and the other with low starch and high sucrose, were designed to lessen this negative correlation (see diets J and K in FIG. 13C).

[0170] Initially, all mice were co-housed and given the diet labeled 'E' in Table 7 (5% fat, 20% protein, 62% carbohydrate). Mice were then individually caged in the gnotobiotic isolator, and every two weeks each animal received another randomly selected diet (second, third and fourth diet periods in Table 7). Mouse 13 received only control diet E to determine if there was any 'drift' in steady state over the 8-week period.

[0171] The steady state mean absolute abundance of each community member was estimated for each of the 36 mouse/diet combinations for the second, third, and fourth diet periods shown in **Table 7**. To do so, DNA was isolated from fecal samples taken from each mouse on days 1, 2, 4, 7, and 14 of a diet and analyzed by COMMUNITY PROFILING by sequencing (COPRO-Seq). This generally applicable method relies on the massive number of short reads generated by the Illumina GA-II instrument during shotgun sequencing of total community DNA. Briefly, "informative" tags are identified that map uniquely to a single location in one species' genome. These tags are then summed to generate raw "counts" of each species' abundance. To account for non-unique matches, species-specific counts are normalized by the "Informative Genome Fraction" of each genome (defined as the fraction of all possible k-mers a genome can produce that are unique). Up to 16 barcoded fecal DNA samples were pooled in each sequencing lane: a minimum of 50,000 reads per sample were generated so that all organisms comprising $\geq 0.02\%$ of the community could be detected (for a mouse colonized at 10^{12} cfu/ml cecal contents or feces, this represents $\sim 10^8$ cfu/ml; at this sequencing depth, all species were detected in all samples). Total DNA yield per fecal pellet was used as a proxy for community biomass and multiplied the relative abundance of each species by the mean total DNA yield per fecal pellet for a particular diet to estimate the absolute abundance of each species in units of nanograms per fecal pellet. The absolute abundance N_{impd} of each species i in mouse m on diet period p on day d was calculated $N_{impd} = F_{impd} T_j$ where F_{impd} is the Informative Genome Fraction adjusted fraction of species i in mouse m on diet period p on day d as measured by COPRO-Seq and T_j is the mean total DNA yield per fecal pellet for all samples taken from mice on diet j . Fecal pellets were used because they reflect overall microbiota composition in the gut and they provide the only means to sample each mouse over time. Mice were weighed during each diet period (**Table 17**). Although there was a trend towards increased weight gain as levels of casein and corn oil were increased (**Table 18**), there were no significant correlations between any of the diet perturbations and weight gain.

Model description and performance evaluation

[0172] Population growth can be modeled as exponential growth with a carrying capacity:

$$\text{Equation 1} \quad \frac{dN}{dt} = rN \left(1 - \frac{N}{K} \right)$$

where r is the growth rate, N is the population size, and K is the carrying capacity. Extending the above equation to include multiple species (i) and multiple diets (j), the model becomes:

$$\text{Equation 2} \quad \frac{dN_i}{dt} = r_{ij} N_i \left(1 - \frac{N_i}{K_{ij}} \right)$$

where K_{ij} is the carrying capacity of species i on diet j (i.e. the steady-state level). We were interested in predicting the steady-state abundance of each species in the synthetic community as a function of the ingredients in the host diet. Thus, we can ignore the time-specific abundance of each community member $N_i(t)$ on each diet and the growth rate r , assuming it is sufficiently large to allow each community member to reach their carrying capacity for each diet within the period that a given mouse was consuming the diet.

[0173] To predict the steady-state levels (K_{ij}) for each community member i given each diet j , we measured the abundance of each community member for each mouse and diet combination (**FIG. S1D,E**). The absolute abundance K_{impd} for each community member (i) in a specific mouse (m) for a specific diet period (p) for a given day (d) was calculated as described above. These abundances were averaged across all available time points for each mouse after the microbiota had reached steady-state (d_s) for a specific diet period (i.e., the cells in the diet/mouse matrices in **Tables 7 and 9**; see below for estimation of steady-state). On average, 2.7 samples were available per mouse per diet period to give the steady-state abundance of each species (i) in the fecal microbiota for a given mouse (m) and diet period (p) combination in **Tables 7 and 9**.

$$\text{Equation 3} \quad y_{imp} = \text{mean}(K_{impd}) \text{ where } d \geq d_s$$

[0174] These y_{imp} values served as the data for the linear model described in the Examples above.

Scoring model performance

[0175] All of the models used in this study were linear. Therefore, model could be scored by using R^2 , which for linear models represent the proportion of variance in the system that is explained by the model. The R^2 was used for each species in the community separately to calculate a weighted mean R^2 , where the weights represent the fraction of total fecal DNA content represented by each species (i.e., the R^2 for abundant taxa are given more weight than those of less abundant taxa). By using this weighted scoring schema, the final R^2 metric represents the amount of the total variation in species DNA content that can be explained by the model. An alternative method is to weight each species' R^2 equally, which produces similar albeit slightly worse results (Table 8).

Estimating steady-state

[0176] Since the model assumes the microbiota is at steady-state, values of species abundance were only included from time points after the microbiota reaches steady-state for a given diet period (i.e. d_s in Equation 3 needs to be define). To determine the time required by the microbiota to reach steady-state after a diet switch, nine 22-week-old C57Bl/6J mice were fed a low protein/low fat diet for 7d, followed by a switch to the high protein/high fat diet for 13 d (FIG. 13B diets A and I respectively). All mice were sampled approximately once every 24 h with twice-daily sampling around the time of the diet switch. To estimate whether the community had stabilized at a given time point, (i) the total microbial community biomass using DNA yield (ng/fecal pellet) as a marker, and (ii) the relative abundance of each species using the Informative Genome Fraction estimated from Illumina DNA sequencing were measured. It was found that the biomass of the gut microbiota was stable at the end of the first diet, highly variable during the initial time points after the diet transition, and then stable again by the fourth day after the diet switch (although even by 24h after the switch, the mean yield largely reflects final steady-state abundance on the new diet). Similar results were found for the

relative abundance data, although the relative abundance of each species appeared to stabilize faster than biomass (**FIG. 20A** compared with **FIG. 20B**). Given the results for both the biomass and relative abundance metrics, the number of days required by the microbiota to reach steady-state was set to four (i.e. $d_s = 4$).

Comparing more complex models

[0177] Although the linear model performed well (see examples above and **Table 8**), more complex linear and nonlinear models could perhaps yield even better predictive ability. Therefore the cross validation procedure was repeated for the casein, corn oil, sucrose, starch diet combinations (see Examples above) using models that allowed for interactions between variables, quadratic terms, and interactions with quadratic terms:

$$\text{Linear } y_i = \beta_0 + \beta_A X_A + \beta_B X_B,$$

$$\text{Interaction } y_i = \beta_0 + \beta_A X_A + \beta_B X_B + \beta_{AB} X_A X_B,$$

$$\text{Quadratic } y_i = \beta_0 + \beta_A X_A + \beta_B X_B + \beta_{AA} X_A X_A + \beta_{BB} X_B X_B,$$

$$\text{Pure Quadratic } y_i = \beta_0 + \beta_A X_A + \beta_B X_B + \beta_{AA} X_A X_A + \beta_{BB} X_B X_B$$

$$+ \beta_{AB} X_A X_B,$$

[0178] Akaike information criterion (AIC) was used as the scoring metric to allow for comparisons between these models with varying numbers of parameters and found the linear model performed best overall (**Table 8**). Given the slightly asymptotic behavior of the microbiota at extremely low and high casein concentrations (**FIG. 14**), it was also attempted to fit a nonlinear logistic function to the data to account for these saturation points. However, it was found that the lack of data at and beyond the asymptote made the nonlinear regression difficult to reliably fit. While it is interesting to question whether the microbiota will reach and maintain an asymptotic behavior with sampling at extreme concentrations of these ingredients, moving beyond the current

maximum and minimum casein values would be unrealistic in terms of modern human eating habits and would be unhealthy for the animals.

Transcriptional responses of the microbiota to host diet perturbations

[0179] To deplete total microbial community RNA of 16S, 23S, 5S rRNA and tRNA species prior to synthesis of cDNA with random hexanucleotide primers, each fecal RNA preparation was subjected to column-based size-selection and hybridization to custom biotinylated oligonucleotides directed at conserved regions of bacterial rRNA genes present in human gut communities, followed by streptavidin-bead based capture of the hybridized RNA sequences. RNA-Seq data were normalized as described previously. After normalization, the list was filtered to remove all transcripts whose total number of counts ($\log_2$) summed across all 36 RNA-Seq expression profiles was < 64 (2^6). This threshold was chosen to be as inclusive as possible while still requiring a sufficient number of reads so that a dynamic range of roughly 5-fold could be detected across the 17 sampled diets. For example, if a transcript linearly increases 5-fold in response to diets with a 20-fold range in their casein concentration, with the lowest concentration yielding a number of reads that was just below level of detection for both replicates and the highest casein concentration yielding 5 reads for that transcript per replicate, 55 reads would be required. After normalization and filtering, a list of 26,643 genes across the 10 species remained ($64 \pm 20\%$ of the annotated genes in each species were detected as 'expressed'). For each of these genes, the correlation and the p-value of the correlation were calculated between (i) each of the four perturbed refined diet ingredients and (ii) the $\log_2$ (gene expression) in reads per million per kilobase (RPKM). Multiple hypotheses correction was performed using the Storey procedure.

Highly expressed transcripts for each species

[0180] the highest 10% expressed genes in each community member were examined (gordonlab.wustl.edu/modeling_microbiota/), as major metabolic activities of gut microbes have consistently been identified among the abundant genes. Among the most highly expressed genes in *B. thetaiotaomicron*, were those encoding components of glycolysis/gluconeogenesis pathways (e.g. BT1658-1660, BT1672,

1691), the pentose phosphate pathway (e.g. BT3946-3950), plus members of polysaccharide utilization loci (PULs), including one PUL predicted to act on O-glycan containing mucins (BT0317-0319; S12), and another PUL involved in the degradation of fructans (BT1757-1763 and BT1765; 26). In addition to several peptidases (BT2522, BT2706, BT3926, BT4583), genes predicted to be involved in the metabolism of glutamate (glutamate dehydrogenase (BT1973); glutamate decarboxylase (BT2570)), glutamine (glutaminase (BT2571)), serine (L-serine dehydrate (BT4678)), aspartate (aspartate ammonia lyase (BT2755)), asparagine (L-asparaginase (BT2757)), and branched-chain amino acids (branched-chain alpha-keto acid dehydrogenase (BT0311-12)) were highly expressed. Similar results were observed in *B. caccae* and *B. ovatus*. Although the ability of colonic *Bacteroides* to access protein has not been extensively explored, there is evidence that members of this genus have extracellular proteolytic activity, and can incorporate amino acids into cellular components other than proteins. This feature, combined with their ability to use complex polysaccharides not accessible to other members in the community (including host glycans), may explain why they benefit from increased levels of dietary protein (casein).

[0181] Among the most highly expressed genes in *C. symbiosum* were components of the hydroxyglutarate pathway for degradation of glutamate (Csym2026-2031), the most abundant amino acid in casein (25.3% w/w), and a sodium/glutamate transporter (Csym3971). This pathway yields crotonyl-CoA, which is metabolized to butyrate, acetate, H₂ and ATP. Genes encoding components of the pathway for butyrate production (Csym1328-1334) were also among the highest expressed.

[0182] Another Firmicute that grows on amino acids is the acetogenic bacterium *B. hydrogenotrophica*. Genes predicted to encode key enzymes of the acetyl-CoA pathway involved in the reductive assimilation of CO₂ were among the most highly expressed in this species (e.g., carbon monoxide dehydrogenase (Rumhyd0314-0320)), as were genes involved in fermentation of aliphatic (Rumhyd0546-0555) and aromatic amino acids (Rumhyd1109-1113), and the metabolism of ribose (Rumhyd2245-2256).

[0183] *E. coli* also benefited from higher levels of protein; among its most highly expressed genes were components of a cytochrome d terminal oxidase involved in the consumption of oxygen (b0733-0734), genes involved in the utilization of simple

sugars (e.g., b2092-2097 (galactitol), b2416-2417 (glucose), b2801-2803 (fucose)) and several genes involved in the metabolism of tryptophan (b3708-b3709), aspartate (b1439) asparagine (b2957), and threonine (b3114-3117).

[0184] *C. aerofaciens* expressed high levels of transcripts encoding proteins predicted to be involved in the catabolism of arginine (COLAER0352-356, COLAER1230), plus components of several phosphotransferase systems (a predicted sucrose-specific PTS (COLAER0919-0921), a predicted mannose/fructose/N-acetylglucosamine-specific PTS (COLAER1259-1260) and a predicted mannitol/fructose PTS (COLAER0058-0061)).

[0185] Levels of *E. rectale* and *M. formatexigens* decreased as protein increased. Inspection of their most highly expressed genes suggested that they focus on catabolism of carbohydrates. For example, among the most highly expressed genes in *M. formatexigens* were components of several ABC transporters with predicted specificities for monosaccharides/oligosaccharides (e.g., BRYFOR5076-BRYFOR5080, BRYFOR06841-06843), and genes encoding key enzymes of the acetyl-CoA pathway (e.g., BRYFOR06355-06360). There was no clear evidence of genes involved in catabolism of amino acids being highly expressed.

[0186] *D. piger* also decreased as casein levels increased. *D. piger* is fairly restricted in its metabolism: it can use a few substrates (e.g., lactate, H₂, succinate) to reduce different forms of sulfur to H₂S and generate energy, and it can oxidize lactate and pyruvate incompletely to acetate. Among its most highly expressed genes were components of the sulfate reducing pathway (DpigGOR12316-18, DpigGOR110789-10794), a C4- dicarboxylate transport system (DpigGOR12113-2115), subunits of a Ni-Fe hydrogenase, and several genes predicted to be involved in lactate metabolism (DpigGOR11071-1075). Three predicted transporters of amino acids were highly expressed, but there was no evidence of further metabolism of these amino acids, which likely indicates that they are used for protein biosynthesis.

Simulation of negatively correlated species with constant behavioral responses

[0187] Using Equation 2 above, a simulated 2-member community was created where one member (species1 = N_1) is casein limited (C) and the second

member (species2 = N_2) is negatively influenced in proportion to the abundance of species1 (α_{12}) (e.g., species1 could consume a limiting resource of species2, produce an inhibitory compound, or act through apparent competition). It was assumed that both species have the same growth constant ($r_1 = r_2$) on all diets and that species1 is able to convert a proportion (s) of the casein into increases in population size ($K_1 = sC$; note that 1.3, 2/3, 0, 2/3, and 15 were used for constants r , s , α_{21} , α_{12} , and K_2 respectively, but this choice of values is arbitrary and similar results can be obtained over a wide-range of stable values):

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1 + \alpha_{21} N_2}{K_1} \right)$$

Equation 4

$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \frac{N_2 + \alpha_{12} N_1}{K_2} \right)$$

Equation 5

[0188] Simulating the above equations, where every ten days we change the amount of casein (C_j) from 2, 5, 10, 20, and 40% respectively, yields the result shown in **FIG. 18** where species1 increases with increased casein while species2 decreases, during which time both maintain the same behaviors.

[0189] This type of behavior at the transcriptional level of a microbial community resembles similar phenomena observed in macro-ecology. For example, if two species of naturally co-occurring grasshoppers, one that eats almost exclusively grasses (*Ageneotettix deorum*) and the other that eats both grasses and forbs (*Melanoplus sanguinipes*), are co-housed to compete in environments with different dietary contexts, the final population size of each grasshopper species is dependent not only on the ability of *A. deorum* to compete for grass (i.e. its essential resource), but also *M. sanguinipes*' ability to utilize both grass and forbs. Thus, if the amount of grass available to the grasshoppers is held constant while the amount of forbs is increased, the population of *A. deorum* decreases even though it maintains the constant behavioral response of exclusively eating grass.

Design, administration, and modeling of complex diets

[0190] The following commercially available eight pureed human baby foods were used as the source ingredients to construct a set of 48 meals: peaches (Gerber 3rd foods®; Gerber Products Company); apple sauce (Gerber 3rd foods); peas (Gerber 2nd foods®), sweet potatoes (Gerber 3rd foods); chicken (Gerber 2nd foods); beef (Gerber 2nd foods), oats (Gerber Single Grain with VitaBlocks®); and rice (Gerber Single Grain with VitaBlocks). Oats and rice were purchased dry and mixed with dH₂O in a 1:5 ratio prior to use (e.g., a meal with 6g of oats contained 1g dried oats and 5g dH₂O). Each meal consisted of four ingredients randomly selected from the set of eight total pureed human foods, with different concentrations of the four ingredients used in different diet periods. Meals were autoclaved and each mouse was fed a sequence of 5 different diets, with each diet provided for 1 week. The order of presentation of the 48 diets to the 8 gnotobiotic mice is described in **Table 15**. The table shows how a 1 week period of consumption of one of the 17 diets composed of refined ingredients was interposed, between each 1 week period of administration of a given pureed baby food meal, to ensure mice obtained adequate amounts of vitamins and minerals.

[0191] Absolute abundance of each bacterial community member was measured on days 1, 5, 6, and 7 of each human baby food diet. As before, the abundance values (y_{imp}) were calculated from the mean of all samples within a given diet period. However, day 1 was excluded in case the microbiota had not yet reached steady state by 24h. To cover more of the potential “meal” space, the schema for the complex diets used less replication than was used for the refined diets, so there were fewer fecal pellets available to estimate the DNA yield data used to calculate absolute abundance of each species. Therefore, to estimate DNA yields for each diet, a nearest neighbor smoothing procedure was used where the DNA yield for each sample was calculated as a weighted average of the ten nearest samples with weights corresponding to the Euclidean distance from the true sample to the Nth-nearest sample (e.g., the nearest samples would be exact replicates and have a weight of 1.0).

[0192] The modeling performance was estimated with a species abundance weighted R² as described above using the following equation:

$$y_i = \beta_0 + \beta_{apple}X_{apple} + \beta_{peach}X_{peach} + \beta_{pea}X_{pea} + \beta_{sweetpotato}X_{sweetpotato} + \beta_{chicken}X_{chicken} + \beta_{beef}X_{beef} + \beta_{oats}X_{oats} + \beta_{rice}X_{rice},$$

[0193] where the variables correspond to the concentration of each pureed ingredient in each meal. The final performance metric was the mean of ten replicates of 10-fold crossvalidation on the 48 samples. The performance when training on the larger set (n~43) and testing on the smaller set (n~5) was similar to training and testing using the larger set only (weighted $R^2=0.62$ and $R^2=0.66$ respectively).

Tables for Examples 10 to 14

Table 6A. Composition of refined diets: seventeen perturbations to casein, sucrose, corn starch, and corn oil concentrations.

Diet ID:	First set of diets										
	A	B	C	D	E	F	G	H	I	J	K
Harlan Teklad Diet Number:	TD.09049	TD.09050	TD.09051	TD.09052	TD.09053	TD.09054	TD.09055	TD.09056	TD.09057	TD.09058	TD.09059
Casein	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg
L-Cystine	69	230	460	69	230	460	69	230	460	230	230
Sucrose	0.9	3	6	0.9	3	6	0.9	3	6	3	3
Sucrose	675.88	514.13	283.08	633.46	471.66	240.66	473.98	312.38	81.38	171.66	571.66
Corn Starch	100	100	100	100	100	100	100	100	100	400	0
Maltodextrin, Lo-Dex 10	50	50	50	50	50	50	50	50	50	50	50
Cellulose (Fiber)	50	50	50	50	50	50	50	50	50	50	50
Corn Oil	10	10	10	50	50	50	200	200	200	50	50
79055 Mineral mix, Ca-P Deficient	12.73	12.73	12.73	13.4	13.4	13.4	16.08	16.08	16.08	13.4	13.4
Calcium Carbonate	2.6	6.25	11.3	2.6	6.4	11.3	3	6.5	11.6	6.4	6.4
Calcium Phosphate	12.5	7.5	0.5	13.4	8.3	1.4	16.4	11.4	4.3	8.3	8.3
40077 Vitamin mix	14.25	14.25	14.25	15	15	15	18	18	18	15	15
Choline Bitartrate	2.1	2.1	2.1	2.2	2.2	2.2	2.6	2.6	2.6	2.2	2.2
Ethoxyquin (Liquid)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Protein (% by weight)	6.1	20.3	40.6	6.1	20.3	40.6	6.1	20.3	40.6	20.3	20.3
Protein (% of kcal)	6.7	22.7	46.5	6.3	21.5	44.1	5.3	18.0	36.8	22.2	21.3
Carbohydrate (% by weight)	82.7	66.6	43.5	78.6	62.4	39.3	62.9	46.8	23.7	59.4	63.4
Carbohydrate (% of kcal)	90.7	74.3	49.7	81.8	66.0	42.6	55.1	41.5	21.4	64.9	66.4
Fat (% by weight)	1.1	1.2	1.5	5.1	5.2	5.5	20.1	20.2	20.5	5.2	5.2
Fat (% of kcal)	2.6	3.1	3.8	11.9	12.5	13.3	39.6	40.4	41.7	12.9	12.3
kcal/g diet	3.6	3.6	3.5	3.8	3.8	3.7	4.6	4.5	4.4	3.7	3.8

Table 6B. Composition of refined diets: seventeen perturbations to casein, sucrose, corn starch, and corn oil concentrations.

Second set of diets					
L	M	N	O	P	Q
TD.09620	TD.09621	TD.09622	TD.09623	TD.09624	TD.09625
g/kg	g/kg	g/kg	g/kg	g/kg	g/kg
138	345	138	345	35	690
1.8	4.5	1.8	4.5	45	9
585.425	377.525	484.52	276.42	667.61	108.11
100	100	100	100	100	0
50	50	50	50	50	50
50	50	50	50	50	50
30	30	125	125	50	50
13.06	13.06	14.74	14.74	13.4	13.4
4.2	8.8	4.5	9	1.9	12.25
10.7	4.3	12.5	6.4	14.4	0
14.625	14.625	16.5	16.5	15	15
2.15	2.15	2.4	2.4	2.2	2.2
0.04	0.04	0.04	0.04	0.04	0.04
12.2	30.5	12.2	30.5	3.1	60.9
13.1	33.5	11.7	29.7	3.2	67.1
73.7	52.9	63.8	43	82	17
79.4	58.1	61.1	41.9	85.1	18.7
3.1	3.4	12.6	12.9	5	5.7
7.5	8.4	27.2	28.3	11.7	14.1
3.7	3.6	4.2	4.1	3.9	3.6

Calcium carbonate and calcium phosphate were used to maintain calcium and phosphorus levels at 0.5% and 0.35%, respectively, across all diets with the exception of the diet with the highest level of protein (TD.09621) where the phosphorus present in casein brings its level to 0.48%. Vitamin and mineral mixes were adjusted based on the caloric density of each diet.

Custom diets designed for this study are now commercial available from Harlan Teklad using the Harlan Teklad Diet Number above.

Table 8. Model performance measurements for individual species in experiments involving variations in dietary casein, corn oil, corn starch and sucrose concentrations.

A. R² performance measurements for linear model			
	cross validation	prediction on new diets	
<i>Eubacterium rectale</i> ATCC 33656	0.25	0.68	
<i>Collinsella aerofaciens</i> ATCC 25986	0.08	0.42	
<i>Blautia hydrogenotrophica</i> DSM 10507	0.70	0.63	
<i>Desulfovibrio piger</i> GOR1	0.08	0.13	
<i>Clostridium symbiosum</i> ATCC 14940	0.70	0.70	
<i>Escherichia coli</i> str. K-12 substr. MG1655	0.22	0.61	
<i>Marvinbryantia formatexigens</i> DSM 14469	0.08	0.43	
<i>Bacteroides ovatus</i> ATCC 8483	0.69	0.54	
<i>Bacteroides thetaiotaomicron</i> VPI-5482	0.69	0.50	
<i>Bacteroides caccae</i> ATCC 43185	0.78	0.80	
mean	0.43	0.54	
weighted mean (weighted by species abundance)	0.61	0.61	
B. Cross validation comparison of AIC scores across different models			
	linear	interaction	pure quadratic
<i>Eubacterium rectale</i> ATCC 33656	121.3	132.9	129.5
<i>Collinsella aerofaciens</i> ATCC 25986	161.5	174.9	170.7
<i>Blautia hydrogenotrophica</i> DSM 10507	137.0	150.8	146.3
<i>Desulfovibrio piger</i> GOR1	193.3	205.9	201.8
<i>Clostridium symbiosum</i> ATCC 14940	169.7	181.2	177.2
<i>Escherichia coli</i> str. K-12 substr. MG1655	200.0	213.3	209.4
<i>Marvinbryantia formatexigens</i> DSM 14469	199.4	210.6	206.7
<i>Bacteroides ovatus</i> ATCC 8483	204.9	216.5	212.4
<i>Bacteroides thetaiotaomicron</i> VPI-5482	211.6	222.0	218.2
<i>Bacteroides caccae</i> ATCC 43185	206.8	220.8	216.3

Table 10. Stepwise regression selection of diet ingredients significantly associated with changes in species abundance.

	Casein	Sucrose	Corn Starch	Corn Oil
<i>Eubacterium rectale</i> ATCC 33656	1.10E-07	0.43	0.35	1.00
<i>Collinsella aerofaciens</i> ATCC 25986	3.13E-03	0.21	0.49	0.26
<i>Blautia hydrogenotrophica</i> DSM 10507	1.51E-08	0.24	0.65	0.18
<i>Desulfovibrio piger</i> GOR1	1.13E-02	0.31	0.90	0.12
<i>Clostridium symbiosum</i> ATCC 14940	2.63E-15	0.64	2.44E-03	0.65
<i>Escherichia coli</i> str. K-12 substr. MG1655	1.57E-07	9.38E-03	0.55	0.56
<i>Marvinbryantia formatexigens</i> DSM 14469	1.31E-03	0.65	0.97	0.48
<i>Bacteroides ovatus</i> ATCC 8483	1.36E-07	0.32	0.06	0.36
<i>Bacteroides thetaiotaomicron</i> VPI-5482	3.29E-12	0.43	0.74	0.38
<i>Bacteroides caccae</i> ATCC 43185	4.48E-19	0.47	0.52	0.76

Significant p-values from regression are shown in boldface

Table 11. Composition of nine diets with combinations of three refined protein sources and two refined fat sources.									
Mouse	m1	m2	m3	m4	m5	m6	m7	m8	m9
Harlan Teklad Diet Number:	TD.10321	TD.10309	TD.10314	TD.10316	TD.10311	TD.10319	TD.10310	TD.10307	TD.10308
Ingredients	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg	g/kg
Isolated Soy Protein	172	17.2	17.2	172	17.2	172	17.2	17.2	17.2
Egg White Solids, spray-dried	187.5	187.5	18.8	187.5	18.8	18.8	18.8	18.8	187.5
Lactalbumin	172	17.2	17.2	17.2	172	172	172	17.2	17.2
Sucrose	15.5	322.3	682.4	264.5	336.2	278.4	528.1	585.3	514.2
Maltodextrin	100	100	100	100	100	100	100	100	100
Corn Starch	50	50	50	50	50	50	50	50	50
Cellulose (Fiber)	50	50	50	50	50	50	50	50	50
79055 Mineral mix, Ca-P Deficient	16.1	16.1	12.7	14.7	16.1	14.7	12.7	14.7	12.7
Calcium Carbonate	6.3	2.3	1.7	5.5	2.6	5.5	2.3	1.8	1.9
Calcium Phosphate Dibasic	10.0	16.8	13.7	9.6	16.5	9.6	12.6	16.1	13.0
40077 Vitamin mix	18.0	18.0	14.3	16.5	18.0	16.5	14.3	16.5	14.3
Choline Bitartrate	2.6	2.6	2.1	2.4	2.6	2.4	2.1	2.4	2.1
Biotin	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004	0.004
Olive Oil	100	100	10	100	100	10	10	10	10
Lard	100	100	10	10	100	100	10	100	10
TBHQ (antioxidant)	0.04	0.04	0.01	0.02	0.04	0.02	0.01	0.02	0.01
Protein, g/kg	451	181	45	316	181	315	181	45	181
CHO, g/kg	177	480	837	423	496	438	684	742	668
Fat, g/kg	215	201	21	118	208	125	28	111	21
Fiber, g/kg	50	50	50	50	50	50	50	50	50

Table 12. Correlation of species abundance with mouse diet casein concentration.

	correlation with casein	p-value
<i>Eubacterium rectale</i> ATCC 33656	-0.61	1.1E-07
<i>Collinsella aerofaciens</i> ATCC 25986	0.37	3.1E-03
<i>Blautia hydrogenotrophica</i> DSM 10507	0.64	1.5E-08
<i>Desulfovibrio piger</i> GOR1	-0.32	1.1E-02
<i>Clostridium symbiosum</i> ATCC 14940	0.77	1.3E-13
<i>Escherichia coli</i> str. K-12 substr. MG1655	0.61	8.5E-08
<i>Marvinbryantia formatexigens</i> DSM 14469	-0.40	1.3E-03
<i>Bacteroides ovatus</i> ATCC 8483	0.61	1.4E-07
<i>Bacteroides thetaiotaomicron</i> VPI-5482	0.74	3.3E-12
<i>Bacteroides caccae</i> ATCC 43185	0.86	4.5E-19

Table 13. Responses of 8-member and 9-member subset communities to low and high casein.**A. Increase in total community DNA from low casein to high casein**

Community	mean total DNA yield (ng/fecal pellet)		percent increase
	diet G (low casein)	diet I (high casein)	
8-member	8568	13800	61%
9-member	12448	16875	36%
10-member	7692	11718	52%

8-member and 9-member community samples were from 13 to 14-week old NMRI mice

10-member community samples were from 10 to 16-week old C57Bl/6J mice

B. Species-level responses to changes in casein concentration

Species	8-member	9-member	10-member
<i>Bacteroides caccae</i> ATCC 43185	p	p	p
<i>Clostridium symbiosum</i> ATCC 14940	p	p	p
<i>Bacteroides thetaiotaomicron</i> VPI-5482	p	p	p
<i>Blautia hydrogenotrophica</i> DSM 10507	-	-	p
<i>Escherichia coli</i> str. K-12 substr. MG1655	p	p	p
<i>Eubacterium rectale</i> ATCC 33656	n	n	n
<i>Bacteroides ovatus</i> ATCC 8483	p	p	p
<i>Marvinbryantia formatexigens</i> DSM 14469	n	n	n
<i>Collinsella aerofaciens</i> ATCC 25986	n	p	p
<i>Desulfovibrio piger</i> GOR1	-	n	n

species are sorted by the p-value of the correlation between casein and species abundance for the 10-member community.

n = negatively correlated with casein concentration.

p = positively correlated with casein concentration.

- = not present in community.

Table 14. Number of expressed genes/species with ≥ 64 sequencing reads.			
Species	Genes Observed	Total Genes in Genome	% Observed Genes
<i>Eubacterium rectale</i> ATCC 33656	453	3621	13%
<i>Collinsella aerofaciens</i> ATCC 25986	1779	2367	75%
<i>Blautia hydrogenotrophica</i> DSM 10507	2612	3869	68%
<i>Desulfovibrio piger</i> GOR1	1660	2487	67%
<i>Clostridium symbiosum</i> ATCC 14940	3141	5128	61%
<i>Escherichia coli</i> str. K-12 substr. MG1655	2969	4132	72%
<i>Marvinbryantia formatexigens</i> DSM 14469	3173	4896	65%
<i>Bacteroides ovatus</i> ATCC 8483	3785	5536	68%
<i>Bacteroides thetaotaomicron</i> VPI-5482	3696	4778	77%
<i>Bacteroides caccae</i> ATCC 43185	3375	3855	88%

Table 15. Composition of and experimental design for complex diets composed of pureed baby foods.

	mouse1	mouse2	mouse3	mouse4	mouse5	mouse6	mouse7	mouse8	g/kg	
week1	apple sauce peaches chicken sweet potatoes	peas peaches chicken oatmeal	sweet potatoes chicken beef peas	peaches peas oatmeal rice	oatmeal rice beef sweet potatoes	sweet potatoes apple sauce beef peaches	rice apple sauce sweet potatoes beef	beef peaches sweet potatoes rice	666.7 222.2 55.6 55.6	BABY FOOD MEAL 1
week2	TD.09052	TD.09623	TD.09621	TD.09622	TD.09620	TD.09624	TD.09625	TD.09054		Refined Diet
week3	apple sauce peaches chicken sweet potatoes	peas peaches chicken oatmeal	sweet potatoes chicken beef peas	peaches peas oatmeal rice	oatmeal rice beef sweet potatoes	sweet potatoes apple sauce beef peaches	rice apple sauce sweet potatoes beef	beef peaches sweet potatoes rice	666.7 222.2 55.6 55.6	BABY FOOD MEAL 2
week4	TD.09620	TD.09623	TD.09621	TD.09622	TD.09625	TD.09624	TD.09052	TD.09054		Refined Diet
week5	sweet potatoes oatmeal peaches peas	chicken peaches beef rice	beef rice peas chicken	apple sauce oatmeal chicken beef	apple sauce peaches chicken rice	apple sauce beef chicken peaches	peas apple sauce sweet potatoes peaches	peas oatmeal sweet potatoes chicken	666.7 222.2 55.6 55.6	BABY FOOD MEAL 3
week6	TD.09049	TD.09053	TD.09050	TD.09056	TD.09058	TD.09053	TD.09051	TD.09059		Refined Diet
week7	peas rice apple sauce beef	chicken sweet potatoes rice oatmeal	peaches peas rice beef	peaches peas apple sauce chicken	peas chicken oatmeal sweet potatoes	chicken peaches rice oatmeal	sweet potatoes rice peaches beef	peas rice sweet potatoes peaches	666.7 222.2 55.6 55.6	BABY FOOD MEAL 4
week8	TD.09053	TD.09051	TD.09059	TD.09053	TD.09050	TD.09049	TD.09058	TD.09056		Refined Diet
week9	oatmeal rice chicken apple sauce	oatmeal apple sauce beef rice	sweet potatoes rice chicken peas	peaches oatmeal sweet potatoes beef	sweet potatoes oatmeal chicken rice	peas chicken rice peaches	peaches beef oatmeal rice	chicken beef peas apple sauce	421.1 421.1 105.3 52.6	BABY FOOD MEAL 5
week10	apple sauce	TD.09053	TD.09053	TD.09055	TD.09057	TD.09055	TD.09057	TD.09053		Refined Diet
week11	peas oatmeal beef chicken	apple sauce rice oatmeal beef	peaches oatmeal rice sweet potatoes	sweet potatoes beef peaches rice	sweet potatoes apple sauce peaches beef	beef peas chicken oatmeal	apple sauce oatmeal rice beef	apple sauce rice peaches chicken	250 250 250 250	BABY FOOD MEAL 6

Custom Harlan Teklad Diet Numbers are provided for the weeks mice were on refined diets.

Table 16. Stepwise regression selection of complex diet ingredients significantly associated with changes in species abundance.

	Apple	Beef	Chicken	Oat	Pea	Peach	Rice	Sweet Potato
<i>Desulfovibrio piger</i> GOR1	0.37	0.18	1.26E-04	6.08E-03	0.81	0.10	2.27E-06	0.17
<i>Collinsella aerofaciens</i> ATCC 25986	0.06	4.49E-05	5.55E-07	0.26	0.10	0.91	0.08	0.19
<i>Blautia hydrogenotrophica</i> DSM 10507	0.14	1.73E-03	5.62E-06	0.99	0.14	0.21	0.50	0.57
<i>Clostridium symbiosum</i>	1.14E-03	7.29E-04	2.30E-04	0.25	0.16	4.96E-04	0.35	0.84
<i>Escherichia coli</i> str. K-12 substr. MG1655	4.30E-04	1.42E-03	8.63E-05	0.85	0.28	0.38	7.74E-05	0.70
<i>Bryantella formatexigens</i> DSM 14469	0.48	0.07	0.87	0.38	0.98	0.10	0.96	0.18
<i>Eubacterium rectale</i> ATCC 33656	3.77E-05	0.27	0.07	6.27E-06	0.13	2.76E-03	0.68	0.58
<i>Bacteroides caccae</i> ATCC 43185	0.28	7.56E-06	2.00E-05	0.37	0.57	0.76	2.46E-03	3.72E-02
<i>Bacteroides thetaiotaomicron</i> VPI-5482	2.31E-05	0.78	0.12	3.51E-03	1.94E-02	6.52E-04	0.31	0.43
<i>Bacteroides ovatus</i> ATCC 8483	0.20	0.46	1.00	1.43E-08	0.92	0.11	0.16	0.34

Table 17. Mean Weight Gain (g/diet period).

mouse	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13
2 nd diet period	0.891	1.400	1.909	2.291	1.527	-1.909	-1.655	-1.145	2.927	0.636	0.255	2.291	0.636
3 rd diet period	0.800	0.000	0.600	1.200	0.300	0.600	0.100	1.400	-0.100	1.900	1.200	-0.700	1.000
4 th diet period	1.400	2.800	0.700	2.300	1.100	2.400	1.300	2.700	3.300	0.300	1.100	1.900	2.700

Table 18. Weight gain as a function casein and corn oil concentration.**A. Weight gain per diet period as a function of Casein concentration**

% Casein	Mean Weight Gain (g/diet period)	SEM
7%	0.589	0.386
23%	1.233	0.295
46%	1.233	0.294

B. Weight gain per diet period as a function of Corn oil concentration

% Corn Oil	Mean Weight Gain (g/diet period)	SEM
1%	0.900	0.180
5%	1.110	0.301
20%	1.211	0.463

Table 19. Fecal DNA yields from diverse protein diets.

mouse	diet	nanograms (mean±SEM)
m1	TD.10321	19716 ± 5689
m2	TD.10309	15488 ± 7023
m3	TD.10314	3781 ± 2906
m4	TD.10316	27989 ± 5462
m5	TD.10311	10378 ± 1515
m6	TD.10319	14283 ± 3203
m7	TD.10310	18604 ± 1007
m8	TD.10307	259 ± 75
m9	TD.10308	19684 ± 9045

Example 15. Intact and cultured gut microbial communities from twins discordant for obesity transplanted into gnotobiotic mice.

[0194] Substantial interpersonal differences in microbial community configurations normally exist between unrelated individuals, creating a challenge in designing surveys of sufficient power to determine whether observed differences between healthy versus disease-associated microbiomes are significantly different from normal interpersonal variation. Microbiome configurations are influenced by early environmental exposures and are generally more similar among family members. In the

case of same-sex twins discordant for a disease phenotype, the genetically related, healthy co-twin provides a valuable reference control for characterizing the disease-associated co-twin's microbiome. However, while each discordant pair can provide a vignette about the potential role of the microbiome in disease pathogenesis, the comparison is fundamentally descriptive and does not establish causality. Transplanting a fecal sample obtained from each co-twin in a discordant pair into multiple recipient mice provides an opportunity to conduct a virtual clinical trial designed to identify structural and functional differences between their communities, to generate and test hypotheses about the impact of these differences on host biology, and to directly test the effects of manipulating the representation of microbial taxa in the community.

[0195] A number of studies of obese and lean humans have revealed compositional differences in their gut microbiomes. Mono- or dizygotic twins discordant for obesity provide an attractive study paradigm for studies of the contributions of the gut microbiome to differences in body mass index (BMI). Two Finnish twin cohort studies have provided much of the published data about BMI discordance for obesity among monozygotic (MZ) twin pairs. In one cohort, with participants aged 35-60 years at the time of data collection, 1.3% of MZ twin pairs were defined as discordant for obesity [body mass index (BMI) difference $\geq 3 \text{ kg/m}^2$ with one twin $>27 \text{ kg/m}^2$ and the other $<25 \text{ kg/m}^2$]. In the other study, with participants aged 22-27 years at data collection, 2.16% of MZ twin pairs had a BMI difference $\geq 4 \text{ kg/m}^2$. Data collected at the 5th wave of assessment from 1539, 21-32 year-old female twin pairs enrolled in the Missouri Adolescent Female Twin Study (MOAFTS) was surveyed. Four discordant twin pairs with a BMI difference $\geq 6 \text{ kg/m}^2$ were recruited for the present study (n=1 MZ; 3 DZ pairs; **Table 20**).

Table 20. Features of the 4 discordant twin pairs.

Twin Pair	1		2		3		4	
BMI (kg/m ²)	23	32	25.5	31	19.5	30.7	24	33
Zygoty	DZ		DZ		DZ		MZ	

[0196] Comparisons of the input human fecal microbiota, and 'output' mouse fecal communities surveyed two weeks after transplantation revealed that $77.8 \pm 7.4\%$ (SD) of genus-level bacterial taxa in the human donor microbiota were represented in the microbiota of gnotobiotic mouse recipients ($n=3-12$ animals analyzed/microbiota; **Table 21**). The UniFrac metric measures the overall degree of phylogenetic similarity of any two bacterial communities by comparing the degree of branch length they share on a Bacterial tree of life. V2-16S rRNA reads sharing $\geq 97\%$ nucleotide sequence identity were considered to represent a given species-level operational taxonomic unit (OTU). Principal Components Analysis (PCoA) of unweighted UniFrac distance matrices based on the 97%ID OTU datasets revealed that transplanted microbial communities achieved a stable configuration in recipients within 3d. This configuration was sustained for at least 38 d. Importantly, the overall phylogenetic architecture of the transplanted community evolved in a reproducible way between singly housed recipient mice within given experiment for a given co-twin microbiota, and between replicate experiments (**FIG. 21B**). Pairwise UniFrac-based comparisons of fecal samples and of communities sampled along the length of the gut of transplant recipients also demonstrated a significantly higher similarity among recipients colonized with the same human donor, and a greater similarity to their human donor compared to mice colonized with unrelated human donor microbiota (**Fig. 21C**; **FIG. 22**).

Table 21. Fecal samples obtained from American female twin pairs discordant for obesity and used as donor samples for gnotobiotic mice. Percent recapitulation at each taxonomical level based on pyrosequencing data from the V2 region of the 16S rRNA gene.									
Twin Pair ID	1		2		3		4		All
BMI of the donor (kg/m ²)	Lean (23)	Obese (32)	Overweight (25.5)	Obese (31)	Lean (19.5)	Obese (30.7)	Lean (24)	Obese (33)	
Twin ID	TSDC17	TSDC16	TSDC19	TSDC20	TSDC22	TSDC23	TSDC7	TSDC8	
Phyla (%)	100	75	100	80	100	100	100	100	94.4±10.5
Class (%)	80	80	100	83.3	85.7	85.7	60	85.7	82.6±11.1
Order (%)	57.1	66.7	80	57.1	87.5	66.6	70	66.6	69.0±10.4
Family (%)	62.5	63.6	83.3	66.7	77.8	83.3	82.3	68.7	73.5±9.1
Genus (%)	69	66.7	88	76	80	84.2	82.9	75.9	77.8±7.4
OTU level (%)	30.80±7.81	44.39±3.81	27.74±2.59	17.13±4.10	16.91±2.09	21.43±1.32	19.41±9.56	28.54±7.92	25.89±10.33

[0197] Transplant recipients not only efficiently captured the organismal features of their human donor's microbiota but also the functions encoded by the donor's microbiome, as judged by shotgun pyrosequencing of cecal DNA samples isolated from mice colonized with each of the 8 human fecal microbiota [n=3-8 mice sampled 15d after transplantation/microbiota; n=45 cecal samples; 90,164 ± 37,526 (mean ± SD) reads per sample; 337±62 (SD) nt/read; 9.43±4.12 Mb/sample]. Shotgun reads were functionally annotated with KEGG orthology groups (KO) and Enzyme Commission (E.C.) numbers (KEGG version 58; see Methods). The results disclosed that 99.69 ± 0.2% of donor ECs were captured in recipients. Remarkably, there was a significant correlation between the proportional representation of reads with given assignable EC and KO in donor and recipient microbiomes (Spearman's correlation, p-value < 0.0001, Spearman's R ≥ 0.88) (**FIG. 23**), highlighting the remarkable ability of a human microbiome to reassemble itself in a mouse gut ecosystem.

[0198] Body composition was analyzed using quantitative magnetic resonance imaging 1d, 15d, and in the case of longer experiments 38d after transplantation. The increased adiposity phenotype of each obese co-twin in a discordant twin pair was transmissible: the differences in adiposity between mice that received an obese co-twins fecal microbiome was statistically greater than the adiposity of mice receiving her lean co-twins microbiome within an experiment, and was reproducible between experiments (One-tail Mann-Whitney U test, p-value<0.001; n=92

recipients phenotyped) (**FIG. 24A, B**). Epididymal fat pad weights were also significantly higher in mice colonized with gut communities from obese co-twins ($p < 0.05$, one-tail Mann-Whitney U test). These differences in adiposity were not associated with statistically significant differences in daily chow consumption among recipients of obese versus lean co-twin microbiomes.

[0199] Supervised machine learning was performed using Random Forests to identify family-level bacterial phylotypes that differentiate gnotobiotic mice harboring a gut community transplanted from all lean versus all obese co-twins. The estimated generalization error of the trained model was 12.6%, indicating that it could be predicted if a sample came from a mouse colonized with a lean or obese human donor microbiota with 87.4% accuracy using family-level taxonomic classifications. Only three family-level taxa were identified as producing a mean decrease in classification accuracy of $\geq 5\%$ when they were ignored; all three families were members of the order Clostridiales: Lachnospiraceae, Ruminococcaceae and Veillonellaceae; two of the three (Ruminococcaceae and Veillonellaceae) were significantly increased in fecal samples obtained from mice colonized with a lean donor's microbiota (**Table 22**).

Table 22. Discriminatory family level taxa between gnotobiotic mice colonized with the microbiota of human co-twins discordant for obesity. Feature importance score was calculated using supervised machine learning (Random Forest algorithm) and represents t

Feature ID	Feature importance scores = Mean decrease in accuracy when the feature is ignored	Relative Abundance (%)		p-adjust (ANOVA, FDR)
		Lean donors (mean $\pm$ SEM)	Obese donors (mean $\pm$ SEM)	
Firmicutes; Clostridia; Clostridiales; Veillonellaceae	10.06%	1.19 $\pm$ 0.47	0.00 $\pm$ 0.00	1.58E-15
Firmicutes; Clostridia; Clostridiales; Ruminococcaceae	9.45%	16.02 $\pm$ 1.68	8.34 $\pm$ 3.19	7.64E-09
Firmicutes; Clostridia; Clostridiales; Lachnospiraceae	7.84%	21.29 $\pm$ 2.29	30.82 $\pm$ 1.35	2.81E-12
Firmicutes; Clostridia; Clostridiales; Eubacteriaceae	3.40%	0.04 $\pm$ 0.03	0.19 $\pm$ 0.10	3.28E-04
Bacteroidetes; Bacteroidia; Bacteroidales; Rikenellaceae	1.62%	0.57 $\pm$ 0.20	0.84 $\pm$ 0.38	7.26E-02
Firmicutes; Clostridia; Clostridiales; Others	0.70%	0.74 $\pm$ 0.18	1.17 $\pm$ 0.26	8.05E-03
Firmicutes; Clostridia; Clostridiales; Clostridiales FamilyXIII. Incertae Sedis	0.46%	0.04 $\pm$ 0.01	0.00 $\pm$ 0.00	8.61E-03
Tenericutes; Erysipelotrichi; Erysipelotrichales; Erysipelotrichaceae	0.42%	1.91 $\pm$ 0.68	2.61 $\pm$ 0.45	5.55E-02
Firmicutes; Clostridia; Clostridiales; Clostridiaceae	0.40%	0.17 $\pm$ 0.11	0.34 $\pm$ 0.22	5.69E-02
Proteobacteria; Deltaproteobacteria; Desulfovibrionales; Desulfovibrionaceae	0.27%	0.09 $\pm$ 0.05	0.16 $\pm$ 0.07	5.10E-02
Actinobacteria; Actinobacteria; Coriobacteriales; Coriobacteriaceae	0.13%	0.11 $\pm$ 0.04	0.11 $\pm$ 0.05	9.63E-01
Actinobacteria; Actinobacteria; Coriobacteriales; Others	0.12%	0.00 $\pm$ 0.00	0.02 $\pm$ 0.01	1
Verrucomicrobia; Verrucomicrobiae; Verrucomicrobiales; Verrucomicrobiaceae	0.12%	1.05 $\pm$ 0.40	0.93 $\pm$ 0.29	7.80E-01
Firmicutes; Clostridia; Clostridiales; Peptococcaceae	0.07%	0.01 $\pm$ 0.00	0.05 $\pm$ 0.07	8.02E-02
Firmicutes; Bacilli; Turicibacterales; Turicibacteraceae	0.01%	0.15 $\pm$ 0.15	0.15 $\pm$ 0.07	9.41E-01

Estimated generalization error = 0.126050

Error estimation method = Out-of-bag prediction of training data

Number of features used = 15

Parameters

values of all non-default parameters

method was "random_forest"

seed:616869664

cross.validation.k:10

[0200] ShotgunFunctionalizerR, a software tool designed for metagenomic analysis and based on a Poisson model, was used to identify genes encoding KEGG KOs and ECs whose proportional representation in cecal microbiomes differed significantly between recipients of transplanted obese versus lean co-twin microbiomes (p -value < 0.0001). Random Forests is an ensemble classifier, that uses multiple decision trees to identify which features are discriminatory among different class labels, rather than features that are over- or underrepresented. This complementary approach to Shotgun FunctionalizerR identified KOs and ECs that best discriminate transplanted obese and lean microbiomes (relevant discriminatory features defined as those with a feature importance score ≥ 0.0001). These predictive KOs and ECs were among the most significantly different KOs and ECs as judged by ShotgunFunctionalizerR.

[0201] This type of DNA-level analysis provides information about functional capacity, but not about expressed functions. Therefore, the same cecal samples were used to prepare RNA for microbial RNA-Seq characterization of the transplanted microbial communities' meta-transcriptomes. Transcripts were mapped to 127 sequenced human gut genomes and assigned to KEGG KOs and ECs (**see Methods**). Significant differences in gene expression between transplanted obese and lean co-microbiomes were defined using ShotgunFunctionalizerR and Random Forests.

[0202] All transplanted lean co-twin microbiomes exhibited increased relative abundance of transcripts encoding components of the KEGG 'Starch and Sucrose' metabolic pathway (cellobiose and pyruvate metabolism) as well as the KEGG 'Mannose and Fructose' metabolic pathway (propanoate and butanoate metabolism). ECs involved in the KEGG pathway for 'Pyrimidine and Purine' metabolism were also enriched in the transcriptomes expressed by lean co-twin microbiomes, consistent with the significantly greater fecal microbial biomass (defined by fecal DNA content) observed in mice colonized with the gut microbial communities of lean versus obese co-twin donors ($p < 0.00X$, ANOVA). Obese microbiomes had a significant increased in the representation of ECs involved in more futile energy metabolic cycles [e.g. KEGG 'Pentose phosphate' pathway] (**FIG. 25**).

[0203] Non-targeted gas chromatography-mass spectrometry (GC/MS) analyses of cecal contents was used to confirm these findings, and to identify other differences in expressed microbiome-associated metabolic activities. A metabolite profile was obtained for each sample using the spectral abundances of all identifiable metabolites (cecal samples from 3-5 mice/microbiome donor). A total of 26 metabolites satisfied a reverse match score cutoff of 65% (**for definition, see Methods**) and were present in at least 50% of samples representing a given transplanted microbiome (**Table 23**); 10 were robust discriminatory biomarkers of lean versus obese microbiome donor transplants in all four pairs (two tail Student's t-test, p-value < 0.01; Random Forests, importance score >0.0007). Five of these 10 metabolites were mono- and disaccharides [cellobiose, mannose, glucose, lactose, and talose (C2-epimer of galactose)]; each was present at significantly lower levels in the cecal metabolomes of mice colonized with the gut communities of lean co-twins and each can be fermented by members of gut microbiota to short chain fatty acids (SCFA). Targeted GC/MS of cecal SCFA revealed significant increases in propionate and butyrate levels in mice harboring transplanted lean co-twin microbiomes (p<0.05, Student's t-test), consistent with increased carbohydrate fermentation (**FIG. 26A, B**). Bomb calorimetry revealed a trend towards increased energy content in the feces of mice containing transplanted lean compared to obese microbiomes (2.4 ± 0.2 (SEM) and 1.9 ± 0.3 kcal/mg dry fecal weight, respectively; p=0.07; two tailed Student's t-test; 2 discordant twin pairs surveyed; 7-12 mice/co-twin microbiome). This trend, coupled with the significantly greater biomass and lower SCFA levels associated with transplanted lean co-twin microbiomes, raise the question of whether the transplanted gut communities from lean co-twin donors may be more 'selfish', consuming energy for their own use, and less altruistic in distributing energy to their gnotobiotic hosts.

Table 23. Metabolites whose levels are significantly different in 29 cecal samples obtained from gnotobiotic recipients of intact fecal samples from obese versus lean co-twins from four discordant pairs. Metabolites were identified using non-targeted GC/M

Metabolite	ttest	p-adjust
tagatose 1	0.165003758	0.4158388
D-lyxosylamine	0.377652896	0.61797747
D-altrose	0.022259913	0.08903965
lactose 2	0.000758951	0.02714539
3-(3-hydroxyphenyl)propionic acid	0.908546205	0.93858256
cellobiose 1	0.001508077	0.02714539
D-allose	0.022259913	0.08903965
gluconic acid lactone 1	0.010717259	0.0654987
talose 1	0.016037437	0.08247825
Phenylalanine	0.529530864	0.65734866
coniferyl alcohol	0.347509677	0.61797747
maleic acid	0.055489479	0.18160193
fumaric acid	0.931223792	0.93858256
beta-alanine 1	0.473120794	0.63082773
acetoacetate 2	0.173266167	0.4158388
cholesterol	0.450002889	0.62308092
oxalic acid	0.444235043	0.62308092
5-Hydroxylysine	0.502472019	0.64603545
succinic acid	0.89644695	0.93858256
L-alanine 1	0.351162848	0.61797747
2-hydroxybutyric acid	0.338703755	0.61797747
glycine	0.006317688	0.0654987
D-malic acid	0.757829475	0.85255816
L-glutamic acid 2	0.145769214	0.40366859
5-aminovaleric acid 1	0.00890099	0.0654987
ribose	0.660979975	0.76758965
D-lyxose	0.429692899	0.62308092
2-amino-1-phenylethanol- ⁻ , Ät	0.325820188	0.61797747
putrescine	0.127190686	0.38157206
Myristic Acid	0.267126773	0.60103524
L- sorbose 1	0.620036783	0.74404414
fructose 1	0.938582558	0.93858256
D-mannose 1	0.033391959	0.12021105
D-glucose 1	0.01091645	0.0654987
Myoinositol	0.364584224	0.61797747
inosine	0.41516831	0.62308092

[0204] GPCRs comprise the largest superfamily of transmembrane signaling proteins encoded in the human genome, and participate in an array of signaling pathways that regulate myriad aspects of host physiology. GPCRs expressed by gut epithelial cell lineages (e.g. enteroendocrine cells) would be in a strategic position to transduce metabolic signals emanating from the microbiota to the host. To determine whether obese versus lean microbiota have differential effects on GPCR signaling, TaqMan assays were used to survey the expression of 350 GPCRs, belonging to 50 subfamilies, in the distal small intestine (ileums) of microbiota transplant recipients (initially 2 discordant pairs; 4 mice/donor microbiota). Three GPCRs satisfied criteria for a consistent > 2-fold difference in expression in the distal small intestines of recipients of lean versus obese co-twin microbiota (p-value < 0.05; Student's t-test). qRT-PCR was used to confirm the differential patterns of expression of these three GPCRs, Gpr15, Gpr3 and Gabbr2, in gnotobiotic recipients of microbiota from members of all four discordant twin pairs (n=4 mice/microbiota). Gpr15/Bob is abundant at the basal surface of the small intestinal epithelium (). In the human enterocyte-like cell line HT-29-D4 cells, activation of Gpr15 leads to a 70% decrease in sodium-dependent glucose and lipid transport. Gpr15 down-regulation in mice harboring an obese microbiome would be expected to result in increased glucose and lipid absorption.

[0205] Procrustes analysis utilizing a Bray-Curtis distance matrix from the different groups of gnotobiotic recipients revealed significant correlations (p-value < 10^{-7}) between taxonomic structure (V2-16S rRNA family-level bacterial phylotype), functional capacity (EC representation in cecal microbiomes), transcriptional profiles (EC representation in cecal mRNA populations), and metabolic profile (non-targeted GC/MS profiles) (Mantel test with 10,000,000 iterations, p-value < 0.001), with separation of groups based on donor microbiota and adiposity phenotype.

[0206] These observations were followed up, generated from studies of transplanted intact (uncultured) donor communities, with a set of experiments involving culture collections produced from the fecal microbiota of one of the discordant twin pairs. The goal was to determine whether cultured bacterial members of the co-twins'

microbiota could transmit the discordant adiposity phenotypes and metabolic profiles of the corresponding uncultured microbiomes to gnotobiotic recipients.

[0207] Collections of cultured anaerobic bacteria were generated from each co-twin in DZ pair 1 (**see Methods**). The bacterial taxa present in the culturable component of each co-twin's microbiota were defined by sequencing V2-16S rRNA amplicons generated DNA isolated from the entire collection harvested directly from plates. Subsequently, each transplanted collection was harvested from plates directly into separate groups of 8-week old germ-free male C57Bl/6J mice (n=3 independent experiments; 4-6 recipient mice/culture collection/experiment). Capture of the cultured taxa was reproducible with $61\pm 2\%$ and $56\pm 7\%$ of the genus-level phylotypes present in the obese and lean co-twin's intact cultured microbiota retained in gnotobiotic recipients of their culture collection as shown (**Table 24**) by unweighted UniFrac analysis of V2-16S rRNA datasets (97%ID OTUs generated from cecal community DNA 15d after gavage of an intact uncultured fecal microbiota or the corresponding non-arrayed culture collection generated from that fecal sample (n=5 mice/treatment group; 4 treatment groups). The culture collections reached a steady state configuration within 3d after transplantation. Shotgun sequencing of the cecal microbiomes of transplant recipients confirmed that the functional features of the intact (non-cultured) donor microbiomes were efficiently captured and their proportional representation was recapitulated. Remarkably, statistically significantly greater increases in adiposity was documented 15d after gavage in recipients of the obese co-twin's compared to the lean co-twin's culture collection ($p<0.02$; Mann Whitney U-test).

Table 24. Efficiency of capture of taxa, present in culture collections prepared from obese and lean co-twins belonging to twin pair 1, in gnotobiotic mouse recipients.				
Twin Pair 1	Shared with mice colonized with an intact community		Shared with intact Human donor sample	
BMI of co-twin donor	Obese	Lean	Obese	Lean
Phyla	89.7 ± 8.1	100 ± 8.7	84.0 ± 8.9	96.0 ± 8.9
Class	88.9 ± 8.0	88.8 ± 11.6	80.0 ± 13.9	74.3 ± 12.0
Order	86.6 ± 7.0	87.2 ± 11.9	68.6 ± 12.0	54.0 ± 8.9
Family	83.2 ± 3.4	86.4 ± 8.2	63.1 ± 3.4	50.9 ± 7.5
Genus	73.9 ± 6.6	82.0 ± 7.3	61 ± 2.2	55.6 ± 6.5
OUT-97%ID	73.8 ± 6.6	69.5 ± 7.4	41.9 ± 9.3	18.1 ± 3.2

[0208] Given that mice are coprophagic, co-housing was used to determine whether exposure of a mouse harboring a culture collection from the lean co-twin could modify or rescue development of an increased adiposity phenotype in a cagemate colonized with the culture collection generated from her obese co-twin, or vice versa. Five days after gavage, a mouse with the lean co-twin's culture collection was co-housed with a mouse with the obese co-twin's culture collection, with or without two age-matched germ-free animals. Control groups consisted of cages of dually-housed recipients of the lean co-twin culture collection, or dually-housed recipients of the obese co-twin's culture collection (n=3-5 cages/experiment; n=2 independent experiments; each cage in each experiment placed a separate gnotobiotic isolator) (FIG. 27A). All mice were fed a low-fat, plant polysaccharide-rich diet. Adiposity phenotypes were measured by quantitative magnetic resonance imaging 1 and 5 days after transplantation, and after 10 days of co-housing. Fecal samples were collected from all mice on days 1, 2, 3, 5, 6, 7, 8, 10 and 15.

[0209] The results revealed that the microbiota of the co-housed mouse harboring the obese co-twin's culture collection (abbreviated Ob^{ch}) was re-configured so that its phylogenetic composition came to resemble that of the animal with the lean co-twin's culture collection (abbreviated Ln^{ch}). In contrast, the microbiota of the Ln^{ch} mouse remained stable (FIG. 27C-F). Moreover, Ob^{ch} mice exhibited a significantly

lower change in adiposity compared to Ob controls that had never been exposed to mice harboring the lean co-twin's culture collection, while Ln^{ch} mice had adiposity phenotypes that were indistinguishable from Ln controls (**FIG. 27B**). Co-housing experiments that included germ-free members revealed that these animals had adiposity phenotypes that were indistinguishable from Ln^{ch} cagemates (**FIG. 27B**).

[0210] Similar to what was observed with complete community transplants, the ceca of mice colonized with the lean co-twin's culture collection exhibited significantly greater levels of short chain fatty acids, particularly acetate, propionate and butyrate than recipients of the obese co-twin's culture collection (**FIG. 27H**). The differences with regards to lactose and cellobiose were even more dramatic than with the complete (uncultured) transplants: These metabolites were undetectable in mice with culture collection from the lean co-twin (**FIG. 27I**). Co-housing not only affected host adiposity phenotypes but also transformed the Ob^{ch} mouse's cecal metabolic profile such that it was indistinguishable from Ln controls (**FIG. 27I**).

[0211] To delineate the contribution of specific members of the lean co-twin's microbiota to these phenotypic changes, limiting dilution was used to produce a clonal arrayed collection of the lean twin's culture collection in replicate 384 well plates (**Methods**). 16S rRNA sequencing was initially used to identify the bacterial taxa present in wells that exhibited growth (**see Methods**) and to confirm that the well contained a clonal population of bacterial cells. The collection contained 54 strains representing 23 phylotypes (**Table 25**). DNA prepared from wells containing a single strain was then sequenced at $\geq 50X$ coverage and their genomes annotated. Then, a pool containing 37 was assembled comprising strains whose genomes were sequenced. This pool contained six strains of the known cellobiose fermentor *Collinsella aerofaciens* (family Coriobacteriaceae) plus 22 members of the family Bacteroidaceae and 11 members of Ruminococcaceae. These latter 33 members were chosen because Random Forests indicated that they discriminated between the transplanted lean and obese co-twin culture collections (feature importance score ≥ 0.1) and because their abundance increased significantly in the Ob^{ch} gut microbiota during the co-housing experiment described above (ANOVA; p-value < 0.05 after Bonferroni correction) (**Table 25**).

Table 25. Components of the arrayed anaerobic bacterial culture collection produced from the lean co-twin in DZ pair 1.

Strain designation	37 Selected taxa	Lactose degrading	Cellobiose degrading
Ruminococcus_bromii_TSDC17.2_V2.2.5	+		
Ruminococcus_bromii_TSDC17.2_V2.2.2	+		
Ruminococcus_bromii_TSDC17.2_V2.1.7	+		
Ruminococcus_albus_TSDC17.2_V2.2.8	+		+
Ruminococcus_albus_TSDC17.2_V2.1.7	+		+
Ruminococcus_albus_TSDC17.2_V2.1.6	+		+
Ruminococcus_albus_TSDC17.2_V2.1.16	+		+
Ruminococcaceae_TSDC17.2_V2.2.3	+		
Ruminococcaceae_TSDC17.2_V2.2.1	+		
Ruminococcaceae_TSDC17.2_V2.2.1	+		
Ruminococcaceae_TSDC17.2_V2.1.1	+		
Peptoniphilus_TSDC17.2_V2.1.1			
Parabacteroides_distasonis_TSDC17.2_V2.1.2			
Odoribacter_splanchnicus_TSDC17.2_V2.1.2			
Odoribacter_splanchnicus_TSDC17.2_V2.1.1			
final_name			
Escherichia_coli_TSDC17.2_V2.1.2			
Escherichia_coli_TSDC17.2_V2.1.1			
Dorea_TSDC17.2_V2.1.1			
Coprococcus_comes_TSDC17.2_V2.1.2			
Coprococcus_comes_TSDC17.2_V2.1.1			
Collinsella_aerofaciens_TSDC17.2_V2.4.22	+	+	
Collinsella_aerofaciens_TSDC17.2_V2.3.23	+	+	
Collinsella_aerofaciens_TSDC17.2_V2.3.20	+	+	
Collinsella_aerofaciens_TSDC17.2_V2.2.24	+	+	
Collinsella_aerofaciens_TSDC17.2_V2.1.9	+	+	
Collinsella_aerofaciens_TSDC17.2_V2.1.10	+	+	

Table 25. Components of the arrayed anaerobic bacterial culture collection produced from the lean co-twin in DZ pair 1.

Strain designation	37 Selected taxa	Lactose degrading	Cellobiose degrading
Clostridiaceae_TSDC17.2_V2.3.1			
Clostridiaceae_TSDC17.2_V2.2.4			
Bifidobacterium_pseudocatenulatum_TSDC17.2_V2.1.5	+		
Bacteroides_vulgatus_TSDC17.2_V2.2.12	+		-
Bacteroides_vulgatus_TSDC17.2_V2.1.5	+		-
Bacteroides_vulgatus_TSDC17.2_V2.1.11	+		-
Bacteroides_thetaiotaomicron_TSDC17.2_V2.3.1	+		-
Bacteroides_thetaiotaomicron_TSDC17.2_V2.2.5	+		-
Bacteroides_thetaiotaomicron_TSDC17.2_V2.2.4	+		-
Bacteroides_thetaiotaomicron_TSDC17.2_V2.1.3	+		-
Bacteroides_ovatus_TSDC17.2_V2.3.1	+		-
Bacteroides_ovatus_TSDC17.2_V2.2.2	+		-
Bacteroides_massiliensis_TSDC17.2_V2.1.3	+		-
Bacteroides_massiliensis_TSDC17.2_V2.1.2	+		-
Bacteroides_intestinalis_TSDC17.2_V2.1.9	+		-
Bacteroides_intestinalis_TSDC17.2_V2.1.7	+		-
Bacteroides_intestinalis_TSDC17.2_V2.1.5	+		-
Bacteroides_finegoldii_TSDC17.2_V2.1.4	+		-
Bacteroides_finegoldii_TSDC17.2_V2.1.2	+		-
Bacteroides_caccae_TSDC17.2_V2.1.7	+		-
Bacteroides_caccae_TSDC17.2_V2.1.6	+		-
Bacteroides_caccae_TSDC17.2_V2.1.3	+		-
Bacteroides_caccae_TSDC17.2_V2.1.1	+		-
Bacteroides_acidifaciens_TSDC17.2_V2.1.8	+		-
Bacteroides_acidifaciens_TSDC17.2_V2.1.3	+		-
Anaerococcus_TSDC17.2_V2.1.1			

[0212] The experimental design is shown in **FIG. 28A, B**. Groups of mice were colonized with one of three culture collections: the non-arrayed collection from the obese co-twin; the non-arrayed collection or the assembled 37-member consortium from lean co-twin. In the case of the 37-member consortium, the abundance of members in the non-arrayed collection was not preserved; rather, equivalent numbers of cells/strain were inoculated into recipient mice. Five days after inoculation, mice harboring the non-arrayed culture collection from the obese co-twin were co-housed with one another (negative control), or with mice containing the non-arrayed culture collection from the lean co-twin (positive control), or were co-housed with mice containing the 'manufactured' 37 member collection (5 mice/treatment group; total of 26 mice). **FIG. 28A** emphasizes how mice with 37-member consortium can be viewed as a

prevention arm of the experiment: i.e., does the presence of these microbes prevent their host from gaining the level of adiposity achieved in mice harboring the obese co-twin microbiota?). These mice can also be viewed as a treatment arm: i.e., can the 37-member consortia ameliorate the increased adiposity phenotype that develops in mice colonized with the obese co-twin's culture collection? Quantitative MR was performed on days 1, 5, 15 and 19 of the 20d long experiment revealed that the non-arrayed lean co-twin's culture collection produced a consistent reduction in adiposity in co-housed mice with the obese co-twin's culture collection. In 4 of 5 cages, co-housed mice with 37-member consortia (Ln37^{ch}) did not develop increased adiposity; in two of these 4 cages, the cage-mate with the obese co-twin's culture collection (Ob^{ch}) exhibited a very marked reduction in adiposity compared to the control group (dually housed mice, each with the obese co-twin's culture collection) (see arrows in **FIG. 28C**).

[0213] A similar experiment may be performed to further identify individual microbes from the 37-member consortium that, when inoculated into gnotobiotic mice and cohoused with mice containing the obese co-twin's culture collection, may induce reduced adiposity in the mice containing the obese co-twin's culture collection, and prevent increased adiposity in the mice containing the individual member of the 37-member consortium. In short, wherein individual members of the 37-member consortium may be used to colonize gnotobiotic mice. The mice may then be cohoused with mice containing the obese co-twin's culture collection, and the change in adiposity of all mice may be measured over time as described above. Such an experiment may identify individual members of the 37-member consortium that may induce reduce adiposity in obese mice, or prevent increased adiposity in mice harboring the individual member of the 37-member consortium.

Example 16. Discordance for obesity among adult female twin pairs in the MOAFTS study cohort.

[0214] There were 3,427 participants in the MOAFTS wave 5 assessment; height and weight data were available for 3,416 (99.7%). The majority of participants (55.8%) were classified as lean (BMI 18.50-24.99 kg/m²), while 21.9% were classified as overweight (25-29.99 kg/m²), 18.3% as obese (≥30 kg/m²), and 3.98% as

underweight ($<18.5 \text{ kg/m}^2$). African-Americans, who comprised 14.4% of the wave 5 sample, had significantly higher rates of overweight and obesity compared to European-Americans (EA) (32.5% and 36.6% vs. 20.1% and 15.2%, respectively $p<0.001$).

[0215] Height and weight data were available for 1,539 complete twin pairs participating in wave 5. 54.3% of the twin pairs were MZ. The mean difference in BMI between co-twins was 3.53 kg/m^2 (SD 3.78 kg/m^2). The mean difference in BMI was greater in DZ compared to MZ twin pairs ($4.65 \pm 4.58 \text{ kg/m}^2$ versus $2.60 \pm 2.57 \text{ kg/m}^2$; $p<0.001$). Using the criteria that one co-twin was obese and the other lean, 5.72% of twin pairs were defined as BMI discordant (mean difference $=11.42 \pm 4.09 \text{ kg/m}^2$). The rate of discordance was substantially lower for MZ pairs compared to DZ pairs (2.3% versus 9.9%; $p<0.001$) and AA twin pairs were more likely to be discordant than EA pairs ($p=0.008$). Alternatively, when BMI discordance was defined as a BMI difference $\geq 8 \text{ kg/m}^2$, 18.3% of DZ pairs and 5.2% of MZ pairs were classified as discordant ($p<0.001$); AA pairs were again more likely to be discordant (21.6% vs. 9.4%; $p<0.001$).

Methods for Examples 15 and 16.

[0216] **Animal husbandry.** All experiments involving mice were performed using protocols approved by the Washington University Animal Studies Committee. Germ-free adult male C57BL/6J mice were maintained in plastic flexible film gnotobiotic isolators under a strict 12 hr light cycle and fed an autoclaved low-fat, polysaccharide-rich chow diet (B&K diet 7378000) ad libitum.

[0217] **Collection of fecal samples from twin pairs discordant for obesity and transplantation of their uncultured fecal microbiota into germ-free mice.** Adult female twin pairs with a BMI discordance ranging from $6\text{--}10 \text{ kg/m}^2$ were recruited for this study. Procedures for obtaining their consent to provide fecal samples were approved by the Washington University Human Studies Committee. A single fecal sample was collected at $t=0$ and another 2 months later from each subject. Each sample was frozen immediately at -20°C , shipped in a frozen state to a biospecimen repository overseen by one of the authors, and then de-identified. All samples were subsequently stored at -80°C until the time of processing.

[0218] A given human fecal sample was homogenized with a mortar and pestle packed in dry ice. A 500 mg aliquot of the pulverized material was diluted in 5 mL of reduced PBS (PBS supplemented with 0.1% Resazurin (w/v), 0.05% L-cysteine-HCl), in an anaerobic Coy chamber (atmosphere, 70% N₂, 25%CO₂, 5% H₂), and then vortexed at room temperature for 5 min. The suspension was allowed to settle by gravity for 5 min, after which time the clarified supernatant was transferred to an anaerobic crimped tube that was then transported to the gnotobiotic mouse facility. The surface of the tube was sterilized by exposure for 20 min to chlorine dioxide in the transfer sleeve attached to the gnotobiotic isolator, transferred into the isolator. A 1 mL syringe was used to obtain a 200 µL aliquot of the suspension and was introduced by gavage into each adult C57BL6/J germ-free recipient. Transplant recipients were maintained in separate cages within an isolator dedicated to mice colonized with the same donor microbiota, except in the case of the co-housing experiments described below.

[0219] **Analysis of body composition by quantitative magnetic resonance imaging (MRI).** Body composition was defined using MRI analysis (EchoMRI-3in1 instrument; EchoMRI, Houston, TX). Mice were transported from the gnotobiotic isolator to the MR instrument using in a HEPA filter capped glass vessel. Fat, lean and tissue-free body water were measured 1 d after gavage, and weekly for up to 5 weeks.

[0220] **Sample collection from gnotobiotic mice.** Fecal samples were collected at defined times after gavage from the mouse. At the time of sacrifice, luminal contents were collected as described in the Examples above at defined positions along the length of the gut (stomach, small intestinal segments 1, 2, 5, 9, 13 and 15 after its division into 16 equal-sized segments, cecum, and proximal and distal halves of the colon). Blood was harvested by retro-orbital phlebotomy into capillary blood collection tubes (BD), which were then centrifuged at 5,000 x g at 4°C for 5 min. The supernatant (serum) was frozen in liquid nitrogen for subsequent GC/MS analyses. Urine, also obtained at the time of sacrifice, was flash frozen in liquid nitrogen for metabolomic analysis. Both epididymal fat pads were recovered from each animal, by dissection, and weighed.

[0221] **Multiplex pyrosequencing of amplicons generated from bacterial 16S rRNA genes.** Genomic DNA was extracted from feces and gut samples using a bead-beating protocol. Briefly, a ~500 mg aliquot of each pulverized frozen human fecal sample, or mouse fecal pellets (~50 mg), or stomach, small intestinal, cecal or colonic contents (~20 mg each); were re-suspended in a solution containing 500µL of extraction buffer [200mM Tris (pH 8.0), 200mM NaCl, 20mM EDTA], 210µL of 20% SDS, 500µL of phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1, Ambion) and 500µL of a slurry of 0.1-mm diameter zirconia/silica beads. Cells were then mechanically disrupted using a bead beater (Biospec, maximum setting; 3 min at room temperature), followed by extraction with phenol:chloroform:isoamyl alcohol and precipitation with isopropanol.

[0222] Amplicons of ~330bp, spanning variable region 2 (V2) of the 16S rRNA gene, were generated by using modified primers 8F and 338R incorporating sample specific barcodes as described in the Examples above and subjected to multiplex pyrosequencing (454 FLX Standard or Titanium chemistry). V2-16S rRNA sequences from Titanium chemistry were trimmed to FLX standard length and, together with the sequenced generated using FLX chemistry, filtered for low quality reads and assigned to a particular pyrosequencing bin according to their sample-specific barcodes. Sequencing errors were corrected using OTUpipeline (QIIME v1.3) and classified into 97% ID OTUs using UCLUST. A representative OTU set was created using the most-abundant OTU from each bin. Reads were aligned using PyNAST. Taxonomy was assigned using RDP classifier.

[0223] Samples were rarefied at a depth of 815 OTUs/sample for time series studies of the fecal microbiota of gnotobiotic recipients of human microbiota and for the donor fecal microbiota) and 800 OTUs/sample in the case of the gut biogeography datasets. Data analysis (beta-diversity calculations, PCoA clustering) was performed using QIIME v1.3 and Vegan R package v 1.17-4 for pairwise distance analysis.

[0224] **Shotgun pyrosequencing of total community DNA.** For multiplex pyrosequencing (454 FLX Titanium chemistry), each of the 45 cecal DNA samples was randomly fragmented by nebulization to 500-800 bp and subsequently

labeled with one of 12 MIDs (Multiplex Identifier; Roche) using the MID manufacturer's protocol (Rapid Library preparation for FLX Titanium). Equivalent amounts of up to 12 MID-labeled samples were pooled prior to each sequencer run. Shotgun reads were filtered to remove all reads <60 nt long, LR70 reads with at least one degenerate base (N), or reads with two continuous and/or three total degenerate bases, plus all duplicates (defined as sequences whose initial 20 nt were identical and shared an overall identity of >97% throughout the length of the shortest read). In the case of human fecal DNAs, all sequences with significant similarity to human reference genomes (BLASTN with e-value < 10^{-5} , bitscore > 50, percent identity > 75%) were removed. Comparable filtering against the mouse genome was performed for reads produced from samples obtained from recipient gnotobiotic animals.

[0225] All resulting filtered sequences were queried against the KEGG database (v58) using BLASTX. Sequences were annotated as the best hit in the database if (i) they had an E-value < 10^{-5} ; (ii) the bit score was > 50; and (iii) the query and subject were at least 50% identical after being aligned. If two entries were assigned as the best BLAST hit, the read was annotated with both entries. KO, E.C., and KEGG Pathway assignments were made using the "ko" file provided by KEGG. A matrix containing the counts for each KEGG annotation for each sample was generated for analysis with ShotgunFunctionalizeR (R package version 1.2-8).

[0226] **Microbial RNA-Seq.** Each fecal pellet (~ 50mg) collected 15 or 17 days after colonization, was suspended while frozen in 1ml of RNeasy Protect Bacteria reagent (Qiagen), vortexed for 5 min at room temperature and centrifuged (10 min; 5,000 X g; 4°C). After decanting the supernatant, pelleted cells were suspended in 500µL of extraction buffer [200 mM NaCl, 20 mM EDTA], 210µl of 20% SDS, 500 µL of phenol:chloroform:isoamyl alcohol (pH 4.5, 125:24:1, Ambion), and 250 µL of acid-washed glass beads (Sigma-Aldrich, 212-300 µm diameter). Microbial cells were lysed by mechanical disruption using a bead beater (Biospec, maximum setting; 5 min at room temperature), followed by phenol:chloroform:isoamyl alcohol extraction and precipitation with isopropanol. RNA was treated with RNase-free TURBO-DNase (Ambion) and 5S rRNA and tRNAs were removed using MEGAClear columns (Ambion). A second DNase treatment was performed (Baseline-ZERO DNase; Epicenter). rRNA

was initially depleted using MICROBexpress kit (Ambion) followed by a second MEGAClear purification. In addition, custom biotinylated oligonucleotides, directed against conserved regions of sequenced human gut bacterial rRNA genes were employed for streptavidin bead-based pulldowns. cDNA was synthesized using SuperScript II (Invitrogen), followed by second strand synthesis with RNaseH, *E. coli* DNA polymerase (NEB) and *E. coli* DNA ligase (NEB). Samples were sheared using a BioRuptor XL sonicator (Diagenode) and 150-200 bp fragments gel selected and prepared for sequencing.

[0227] Multiplexed microbial cDNA sequencing was performed using Illumina Hi-Seq2000 instruments to generate 23.7 ± 6.5 million unidirectional 101 nt reads/sample. Reads were split according to 4-bp barcodes used to label each of four samples pooled together. After dividing sequences by barcode, reads were mapped to genes in a custom database of 127 sequenced human gut microbial genomes using the ssaha2 algorithm. A minimum score threshold of 42 was selected for ssaha, based on the distribution of scores for 101 nt barcoded reads. If a read mapped to more than one location in a genome or multiple genomes, the counts for each gene were added to the gene according to the gene's fraction of unique-match counts. Pseudo counts were added (i.e. added 1 count) to each gene prior to normalization to account for different sampling depths (i.e. normalized to reads/kb/million mapped reads).

[0228] **Culturing fecal microbiota.** Each human fecal sample was pulverized while frozen and resuspended in pre-reduced PBS (0.1% Resazurin, 0.05% Cysteine/HCl; 15mL/g feces). Samples were subsequently vortexed for 5 min and allowed to settle by gravity for 5 min to permit large, insoluble particles to settle. The supernatant was diluted 1000-fold in pre-reduced PBS and plated on 150 mm diameter plates containing pre-reduced, non-selective Gut Microbiota Medium (GMM, Goodman et al., 2011). Plates were incubated in a Coy chamber under anaerobic conditions for 7d at 37°C. Colonies were subsequently harvested en masse from six plates by scraping (10 mL of pre-reduced PBS/plate). Glycerol (30%)/PBS stocks were stored in anaerobic glass vials at -80°C. 200 μ L of the non-arrayed culture collection was used to gavage each germ-free recipient.

[0229] **Creating a clonally arrayed taxonomically defined sequenced culture collection.** Methods for creating clonally arrayed culture collections from frozen fecal samples were as described in the Examples above. A set of interfaces was also created for a Precision XS robot (BioTek) so that picking, arraying, and archiving of fecal bacterial culture collections can be done with speed and economy within a Coy anaerobic chamber. Taxonomies were assigned to each strain in an arrayed collection by 454 Titanium V2-16S rRNA pyrosequencing.

[0230] For a given culture collection, most strains (unique V2-16S rRNA sequence) are found in more than one well across the arrayed library. Therefore, several replicate wells of each strain were picked robotically from the 384-well plate, and streaked onto 8-well TYGS-agar plates. Plates were incubated under anaerobic conditions for 3 d at 37°C in a Coy chamber. A single colony from each agar well was picked, grown in TYGS and archived as a TYGs/15% glycerol stock at -80°C. A small aliquot of each stock was taken for DNA extraction and subjected to multiplex genome sequencing with an Illumina HiSeq 2000 instrument [fold coverage 119 ± 66 (mean $\pm$ SD) coverage; range, 35-289].

CLAIMS

What is claimed is:

1. A model, the model comprising a gnotobiotic animal comprising a cultured gut microbial community.
2. The model of claim 1, wherein the gnotobiotic animal is a mouse.
3. The model of claim 1, wherein the gnotobiotic animal is a pig.
4. The model of claim 2, wherein the cultured gut microbial community is from an animal other than the gnotobiotic animal.
5. The model of claim 1, wherein the cultured gut microbial community is from a human.
6. The model of claim 1, wherein the cultured gut microbial community is an arrayed culture collection of a gut microbial community.
7. The model of claim 1, wherein the cultured gut microbial community has at least 98.0% of the order-level phylotopic composition of the original gut microbial community.
8. The model of claim 1, wherein the cultured gut microbial community has at least 98.0% of the metagenome, transcriptome, or proteome composition of the original gut microbial community.
9. A composition, the composition comprising an in vitro cultured collection of a gut microbial community.
10. The composition of claim 9, wherein the gut microbial community is from a human.
11. The composition of claim 9, wherein the cultured microbial members were cultured on gut microbiota medium.

12. The composition of claim 9, wherein the cultured microbial members are from an arrayed culture collection of a gut microbial community.
13. The composition of claim 9, wherein the cultured gut microbial community has at least 98.0% of the order-level phylotopic composition of the original gut microbial community.
14. The composition of claim 9, wherein the cultured gut microbial community has at least 98.0% of the metagenome, transcriptome, or proteome composition of the original gut microbial community.
15. The composition of claim 9, wherein the gut microbial community is from a germfree mouse colonized with a gut microbial community.
16. A composition, the composition comprising a clonally arrayed culture collection of a gut microbial community.
17. The composition of claim 14, wherein each member of the collection is assigned a barcode.
18. The composition of claim 14, wherein the barcode is generated using a two step barcoded pyrosequencing scheme.
19. The composition of claim 14, wherein the gut microbial community is from a human.
20. The composition of claim 14, wherein the gut microbial community is from a germfree mouse colonized with a gut microbial community.
21. The composition of claim 14, wherein the clonally arrayed culture collection was prepared without colony picking.
22. The composition of claim 14, wherein the clonally arrayed culture collection was prepared using a most probable number (MPN) technique.

23. The composition of claim 14, wherein the clonally arrayed culture collection has at least 98.0% of the phylotopic composition of the original gut microbial community.
24. The composition of claim 14, wherein the cultured gut microbial community has at least 98.0% of the metagenome, transcriptome, or proteome composition of the original gut microbial community.
25. A method of determining the effect of a perturbation on a gut microbial community, the method comprising applying the perturbation to a cultured collection of a gut microbial community and determining the difference in the community before and after the application of the perturbation, wherein the difference in the cultured collection represents the effect of the perturbation on the original gut microbial community.
26. The method of claim 25, wherein the perturbation is a diet related perturbation, an environmental perturbation, a genetic perturbation or a pharmaceutical perturbation.
27. A method of predicting variations in the abundance of a member of a gut microbiome of a host in response to a proposed diet, the method comprising
 - a. determining the abundance of a member of a gut microbime in a host,
 - b. determining the amount of the diet ingredients protein, fat, polysaccharide and simple sugar in a proposed diet,
 - c. determining the linear coefficient for a particular gut microbiome member in relation to protein, fat, polysaccharide, and simple sugar,
 - d. predicting the absolute abundance of the member of the gut microbiome if the host were to be fed the proposed diet in (b) based on the linear coefficient from (c) for a particular diet ingredient and the amount of the diet ingredient, and

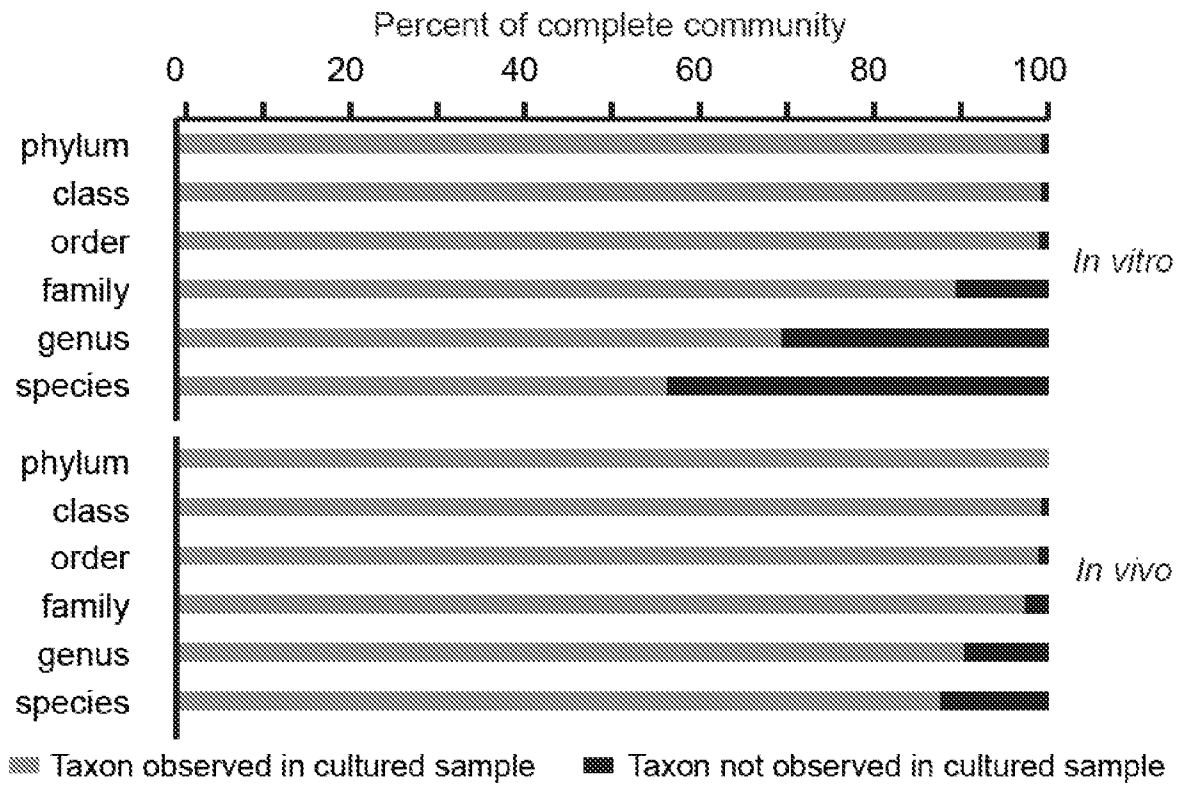
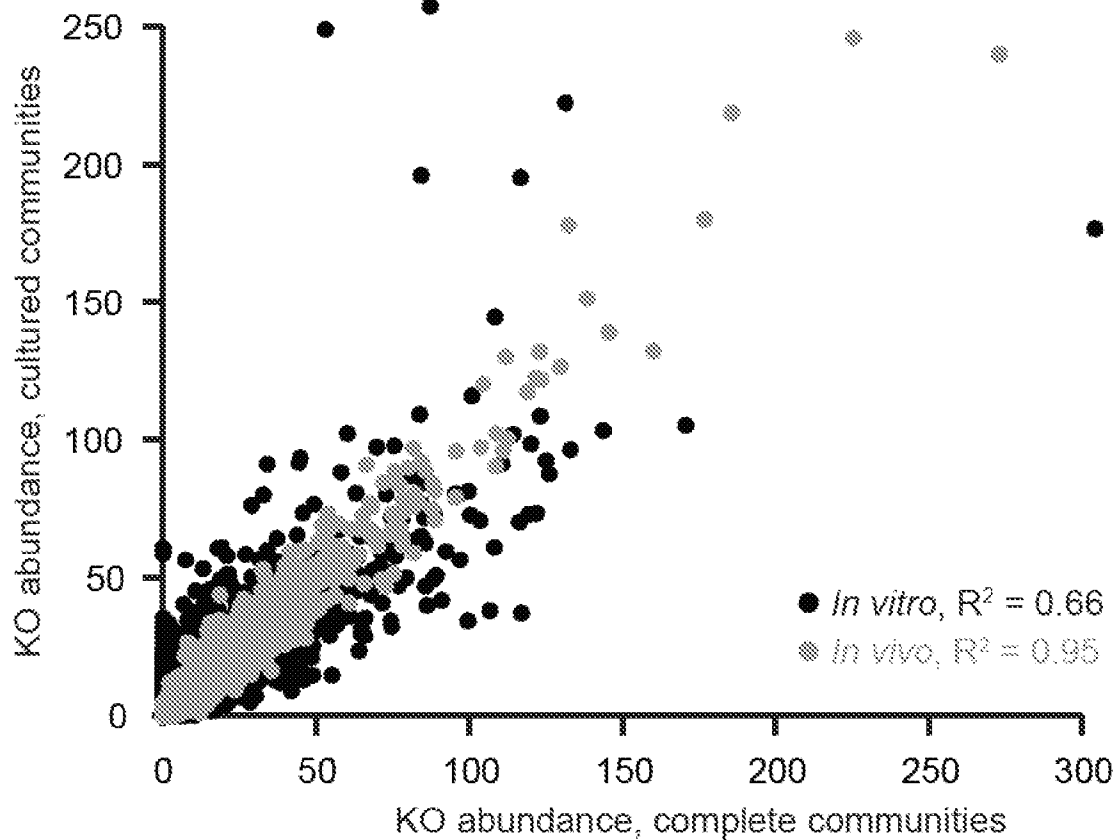
- e. determining the difference between (a) and (d), wherein the difference is the predicted variation in the abundance of a gut microbiome member in response to the proposed diet.
28. A method of predicting the abundance of a member of a gut microbiome of a host in response to a proposed diet, the method comprising
- a. determining the amount of the diet ingredients protein, fat, polysaccharide and simple sugar in a proposed diet,
 - b. determining the linear coefficient for a particular gut microbiome member in relation to protein, fat, polysaccharide, and simple sugar,
 - c. predicting the absolute abundance of the member of the gut microbiome if the host were to be fed the proposed diet in (b) based on the linear coefficient from (c) for a particular diet ingredient and the amount of the diet ingredient.
29. A method of specifically manipulating the abundance of a member of a gut microbiome of a host to a target level by changing the diet of the host, the method comprising
- a. determining the linear coefficient for a particular gut microbiome member in relation to protein, fat, polysaccharide, and simple sugar,
 - b. determining the amount of protein, fat, polysaccharide and sugar in a diet necessary to achieve the target level of the gut microbiome member based on the linear coefficients from (a), and
 - c. feeding a diet to the host that contains the amount of protein, fat, polysaccharide and sugar determined in (b).
30. The method of claims 27-29, wherein the linear coefficient for a particular gut microbiome member for a particular food ingredient is determined using a gnotobiotic mouse model of a human gut microbiome community.

31. The method of claims 27-29, wherein the abundance of a member of a gut microbiome may be calculated with the equation

$$y_i = \beta_0 + \beta_{\text{protein}} X_{\text{protein}} + \beta_{\text{polysaccharide}} X_{\text{polysaccharide}} + \beta_{\text{sucrose}} X_{\text{sucrose}} + \beta_{\text{fat}} X_{\text{fat}}$$

where y_i is the abundance of the member of the gut microbiome, β_0 is the calculated parameter for the intercept, X is the amount in g/(kg of total diet) of the diet ingredient, and β_{protein} , $\beta_{\text{polysaccharide}}$, β_{sucrose} , and β_{fat} are the linear coefficients for a particular gut microbiome member for each of the diet components.

32. The method of claim 31, wherein β_0 for a particular gut microbiome member for a particular food ingredient is determined using a gnotobiotic mouse model of a human gut microbiome community.
33. The method of claims 27-29, wherein the host is human.

A**B****FIG. 1**

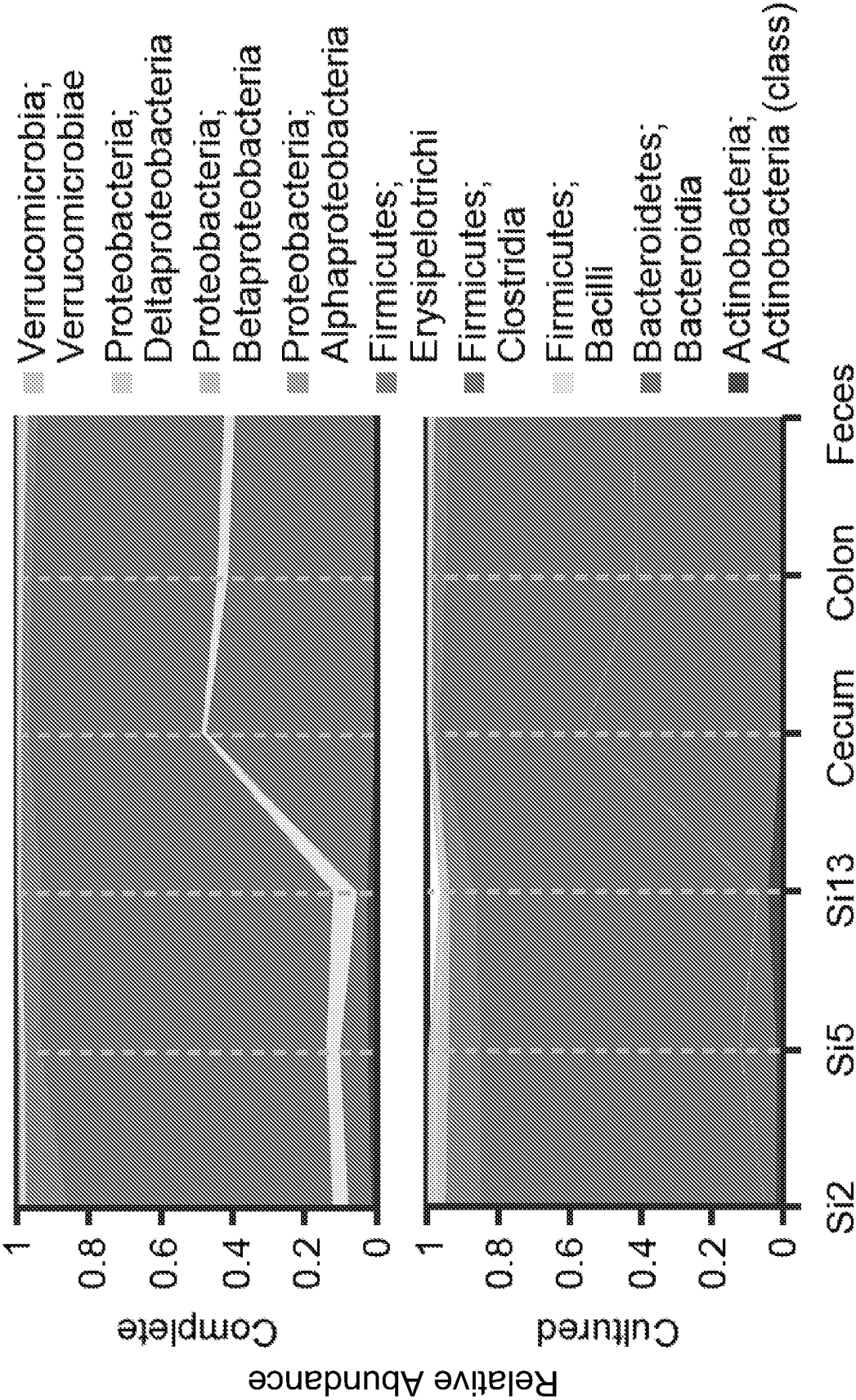
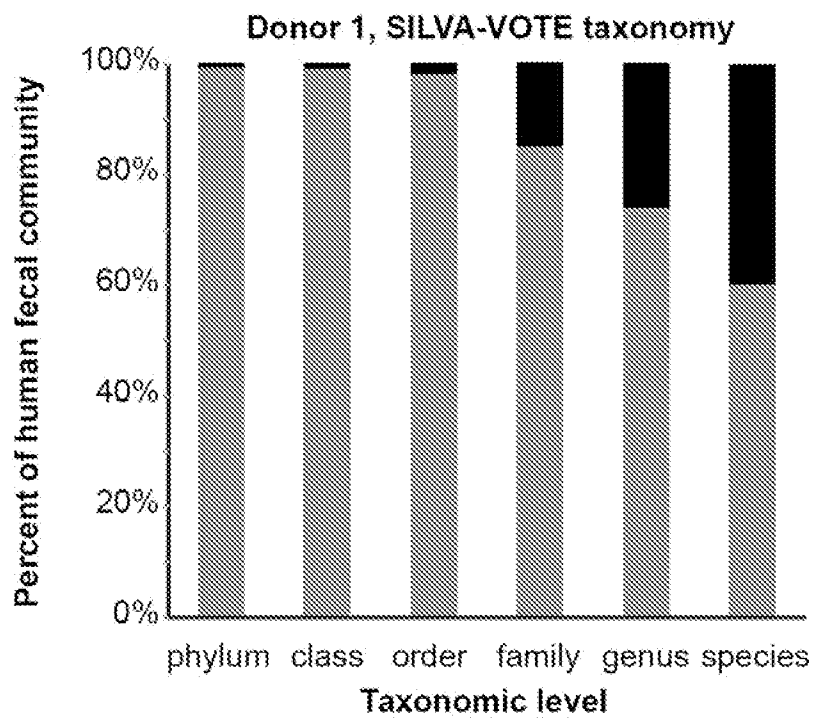
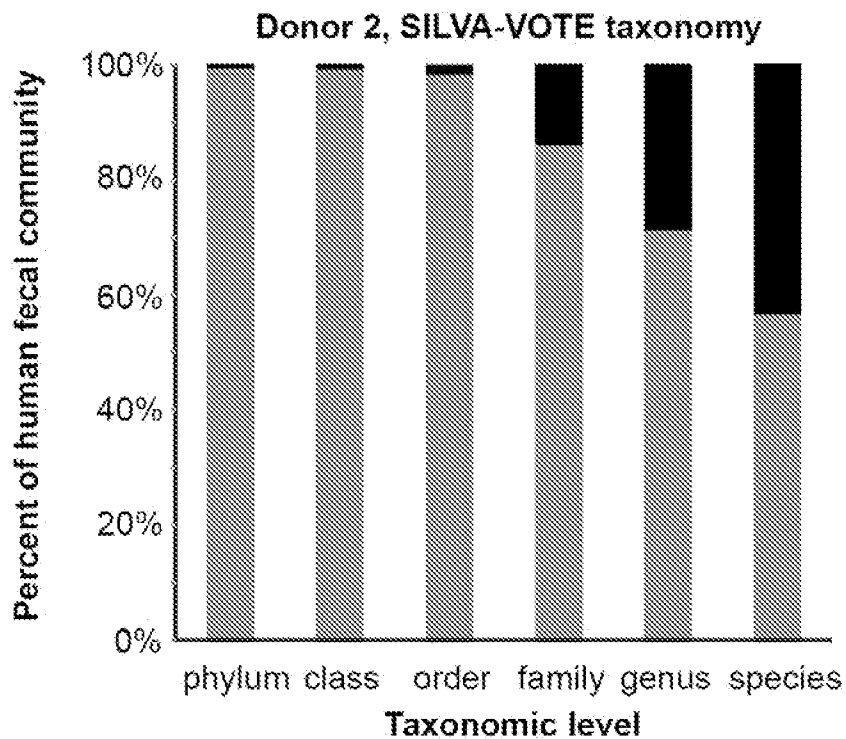
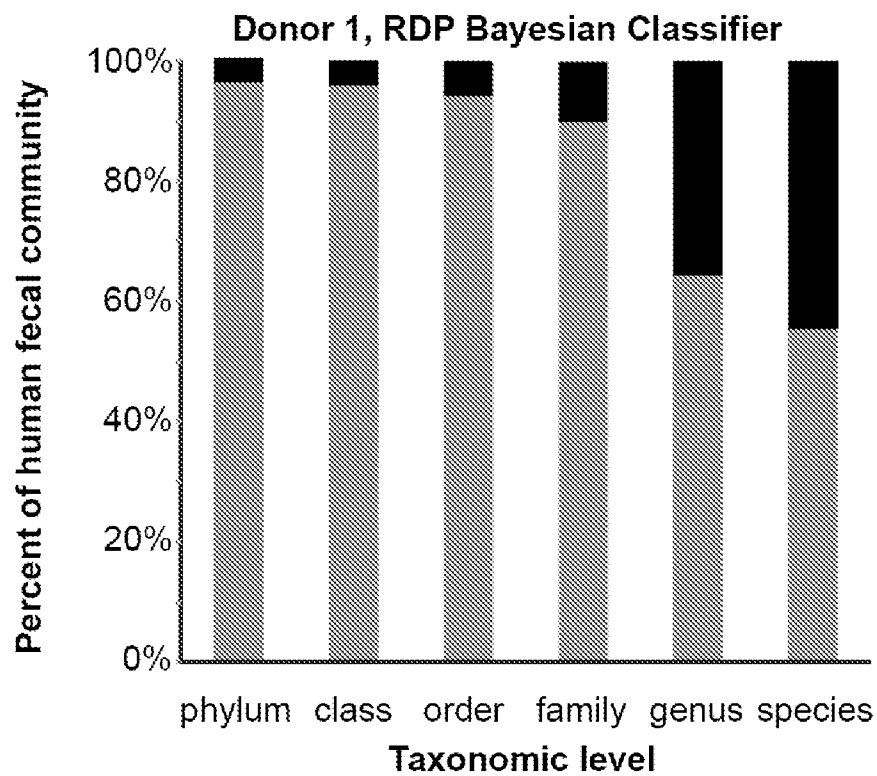
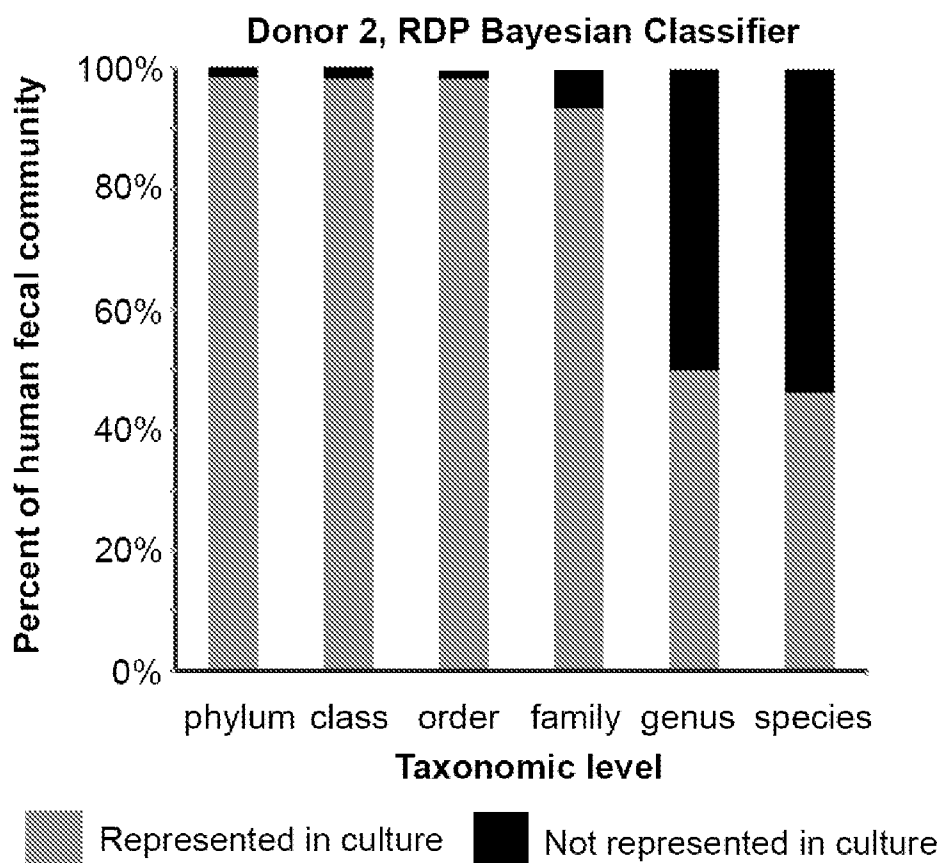


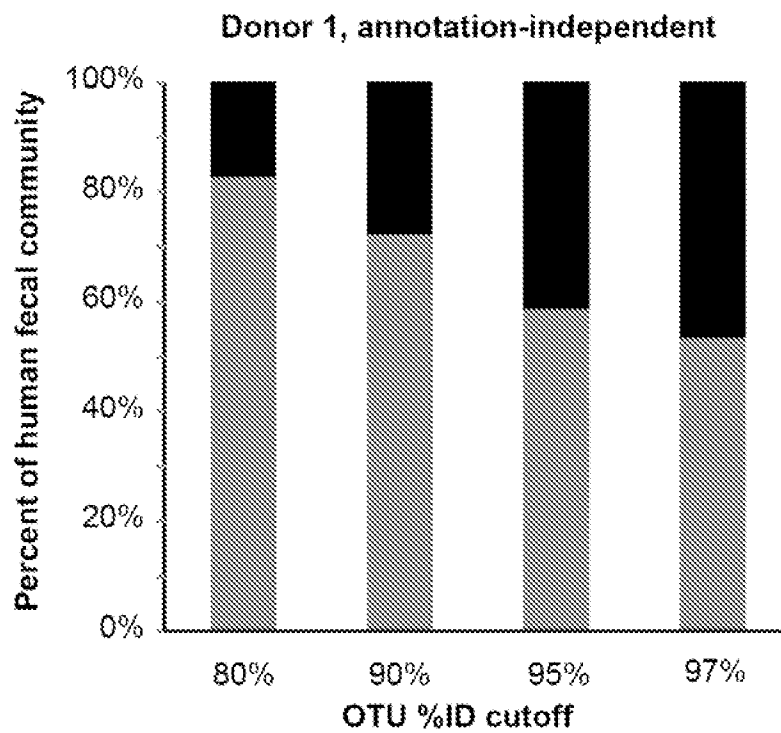
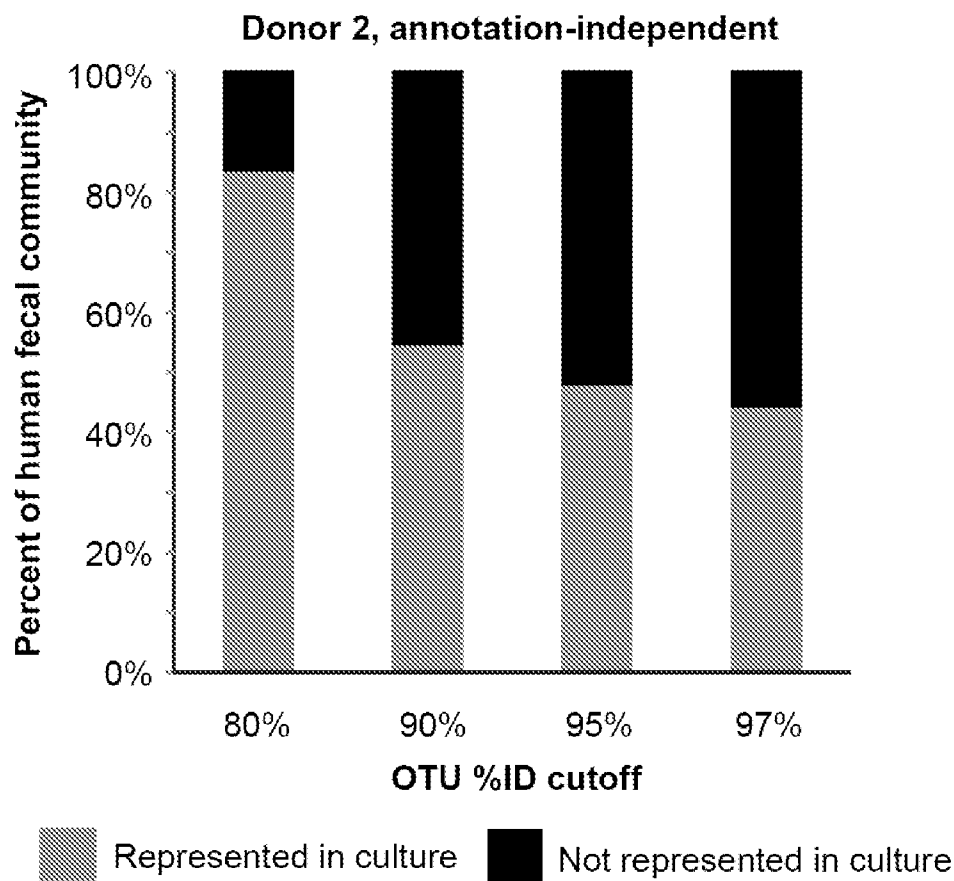
FIG. 1C

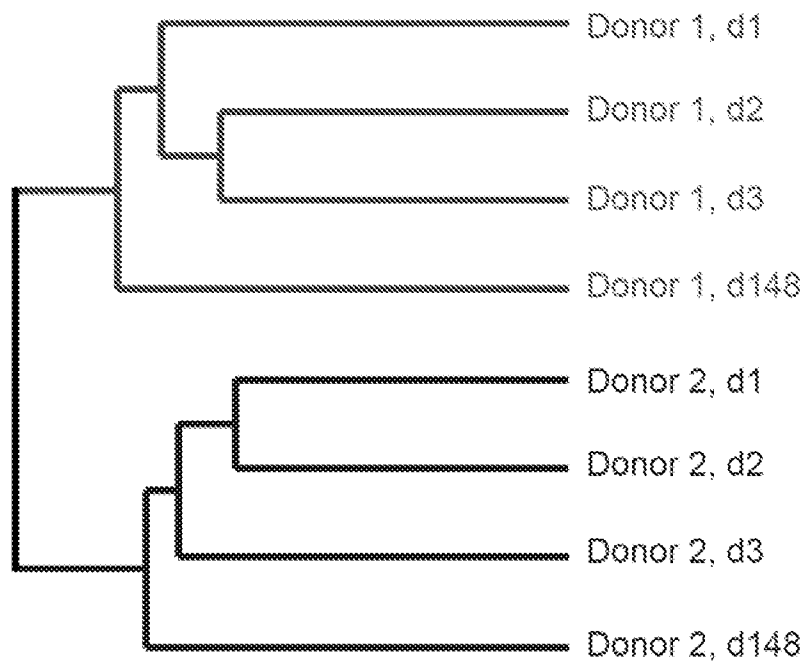
A**B**

 Represented in culture  Not represented in culture

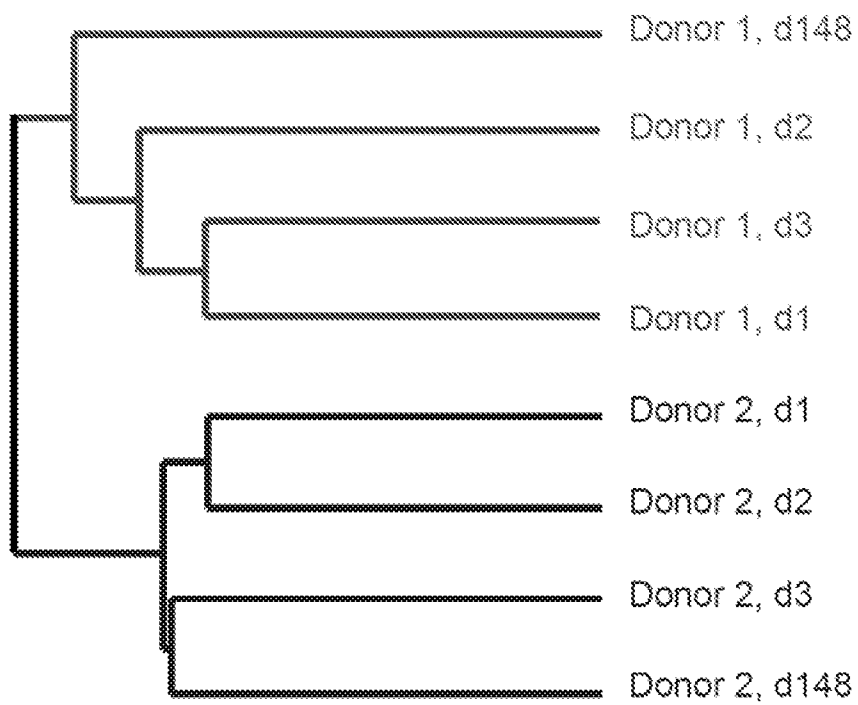
FIG. 2

C**D****FIG. 2**

E**F****FIG. 2**

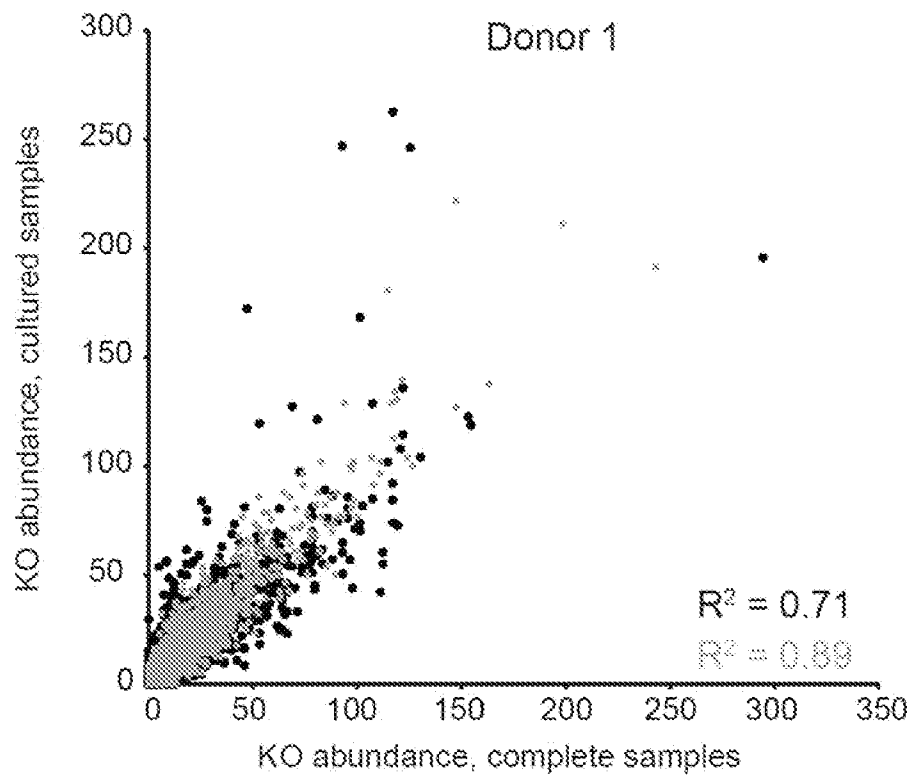
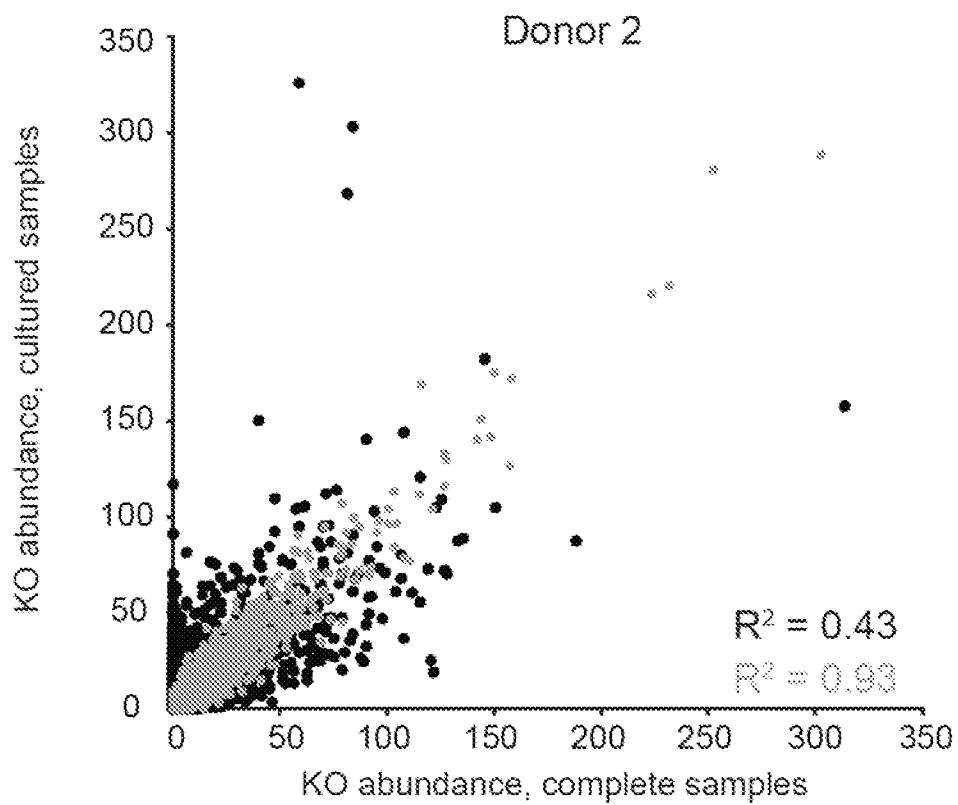
G

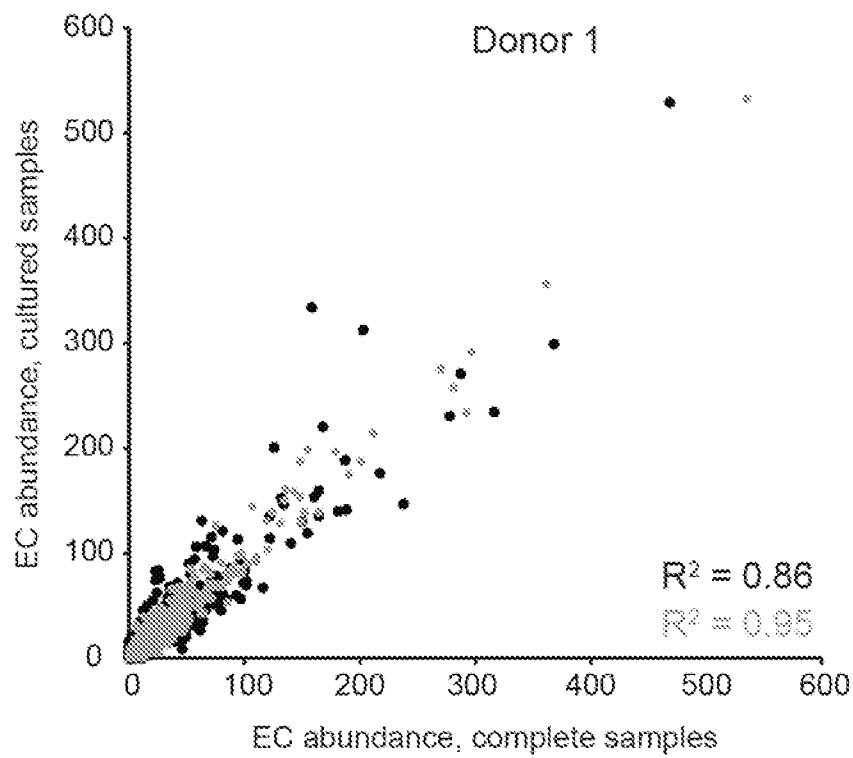
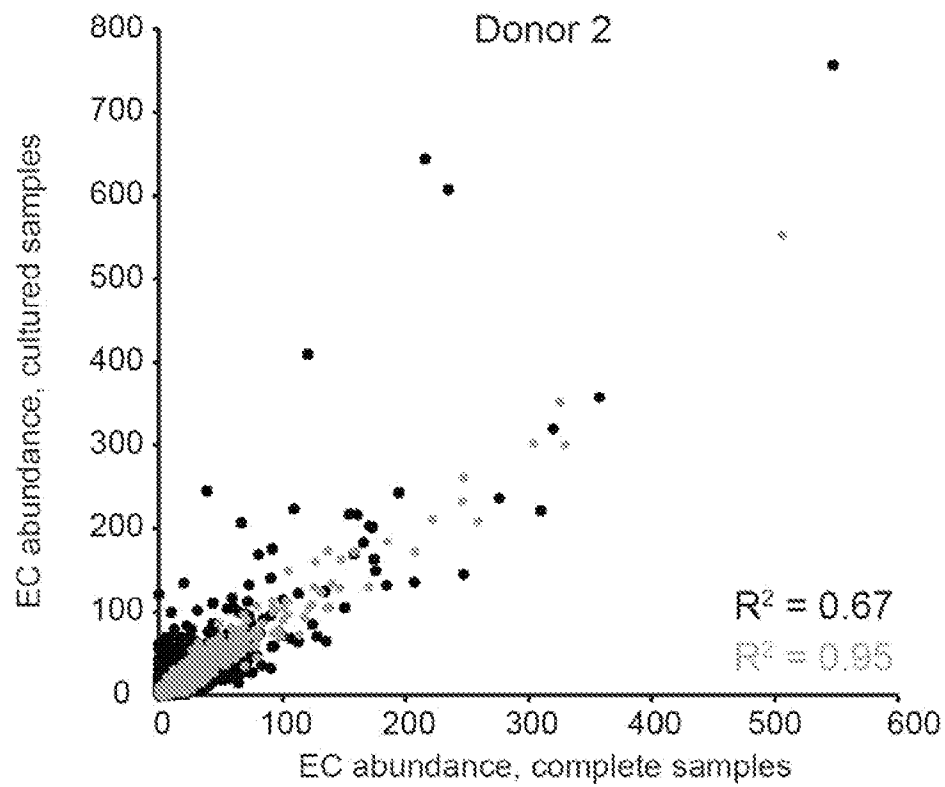
Complete

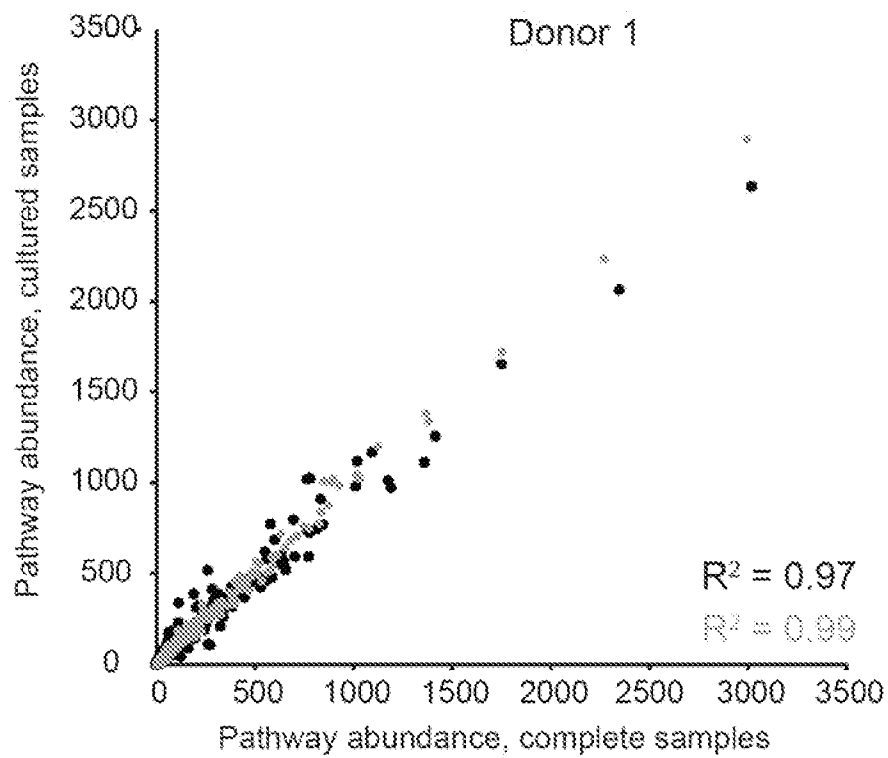
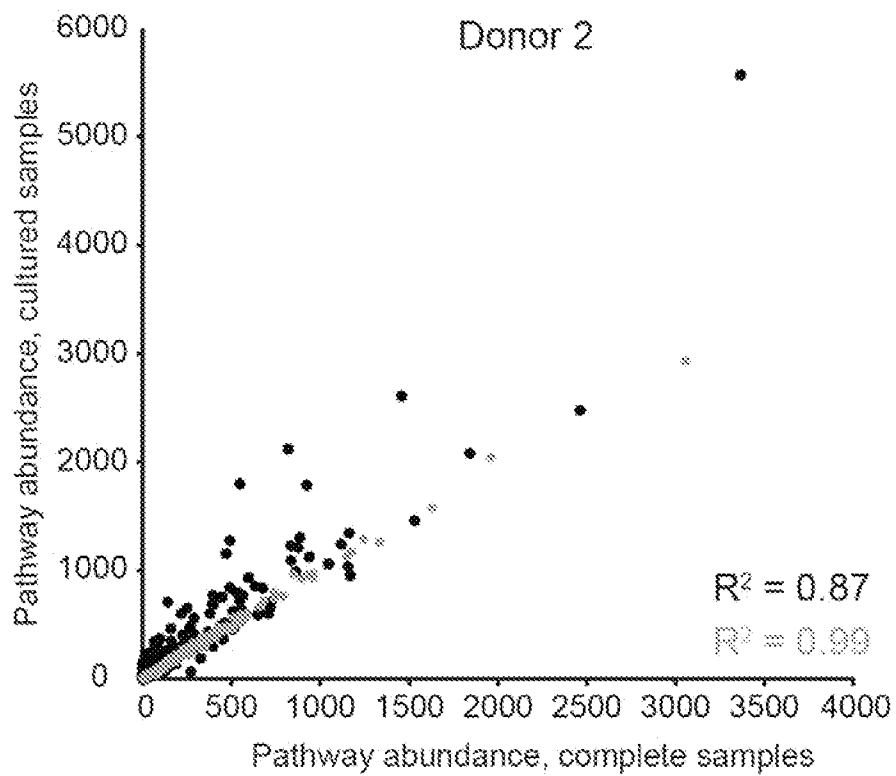
H

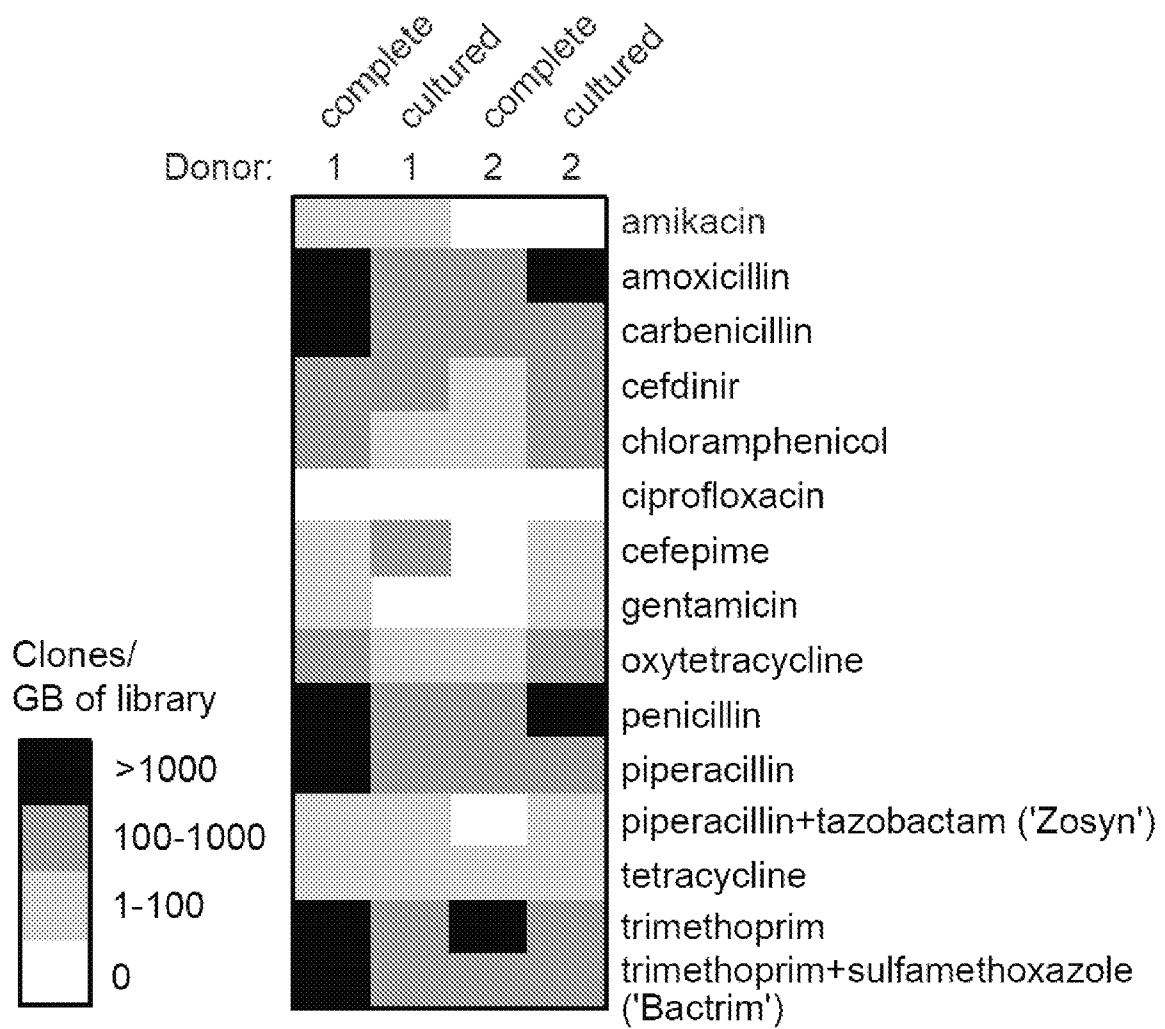
Cultured

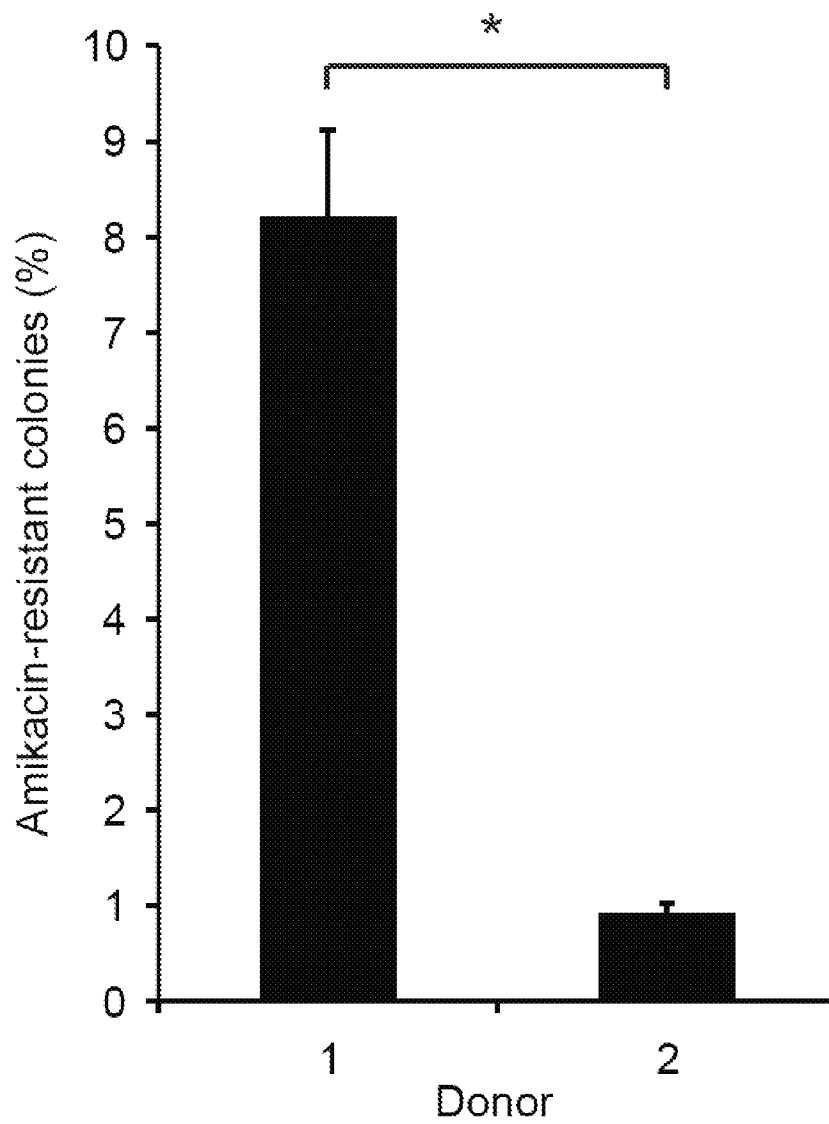
FIG. 2

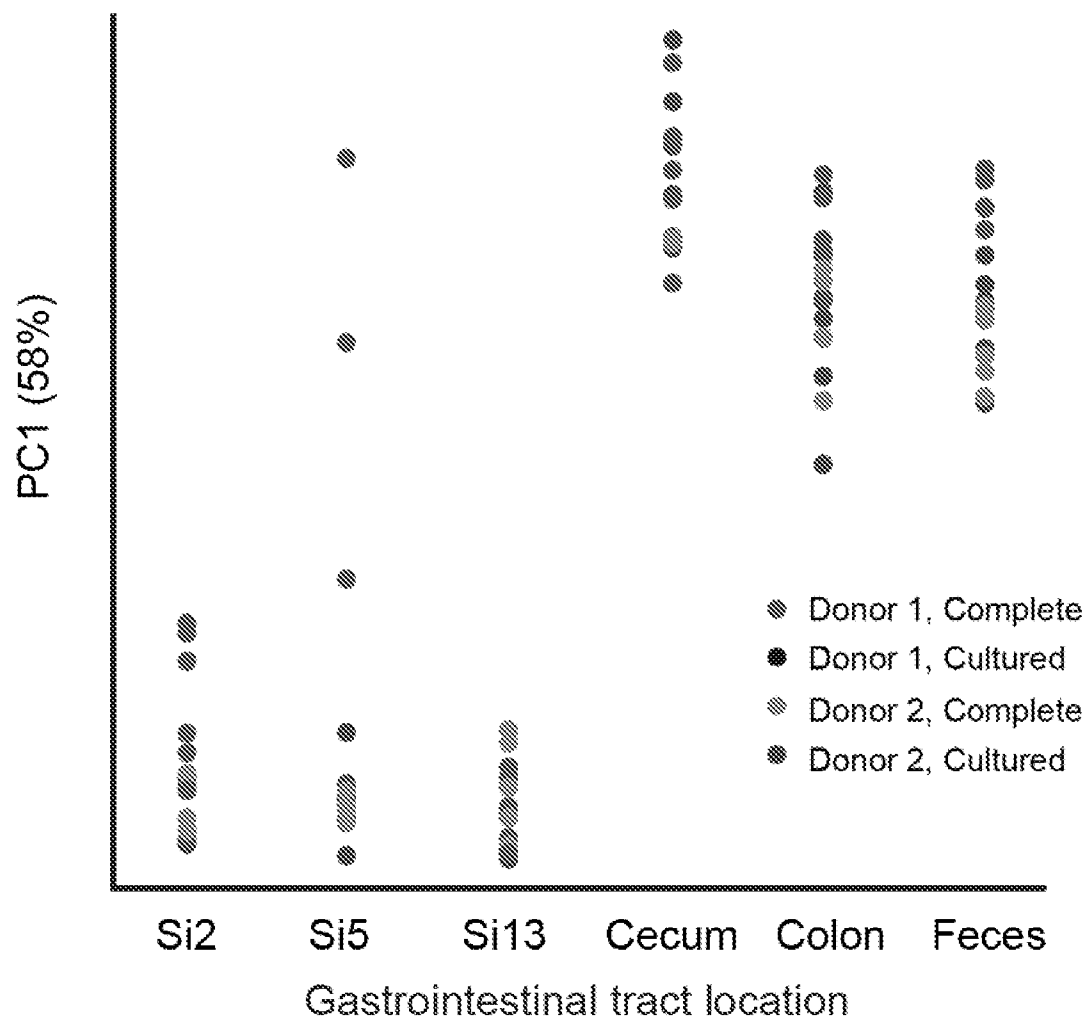
A**B****FIG. 3**

C**D****FIG. 3**

E**F****FIG. 3**

**FIG. 3G**

**FIG. 3H**

**FIG. 4A**

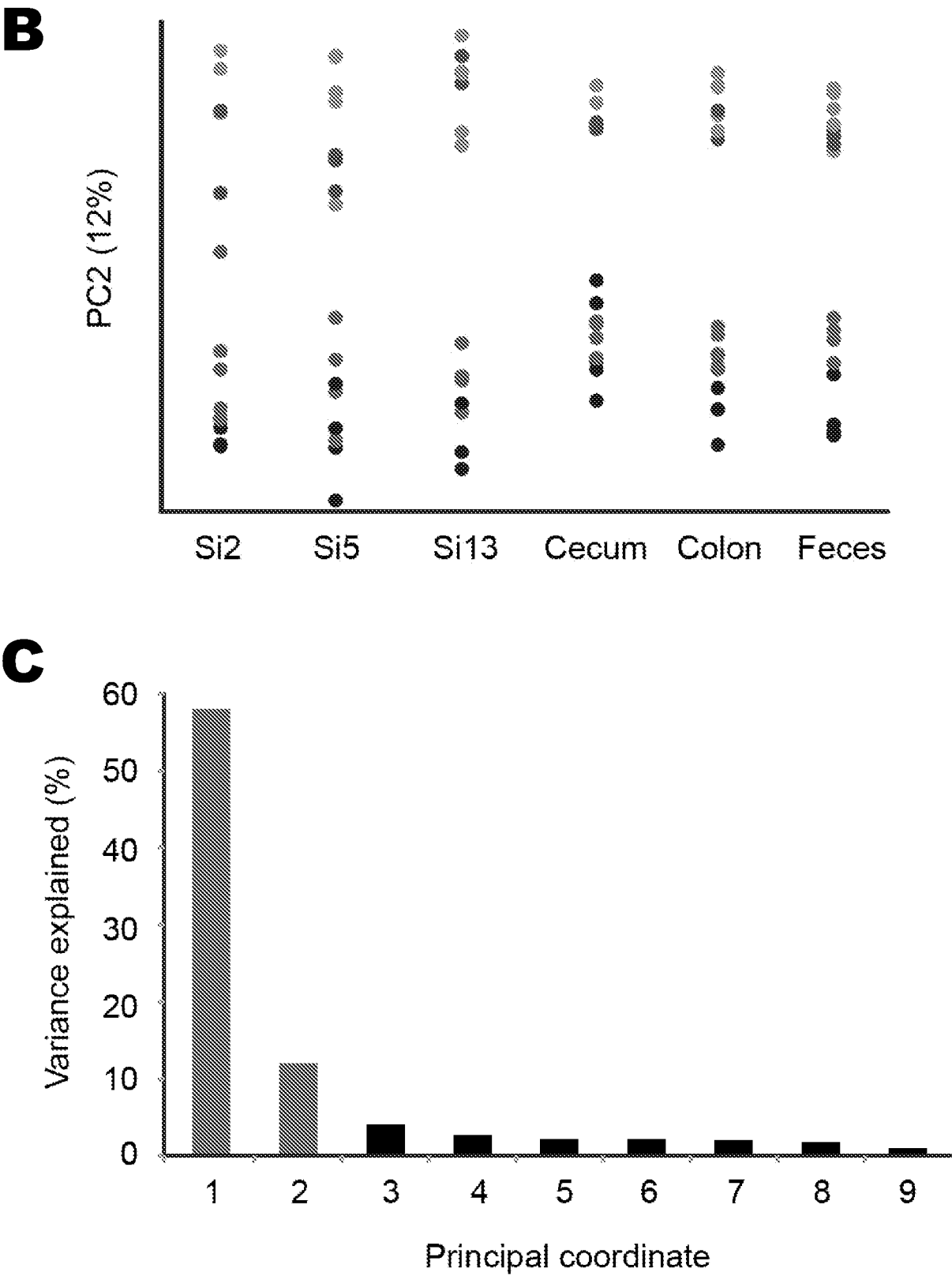


FIG. 4

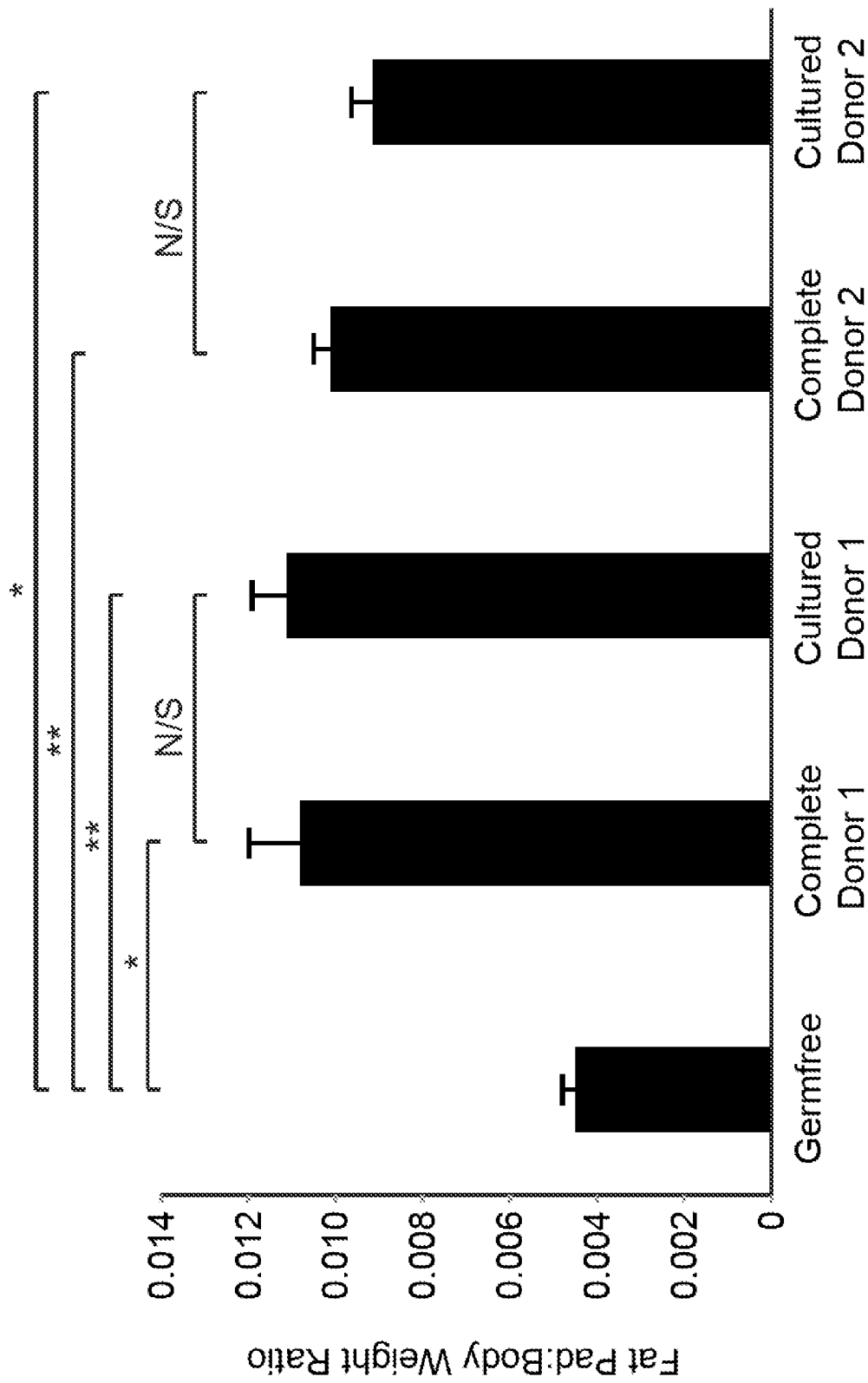
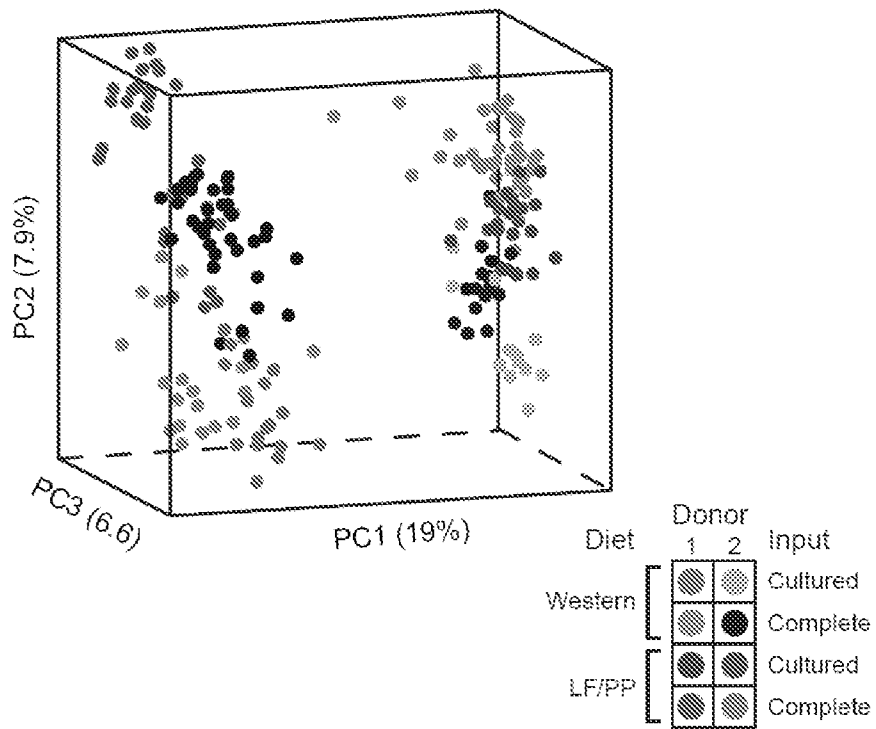
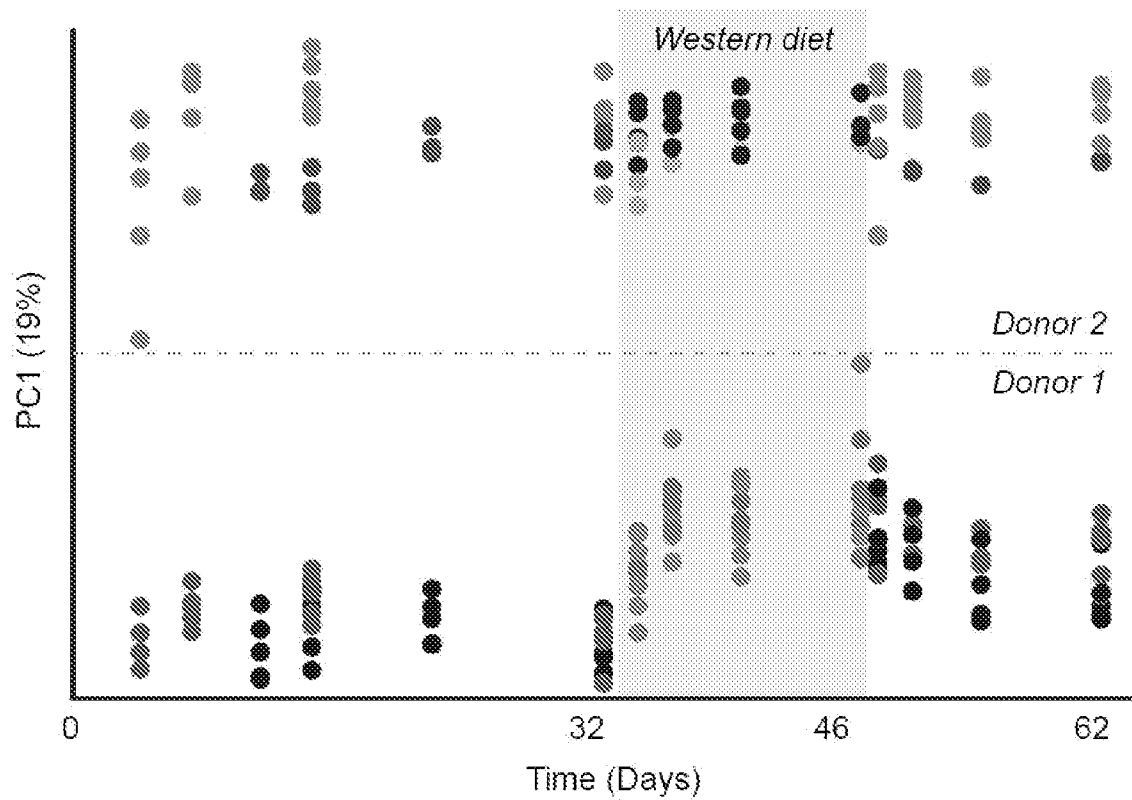
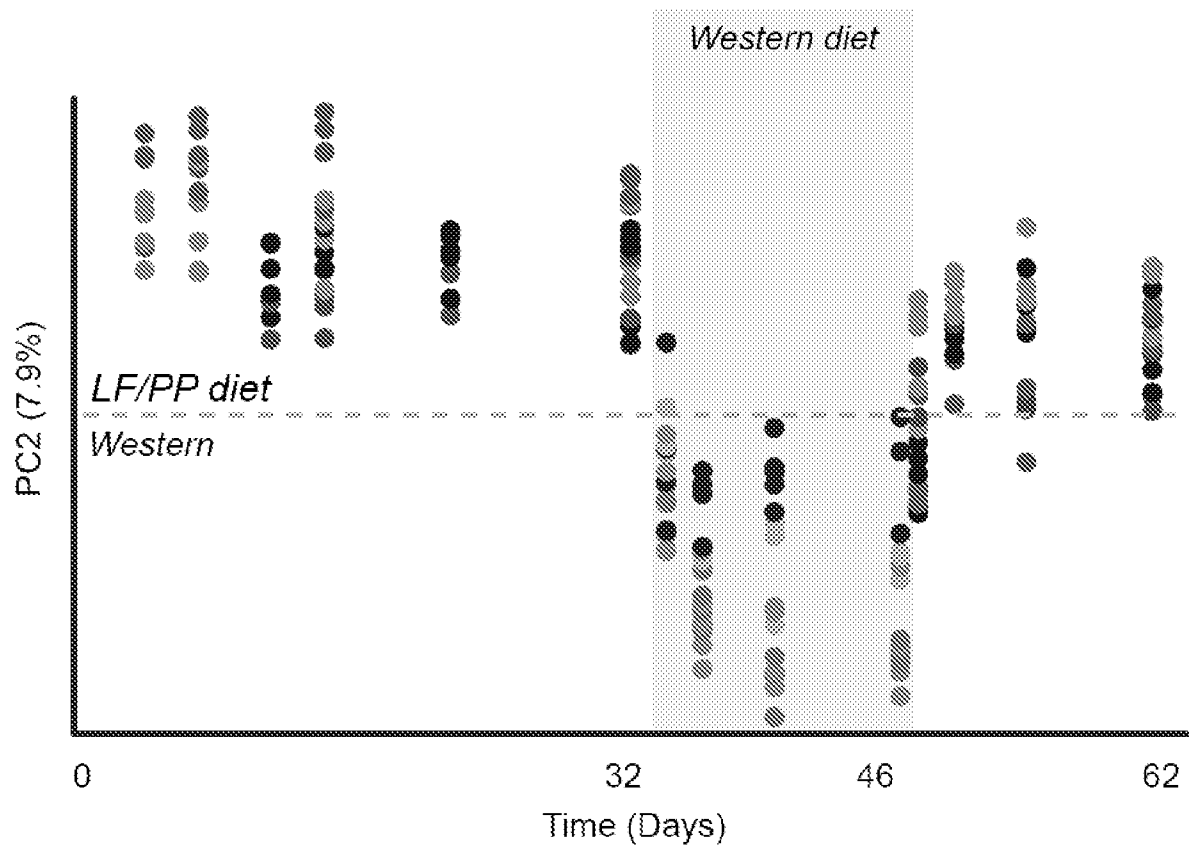


FIG. 4D

A**B****FIG. 5**

**FIG. 5C**

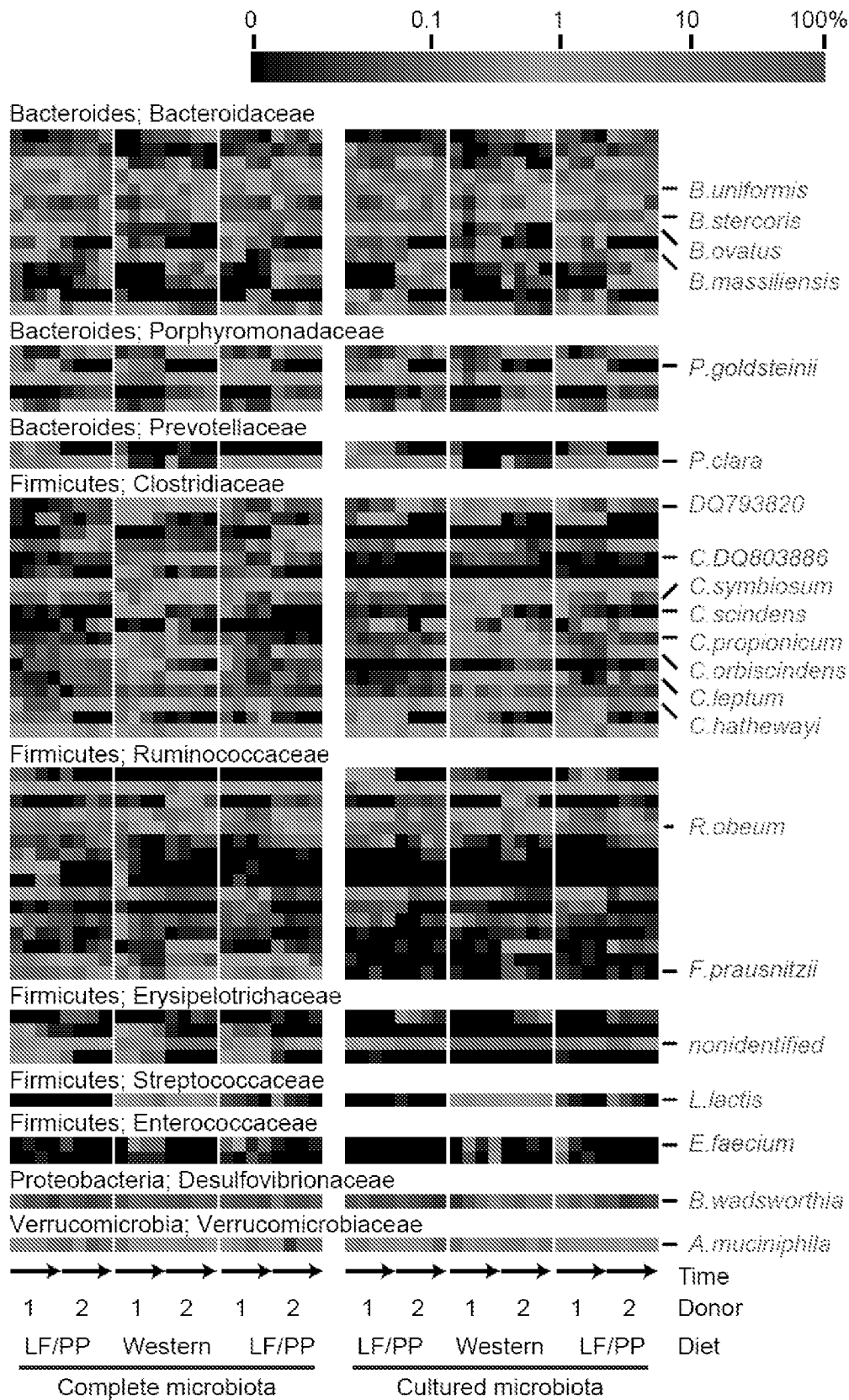
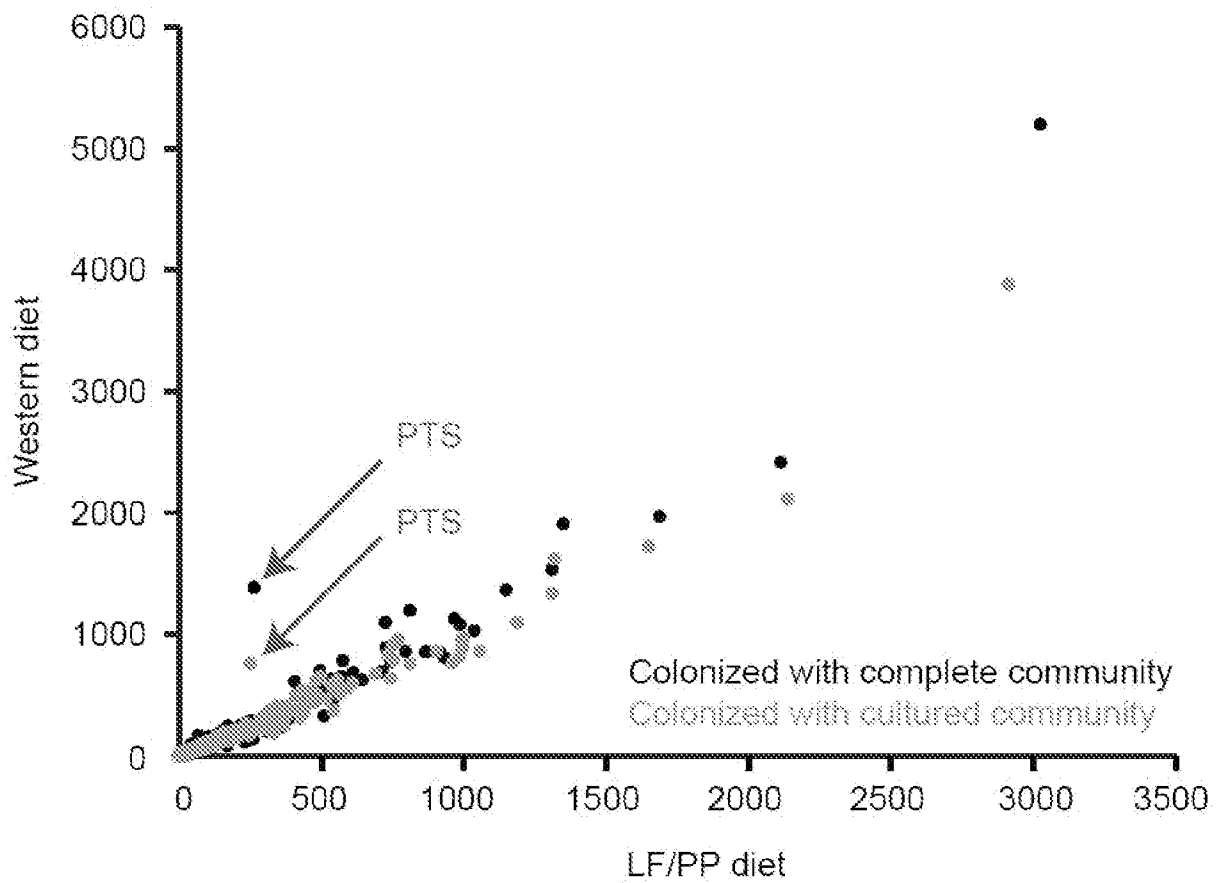
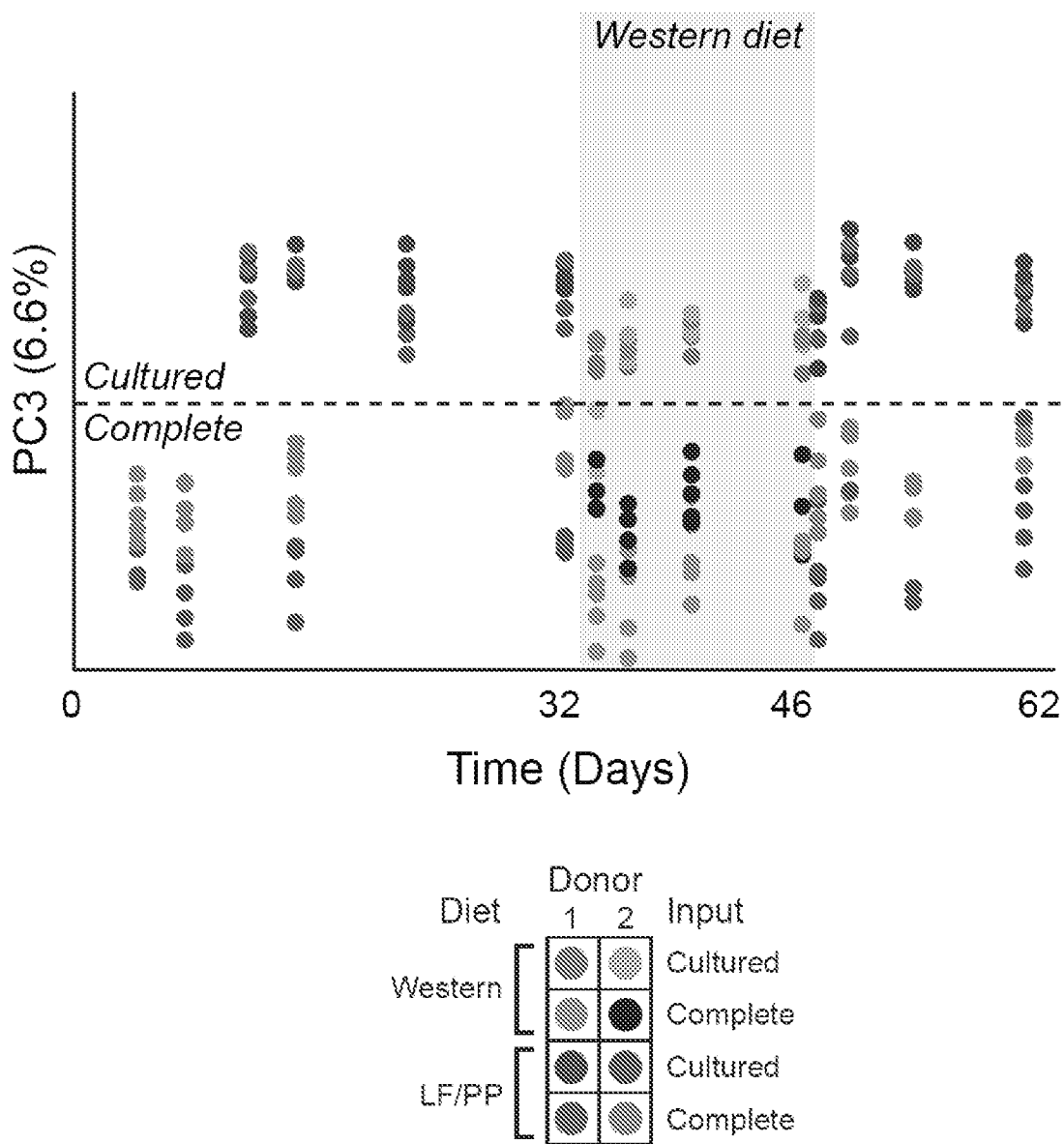


FIG. 5D

**FIG. 5E**

**FIG. 6A**

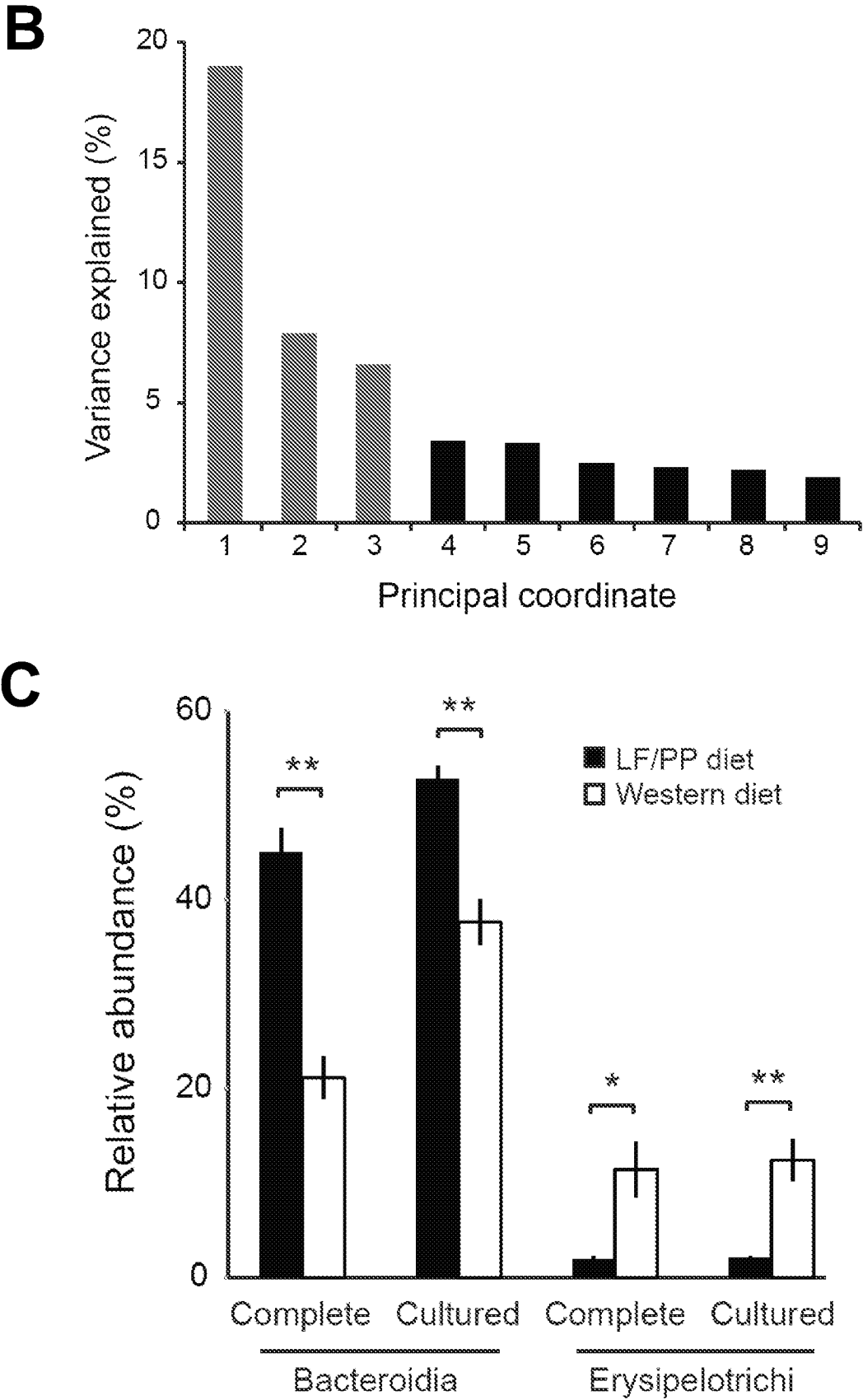
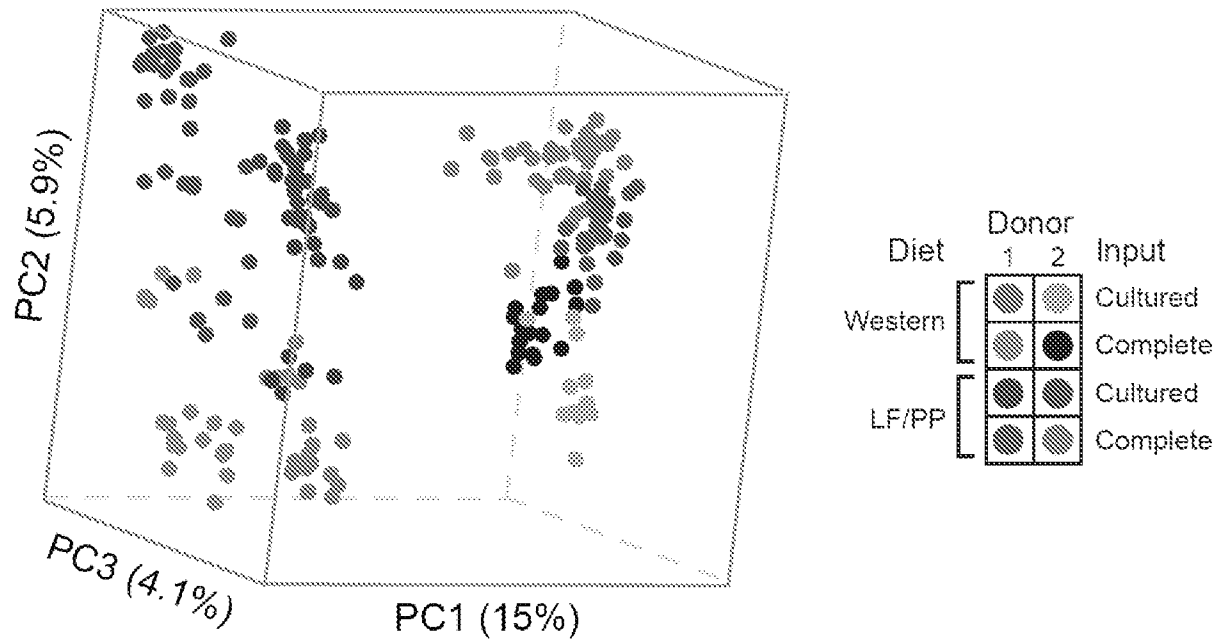
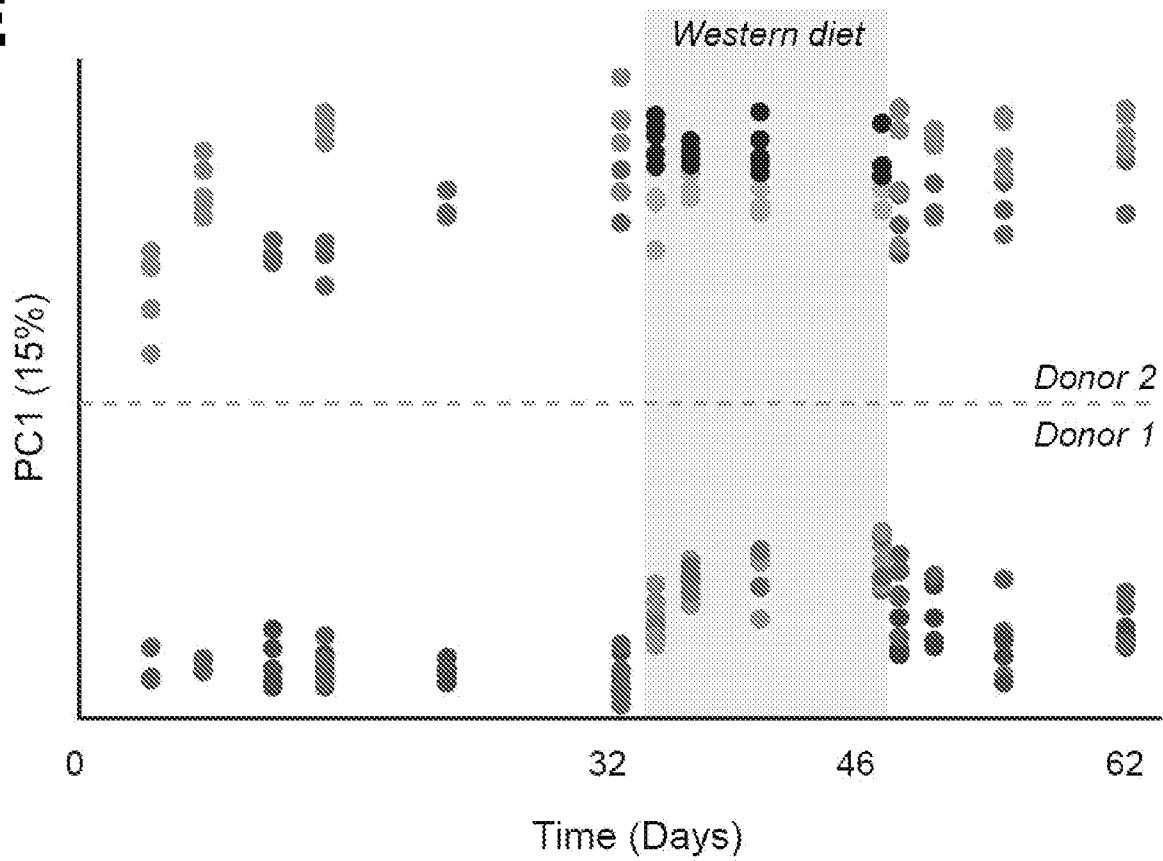
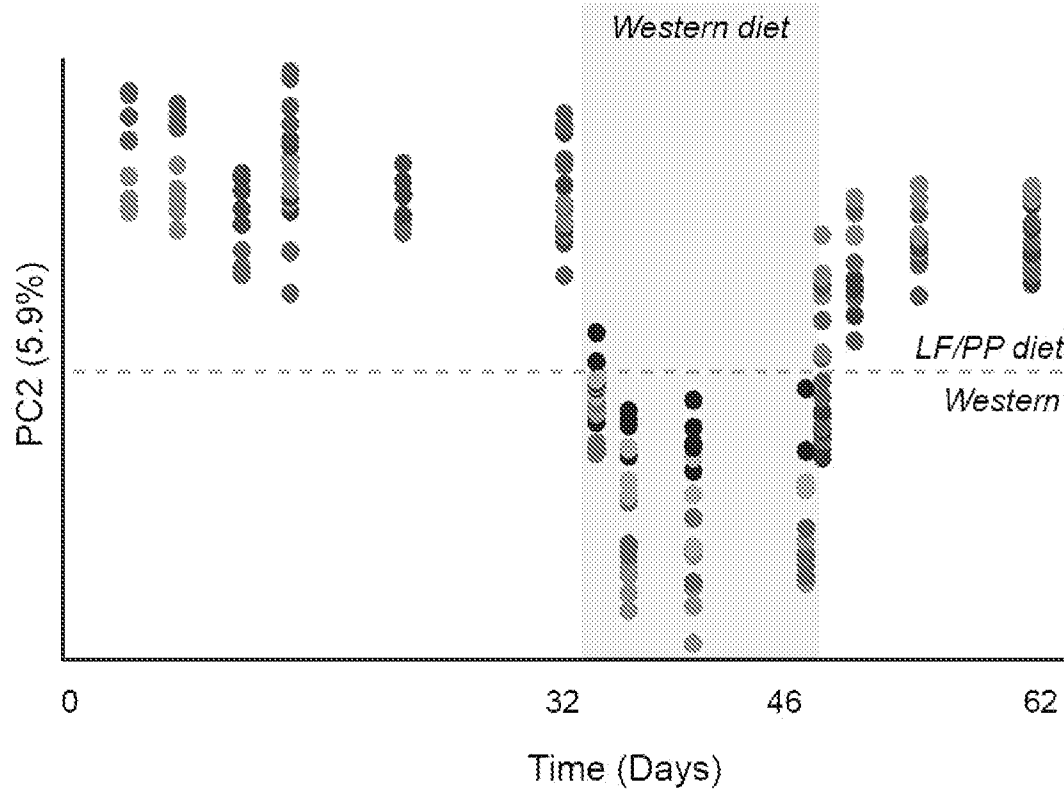


FIG. 6

D**E****FIG. 6**

F



G

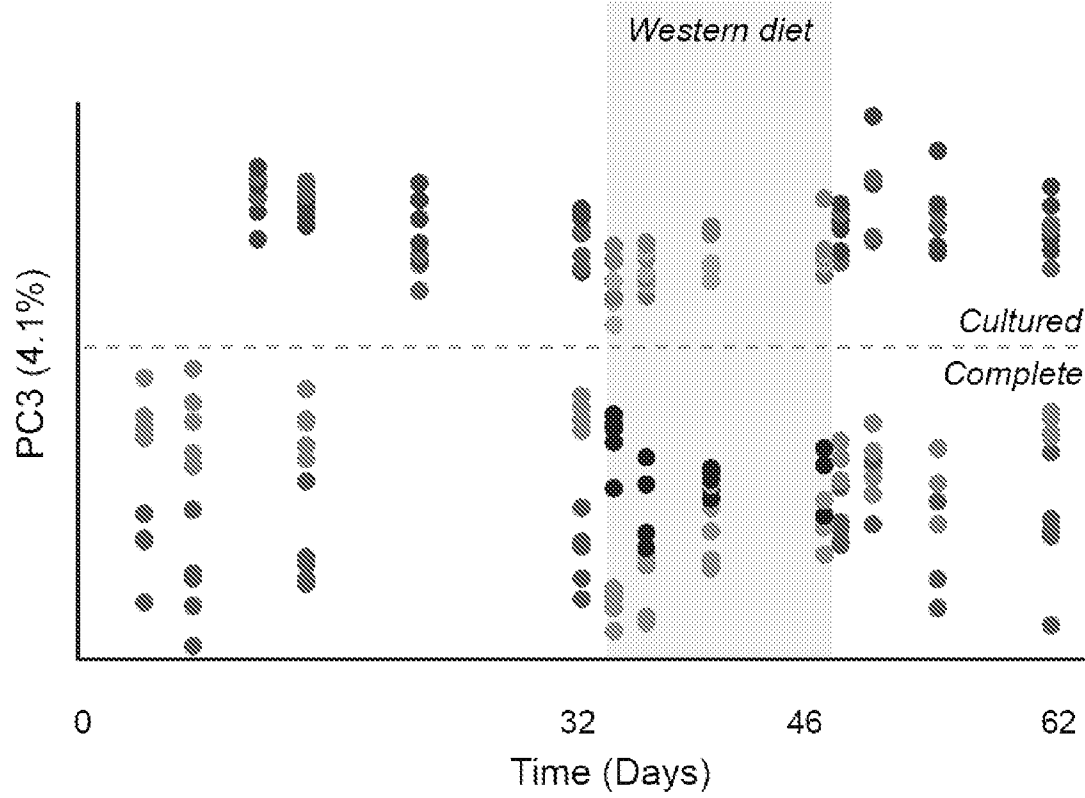
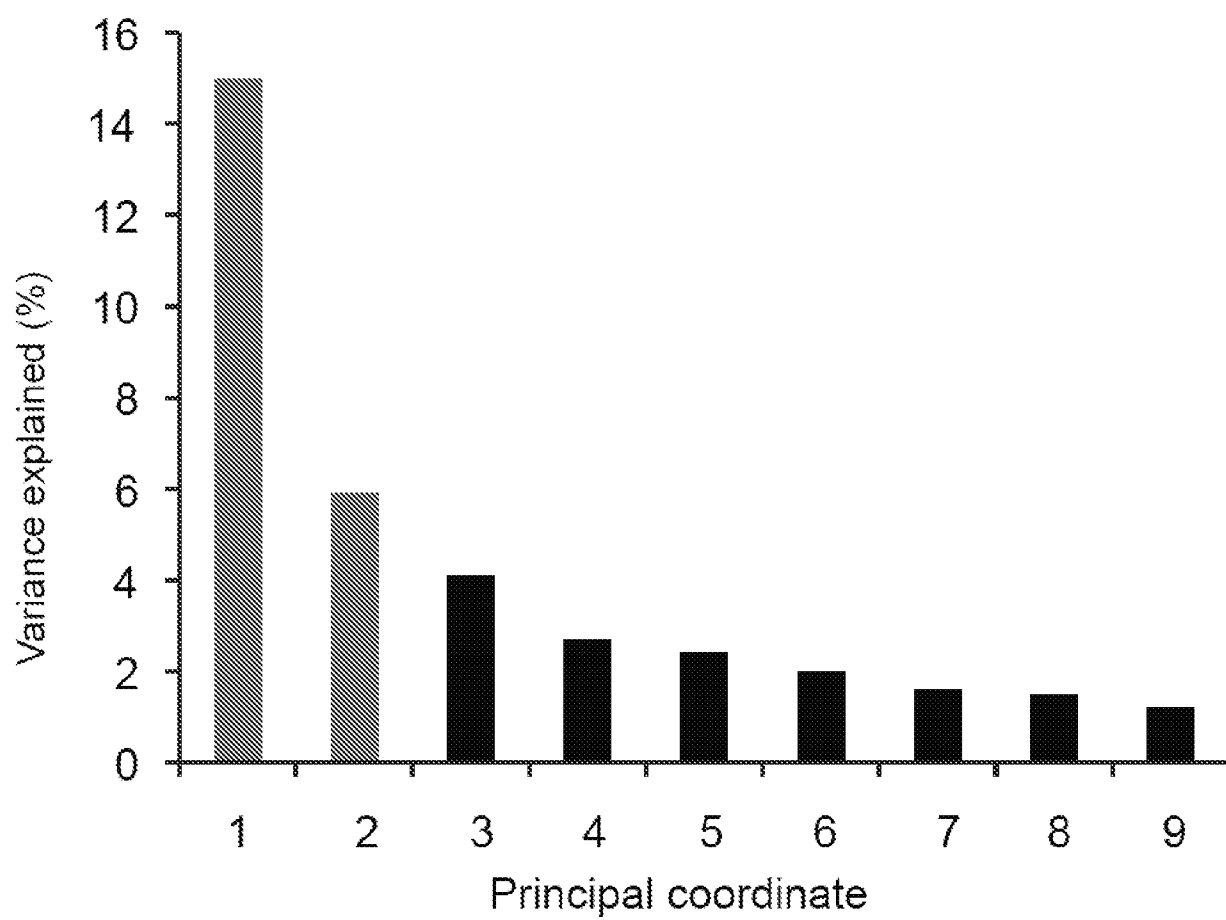


FIG. 6

**FIG. 6H**

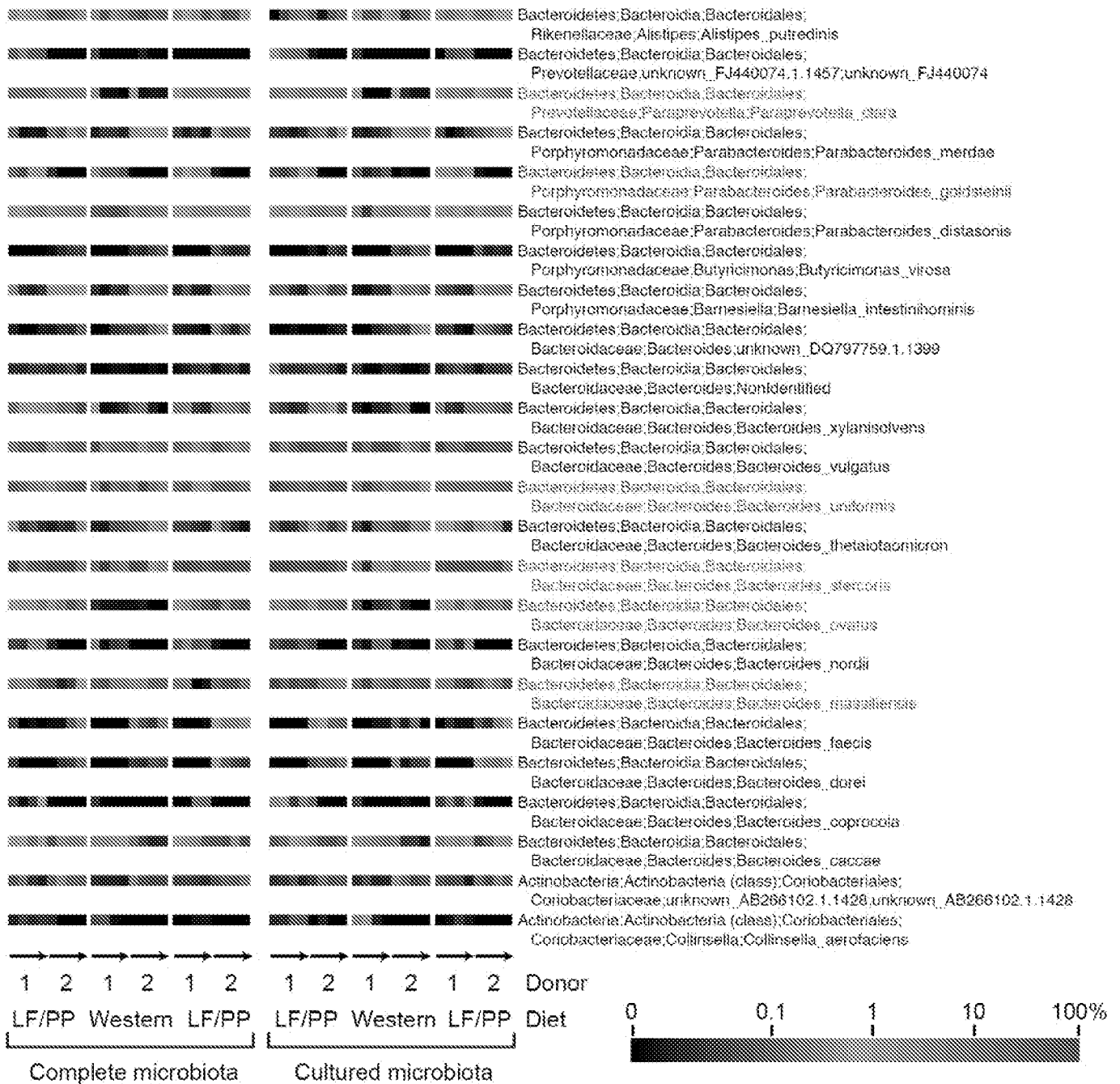


FIG. 7A



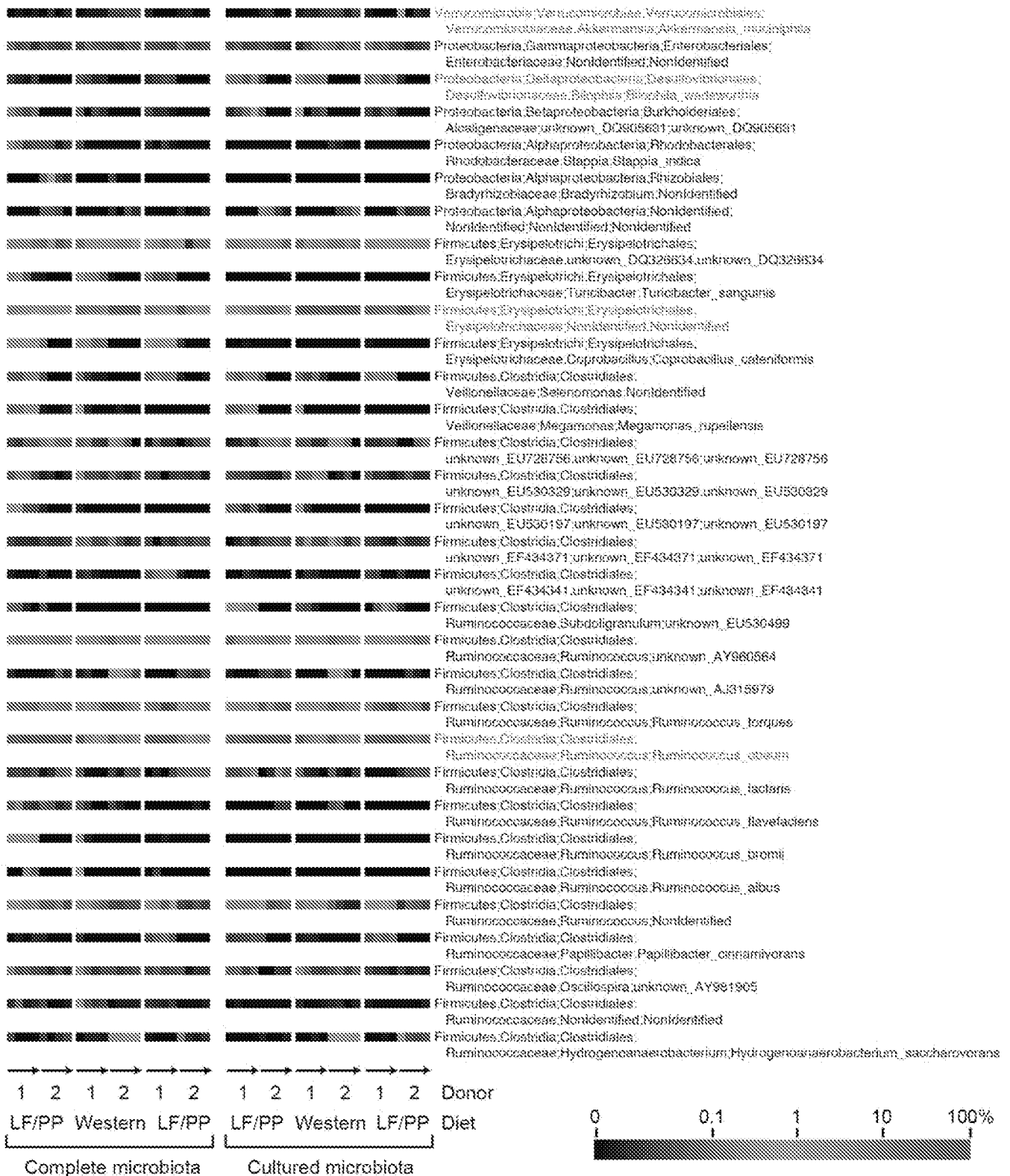
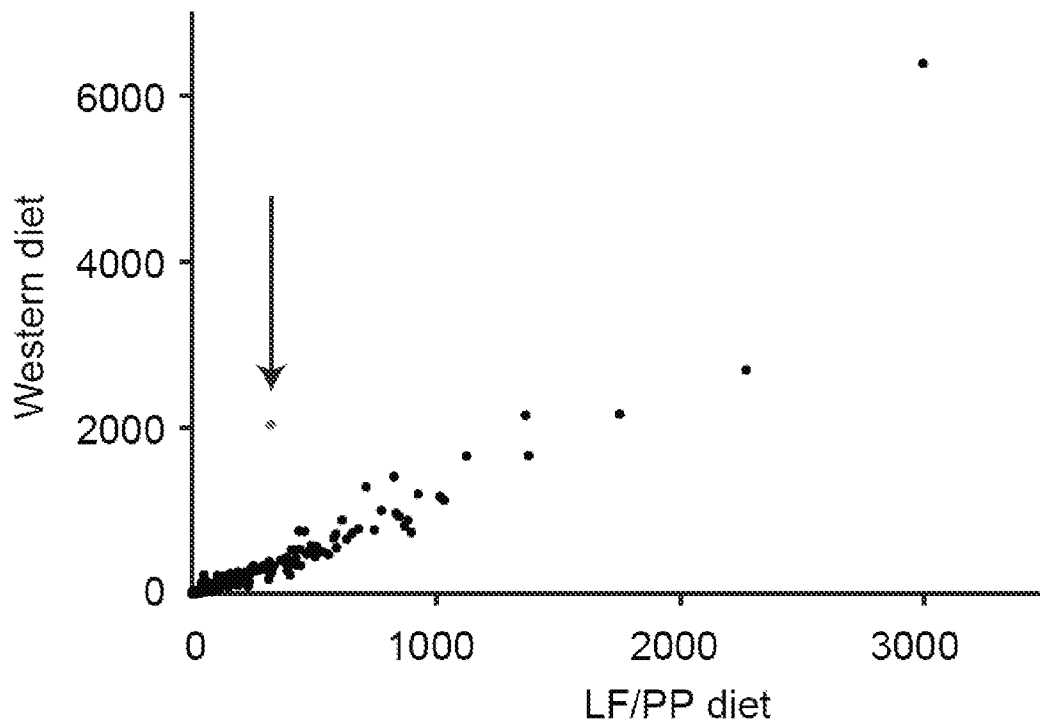


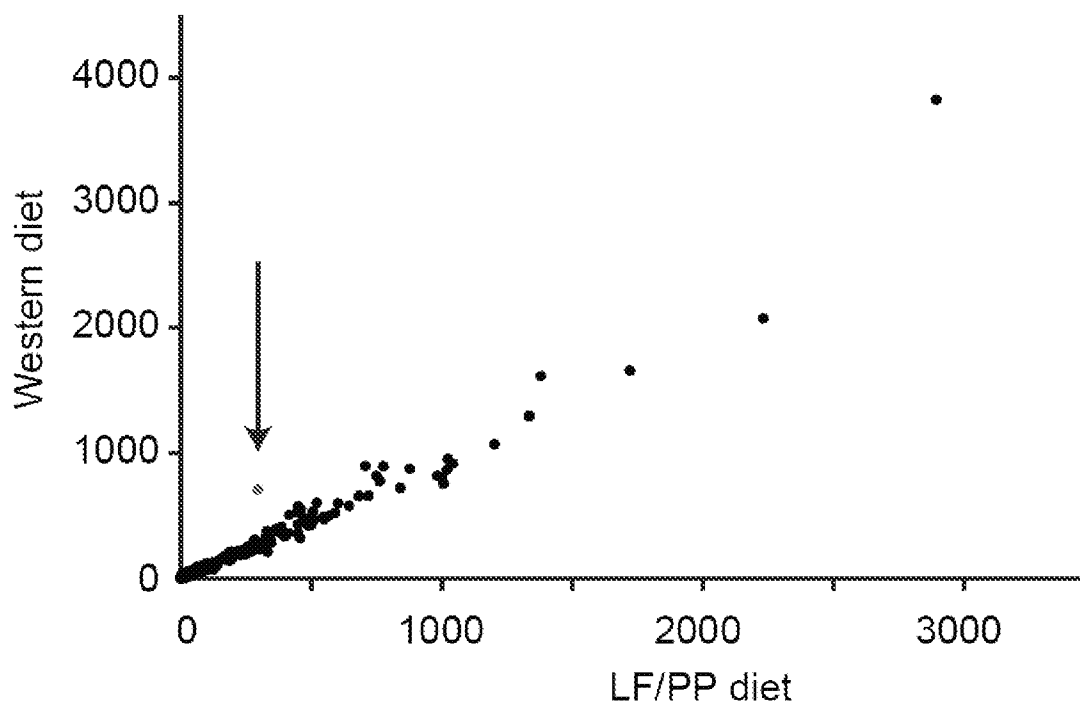
FIG. 7C

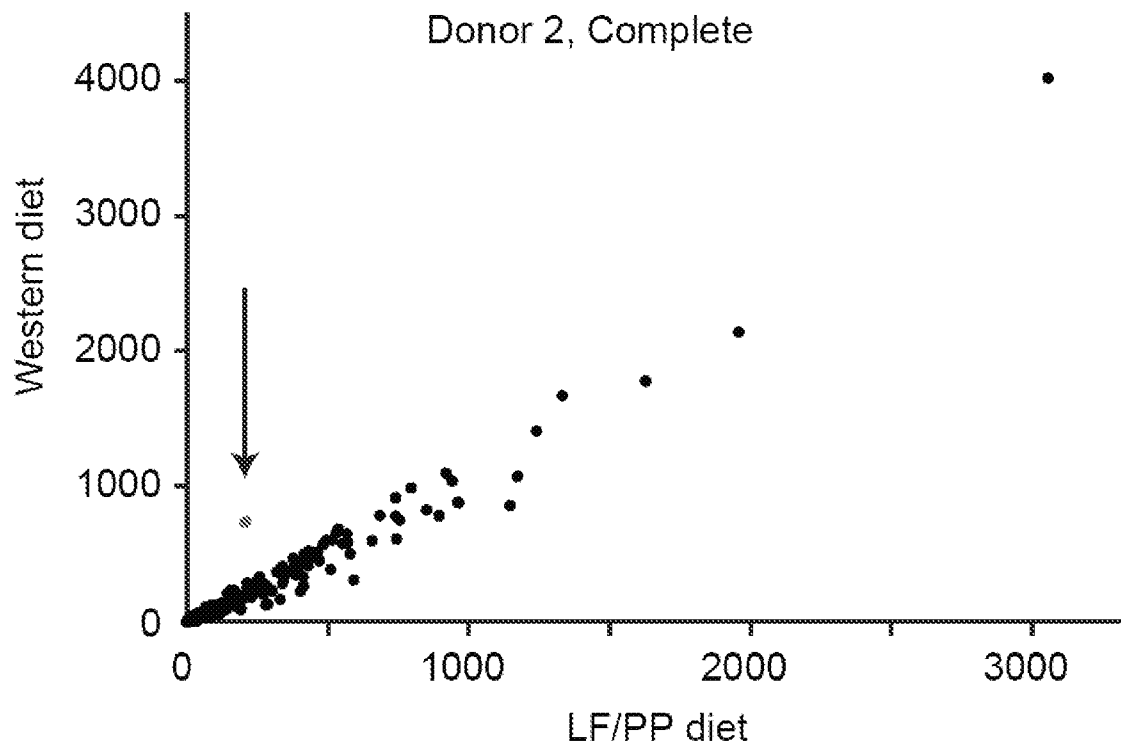
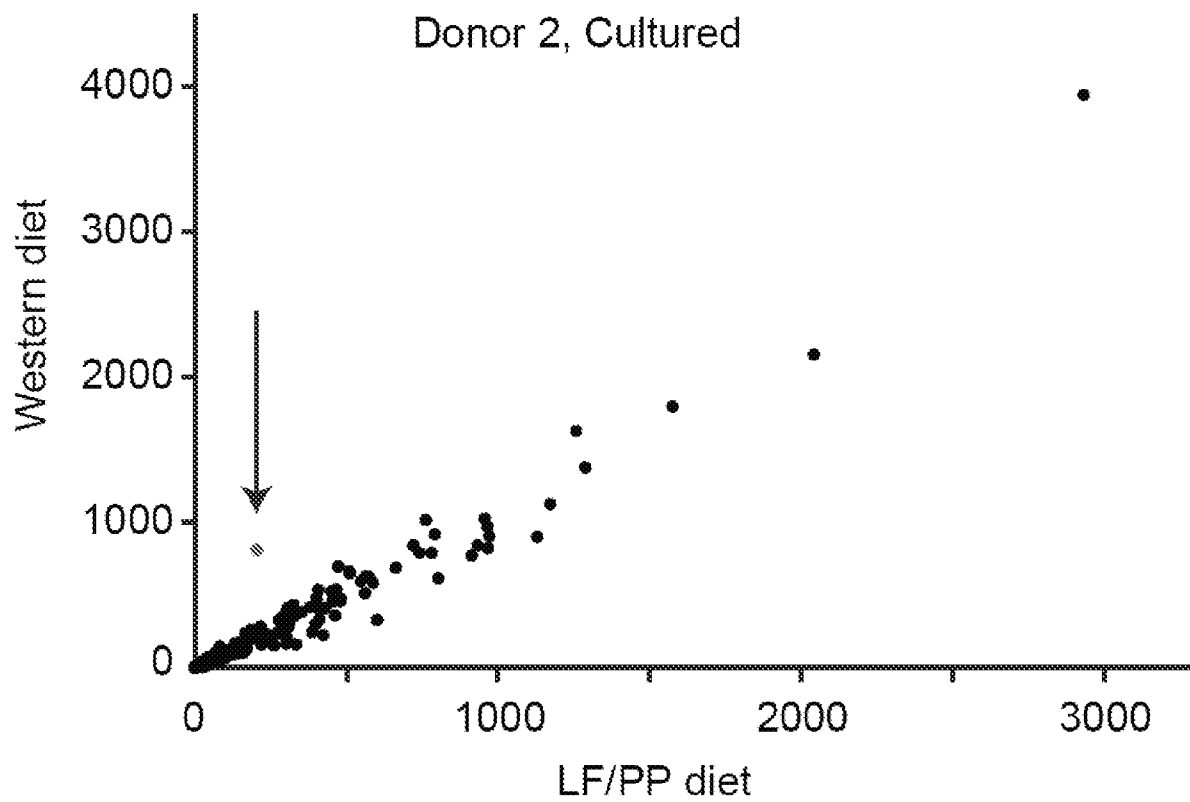
A

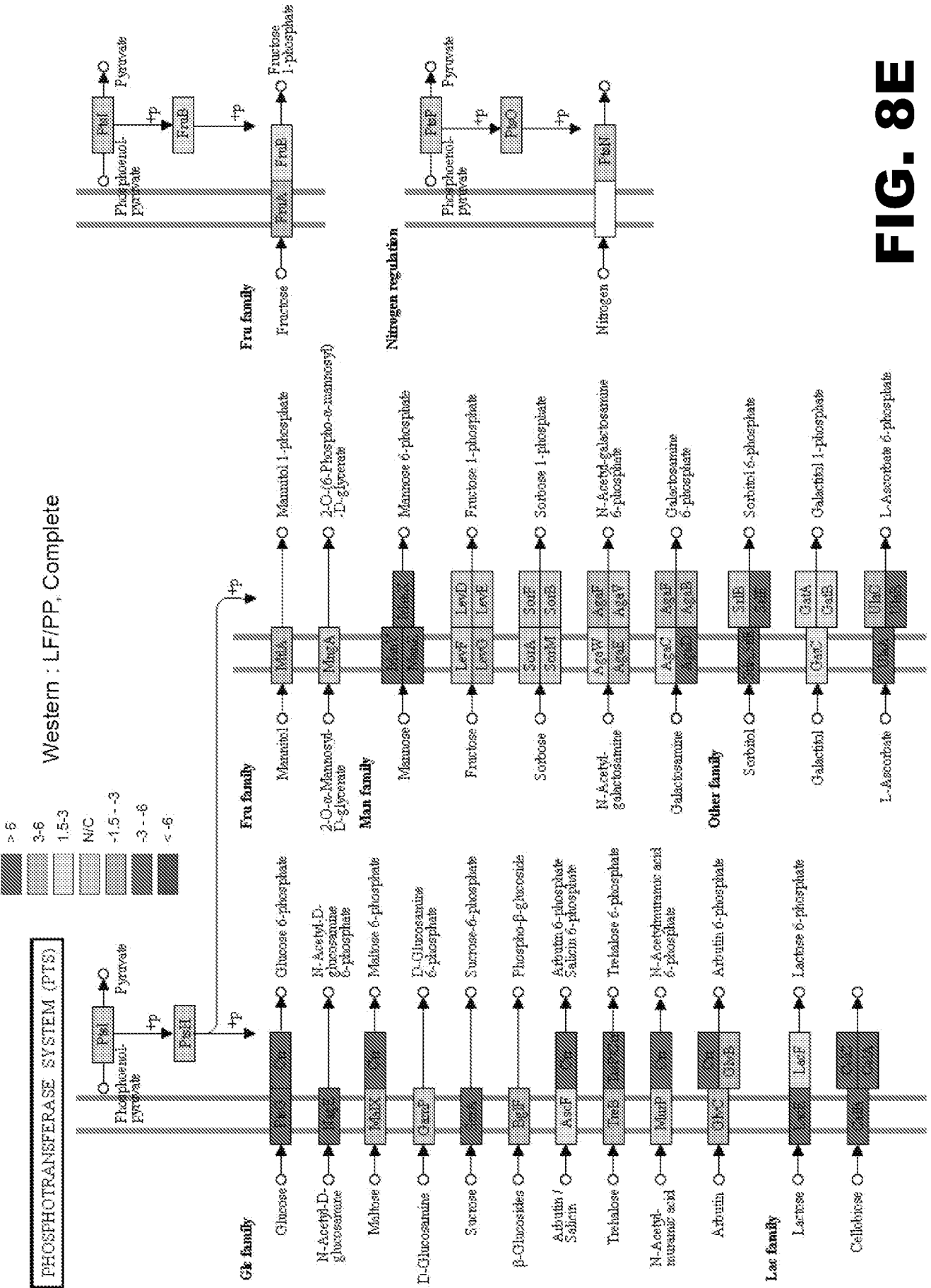
Donor 1, Complete

**B**

Donor 1, Cultured

**FIG. 8**

C**D****FIG. 8**



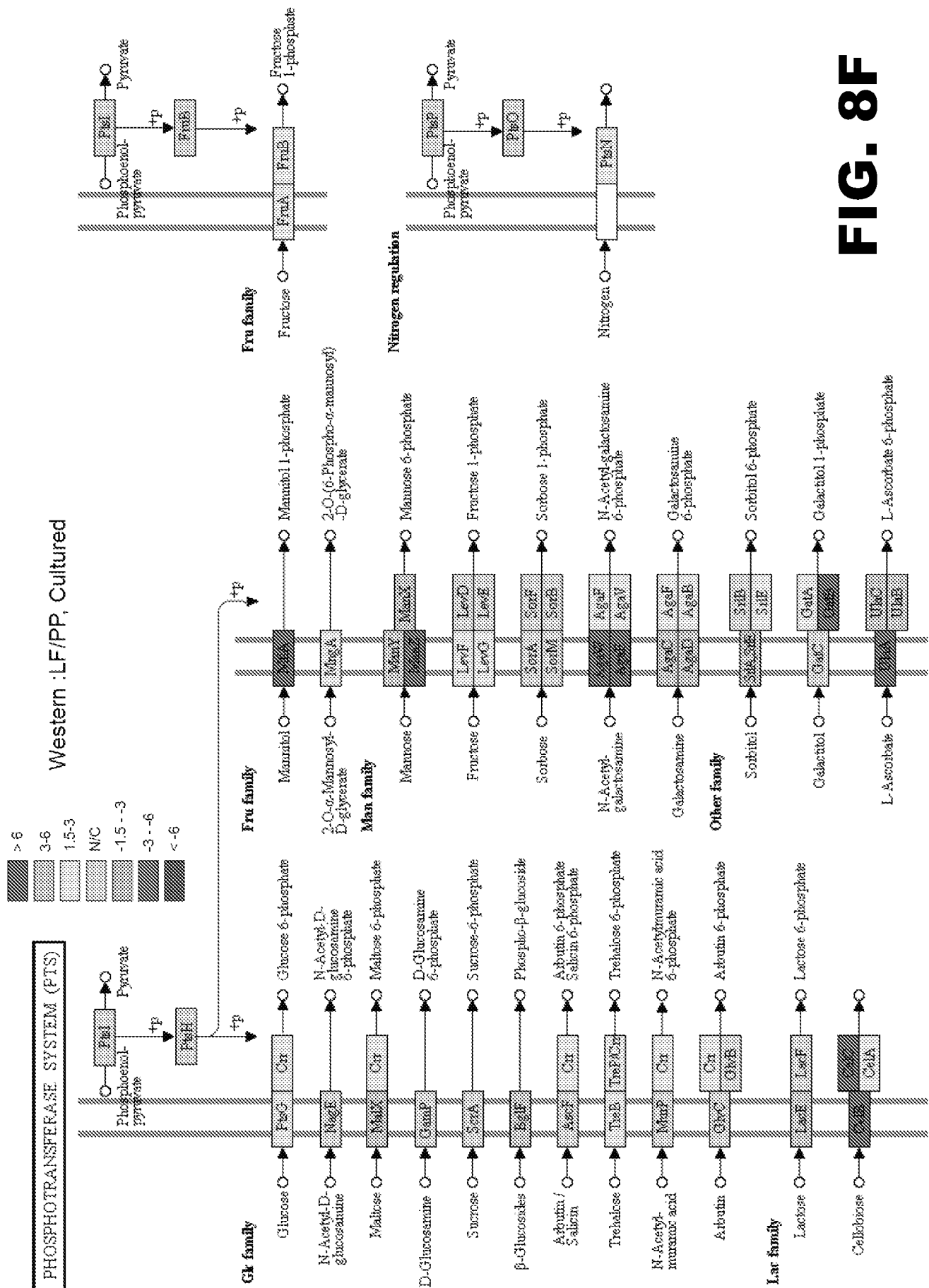
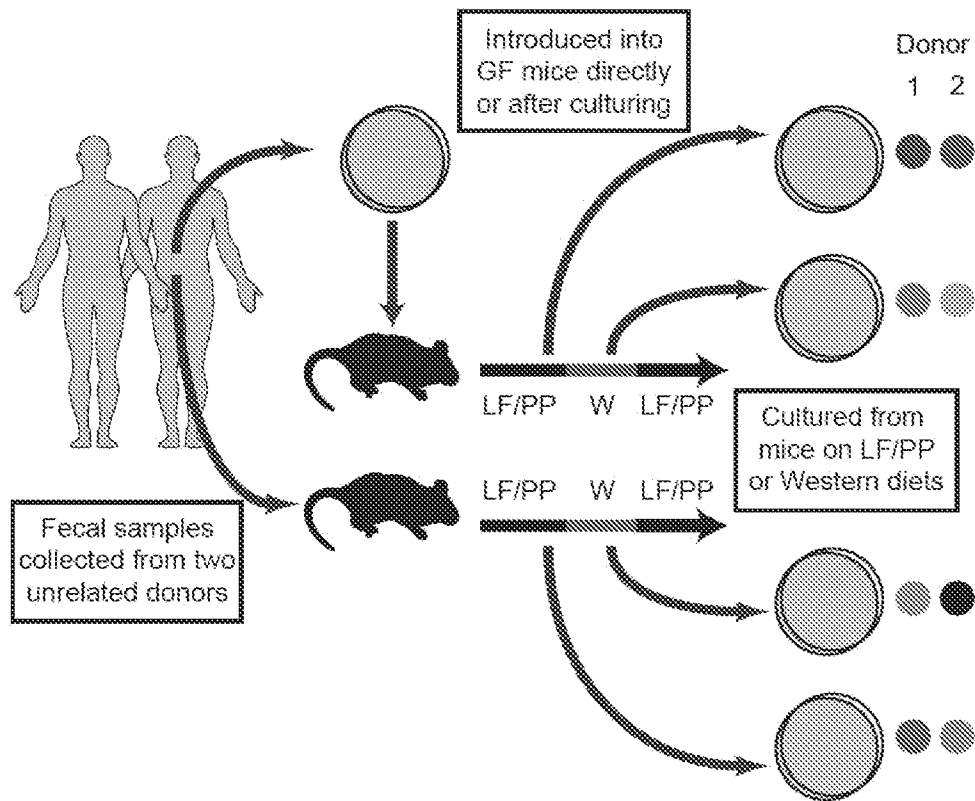
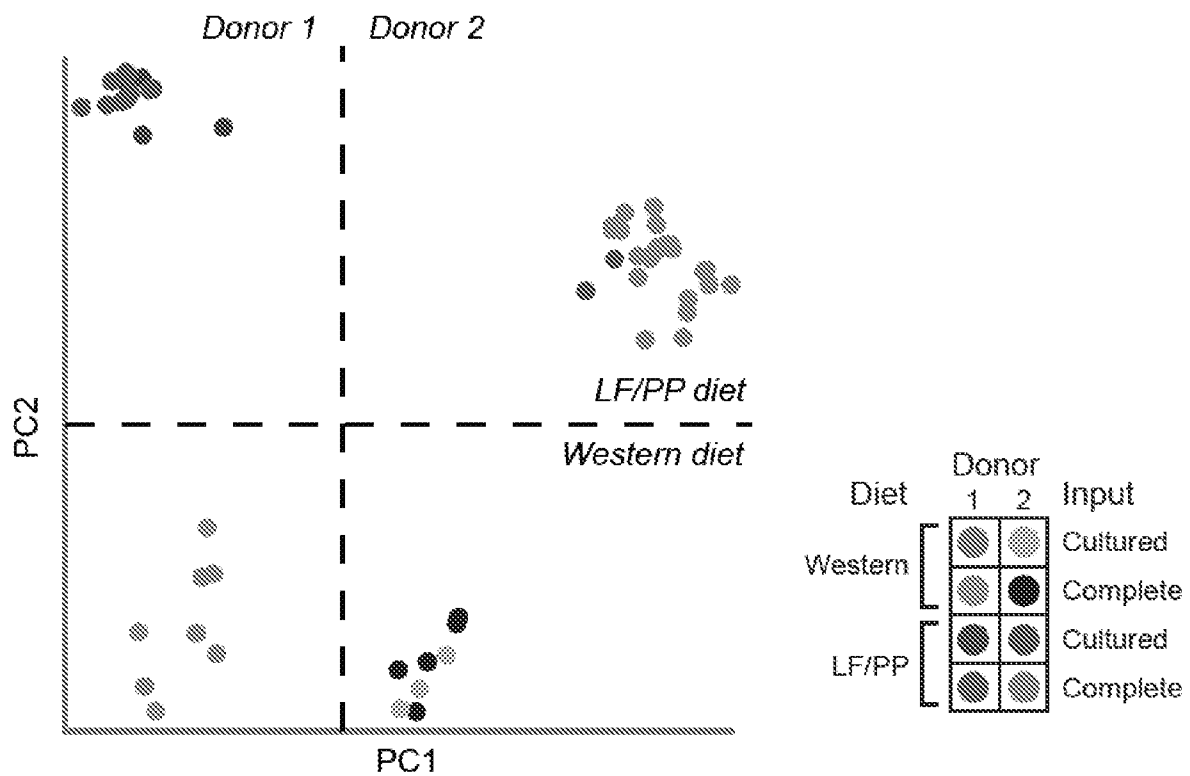
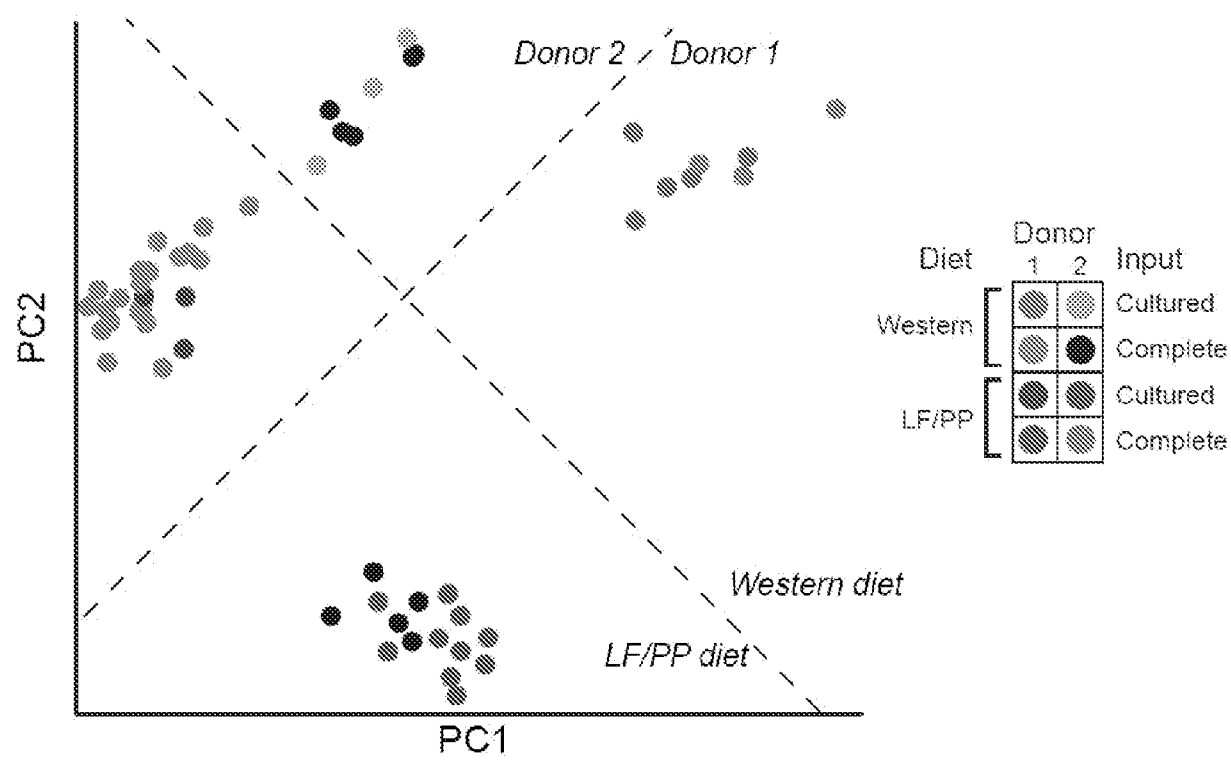


FIG. 8F

A**B****FIG. 9**

A



B

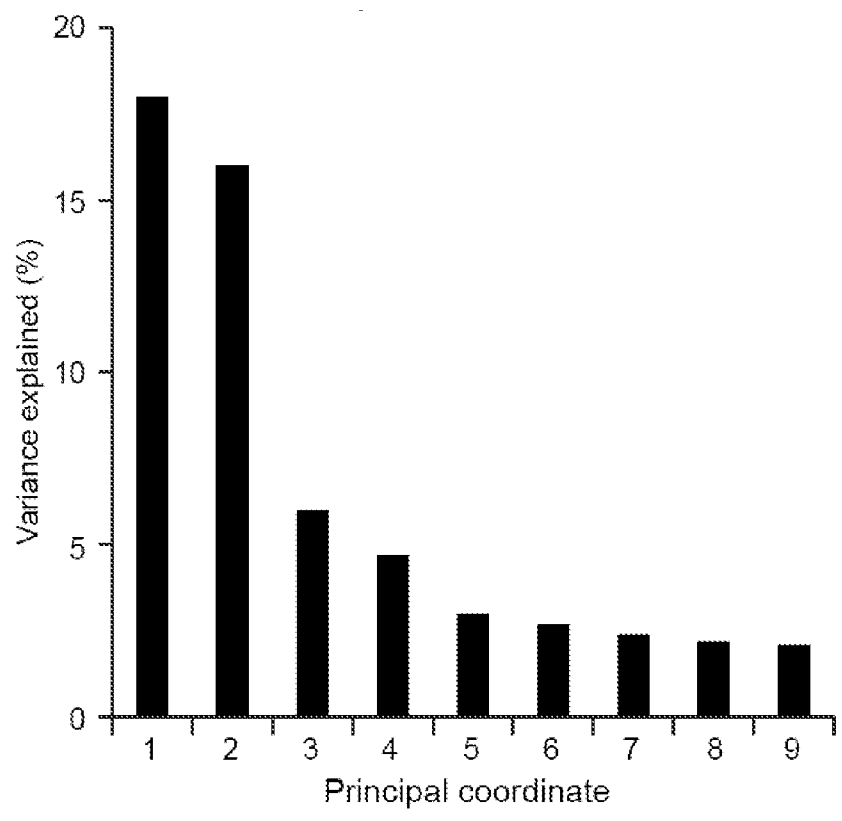
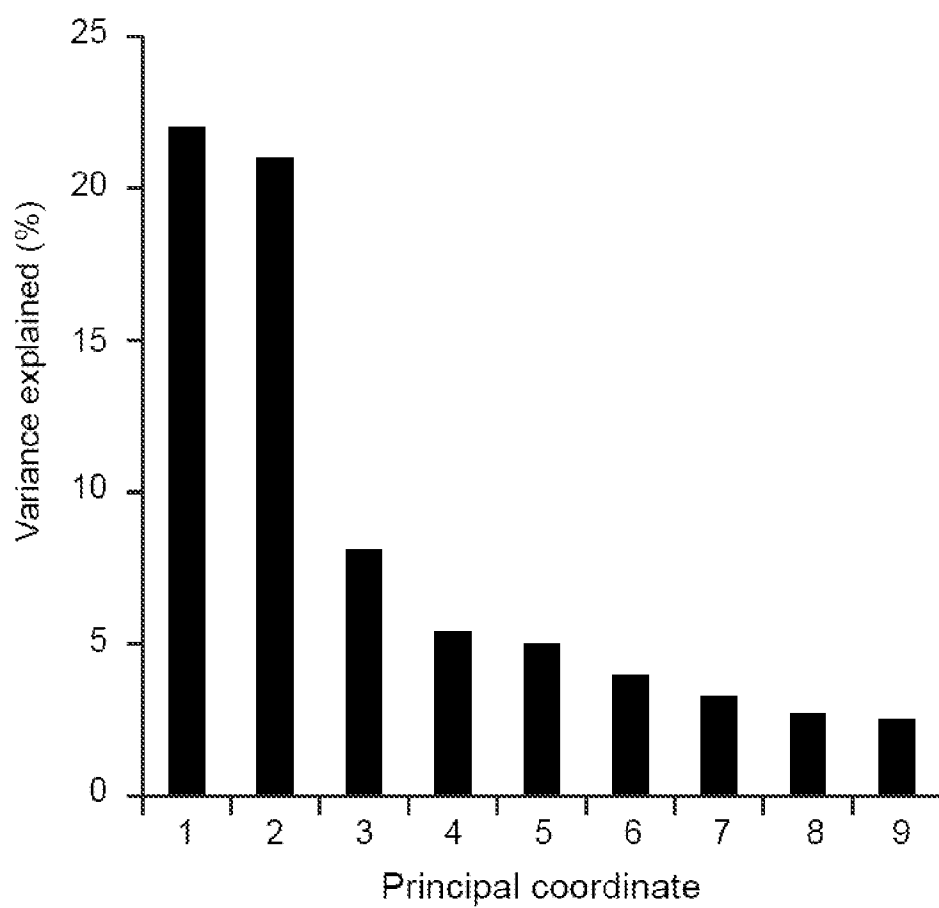
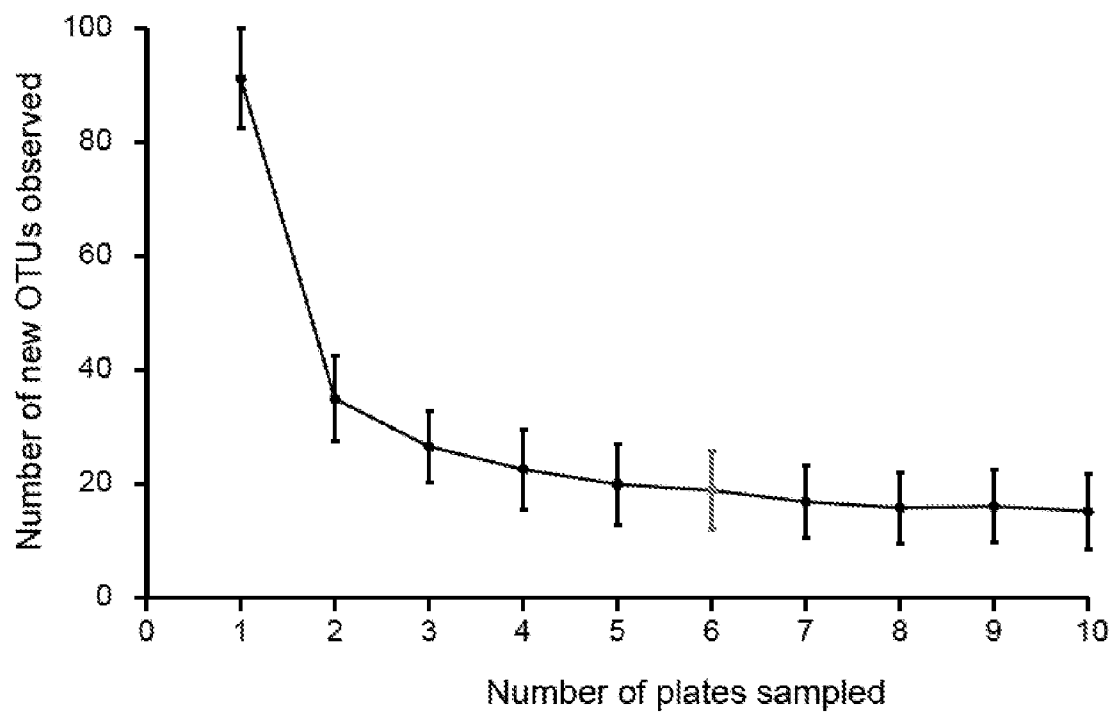
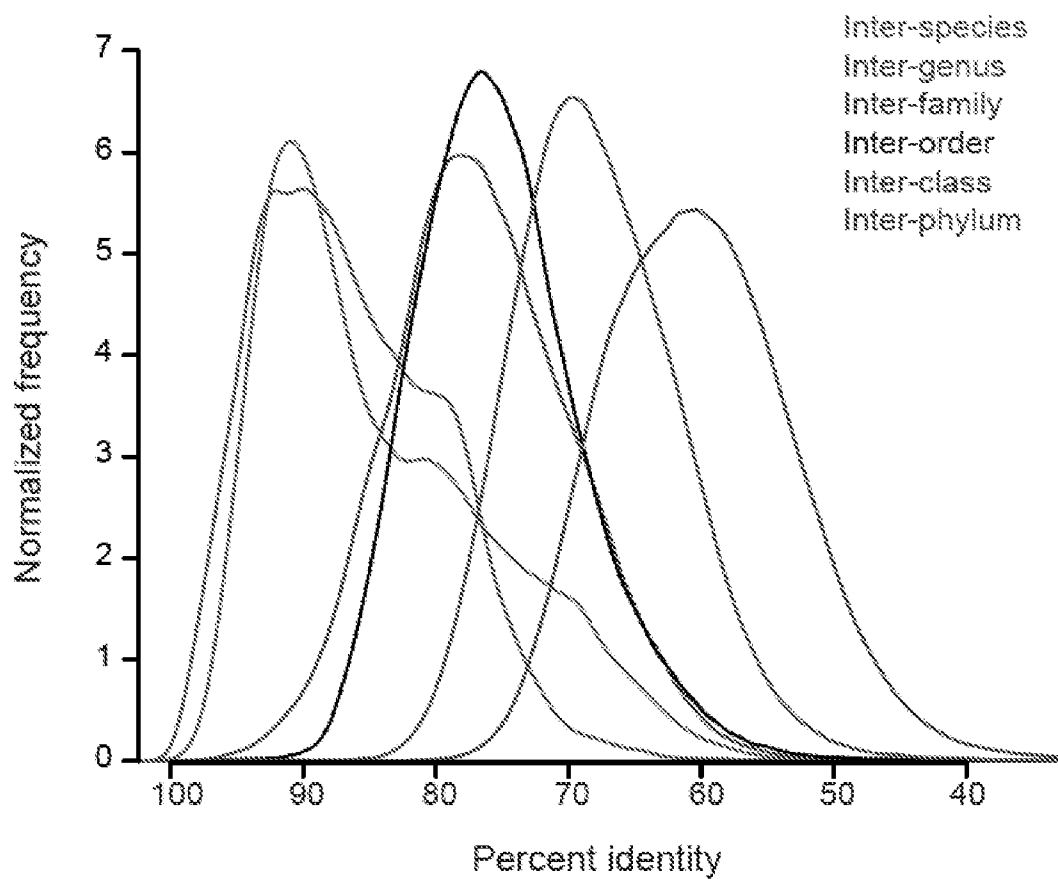
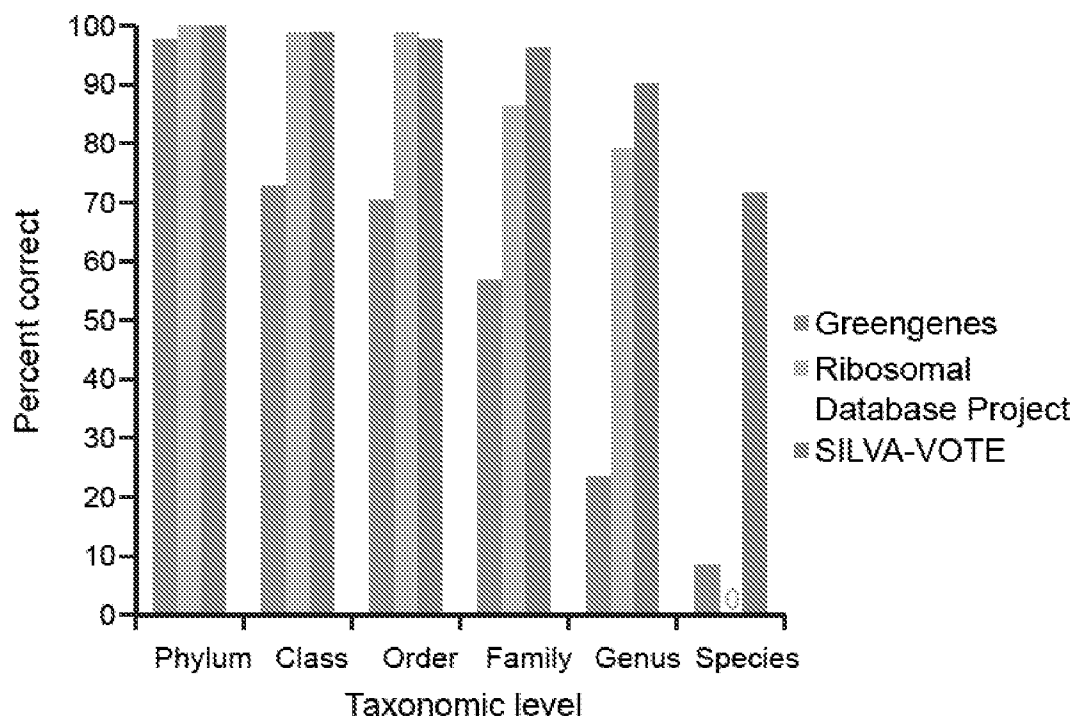
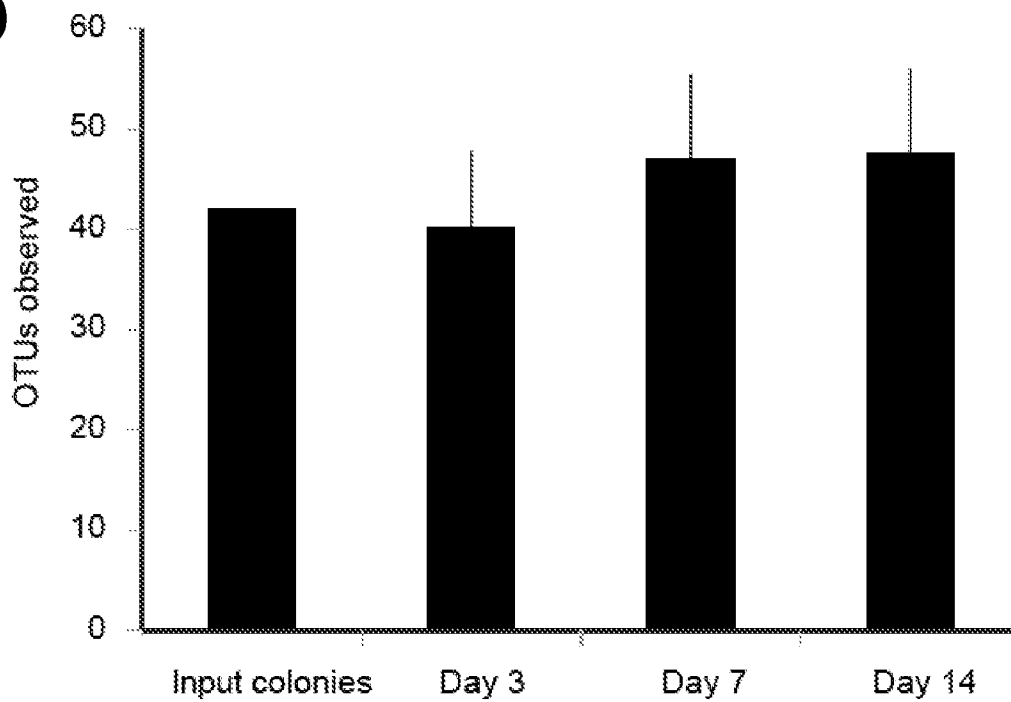
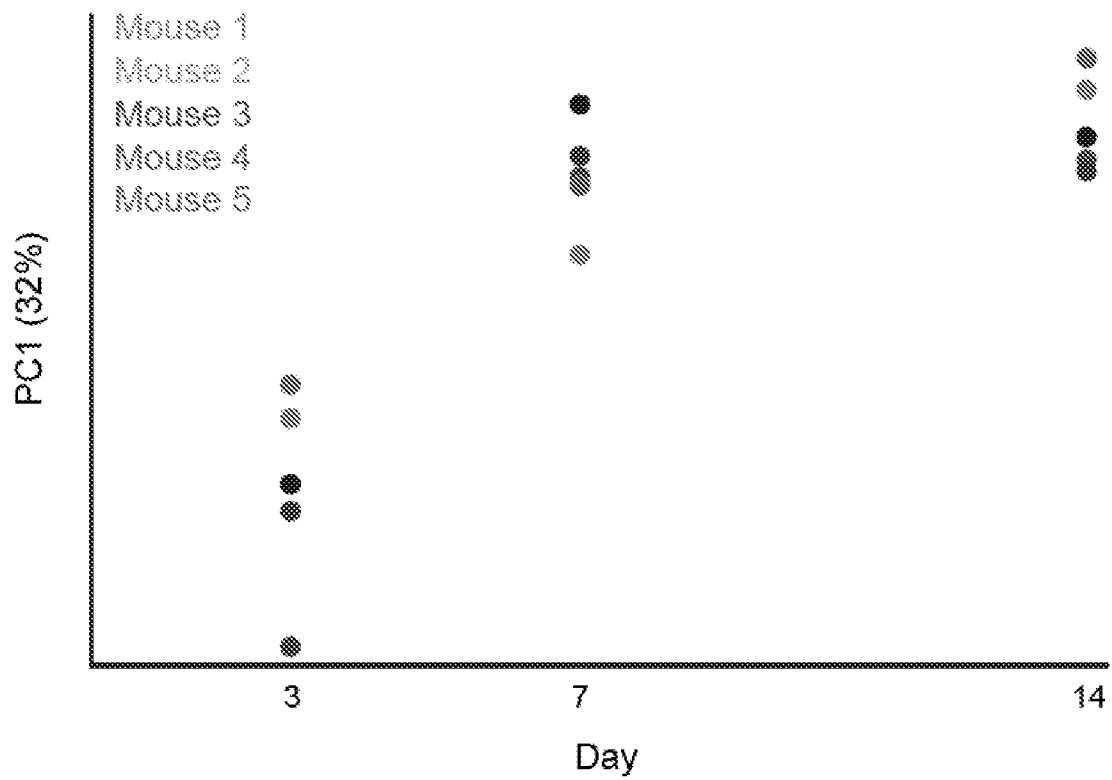
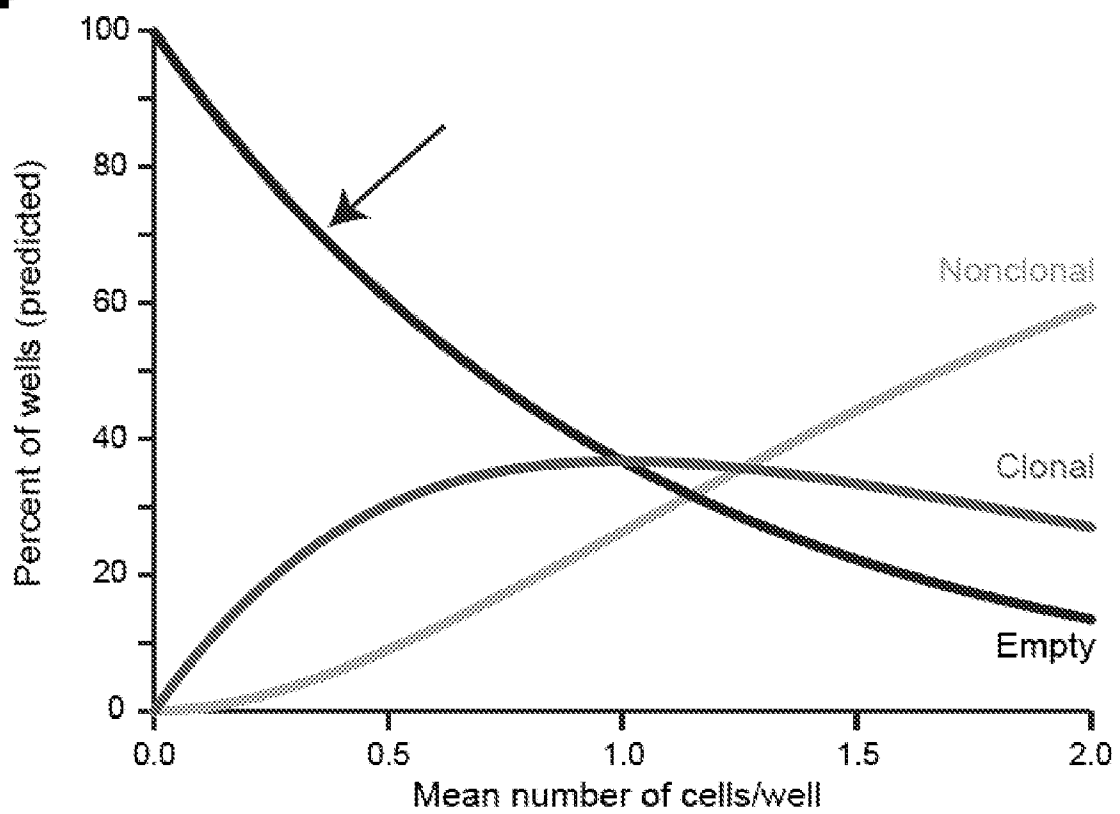


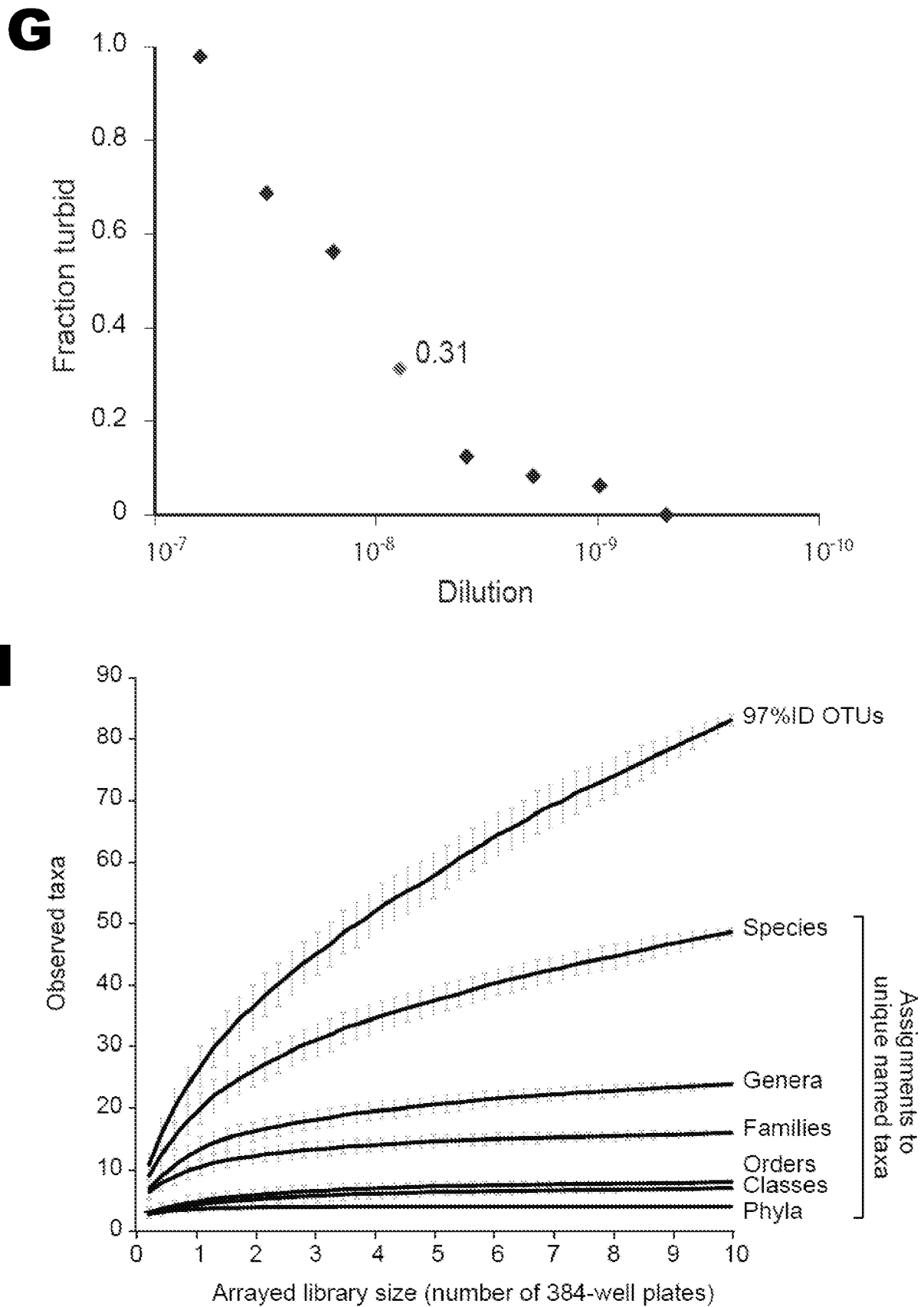
FIG. 10

**FIG. 10C**

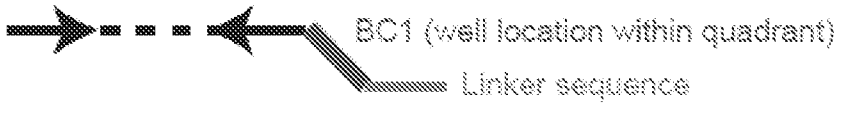
A**B****FIG. 11**

C**D****FIG. 11**

E**F****FIG. 11**

**FIG. 11**

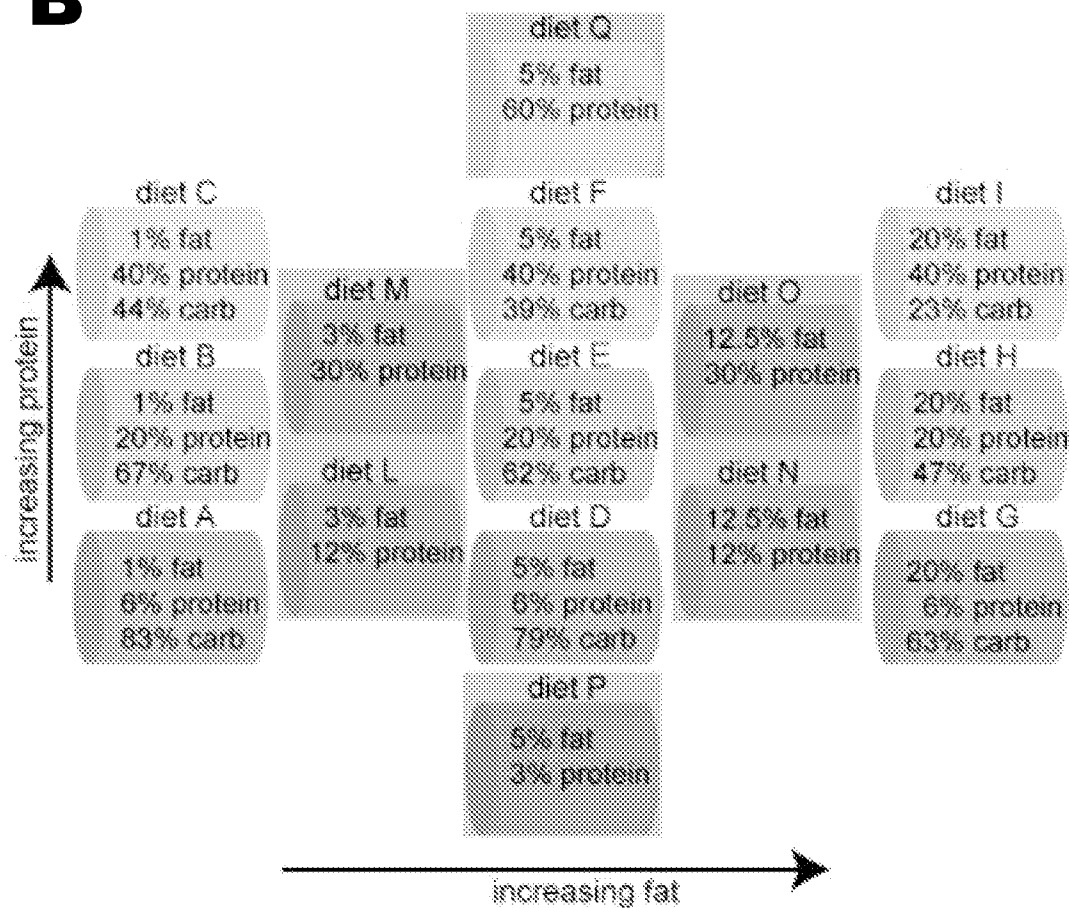
A

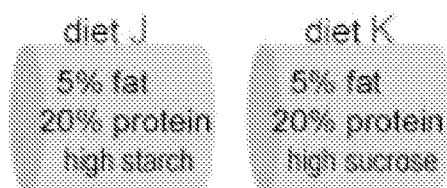
1. Anaerobic limiting dilution, 384-well culturing, quantification, and lysis
- 16S rRNA gene 
2. PCR #1 
 3. Pool all samples within each quadrant
 4. PCR #2 
 5. All samples uniquely barcoded; combine into a single tube and sequence

B**FIG. 12**

A

B. thetaiotaomicron
B. ovatus
B. caccae
E. coli
D. piger
C. aerofaciens
C. symbiosum
B. hydrogenotrophica
E. rectale
M. formatexigens

B**FIG. 13**

**FIG. 13C**

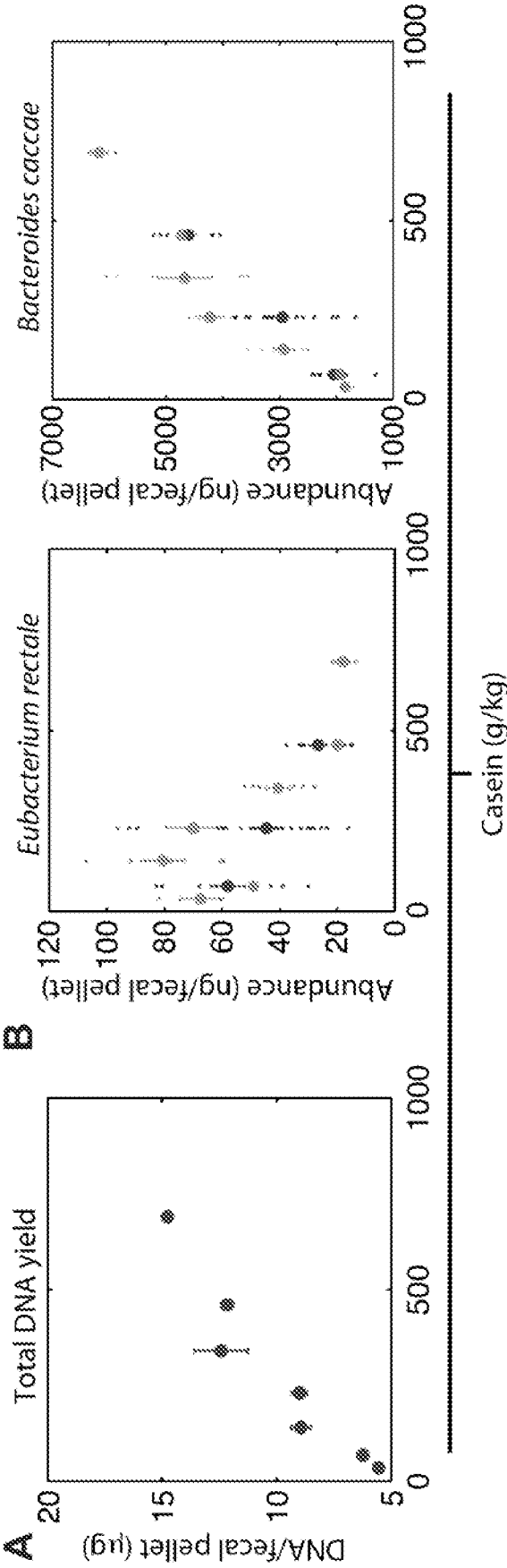
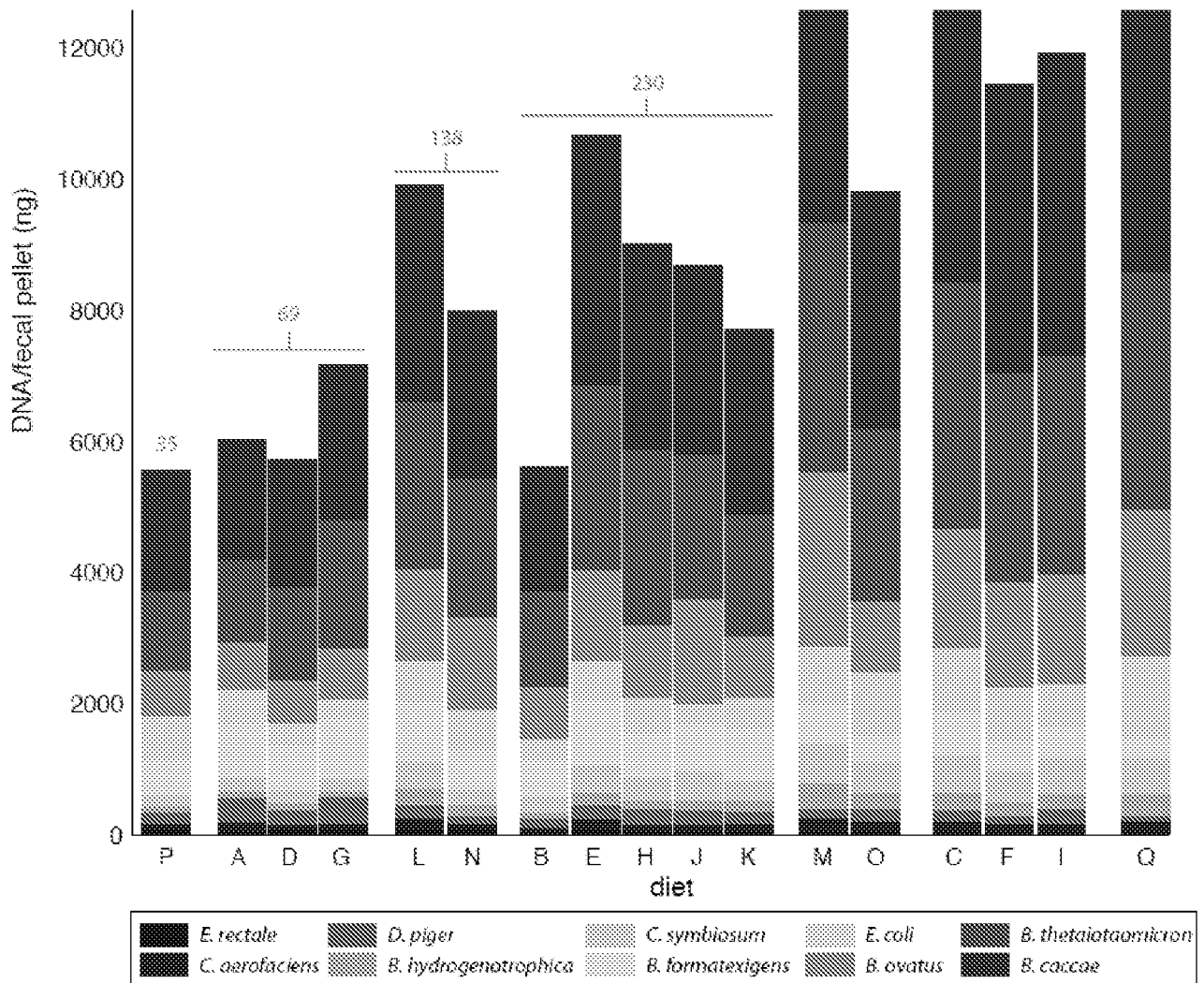
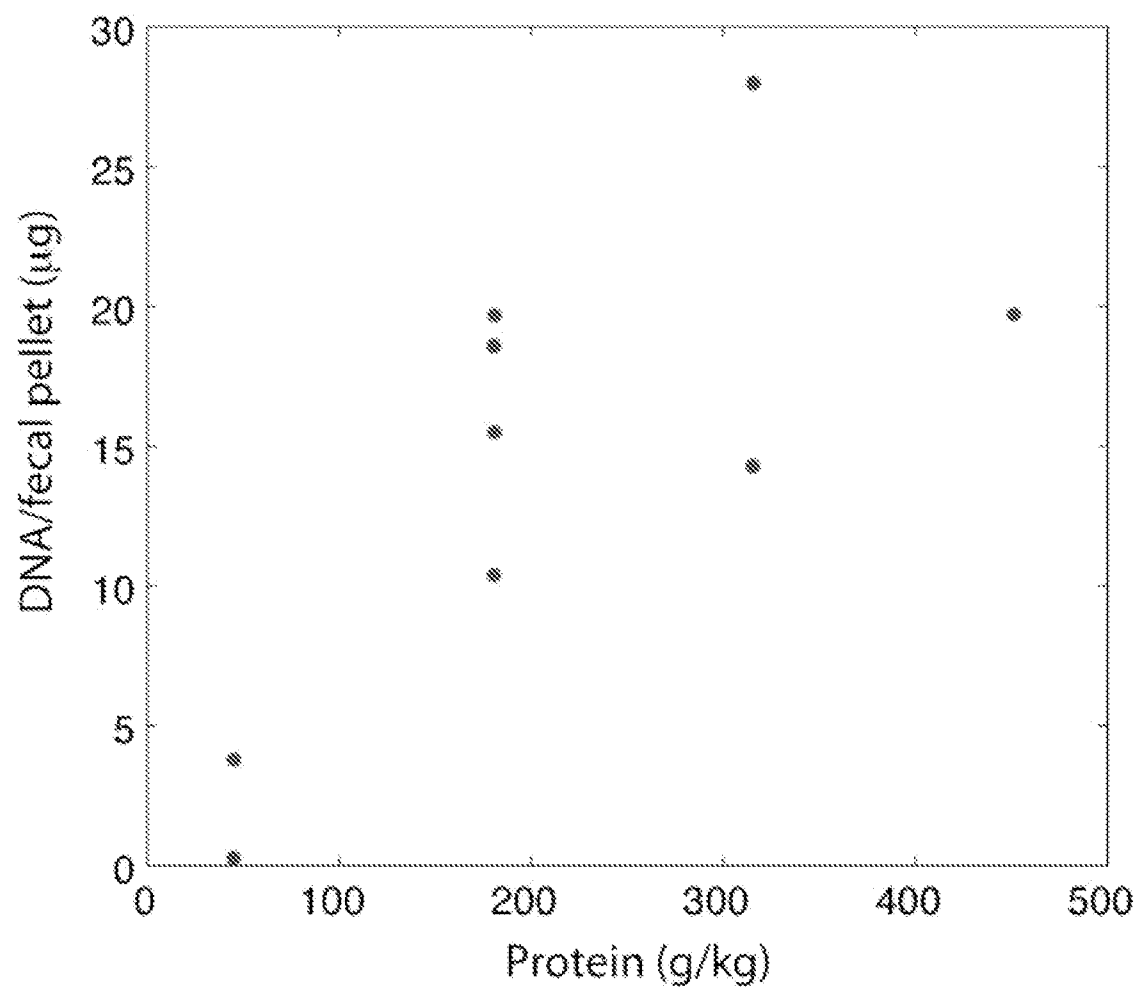


FIG. 14

**FIG. 15**

**FIG. 16**

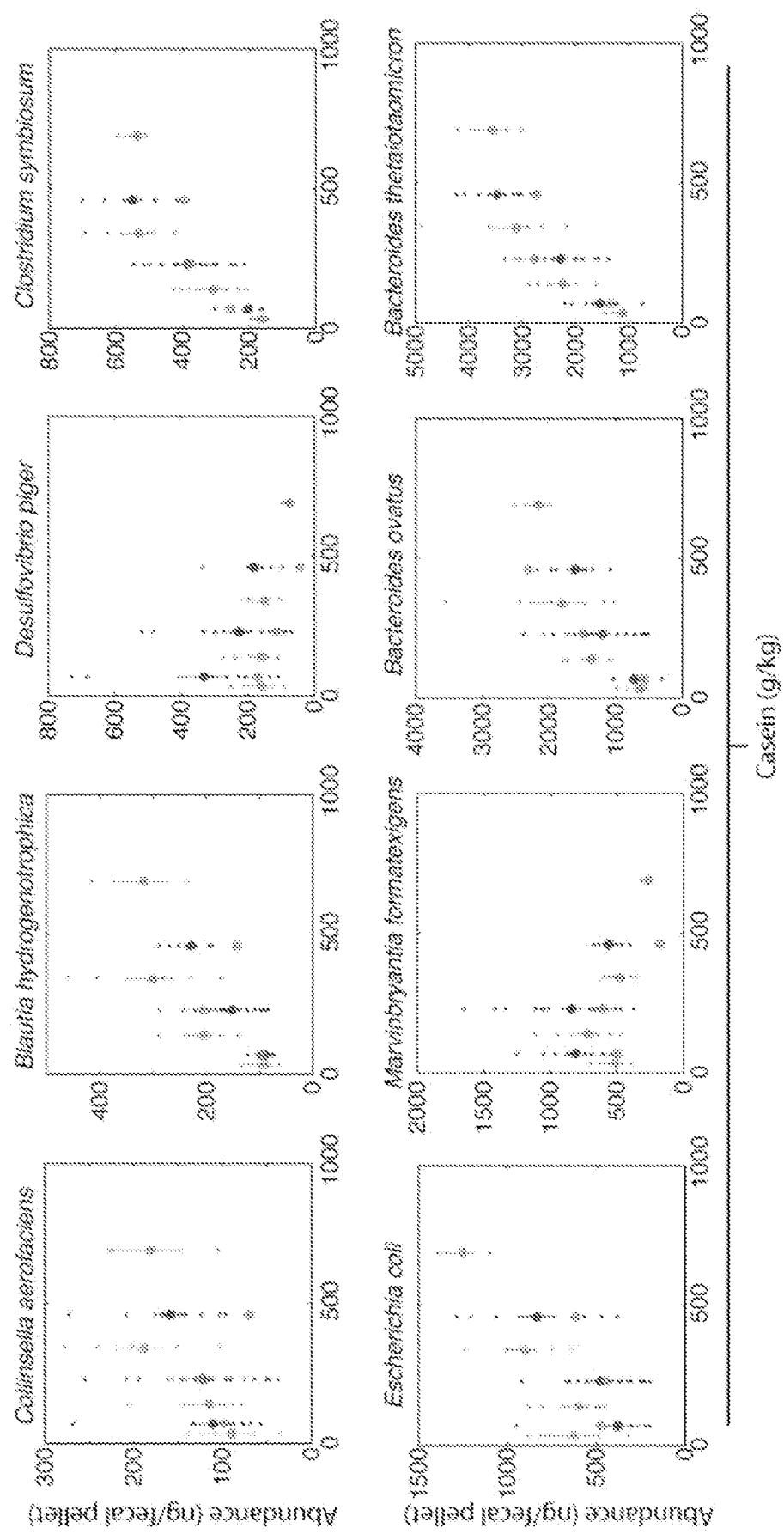
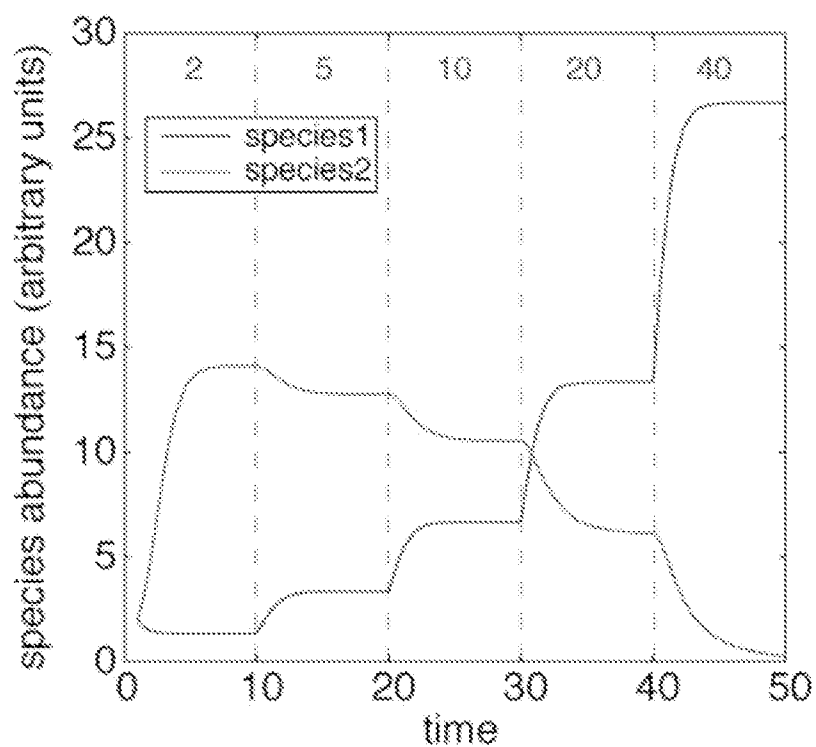
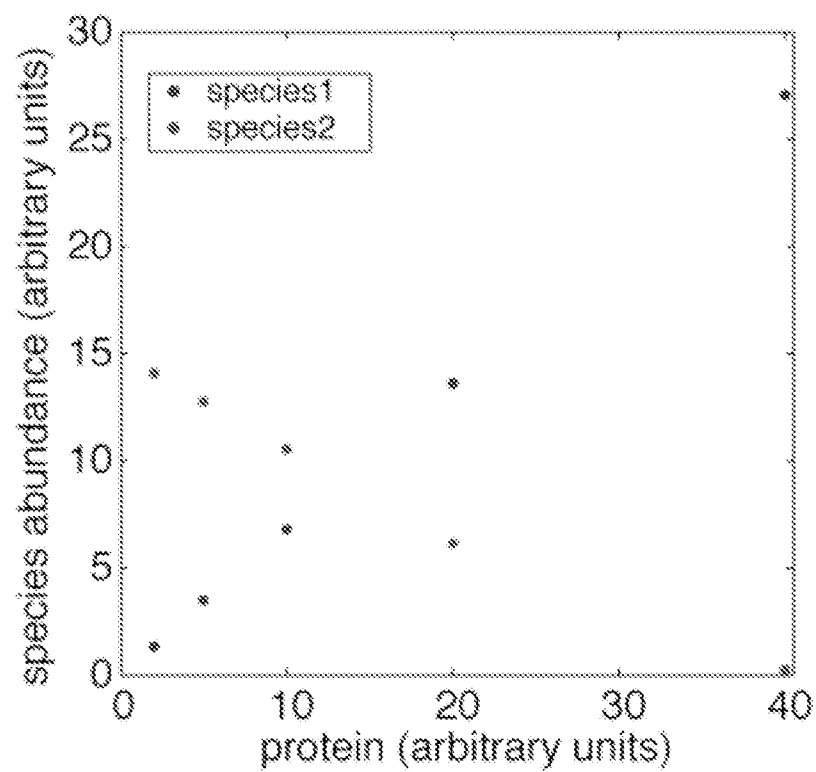
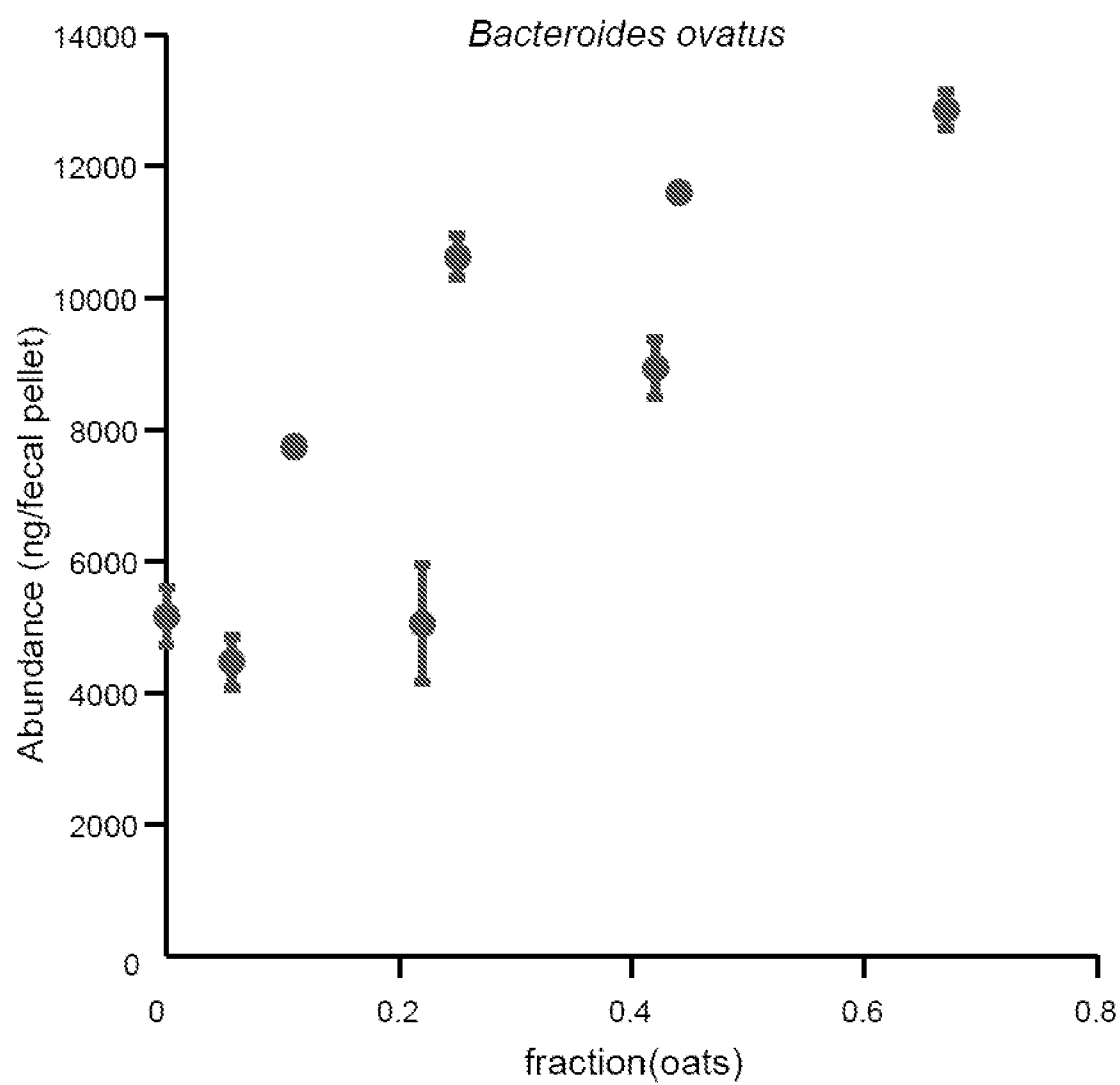
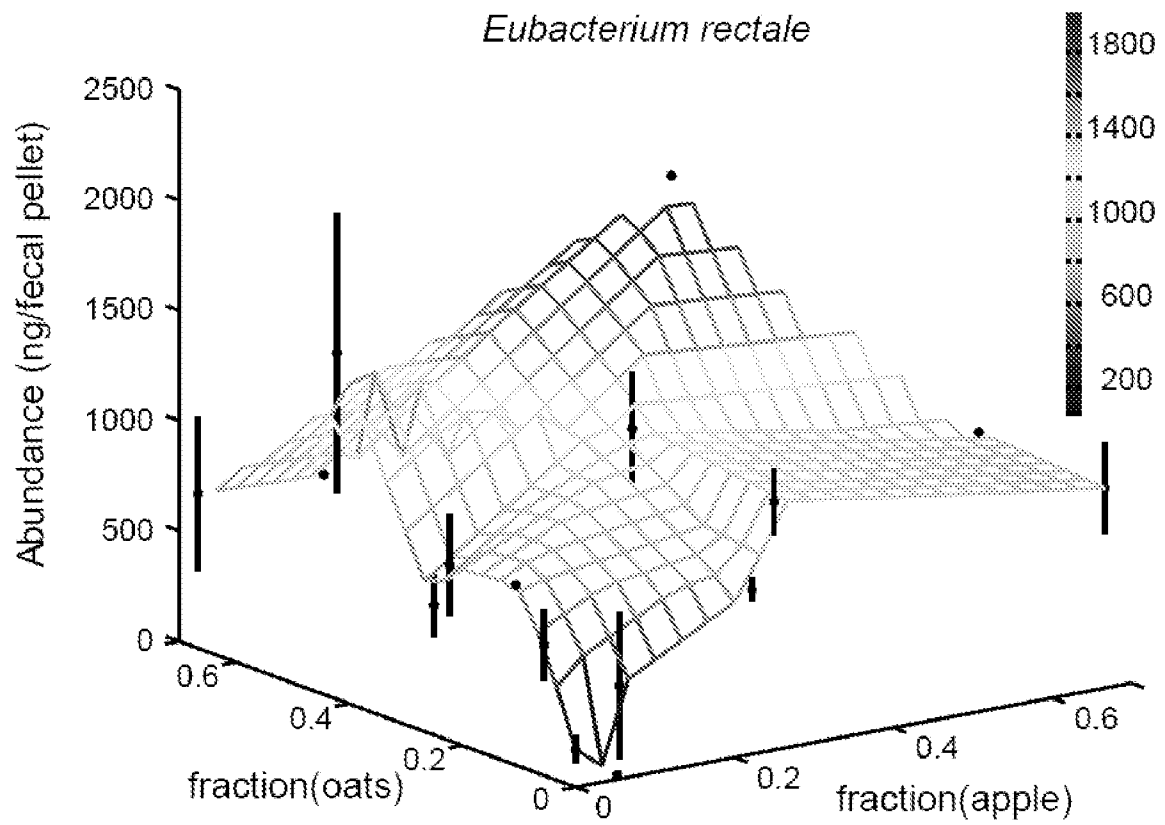
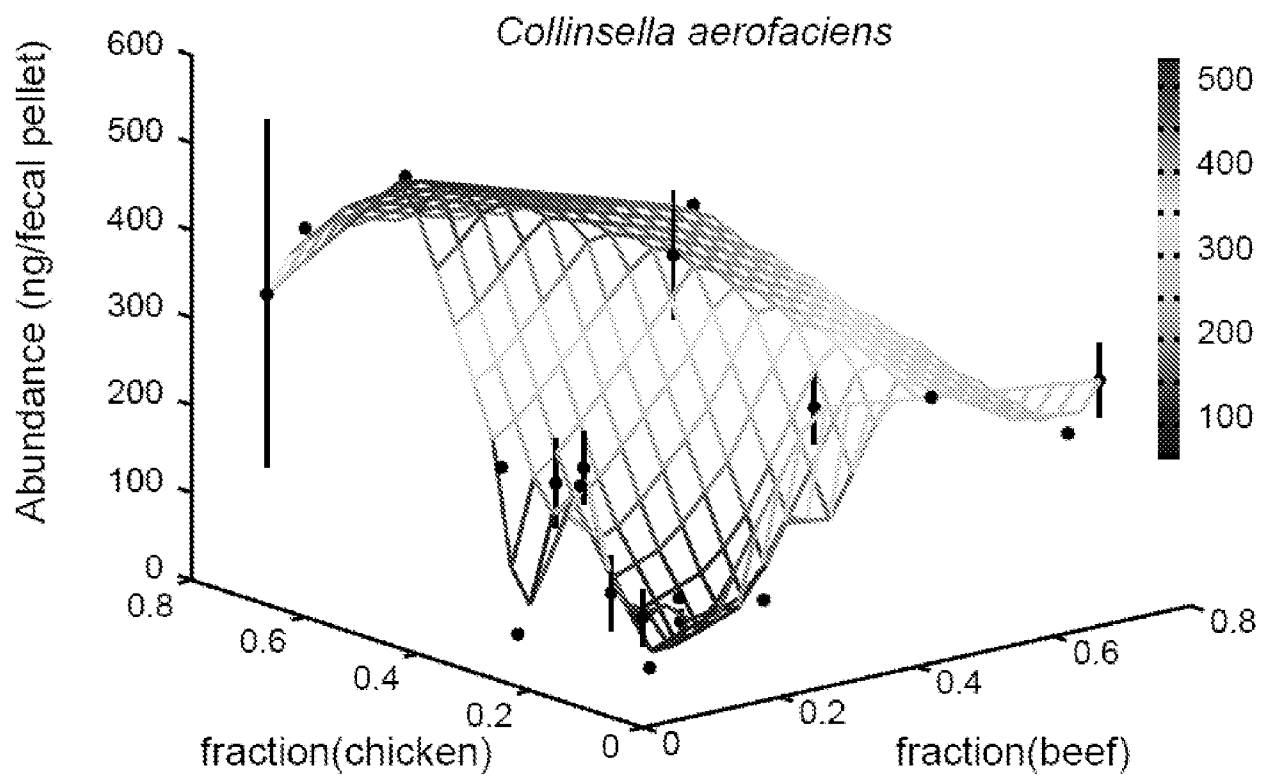


FIG. 17

A**B****FIG. 18**

**FIG. 19A**

**FIG. 19B**

**FIG. 19C**

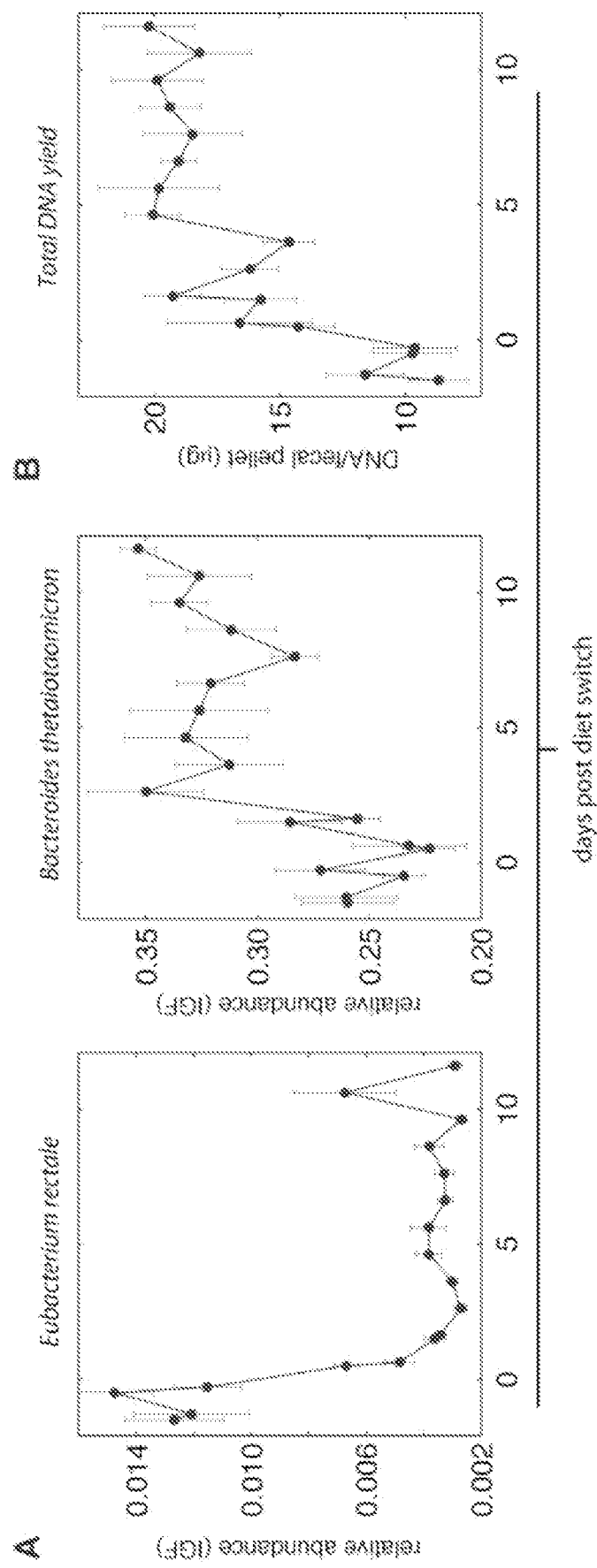


FIG. 20

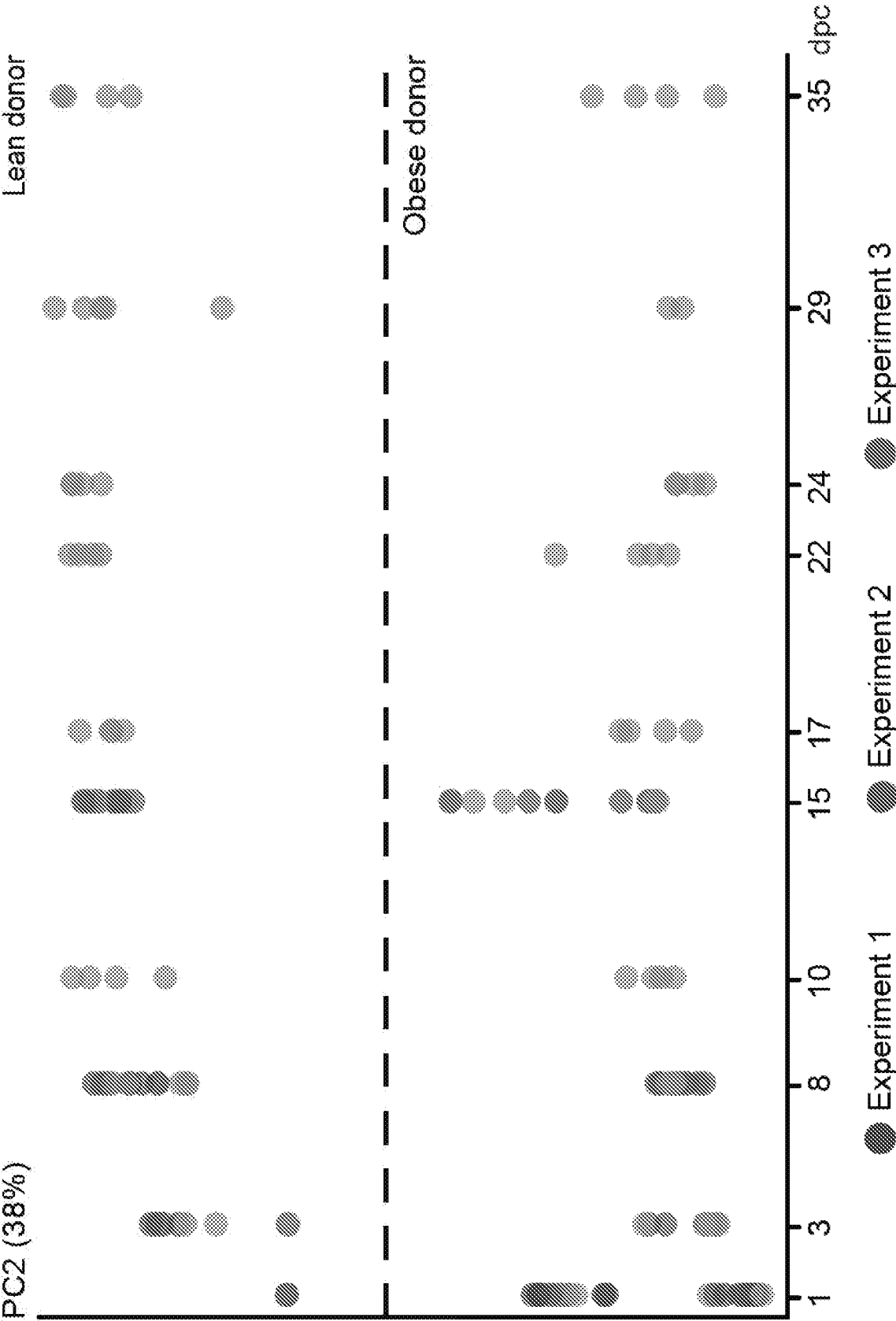


FIG. 21A

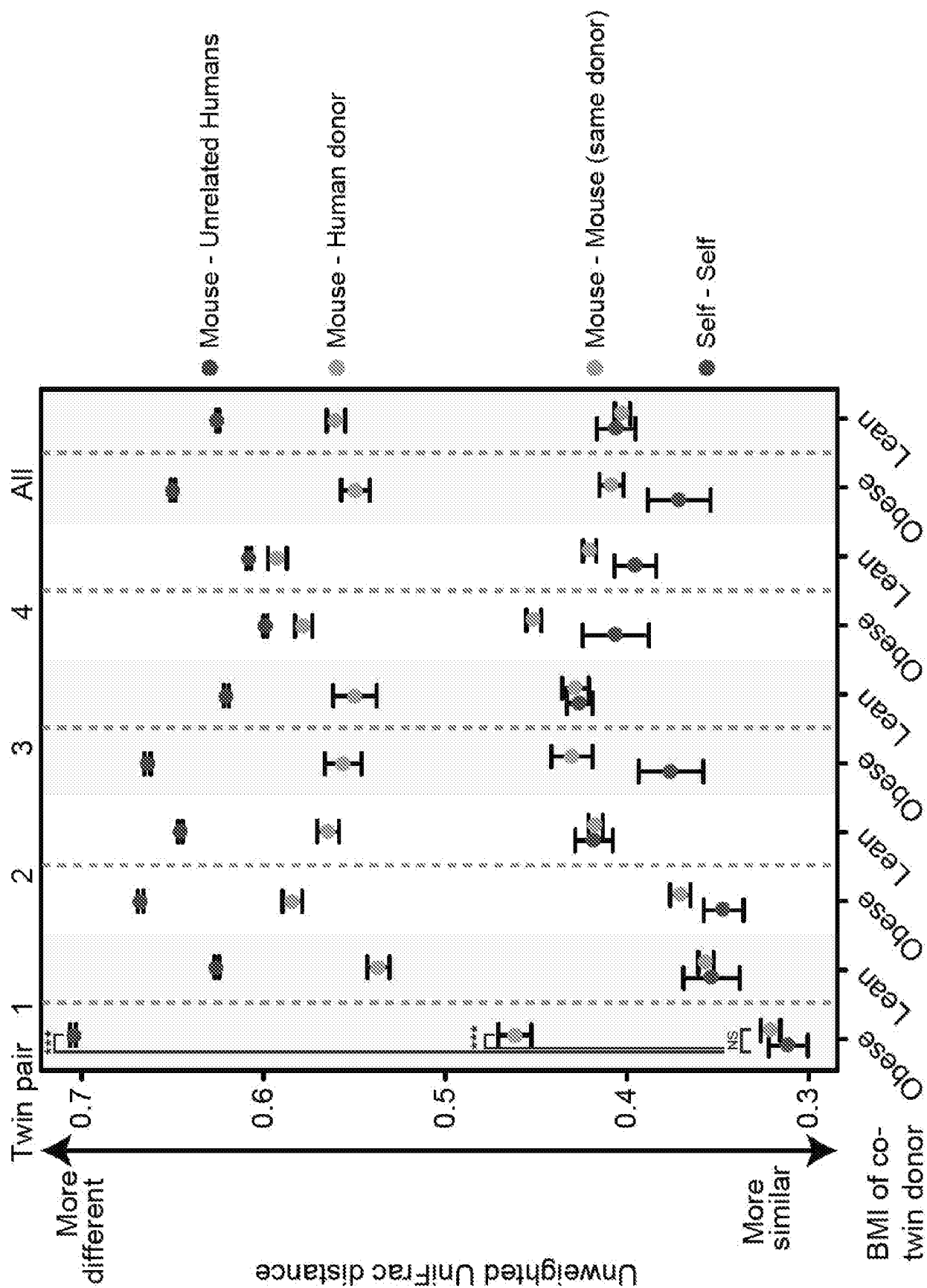


FIG. 21B

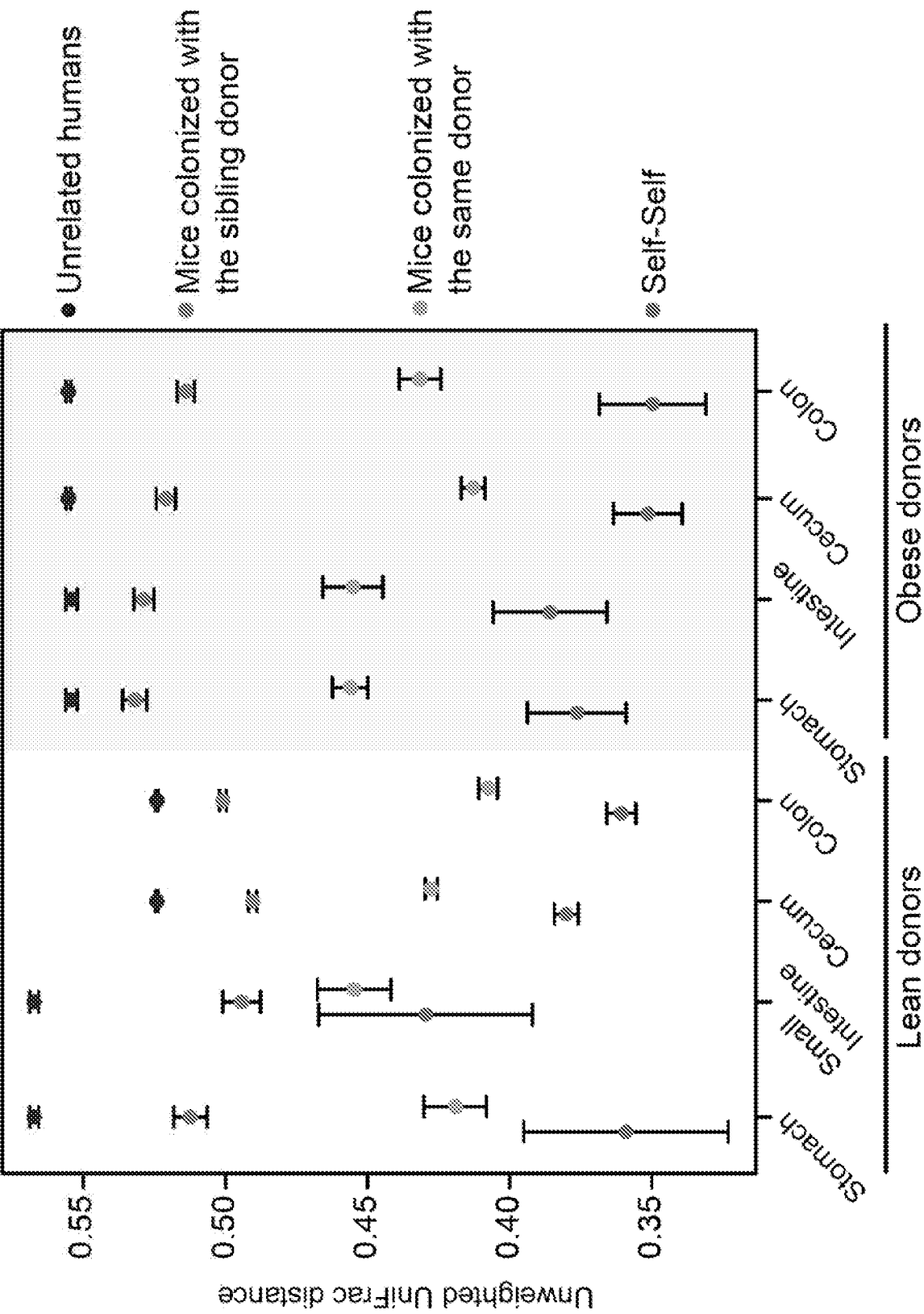
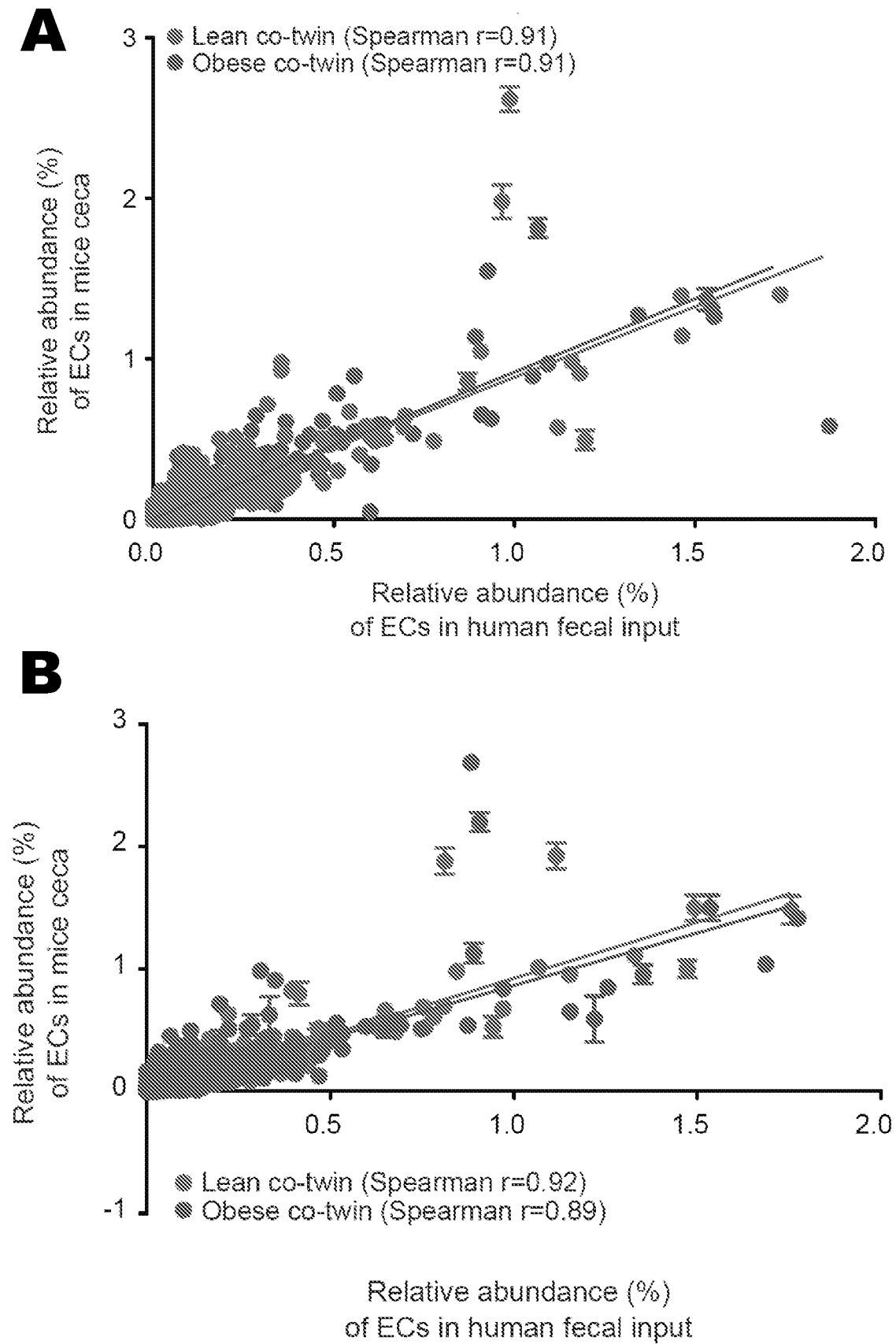
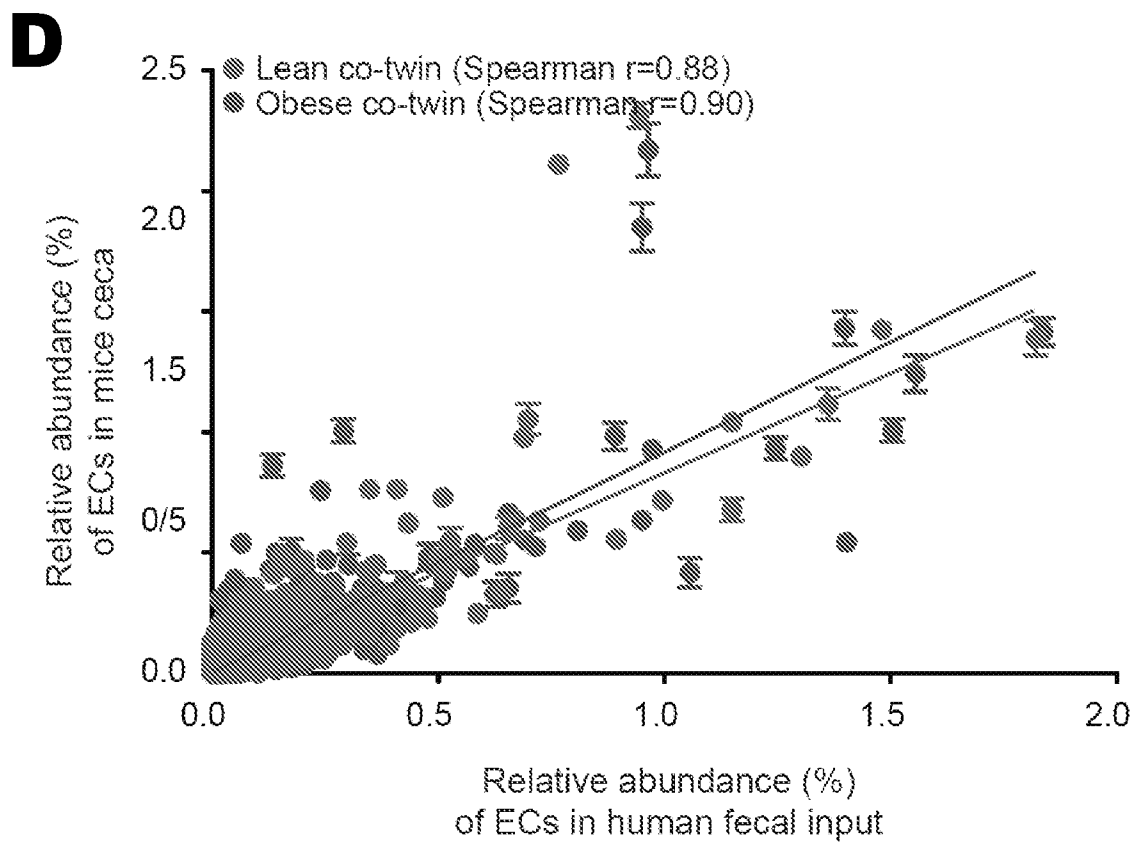
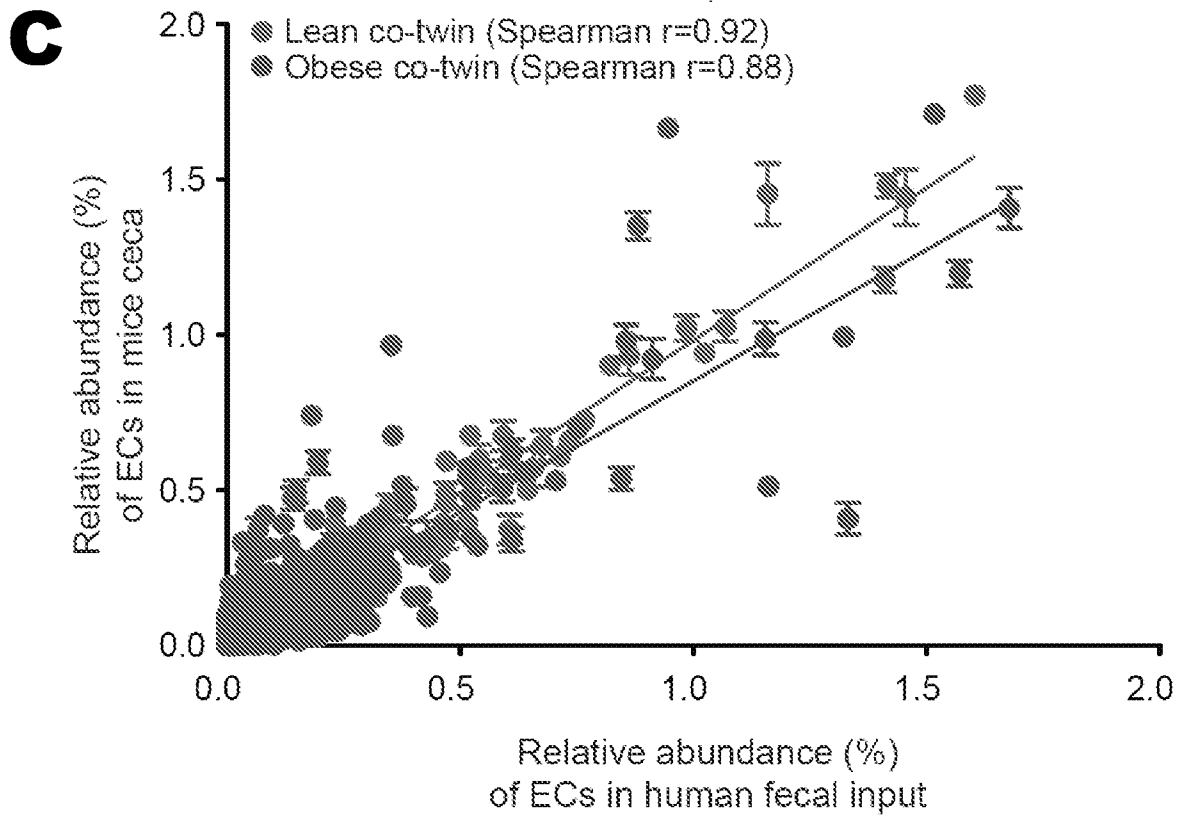
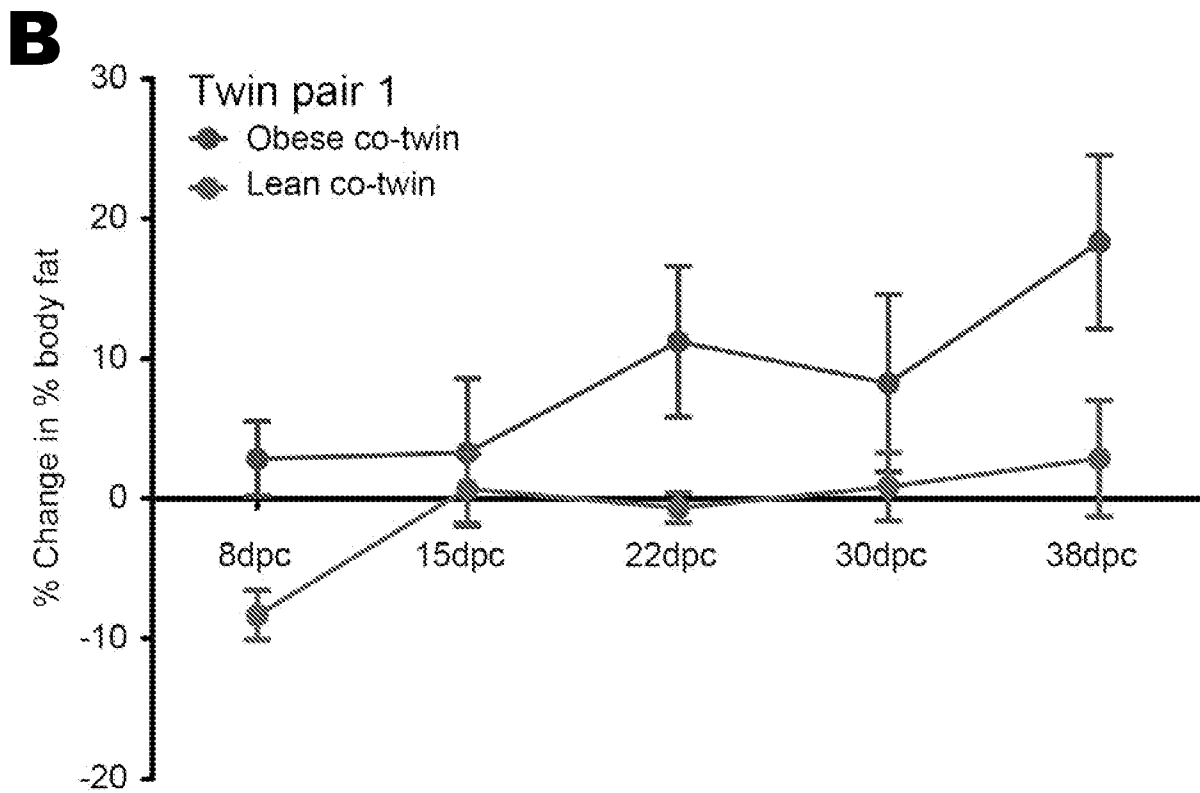
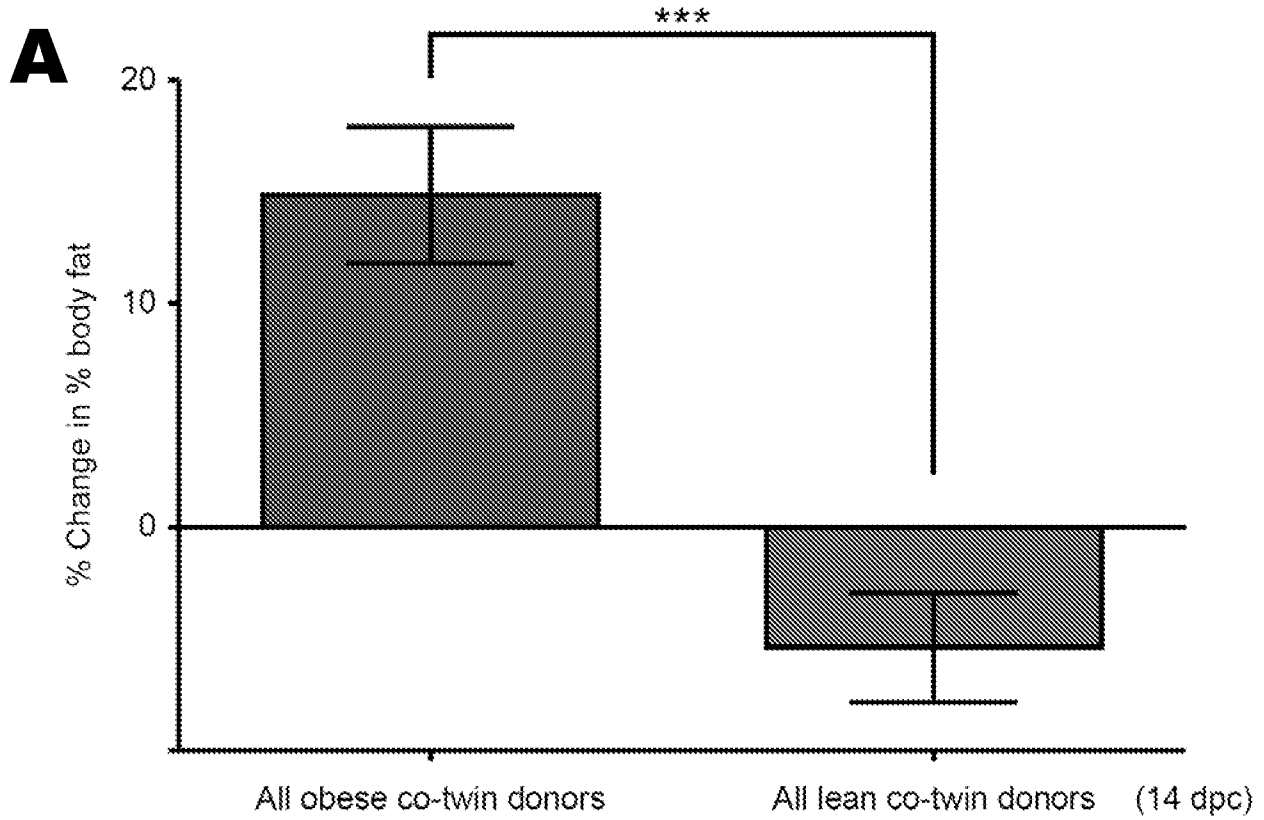


FIG. 22

**FIG. 23**

**FIG. 23**

**FIG. 24**

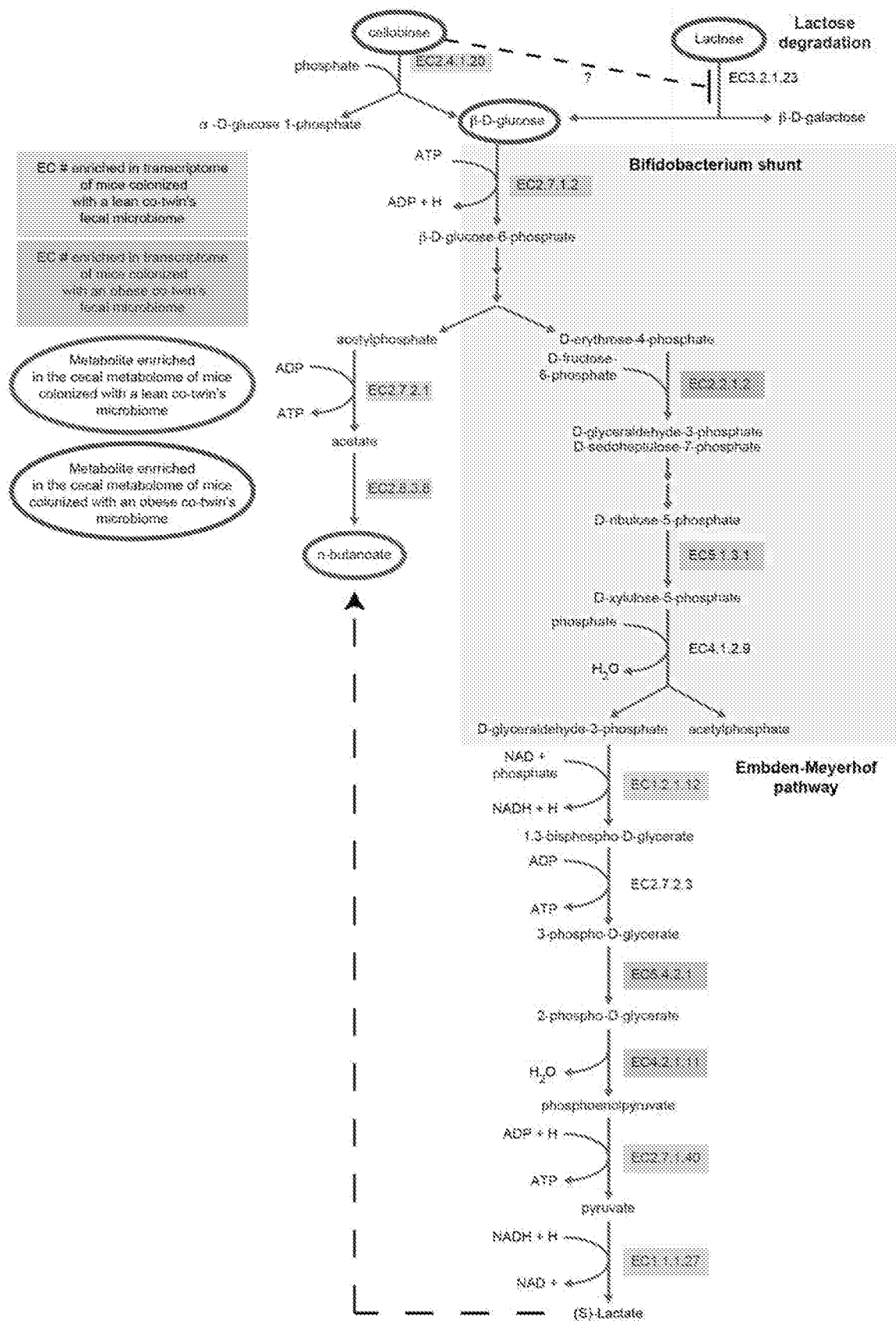
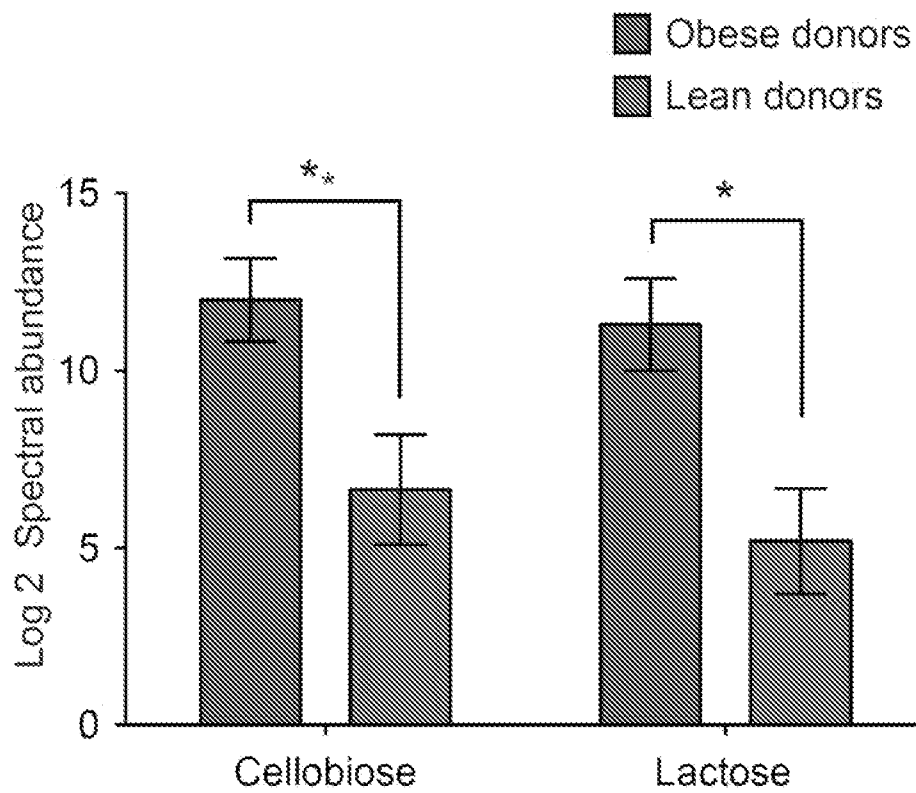
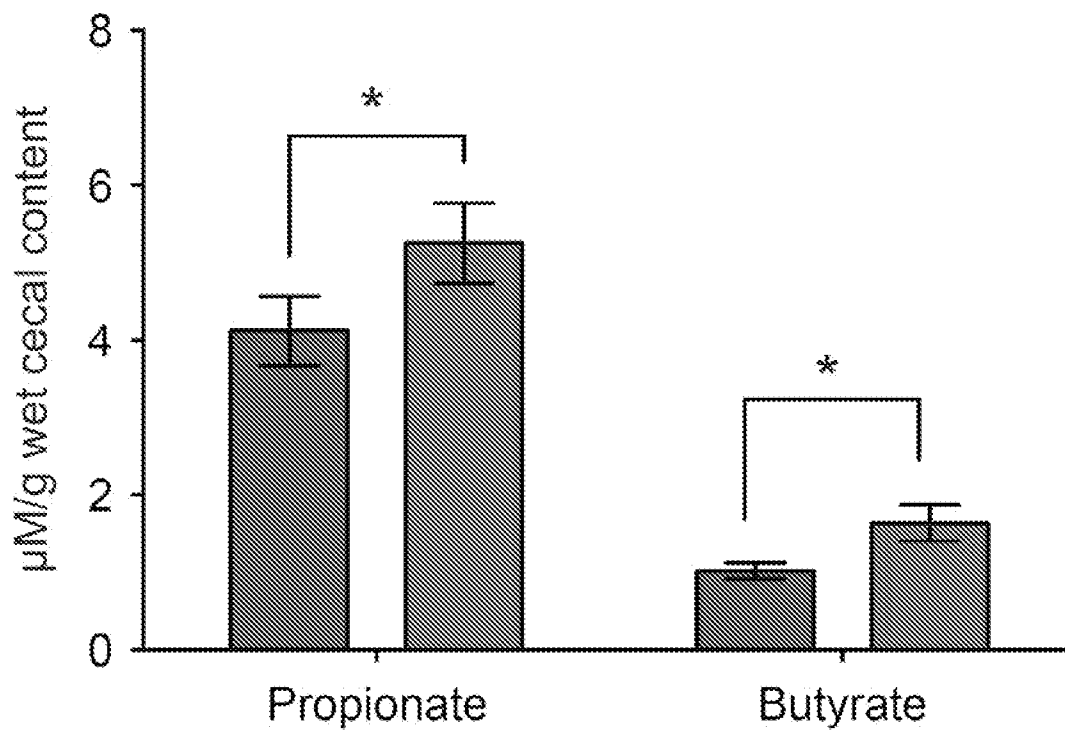
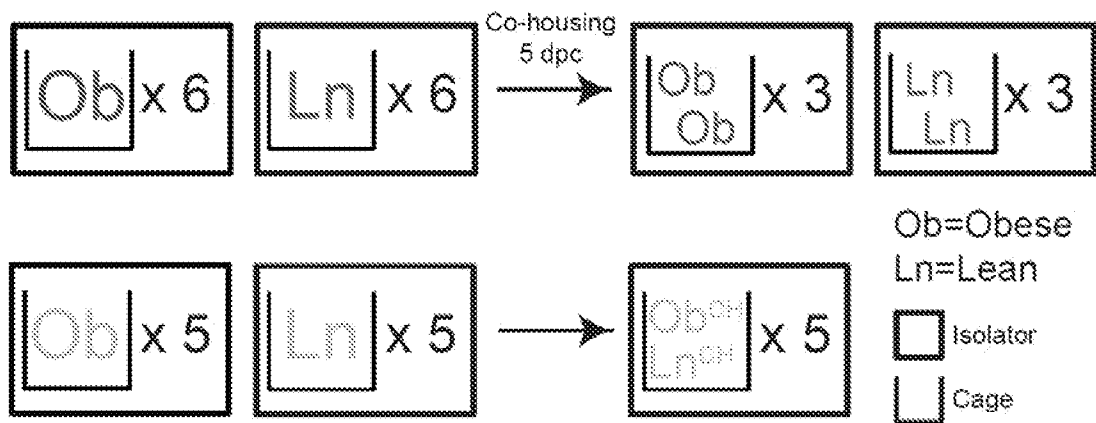
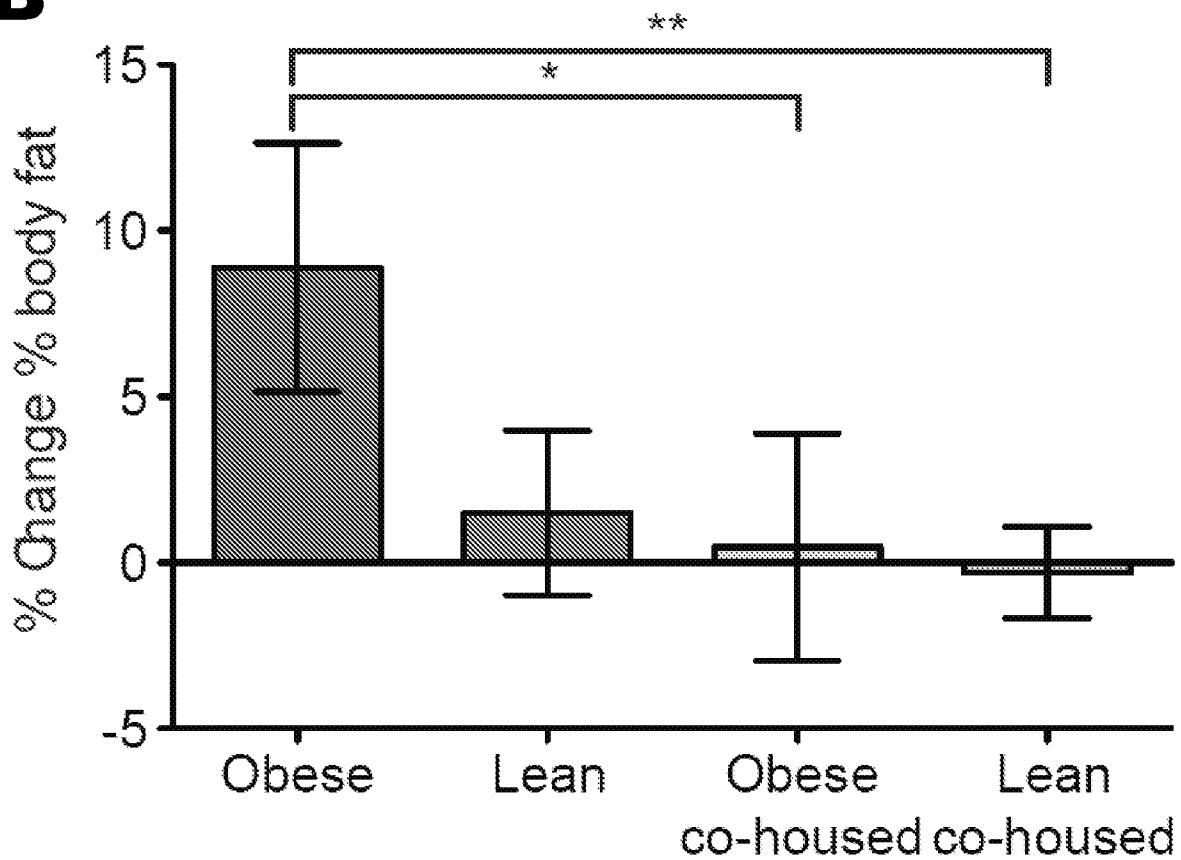
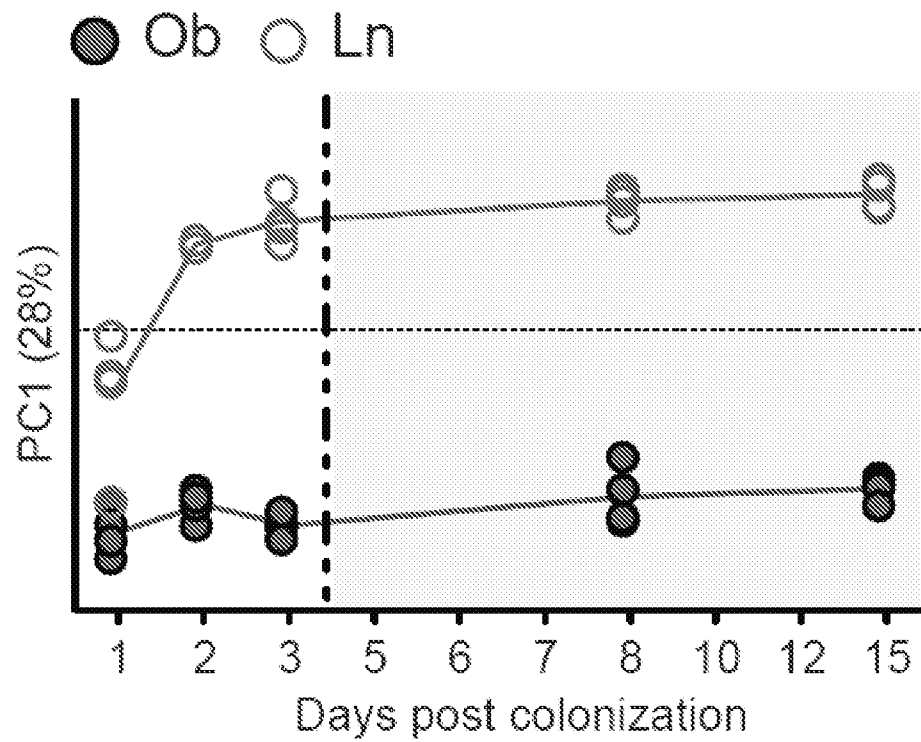
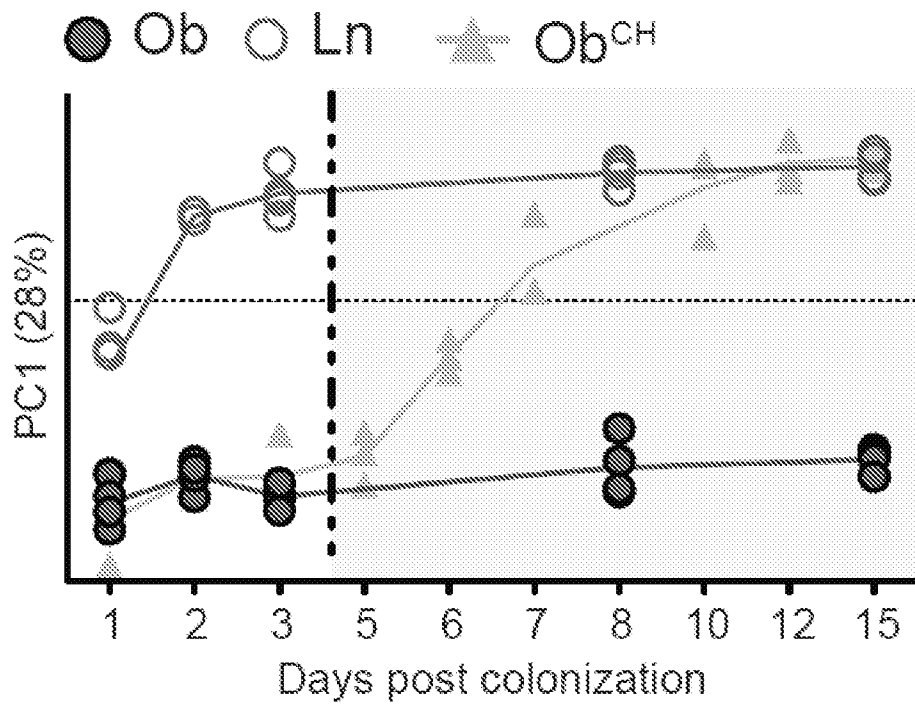
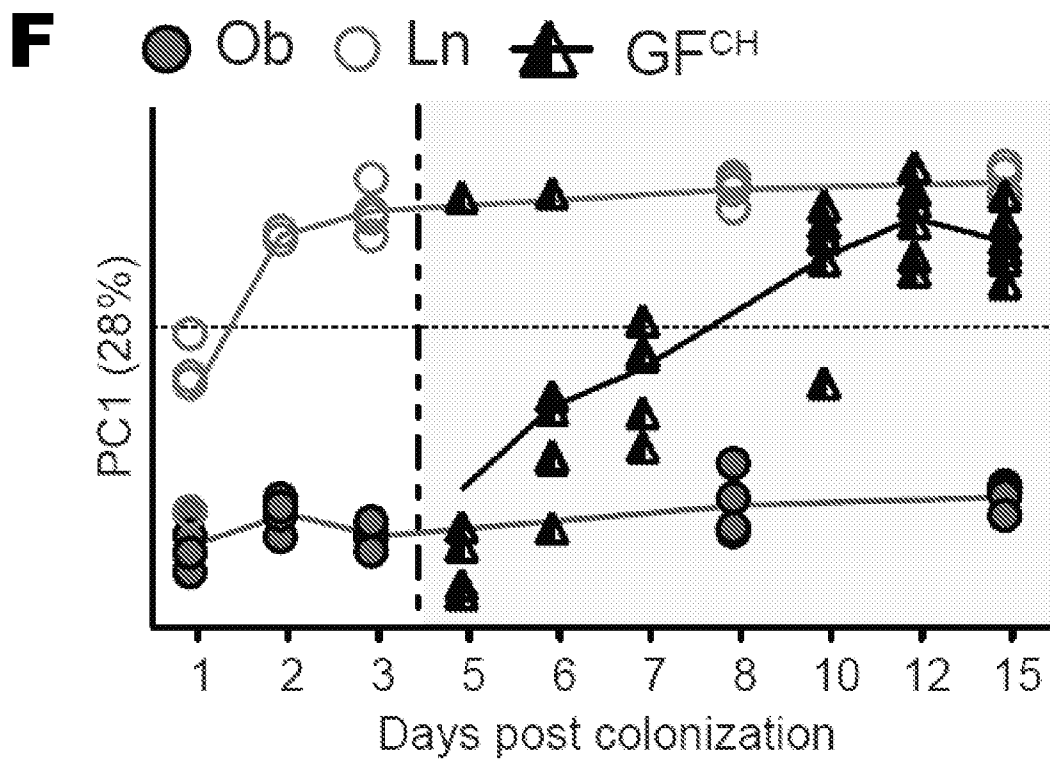
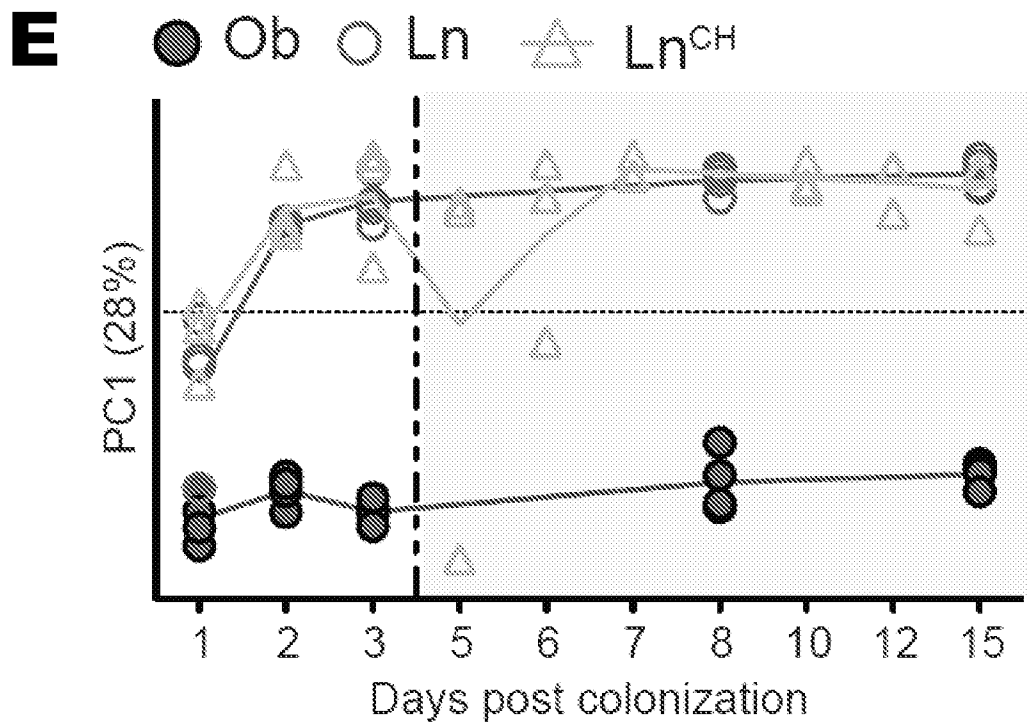


FIG. 25

A**B****FIG. 26**

A**B****FIG. 27**

C**D****FIG. 27**

**FIG. 27**

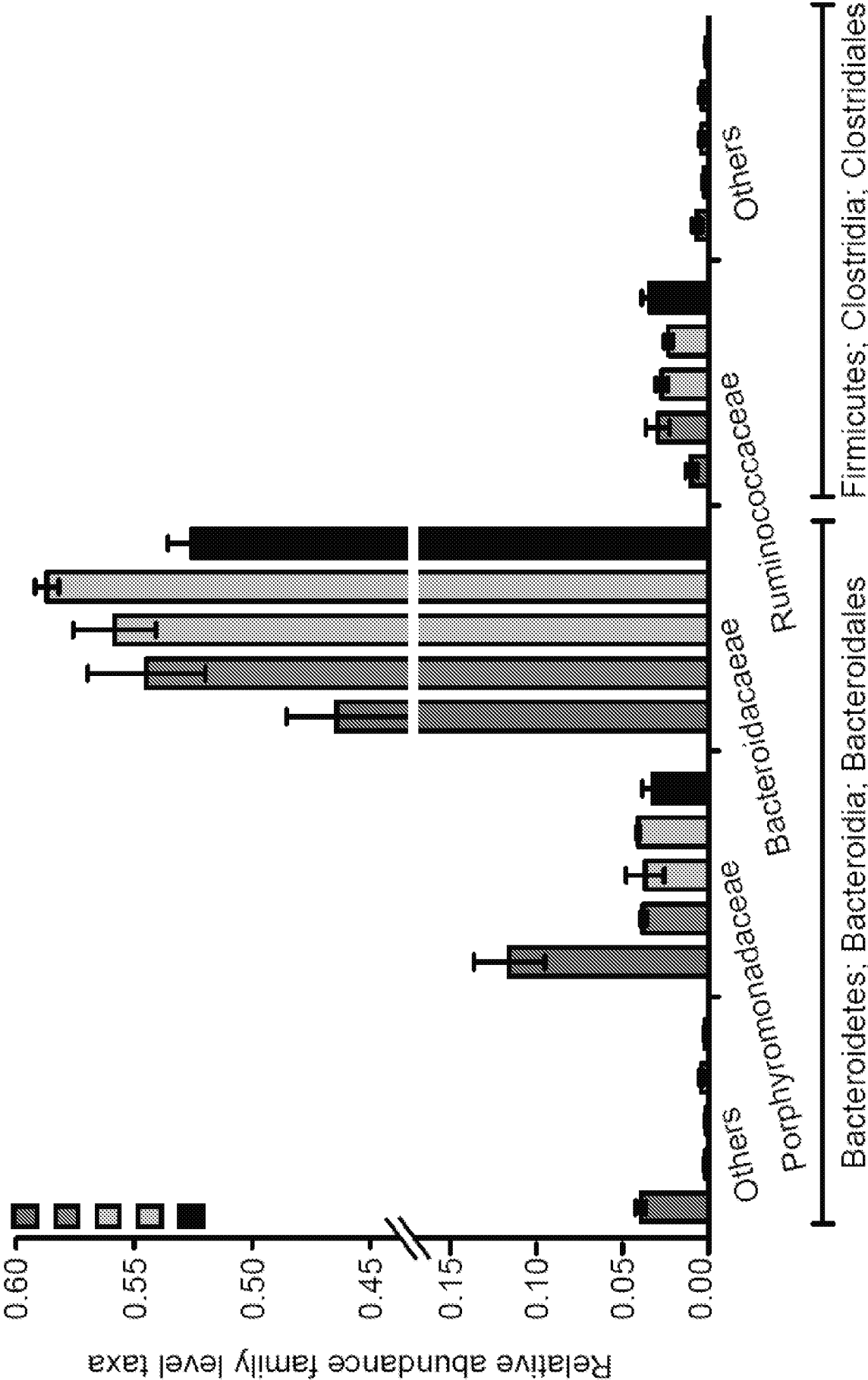
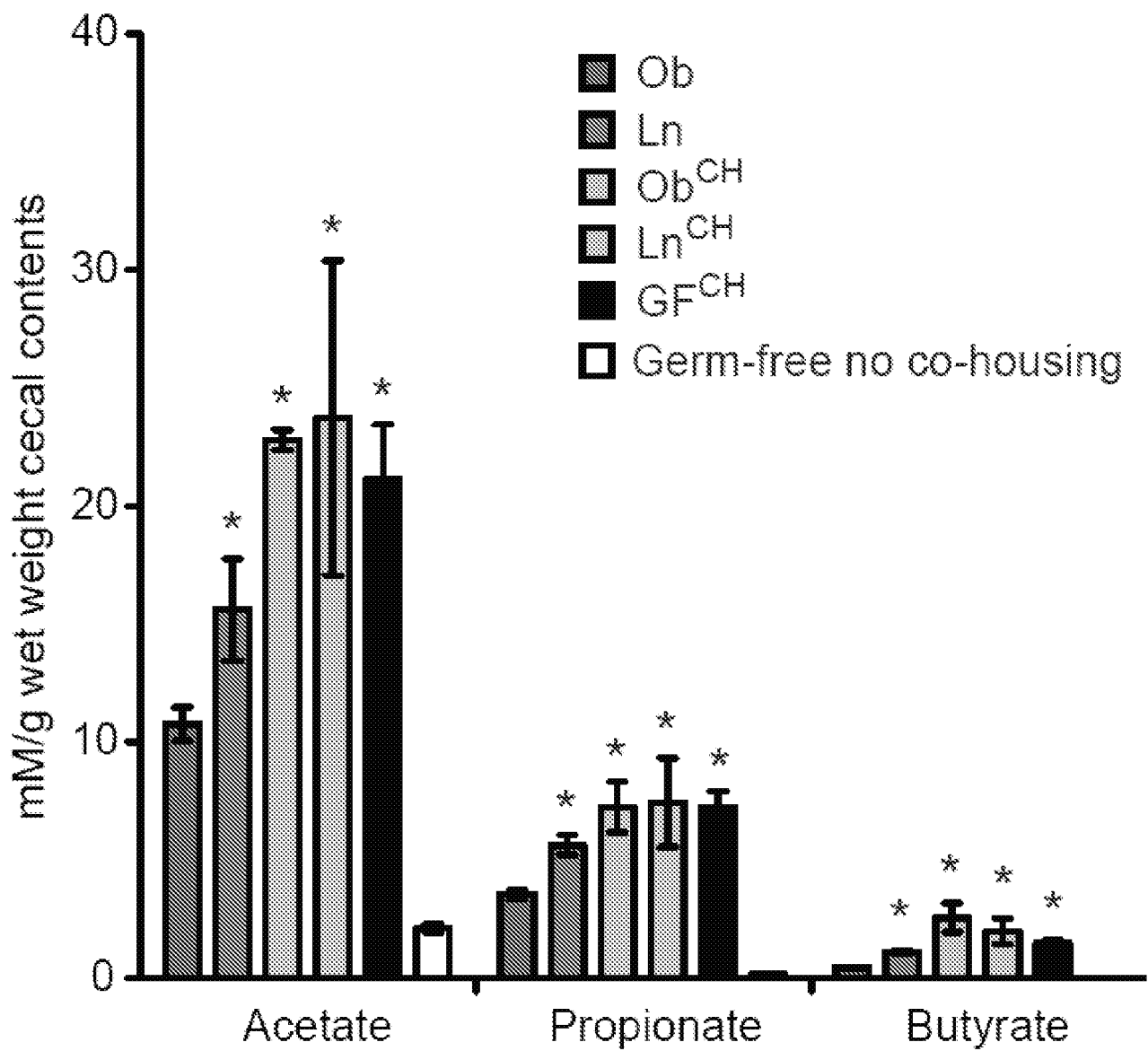
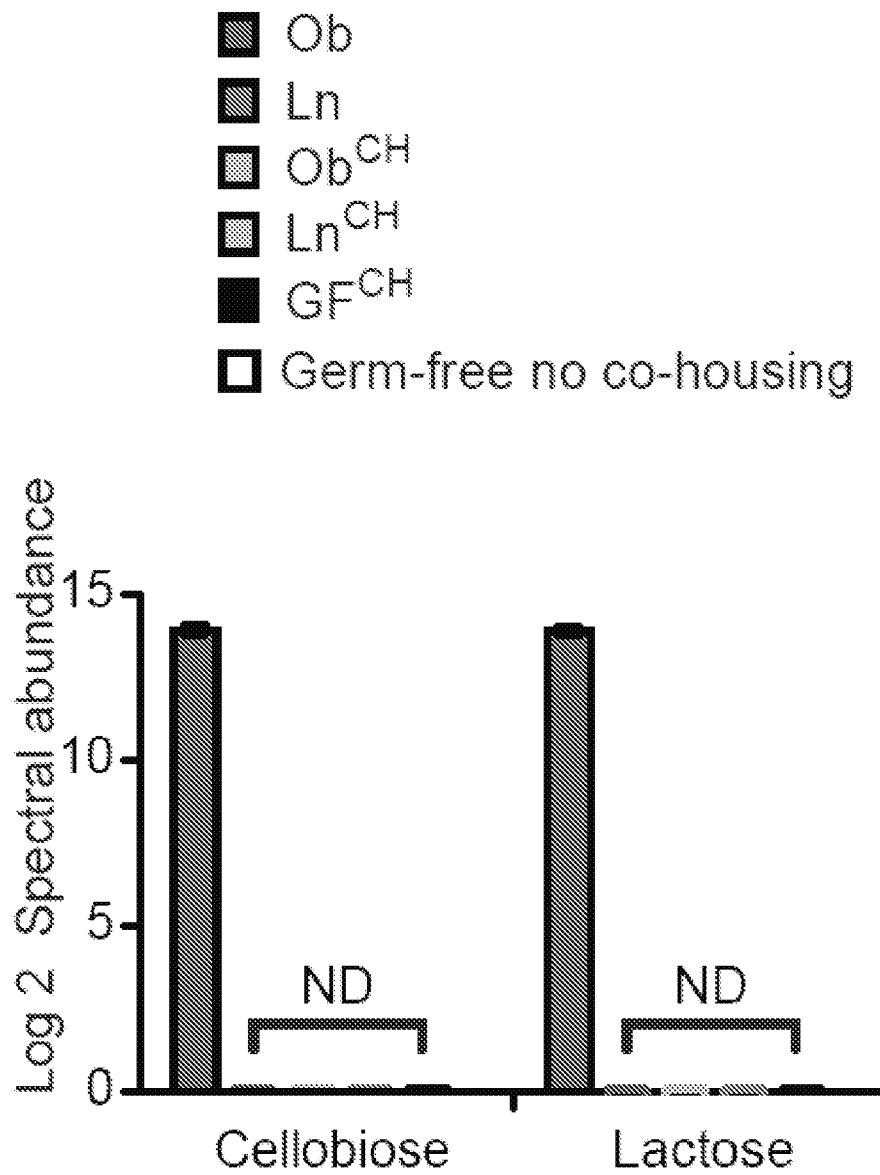


FIG. 27G

**FIG. 27H**

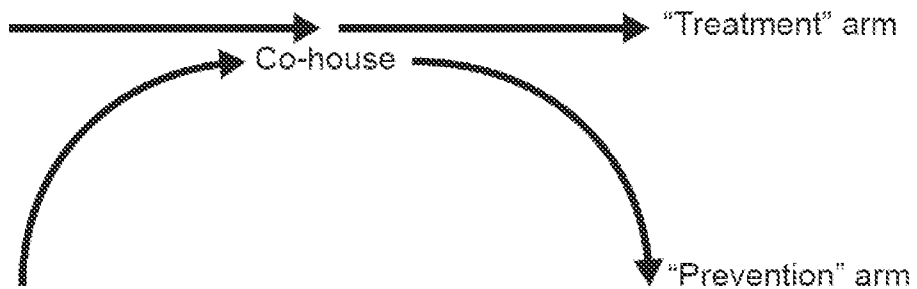
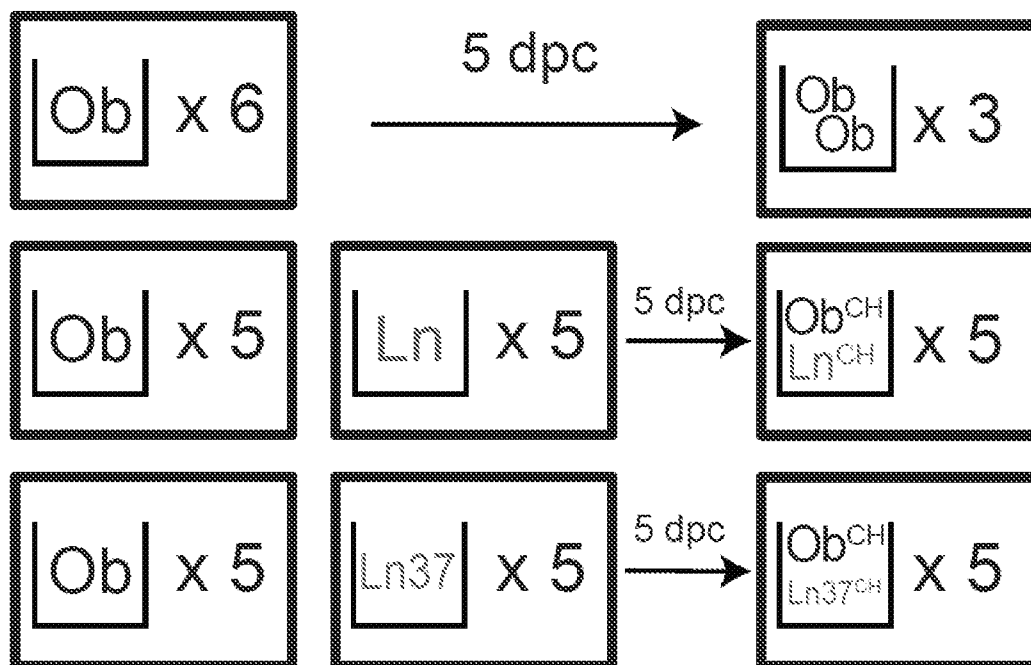
**FIG. 27I**

A

Mouse with obese
co-twin microbiota



Mouse with
"manufactured"
consortium

**B**

Key



Isolator



Cage

Ob = Obese co-twin's culture collection

Ln = Lean co-twin's culture collection

Ln37 = 37 members of lean co-twin's arrayed culture collection

FIG. 28

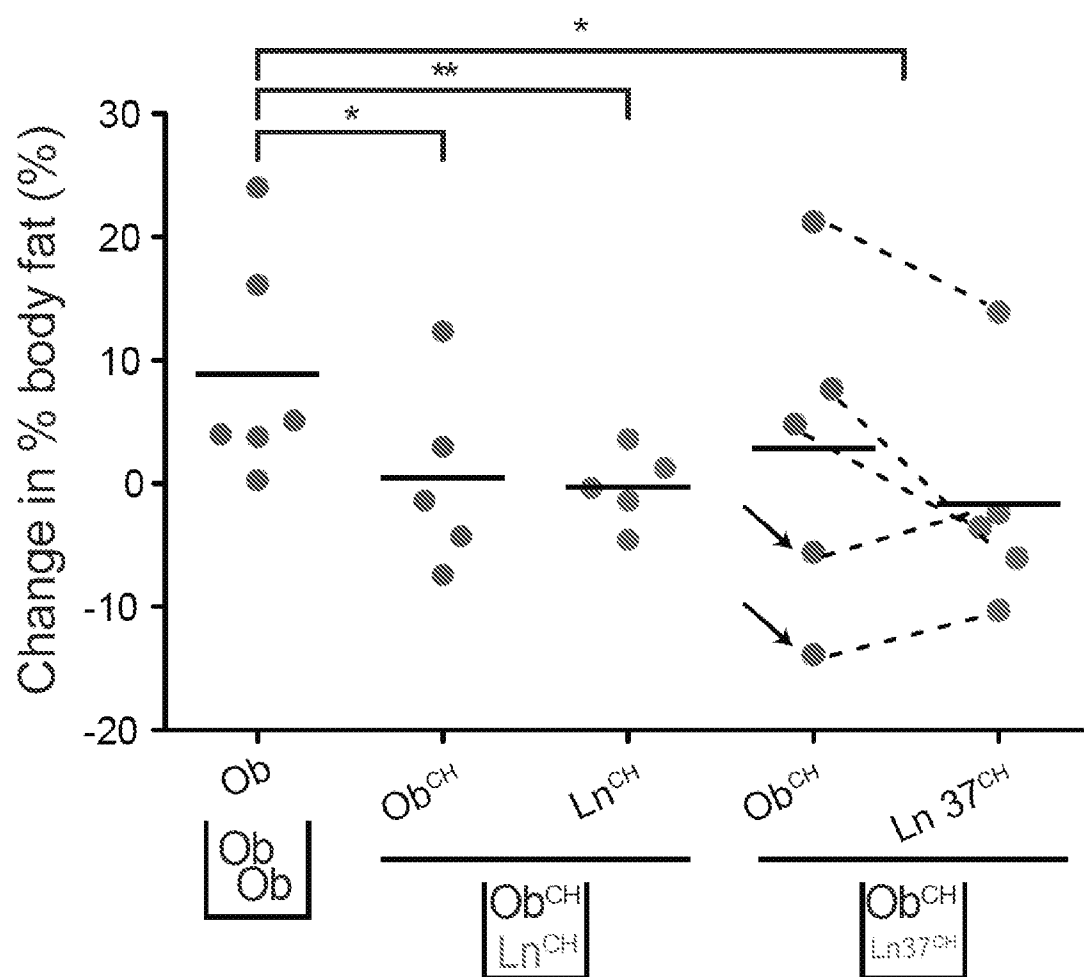


FIG. 28C