Supplementary Material for

Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children

Laura V. Blanton, Mark R. Charbonneau, Tarek Salih, Michael J. Barratt, Siddarth Venkatesh, Olga Ilkaveya, Sathish Subramanian, Mark J. Manary, Indi Trehan, Josh M. Jorgensen, Yuemei Fan, Bernard Henrissat, Semen A. Leyn, Dmitry A. Rodionov, Andrei L. Osterman, Kenneth M. Maleta, Christopher B. Newgard, Per Ashorn, Kathryn G. Dewey, and Jeffrey I. Gordon*

*Correspondence to: jgordon@wustl.edu

This PDF file includes:

Materials and Methods

Human studies

Details of the 317 infant twin pairs and three sets of triplets enrolled in the Malawi Twin Cohort study from birth to 32 months of age are described in an earlier publication (*13*). Enrollment for the iLiNS-DYAD-M study [clinicaltrials.gov #NCT01239693] was open to consenting women over the age of 15 years, with ultrasound confirmation of a pregnancy that was <20 weeks, who had come for prenatal care at one of three study clinics located in Mangochi District of southern Malawi (one public district hospital in Mangochi, one rural semi-private hospital in Malindi, and one rural public health center in Lungwena). This randomized, controlled clinical trial tested the effects of providing Small Quantity Lipid-based Nutrient Supplements (SQ-LNS) to pregnant and lactating women through 6 months postpartum and to their infants from 6-18 months of age (*16*).

Bacterial 16S rRNA sequencing

DNA was extracted by suspending pulverized human fecal samples (~50mg), or in the case of mouse studies resuspending fecal pellets, in a solution containing 500µL of extraction buffer [200 mM Tris (pH 8.0), 200 mM NaCl, 20 mM EDTA], 210µL of 20% SDS, 500µL phenol:chloroform:isoamyl alcohol (pH 7.9, 25:24:1, Ambion), and 500µL of 0.1-mm diameter zirconia/silica beads. Cells were mechanically disrupted using a bead beater (BioSpec Products, Bartlesville, OK; maximum setting for 4 minutes at room temperature), followed by extraction with phenol:chloroform:isoamyl alcohol and precipitation with isopropanol.

PCR was used to generate amplicons (~365bp) of variable region 4 (V4) of the bacterial 16S rRNA gene using primers and cycling conditions described in a previous report (*11*). PCR primers incorporated sample-specific barcodes allowing samples to be subjected to multiplex sequencing using the Illumina MiSeq instrument (paired-end 250nt reads). Paired V4-16S rRNA sequences were trimmed to 200bp and merged into a single sequence with Flash software (*24*). Merged sequences were filtered for low quality reads and binned according to their sample-specific barcodes. Reads were clustered into 97%ID OTUs using UCLUST (*14*) and the Greengenes OTU reference database (version May 2013). Reads that failed to hit the reference dataset were clustered *de novo* using UCLUST. A representative OTU set was created using the most abundant OTU from each bin. Reads were aligned using PyNAST. A custom dataset of manually modified NCBI bacterial taxonomy (*18*) was used to train the Ribosomal Database Project (RDP) version 2.4 classifier (*15*) and to assign taxonomy to picked OTUs. Validation of this assignment strategy is described in a previous publication (*18*). For analysis of bacterial abundance and colonization efficiency, rare OTUs (<0.1% relative abundance within each sample) were removed.

Random Forests-derived models and statistical analysis

Models were generated using the randomForests package in R. 1000 trees were generated per iteration, and 100 iterations were performed to generate average MSE values. The rfcv function was utilized for cross-validation.

Univariate analyses (t-tests and ANOVA) were performed using Microsoft Excel or Prism 6.0 (GraphPad Software, Inc.). Heatmaps were generated using Gene-E software (http://www.broadinstitute.org/cancer/software/GENE-E/index.html). Spearman's rank correlation analyses were performed in the R environment using the ltm package and rcor.test function.

Gnotobiotic mouse studies

All gnotobiotic mouse experiments were performed using protocols approved by the Washington University Animal Studies Committee. Male germ-free C57BL/6J mice were maintained in sterile, flexible, plastic gnotobiotic isolators (Class Biologically Clean Ltd., Madison, WI) under a strict 12-hour cycle (lights on at 0600h, off at 1800h). Mice were fed an autoclaved low-fat, polysaccharide-rich chow (LF/HPP) diet (B&K University, East Yorkshire, U.K.; diet 7378000) from weaning until 3 days prior to the beginning of an experiment. At that time, 4.5-week-old male mice were switched to the M8 diet for the remainder of the experiment. All animals were euthanized without prior fasting.

Preparation of the Malawi-8 (M8) Diet –The macro- and micronutrient content of components of this prototypic Malawian diet was computed using the USDA Nutrient Database. Linear programming was then employed to calculate combinations of ingredients that resembled the mean energy and nutrient values for foods identified from the diet survey.

The M8 diet was prepared as follows. Maseca® corn flour was obtained from Restaurant Depot (College Point, NY). The remaining ingredients were purchased from Whole Foods Supermarkets. A relish containing 2kg mustard greens, 1.5kg onions, and 1.5kg tomatoes was first prepared by pureeing them in a food processor (Robot Coupe Model R23, Jackson, MS) and cooking the puree in 1L water for 60 minutes on a Corning stirrer/hot plot (high setting, until browned). After cooking, the relish was combined with a pureed mixture of 1kg ground peanuts, 700g soaked red kidney beans, 1kg canned pumpkin, and 2.5kg peeled bananas in an industrial mixer (Globe SP30P 30-quart pizza mixer; gear speed 1; Globe Food Equipment Company, Dayton, OH). Corn flour (5kg) and hot, freshly autoclaved water (5L) were then added slowly and mixed using the industrial mixer for 5 minutes to ensure uniformity of the food. Dry pellets of the M8 diet were then generated (extruded diameter, 0.5-inch) (Dyets, Inc. Bethlehem, PA), placed in FDA/USDA-compliant poly-nylon vacuum pouches (#S-7556; Uline, Pleasant Prairie, WI; 500 g aliquots), double-bagged and sterilized by irradiation (20-50 kGy) within 24 hours of production (Steris Co; Chicago, IL). The nutritional content of all cooked and irradiated custom diets was defined by N.P. Analytical Laboratories (St Louis, MO) (**Table S5**). Irradiated food was stored at 4°C for up to six months. Sterility was determined by resuspending a small aliquot of each batch of food in pre-reduced Gut Microbiota Medium (GMM; (*25*)) under anaerobic conditions and incubating the suspension for 3 days at 37°C. Sterility was further verified by subculture on pre-reduced anaerobic GMM agar plates.

Generation of clarified stool for gavage into gnotobiotic mice – An aliquot of the frozen sample was pulverized in a Biosafety Class II hood with a ceramic mortar and pestle filled with liquid nitrogen. An aliquot (1g) of the pulverized material, sealed in a sterile screw-capped tube (Axygen SCT-200-C-S), was brought into an anaerobic Coy

chamber (atmosphere: 20% CO_2, 5% H_2, and 75% N_2), immediately suspended in 15mL GMM and vortexed on maximum speed (four cycles of blending for 20 seconds followed by a 30 second pause). The sample was allowed to stand for 5 minutes so that particulate matter could settle by gravity; the resulting supernatant was passed through a 100μm pore diameter filter (BD systems, Inc. Franklin Lakes, NJ) to remove remaining particulate material, mixed with an equal volume of pre-reduced GMM containing 30% glycerol (final concentration 15% glycerol) and placed in Wheaton crimp top tubes for storage at -80°C. Mice received a single oral gavage (200mL) of a clarified uncultured human fecal sample 3 days after switching to the M8 diet.

Quantitative magnetic resonance (qMR) analysis of body composition – Body composition was defined using an EchoMRI-3in1 instrument (EchoMRI, Houston, TX). Each mouse was transported from a gnotobiotic isolator to the MR instrument in a 9.5" long, 2" diameter, HEPA filter-capped glass vessel. At each time point surveyed, each mouse was scanned 2-3 times in the instrument over a 5 minute period and the values from each scan were averaged.

Sample collection – Fecal samples were collected at defined times after gavage. At sacrifice, cecal contents, liver, brain, and skeletal muscle (gastrocnemius) were collected and immediately frozen in liquid nitrogen. Gastrocnemius muscle was freeze-clamped immediately after sacrifice using pre-chilled forceps.

Defining the efficiency of microbiota transplantation – Transplantation efficiency was assessed by comparing 97%ID OTUs represented in the input human donor sample, weighted to relative abundance, and in the fecal microbiota of recipient mice sampled throughout the course of the experiment [**Table S4B;** note that microbiota from four 6-month-old singleton infants from iLiNS-DYAD-M (two healthy, two undernourished) and four 18-month-old co-twins from the Malawian twin study (one healthy, three undernourished) had >50% transplantation efficiency]. Bacterial viability, quantified by the number of colony forming units 24 hours after plating onto GMM medium, was significantly greater ($p<3.2x10^{-6}$; Student's t-test) in the case of fecal samples from the twins, which had been flash frozen in liquid nitrogen within 5-10 minutes after they were produced. In iLiNS-DYAD-M, the time to freezing was often several hours.

Micro-computed tomography (micro-CT) – Femurs were harvested from mice at time of sacrifice, cleaned of soft tissue and stored at 4°C in 70% ethanol solution until scanned. Micro-CT was performed using a μCT 40 desktop, cone-beam instrument (ScanCO Medical, Brüttisellen, Switzerland). For cortical analyses, 200-300 slices were taken of each femur in the transverse plane with a 6μm voxel size (high resolution). For all cortical scans, slices began at the midpoint of the femur and extended toward the distal femur. Boundaries of and thresholds for bone were drawn manually using the μCT 40 software, and volumetric parameters (bone volume/tissue volume, bone mineral density, and cortical thickness) were calculated using custom scripts.

Cohousing experiments – Four days after gavage of donor microbiota, all mice were placed into HEPA filter-capped glass transport tubes for qMR measurements. Following qMR, control mice were returned to their original gnotobiotic isolator and housed with a cagemate of similar body weight colonized with the same donor microbiota (H-H controls; Un-Un controls). For pairing of Un^{CH} and H^{CH} mice, animals were transported to a new gnotobiotic isolator. All mice received fresh autoclaved

bedding (Aspen wood shavings, NEPCO) upon initiation of cohousing. Bedding was changed weekly in all cages in each gnotobiotic isolator.

We defined the direction of invasion by calculating the log odds ratio of the probability of a H origin (*PH*) or an Un origin (*PUn*) for each species-level taxon or 97%ID OTU, *i*:

$$\log_2[PH_i/PUn_i]$$

A positive log odds ratio indicates that a species or 97%ID OTU is derived from a H source, while a negative log odds ratio indicates a Un source. An invasion score was calculated to quantify the success of invasion of each species or 97%ID OTU, *i*, into each cohousing group, *j*:

$$Invasion\ Score_{ij} = \log_2\left[\overline{A_{ij}}/\overline{B_{ij}}\right]$$

where $\overline{A_{ij}}$ is the average relative abundance of taxon *i* in all fecal samples collected from group *j* after cohousing, and $\overline{B_{ij}}$ is its relative abundance in all samples from that group prior to initiation of cohousing. The mean of the distribution of invasion scores for Un^{CH} animals was significantly higher than that for Un-Un controls ($P \leq 0.0085$, Welch's two-sample t-test). This was not the case for H^{CH} animals compared to H-H controls ($P > 0.05$).

Targeted mass spectrometry

Liver, brain, and muscle tissue samples were homogenized in 50% aqueous acetonitrile containing 0.3% formic acid (50mg wet weight tissue/mL solution) using a high-speed homogenizer (IKA Model #EW-04739-21) set at maximum speed for 30-45 seconds. Samples were maintained on ice or dry ice throughout homogenization.

Amino acids, acylcarnitines, organic acids, acyl CoAs and ceramides were analyzed using stable isotope dilution techniques. Amino acids and acylcarnitine measurements were made by flow injection tandem mass spectrometry using sample preparation methods described previously (*26*, *27*). Data were acquired using a Waters Acquity™ UPLC system equipped with a TQ (triple quadrupole) detector and a data system controlled by MassLynx 4.1 operating system (Waters, Milford, MA). Organic acids were quantified according to a previously published protocol (*28*) using Trace Ultra GC coupled to ISQ MS operating under Xcalibur 2.2 (Thermo Fisher Scientific, Austin, TX). Acyl CoAs were extracted and purified (*29-31*), and analyzed by flow injection analysis using positive electrospray ionization on a Xevo TQ-S triple quadrupole mass spectrometer (Waters, Milford, MA). Heptadecanoyl CoA was employed as an internal standard. Ceramides were extracted (*32*) and analyzed by flow injection tandem mass spectrometry using a Xevo TQS spectrometer (Waters Milford, MA) for precursors of m/z 264.

Culturing and characterization of age- and/or growth- discriminatory bacterial strains from the microbiota of healthy Malawian children

Culturing – Fecal samples were resuspended in 10ml of GMM under anaerobic conditions, diluted 1:1000, and then plated onto GMM and incubated at 37°C under anaerobic conditions overnight. Isolated colonies were picked into single wells of a deep 96-well plate, with 600µl liquid GMM per well, and grown anaerobically at 37°C overnight. Cultures were then mixed with an equal volume of pre-reduced GMM containing 30% glycerol (final concentration 15% glycerol) for storage at -80°C. To assign taxonomy, DNA was extracted from each well and V4-16S rRNA gene amplicons were generated and sequenced.

Taxonomies were confirmed by sequencing full-length 16S rRNA amplicons produced using primers 8F and 1391R.

Genome sequencing of two strains obtained from the healthy infant Malawian donor 8243C – DNA libraries for sequencing *R. gnavus* TS8243C and *C. symbiosum* TS8243C on Illumina MiSeq were prepared using a multistep protocol of end-repair, A-tailing, and ligation to customized sample-specific barcoded Illumina adapters. The resulting adapter-ligated DNA was size-selected (250-350bp) using agarose gel electrophoresis and then used as a template for enrichment PCR (cycling conditions: 98°C for 30 sec, followed by 18 cycles of 98°C for 10 sec, 65°C for 30 sec, 72°C for 45 sec, and then 72°C for 5 minutes). Multiplex sequencing was performed (paired-end 250nt reads). Genomes were assembled with the mira software package (*33*). For *R. gnavus* TS8243C, N50 contig length was 2,263,429bp with an aggregate genome size of 3.7Mbp. For *C. symbiosum* TS8243C, N50 contig length was 328,298bp with a 5.3Mbp genome size. Annotation was initially performed using PROKKA (*34*).

Genome-based reconstruction of selected metabolic subsystems in R. gnavus and C. symbiosum strains – Reconstruction of selected metabolic pathways (also termed as subsystems) was based on functional gene annotation and prediction using two comparative genomics techniques: (i) homology-based methods and (ii) genome context analysis. First, predicted proteomes were scanned against the KEGG Orthology, Pfam, CAZY, TCDB, Uniprot and SEED databases to identify proteins potentially involved in the carbohydrate utilization, amino acid metabolism and vitamin and cofactor biosynthesis and salvage pathways. The genomic and functional contexts of loci encoding the proteins were then analyzed and the respective metabolic pathways and transcriptional regulons were reconstructed. Additional closely related bacterial genomes available in the Integrated Microbial Genomes (*35*) and SEED databases (*36*) were used for comparative genomics-enabled pathway and regulon inference. Transcriptional regulons (sets of genes co-regulated via shared transcription factors (TFs) and transcription factor binding sites (TFBS) or shared RNA elements, riboswitches, responding to respective pathway metabolites) were predicted and reconstructed using the comparative genomics approach (*37*) implemented in the RegPrecise database (*38*). This integrative subsystems-based approach was used previously for reconstruction of carbohydrate utilization pathways and regulons in members of *Shewanella*, *Thermotoga*, and *Bacteroides* lineages (*39-41*).

Supplementary Text

Effects of gut microbiota immaturity on host metabolism

As noted in the main text, we cross-referenced the results of metabolic profiling conducted in mice treated with *R. gnavus* and *C. symbiosum* and untreated controls, with those obtained from mice harboring the same undernourished donor's microbiota and the healthy donor's microbiota used in the cohousing experiments. For the latter experiments, serum, liver, gastrocnemius muscle, and brain were obtained at the time of euthanasia three weeks after gavage of these microbial communities, and analyzed by targeted mass spectrometry (N = 5-6 mice/treatment group). The two donors' microbiota conferred distinct metabolic phenotypes with the major differences involving fatty acid and amino acid metabolism; mice colonized with the healthy donor's microbiota had significantly higher concentrations of long chain acylcarnitines in their serum and liver

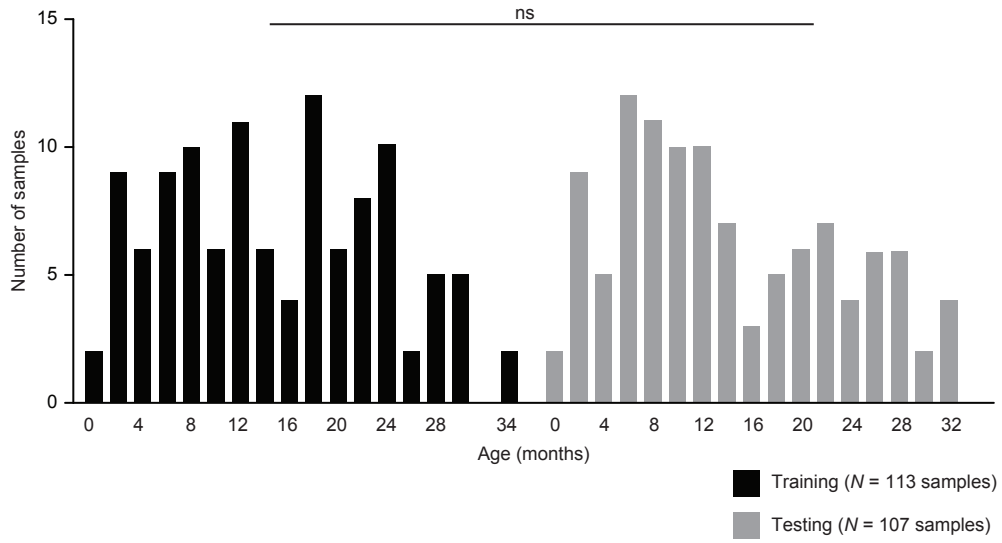
and significantly decreased concentrations of a number of amino acids in their gastrocnemius muscle (and cecal contents) ($P < 0.05$; Student's t-test; **Fig. S7, Table S16**). These data support the notion that mice colonized with the healthy donor microbiota funnel amino acids into new protein synthesis and lean body mass, whereas those colonized with the undernourished donor microbiota preserve the amino acid pools for oxidation and energy production. Consistent with this idea, increased levels of the branched-chain amino acids Val and Leu/Ile in the gastrocnemius muscle of mice colonized with undernourished donor microbiota were accompanied by an increase in a metabolite produced by their oxidation, C3 (propionyl) acylcarnitine (**Table S16**).

These differences in metabolic profiles extended to the brain (**Fig. S8**; **Table S16**). Echoing results from serum and liver, levels of long chain acylcarnitines were significantly elevated in the cerebral cortex of mice harboring the healthy donor microbiota compared to the undernourished donor microbiota ($P < 0.05$; Student's t-test). Amino acid metabolism was also impacted with concentrations of histidine, proline, glycine, arginine, and citrulline being significantly higher in the brains of mice with the healthy donor's microbiota. Additionally, there was evidence of perturbations in the TCA cycle, with citrate levels being significantly higher in animals harboring the undernourished infant's gut community.

There were strong correlations between the relative abundances of bacterial species in the fecal microbiota of gnotobiotic mice and serum C10, C12, and C14 acylcarnitine levels, as well as levels of amino acids in gastrocnemius muscle. Species enriched in the healthy community that significantly correlated to serum acylcarnitines and amino acids included taxa that were also weight- and/or lean mass gain-discriminatory: *R. gnavus* (C10:2, C12, C12-OH/C10-DC acylcarnitines; valine, leucine/isoleucine, and methionine), *Dorea formicigenerans* (C10, C10:1, C12, C12-OH/C10-DC , C14, C14:2 acylcarnitines; methionine), *F. prausnitzii* (C12-OH/C10-DC, C14 acylcarnitines, leucine/isoleucine), and *Clostridium symbiosum* (C10:1, C12, C12-OH/C10-DC, C14 acylcarnitines; and 11 amino acids including valine, leucine/isoleucine, and methionine) ($P < 0.05$, see **Figs. 2, S2, S9A,B**; **Table S17** for a complete list with Spearman's ranked correlation ρ-values). *B. longum* and *F. prausnitzii*, the two age- and growth-discriminatory species, were also correlated with hepatic levels of the TCA cycle intermediate citrate ($P < 0.05$; **Fig. S9C, Table S17**).

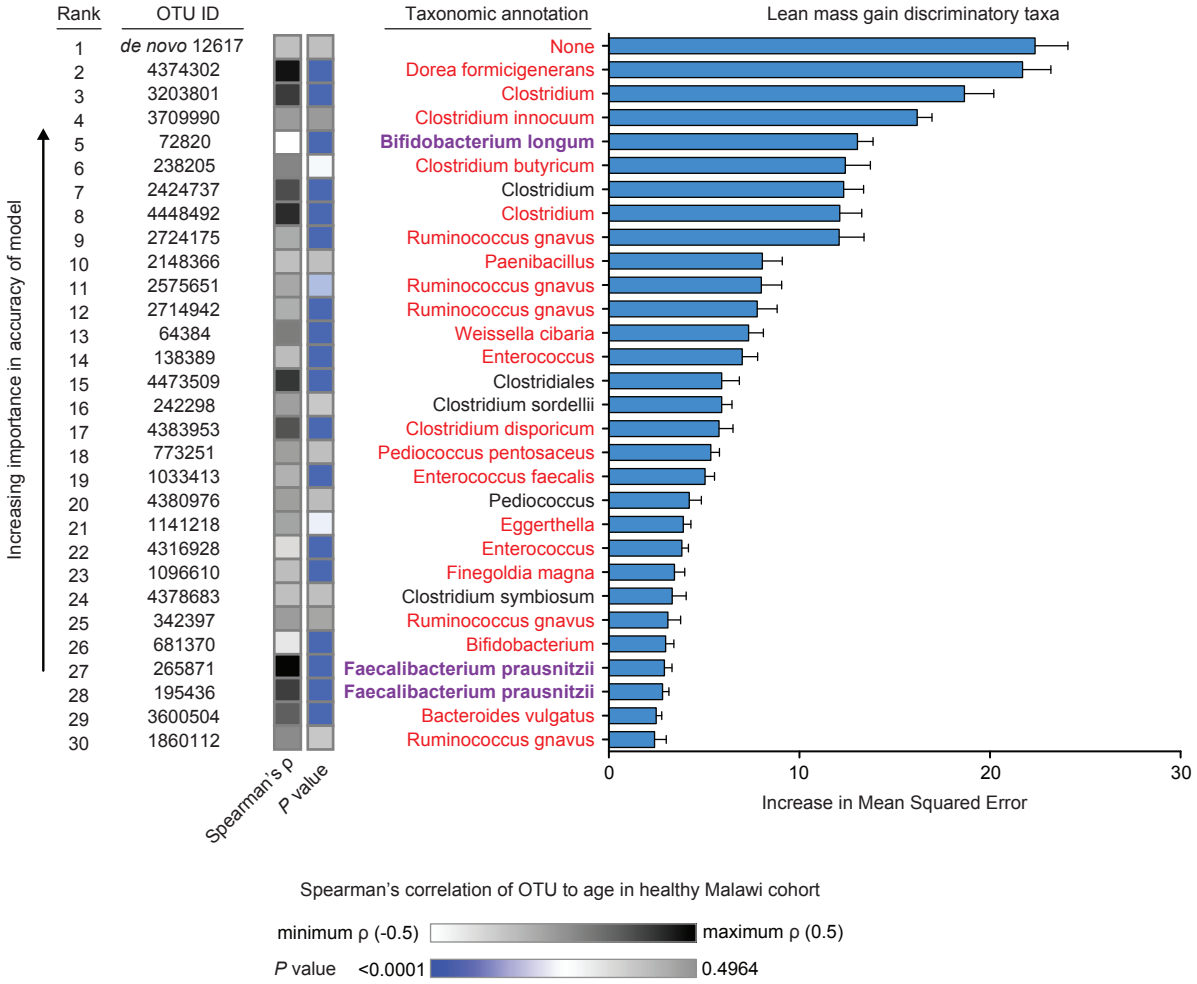
While these two donors' microbiota transmitted discordant weight, lean mass gain and metabolic phenotypes, they did not transmit discordant bone morphometric phenotypes ($P > 0.05$ on all metrics; Student's t-test). This implies, perhaps not surprisingly, that 'anthropometrically undernourished' and 'anthropometrically healthy' donor microbiota do not transmit discrete sets of effects, but may produce overlapping phenotypes, each component of which may vary in its degree of discordance from donor to donor. A corollary is that different host phenotypes need to be assayed to obtain a more robust functional classification of 'healthy' versus 'undernourished' microbiota.

Fig. S1.



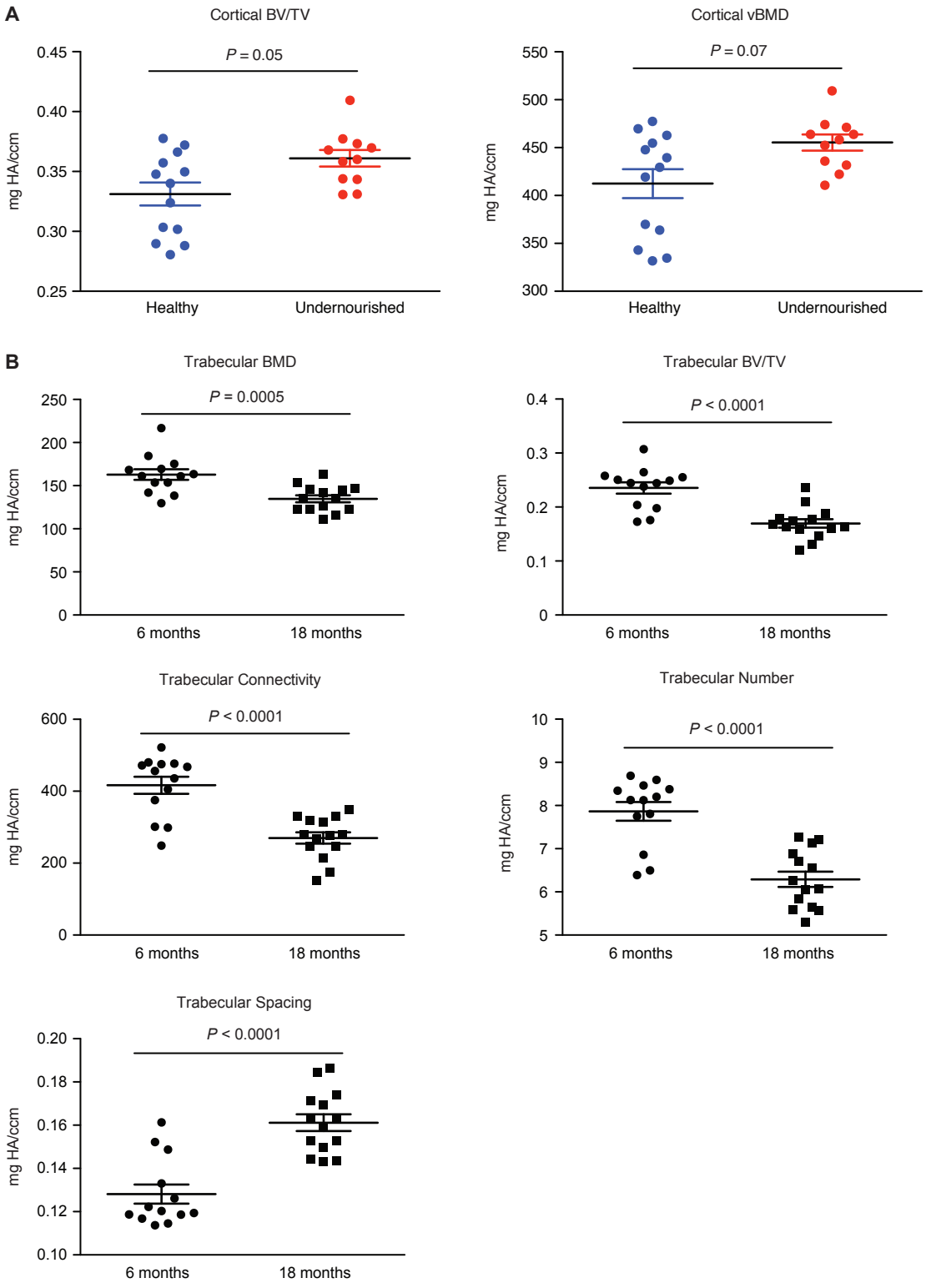
Distribution of ages at the time of sample collection of healthy microbiota donors that were used to generate the Random Forests-derived model of age-discriminatory bacterial 97%ID OTUs. Histogram of donor ages in the training set cohort (N = 113 samples from 14 twin pairs and 1 set of triplets; minimum age 0.6 months, maximum age 33.51 months, median age 13.6 months) and testing set cohort (N = 107 samples from 13 twin pairs and 1 set of triplets; minimum age 0.98 months, maximum age 31.8 months, median age 12.02 months). ns, no significant difference based on Kolmogorov-Smirnov test (P = 0.58).

Fig. S2.



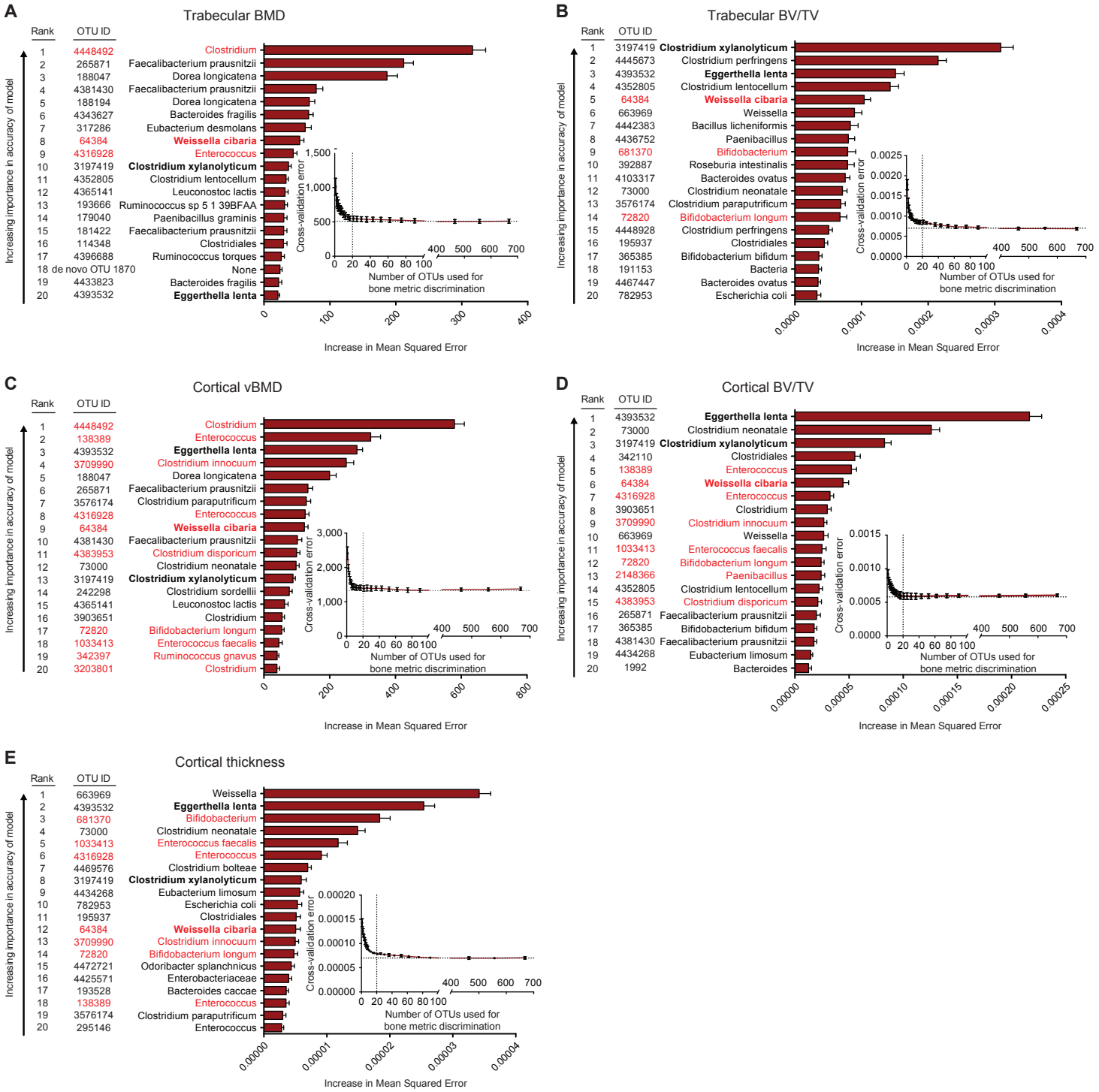
Random Forests-derived model of lean body mass gain discriminatory OTUs. The 30 most lean mass gain discriminatory OTUs with their taxonomic assignments, ranked by feature importance as determined by increase in mean squared error. The lean body mass gain model explained ~58% of the observed phenotypic variation ($P < 0.0001$, permutation test; 999 permutations). Models were iterated 100 times, and average mean squared error values were used to generate the final rankings. Red indicates OTUs that appear within the 30 most discriminatory OTUs for both the weight and lean mass gain models. Purple indicates species that also appear in the 25-taxa sparse model of Malawian microbiota maturation. Colors to the right of the OTU ID numbers represent Spearman's rank correlation of the same OTU ID to chronological age in the healthy Malawian infant/child cohort (**Table S9)**.

Fig. S3.


A
Cortical BV/TV
P = 0.05
mg HA/ccm
0.45
0.40
0.35
0.30
0.25
Healthy
Undernourished
Cortical vBMD
P = 0.07
mg HA/ccm
550
500
450
400
350
300
Healthy
Undernourished
B
Trabecular BMD
P = 0.0005
mg HA/ccm
250
200
150
100
50
0
6 months
18 months
Trabecular BV/TV
P < 0.0001
mg HA/ccm
0.4
0.3
0.2
0.1
0
6 months
18 months
Trabecular Connectivity
P < 0.0001
mg HA/ccm
600
400
200
0
6 months
18 months
Trabecular Number
P < 0.0001
mg HA/ccm
10
9
8
7
6
5
6 months
18 months
Trabecular Spacing
P < 0.0001
mg HA/ccm
0.20
0.18
0.16
0.14
0.12
0.10
6 months
18 months

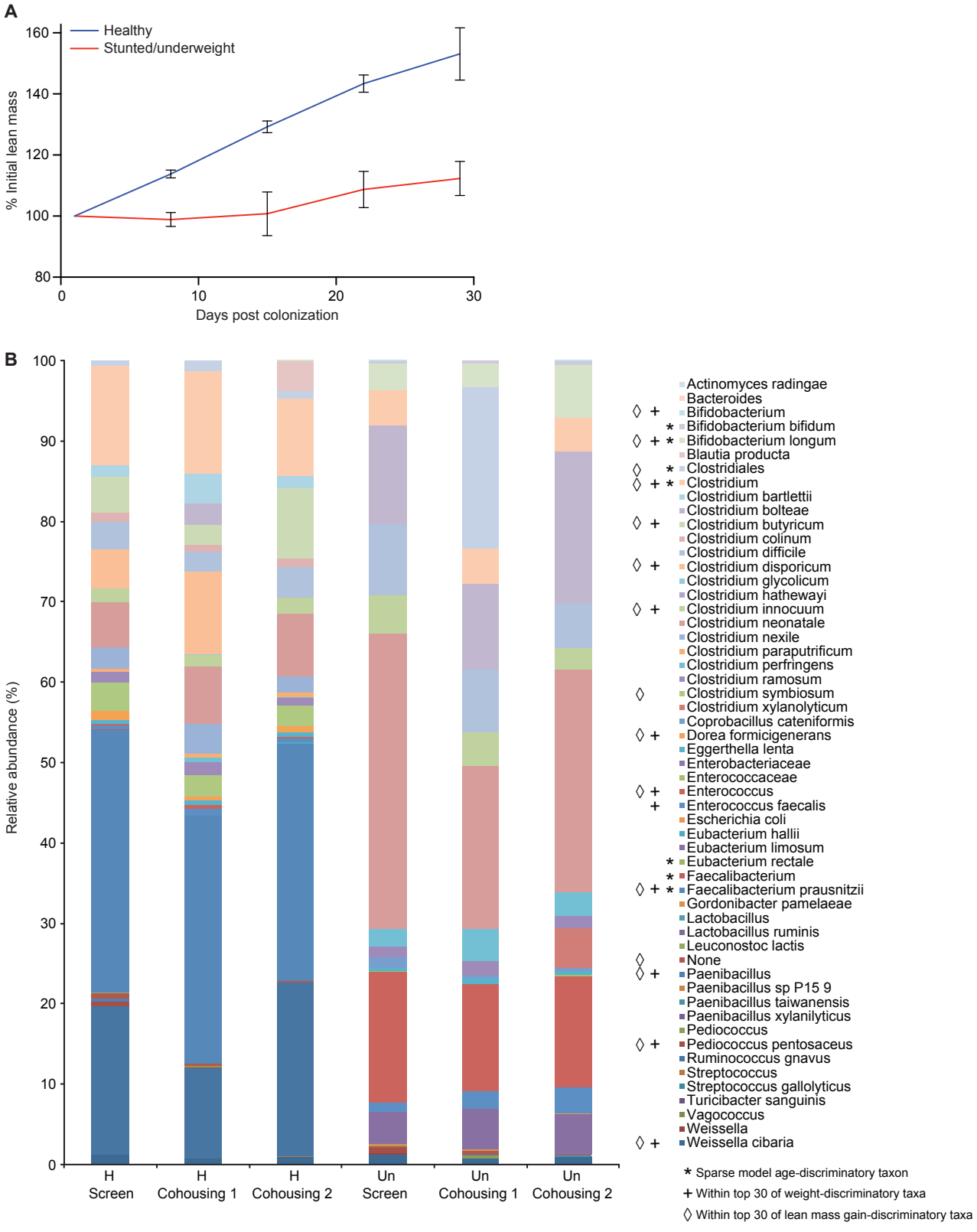
Effects of microbiota donor health status and age on femoral bone morphology. Micro-CT was performed on femurs harvested from gnotobiotic mice colonized with microbiota from 6- and 18-month-old healthy or undernourished donors (in all cases, the efficiency of transplantation was >50%). Each circle represents a single mouse. **(A)** Bone morphometric phenotypes expressed as a function of donor nutritional status irrespective of donor age. Differences in the ratio of bone volume to tissue volume (BV/TV) and volumetric bone mineral density (vBMD) in the cortical region of femurs are shown. **(B)** Bone morphometric phenotypes expressed as a function of age, irrespective of donor nutritional status. Differences in trabecular bone mineral density (BMD), BV/TV, connectivity, number, and spacing are shown. *P* values are based on a Mann-Whitney test. Y-axis abbreviation; HA/ccm, hydroxyapatite per cubic centimeter.

Fig. S4.


A
Trabecular BMD
Rank
OTU ID
1 4448492 Clostridium
2 265871 Faecalibacterium prausnitzii
3 188047 Dorea longicatena
4 4381430 Faecalibacterium prausnitzii
5 188194 Dorea longicatena
6 4343627 Bacteroides fragilis
7 317286 Eubacterium desmolans
8 64384 Weissella cibaria
9 4316928 Enterococcus
10 3197419 Clostridium xylanolyticum
11 4352805 Clostridium lentocellum
12 4365141 Leuconostoc lactis
13 193666 Ruminococcus sp 5 1 39BFAA
14 179040 Paenibacillus graminis
15 181422 Faecalibacterium prausnitzii
16 114348 Clostridiales
17 4396688 Ruminococcus torques
18 de novo OTU 1870 None
19 4433823 Bacteroides fragilis
20 4393532 Eggerthella lenta
Increasing importance in accuracy of model
Cross-validation error
Number of OTUs used for bone metric discrimination
Increase in Mean Squared Error
B
Trabecular BV/TV
1 3197419 Clostridium xylanolyticum
2 4445673 Clostridium perfringens
3 4393532 Eggerthella lenta
4 4352805 Clostridium lentocellum
5 64384 Weissella cibaria
6 663969 Weissella
7 4442383 Bacillus licheniformis
8 4436752 Paenibacillus
9 681370 Bifidobacterium
10 392887 Roseburia intestinalis
11 4103317 Bacteroides ovatus
12 73000 Clostridium neonatale
13 3576174 Clostridium paraputrificum
14 72820 Bifidobacterium longum
15 4448928 Clostridium perfringens
16 195937 Clostridiales
17 365385 Bifidobacterium bifidum
18 191153 Bacteria
19 4467447 Bacteroides ovatus
20 782953 Escherichia coli
C
Cortical vBMD
1 4448492 Clostridium
2 138389 Enterococcus
3 4393532 Eggerthella lenta
4 3709990 Clostridium innocuum
5 188047 Dorea longicatena
6 265871 Faecalibacterium prausnitzii
7 3576174 Clostridium paraputrificum
8 4316928 Enterococcus
9 64384 Weissella cibaria
10 4381430 Faecalibacterium prausnitzii
11 4383953 Clostridium disporicum
12 73000 Clostridium neonatale
13 3197419 Clostridium xylanolyticum
14 242298 Clostridium sordellii
15 4365141 Leuconostoc lactis
16 3903651 Clostridium
17 72820 Bifidobacterium longum
18 1033413 Enterococcus faecalis
19 342397 Ruminococcus gnavus
20 3203801 Clostridium
D
Cortical BV/TV
1 4393532 Eggerthella lenta
2 73000 Clostridium neonatale
3 3197419 Clostridium xylanolyticum
4 342110 Clostridiales
5 138389 Enterococcus
6 64384 Weissella cibaria
7 4316928 Enterococcus
8 3903651 Clostridium
9 3709990 Clostridium innocuum
10 663969 Weissella
11 1033413 Enterococcus faecalis
12 72820 Bifidobacterium longum
13 2148366 Paenibacillus
14 4352805 Clostridium lentocellum
15 4383953 Clostridium disporicum
16 265871 Faecalibacterium prausnitzii
17 365385 Bifidobacterium bifidum
18 4381430 Faecalibacterium prausnitzii
19 4434268 Eubacterium limosum
20 1992 Bacteroides
E
Cortical thickness
1 663969 Weissella
2 4393532 Eggerthella lenta
3 681370 Bifidobacterium
4 73000 Clostridium neonatale
5 1033413 Enterococcus faecalis
6 4316928 Enterococcus
7 4469576 Clostridium bolteae
8 3197419 Clostridium xylanolyticum
9 4434268 Eubacterium limosum
10 782953 Escherichia coli
11 195937 Clostridiales
12 64384 Weissella cibaria
13 3709990 Clostridium innocuum
14 72820 Bifidobacterium longum
15 4472721 Odoribacter splanchnicus
16 4425571 Enterobacteriaceae
17 193528 Bacteroides caccae
18 138389 Enterococcus
19 3576174 Clostridium paraputrificum
20 295146 Enterococcus

Random Forests-derived models of OTUs that discriminate femoral bone metrics defined by micro-CT. Top 20 most discriminatory OTUs, ranked by increase in mean squared error, for **(A)** trabecular bone mineral density, **(B)** trabecular bone volume per tissue volume, **(C)** cortical volumetric bone mineral density, **(D)** cortical bone volume per tissue volume, and **(E)** cortical thickness. Insets show cross-validation of each model; error decreases as OTUs are added to the model. Red indicates that the OTU is also among the top 30 ranked OTUs the Random Forests-derived models of both weight- and lean mass gain-discriminatory taxa. Boldface indicates that the OTU appears in all five bone metric models.

Fig. S5.


A
Healthy
Stunted/underweight
% Initial lean mass
160
140
120
100
80
0
10
20
30
Days post colonization
B
Relative abundance (%)
100
90
80
70
60
50
40
30
20
10
0
H Screen
H Cohousing 1
H Cohousing 2
Un Screen
Un Cohousing 1
Un Cohousing 2
Actinomyces radingae
Bacteroides
◊ + Bifidobacterium
* Bifidobacterium bifidum
◊ + * Bifidobacterium longum
Blautia producta
◊ * Clostridiales
◊ + * Clostridium
Clostridium bartlettii
Clostridium bolteae
◊ + Clostridium butyricum
Clostridium colinum
Clostridium difficile
◊ + Clostridium disporicum
Clostridium glycolicum
Clostridium hathewayi
◊ + Clostridium innocuum
Clostridium neonatale
Clostridium nexile
Clostridium paraputrificum
Clostridium perfringens
Clostridium ramosum
◊ Clostridium symbiosum
Clostridium xylanolyticum
Coprobacillus cateniformis
◊ + Dorea formicigenerans
Eggerthella lenta
Enterobacteriaceae
Enterococcaceae
◊ + Enterococcus
+ Enterococcus faecalis
Escherichia coli
Eubacterium hallii
Eubacterium limosum
* Eubacterium rectale
* Faecalibacterium
◊ + * Faecalibacterium prausnitzii
Gordonibacter pamelaeae
Lactobacillus
Lactobacillus ruminis
Leuconostoc lactis
◊ None
◊ + Paenibacillus
Paenibacillus sp P15 9
Paenibacillus taiwanensis
Paenibacillus xylanilyticus
Pediococcus
◊ + Pediococcus pentosaceus
Ruminococcus gnavus
Streptococcus
Streptococcus gallolyticus
Turicibacter sanguinis
Vagococcus
Weissella
◊ + Weissella cibaria
* Sparse model age-discriminatory taxon
+ Within top 30 of weight-discriminatory taxa
◊ Within top 30 of lean mass gain-discriminatory taxa

Discordant growth phenotypes transmitted by two exemplary iLiNS-DYAD-M donor gut communities: one from a healthy 6-month-old donor with normal gut microbiota maturity and the other from a severely stunted/underweight 6-month-old donor with significant microbiota immaturity. (A) Gain of lean body mass ($N = 5$ recipients/donor microbiota). Mean±SEM values are shown. $P < 0.0001$ (2-way ANOVA). **(B)** Bacterial composition of the fecal microbiota of gnotobiotic mice in the cohousing experiments (relative abundances, averaged over all time points, of taxa from all animals in each treatment group). Abbreviations: 'screen', initial survey of mice harboring donor microbiota; 'H', cagemates colonized with the healthy donor's microbiota in each cohousing experiment; 'Un', cagemates colonized with the severely stunted/underweight donor's microbiota.

Fig. S6A.

Central carbohydrate metabolism and fermentation

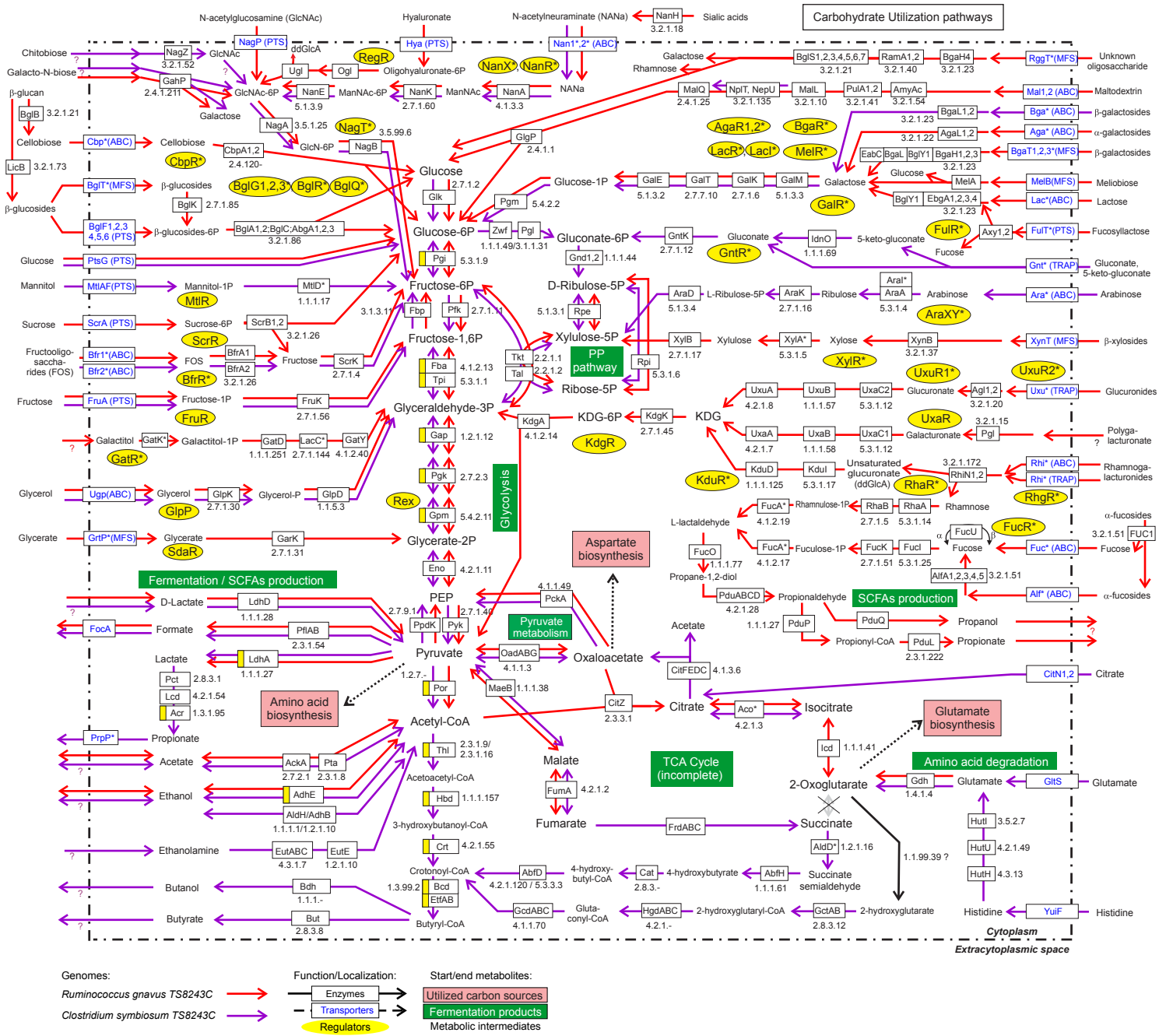


Fig. S6B.

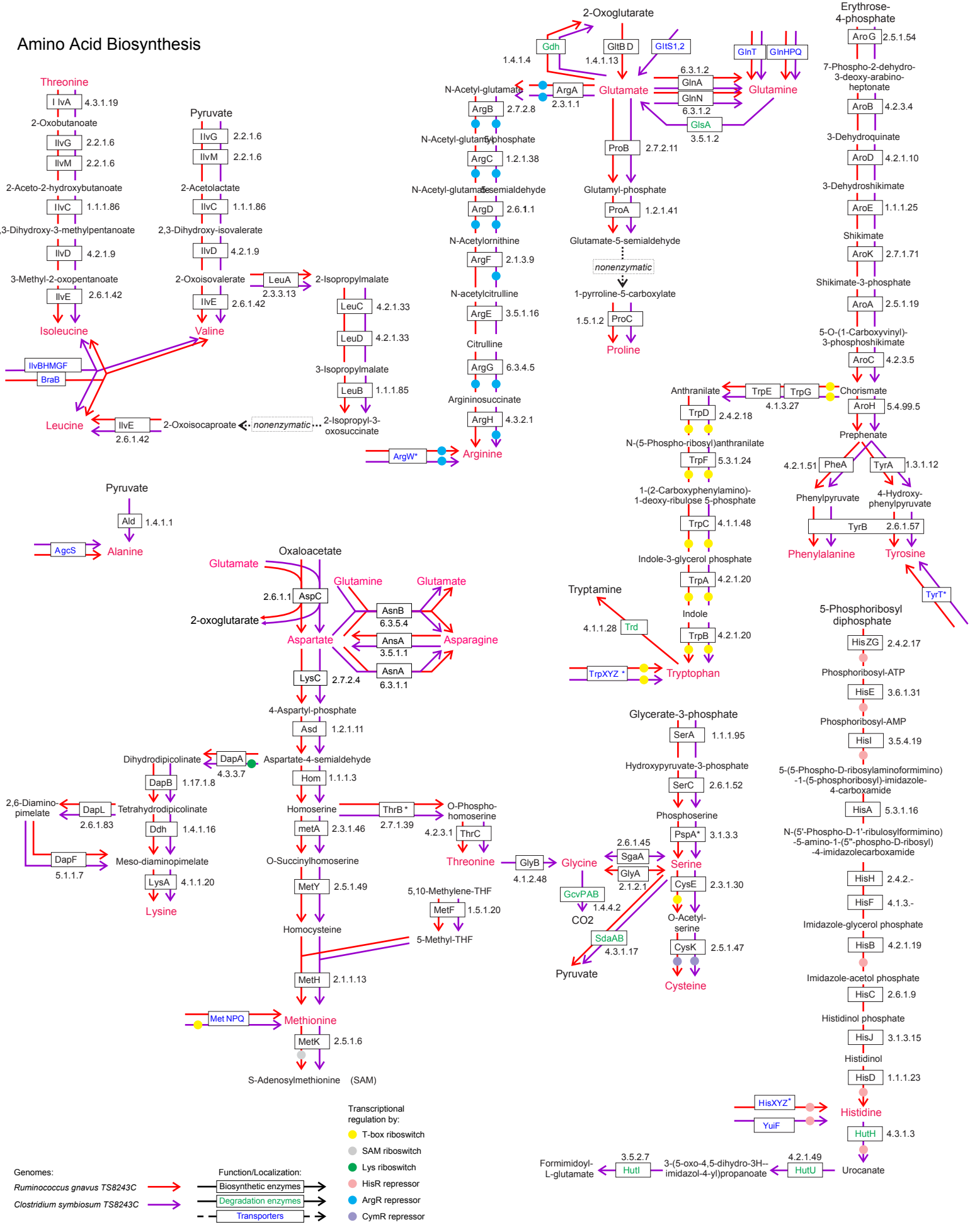
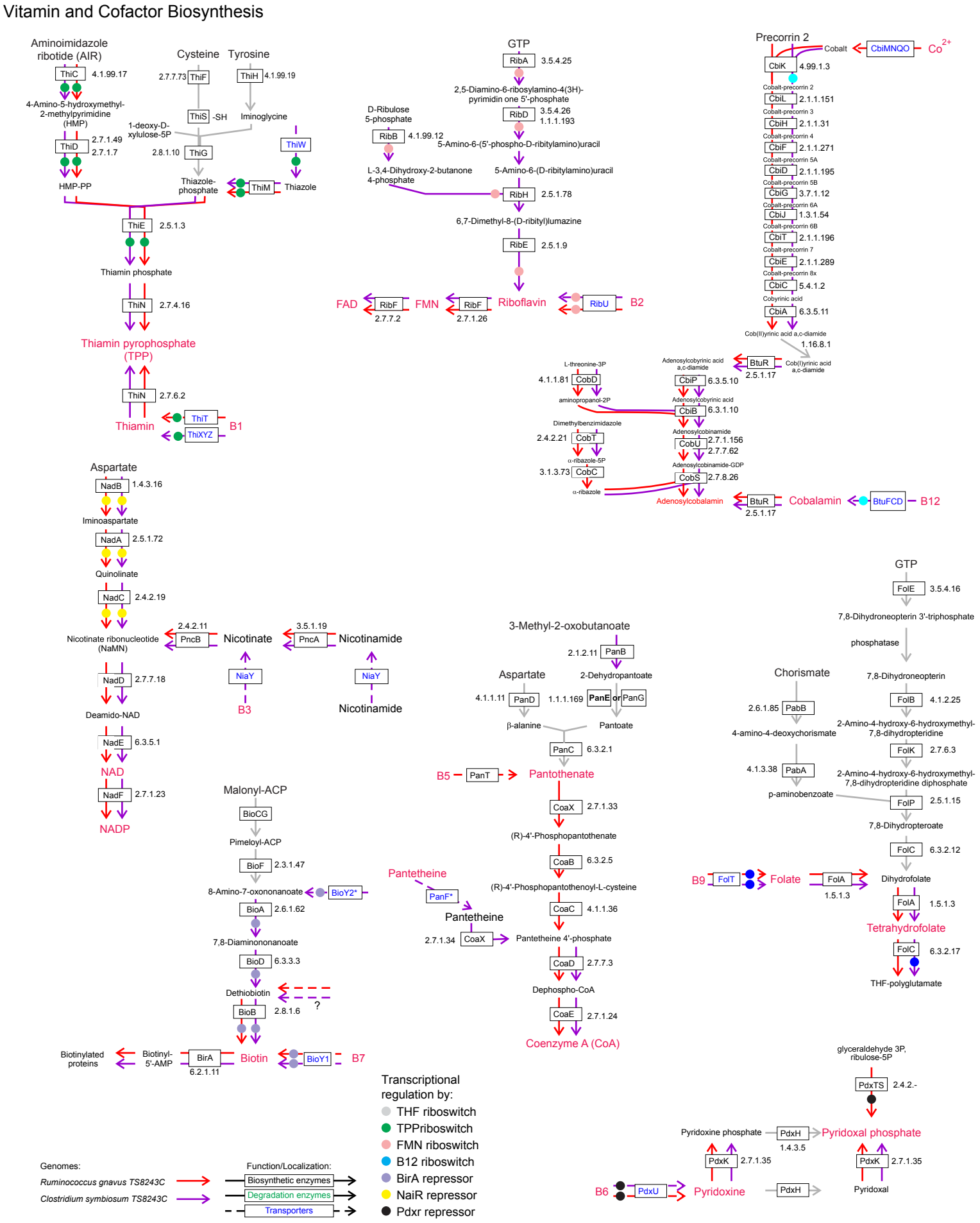

Amino Acid Biosynthesis
Threonine
I lvA 4.3.1.19
2-Oxobutanoate
IlvG 2.2.1.6
IlvM 2.2.1.6
2-Aceto-2-hydroxybutanoate
IlvC 1.1.1.86
2,3-Dihydroxy-3-methylpentanoate
IlvD 4.2.1.9
3-Methyl-2-oxopentanoate
IlvE 2.6.1.42
Isoleucine
Pyruvate
IlvG 2.2.1.6
IlvM 2.2.1.6
2-Acetolactate
IlvC 1.1.1.86
2,3-Dihydroxy-isovalerate
IlvD 4.2.1.9
2-Oxoisovalerate
IlvE 2.6.1.42
Valine
LeuA 2.3.3.13
2-Isopropylmalate
LeuC 4.2.1.33
LeuD 4.2.1.33
3-Isopropylmalate
LeuB 1.1.1.85
2-Isopropyl-3-oxosuccinate
IlvBHMGF
BraB
Leucine
IlvE 2.6.1.42
2-Oxoisocaproate
nonenzymatic
2-Oxoglutarate
Gdh 1.4.1.4
GltB D 1.4.1.13
GltS1,2
Glutamate
GlnT
GlnHPQ
Glutamine
6.3.1.2
GlnA
GlnN
6.3.1.2
GlsA 3.5.1.2
N-Acetyl-glutamate
ArgA 2.3.1.1
ArgB 2.7.2.8
N-Acetyl-glutamyl-phosphate
ArgC 1.2.1.38
N-Acetyl-glutamate-5-semialdehyde
ArgD 2.6.1.11
N-Acetylornithine
ArgF 2.1.3.9
N-acetylcitrulline
ArgE 3.5.1.16
Citrulline
ArgG 6.3.4.5
Argininosuccinate
ArgH 4.3.2.1
Arginine
ArgW*
ProB 2.7.2.11
Glutamyl-phosphate
ProA 1.2.1.41
Glutamate-5-semialdehyde
nonenzymatic
1-pyrroline-5-carboxylate
1.5.1.2 ProC
Proline
Erythrose-4-phosphate
AroG 2.5.1.54
7-Phospho-2-dehydro-3-deoxy-arabino-heptonate
AroB 4.2.3.4
3-Dehydroquinate
AroD 4.2.1.10
3-Dehydroshikimate
AroE 1.1.1.25
Shikimate
AroK 2.7.1.71
Shikimate-3-phosphate
AroA 2.5.1.19
5-O-(1-Carboxyvinyl)-3-phosphoshikimate
AroC 4.2.3.5
Chorismate
AroH 5.4.99.5
Prephenate
4.2.1.51 PheA
TyrA 1.3.1.12
Phenylpyruvate
4-Hydroxy-phenylpyruvate
TyrB 2.6.1.57
Phenylalanine
Tyrosine
TyrT*
Anthranilate
TrpE
TrpG
4.1.3.27
TrpD 2.4.2.18
N-(5-Phospho-ribosyl)anthranilate
TrpF 5.3.1.24
1-(2-Carboxyphenylamino)-1-deoxy-ribulose 5-phosphate
TrpC 4.1.1.48
Indole-3-glycerol phosphate
TrpA 4.2.1.20
Indole
TrpB 4.2.1.20
Tryptophan
TrpXYZ *
Tryptamine
4.1.1.28 Trd
Pyruvate
Ald 1.4.1.1
AgcS
Alanine
Oxaloacetate
Glutamate
2.6.1.1 AspC
2-oxoglutarate
Aspartate
Glutamine
AsnB 6.3.5.4
Asparagine
Glutamate
AnsA 3.5.1.1
AsnA 6.3.1.1
LysC 2.7.2.4
4-Aspartyl-phosphate
Asd 1.2.1.11
Aspartate-4-semialdehyde
DapA 4.3.3.7
Dihydrodipicolinate
DapB 1.17.1.8
Tetrahydrodipicolinate
DapL 2.6.1.83
2,6-Diamino-pimelate
Ddh 1.4.1.16
DapF 5.1.1.7
Meso-diaminopimelate
LysA 4.1.1.20
Lysine
Hom 1.1.1.3
Homoserine
ThrB * 2.7.1.39
O-Phospho-homoserine
metA 2.3.1.46
4.2.3.1 ThrC
O-Succinylhomoserine
Threonine
GlyB 4.1.2.48
Glycine
MetY 2.5.1.49
Homocysteine
5,10-Methylene-THF
MetF 1.5.1.20
5-Methyl-THF
MetH 2.1.1.13
Met NPQ
Methionine
MetK 2.5.1.6
S-Adenosylmethionine (SAM)
Glycerate-3-phosphate
SerA 1.1.1.95
Hydroxypyruvate-3-phosphate
SerC 2.6.1.52
Phosphoserine
PspA* 3.1.3.3
Serine
2.6.1.45 SgaA
GlyA 2.1.2.1
CysE 2.3.1.30
O-Acetyl-serine
CysK 2.5.1.47
Cysteine
GcvPAB 1.4.4.2
CO2
SdaAB 4.3.1.17
Pyruvate
5-Phosphoribosyl diphosphate
HisZG 2.4.2.17
Phosphoribosyl-ATP
HisE 3.6.1.31
Phosphoribosyl-AMP
HisI 3.5.4.19
5-(5-Phospho-D-ribosylaminoformimino)-1-(5-phosphoribosyl)-imidazole-4-carboxamide
HisA 5.3.1.16
N-(5'-Phospho-D-1'-ribulosylformimino)-5-amino-1-(5"-phospho-D-ribosyl)-4-imidazolecarboxamide
HisH 2.4.2.-
HisF 4.1.3.-
Imidazole-glycerol phosphate
HisB 4.2.1.19
Imidazole-acetol phosphate
HisC 2.6.1.9
Histidinol phosphate
HisJ 3.1.3.15
Histidinol
HisD 1.1.1.23
Histidine
HisXYZ*
YuiF
HutH 4.3.1.3
Urocanate
4.2.1.49 HutU
3-(5-oxo-4,5-dihydro-3H-imidazol-4-yl)propanoate
3.5.2.7 HutI
Formimidoyl-L-glutamate
Transcriptional regulation by:
T-box riboswitch
SAM riboswitch
Lys riboswitch
HisR repressor
ArgR repressor
CymR repressor
Genomes:
Ruminococcus gnavus TS8243C
Clostridium symbiosum TS8243C
Function/Localization:
Biosynthetic enzymes
Degradation enzymes
Transporters

Fig. S6C.


Vitamin and Cofactor Biosynthesis
Aminoimidazole ribotide (AIR)
Cysteine
Tyrosine
ThiC 4.1.99.17
4-Amino-5-hydroxymethyl-2-methylpyrimidine (HMP)
ThiD 2.7.1.49 2.7.1.7
HMP-PP
ThiF 2.7.7.73
ThiS -SH
1-deoxy-D-xylulose-5P
ThiG 2.8.1.10
Thiazole-phosphate
ThiH 4.1.99.19
Iminoglycine
ThiW
ThiM
Thiazole
ThiE 2.5.1.3
Thiamin phosphate
ThiN 2.7.4.16
Thiamin pyrophosphate (TPP)
ThiN 2.7.6.2
Thiamin
ThiT
ThiXYZ
B1
GTP
RibA 3.5.4.25
2,5-Diamino-6-ribosylamino-4(3H)-pyrimidin one 5'-phosphate
RibD 3.5.4.26 1.1.1.193
D-Ribulose 5-phosphate
RibB 4.1.99.12
L-3,4-Dihydroxy-2-butanone 4-phosphate
5-Amino-6-(5'-phospho-D-ribitylamino)uracil
5-Amino-6-(D-ribitylamino)uracil
RibH 2.5.1.78
6,7-Dimethyl-8-(D-ribityl)lumazine
RibE 2.5.1.9
Riboflavin
RibF 2.7.7.2
FAD
RibF 2.7.1.26
FMN
RibU
B2
Precorrin 2
Cobalt
CbiMNQO
Co2+
CbiK 4.99.1.3
Cobalt-precorrin 2
CbiL 2.1.1.151
Cobalt-precorrin 3
CbiH 2.1.1.31
Cobalt-precorrin 4
CbiF 2.1.1.271
Cobalt-precorrin 5A
CbiD 2.1.1.195
Cobalt-precorrin 5B
CbiG 3.7.1.12
Cobalt-precorrin 6A
CbiJ 1.3.1.54
Cobalt-precorrin 6B
CbiT 2.1.1.196
Cobalt-precorrin 7
CbiE 2.1.1.289
Cobalt-precorrin 8x
CbiC 5.4.1.2
Cobyrinic acid
CbiA 6.3.5.11
Cob(II)yrinic acid a,c-diamide
1.16.8.1
L-threonine-3P
4.1.1.81 CobD
aminopropanol-2P
Dimethylbenzimidazole
2.4.2.21 CobT
α-ribazole-5P
3.1.3.73 CobC
α-ribazole
Adenosylcobyrinic acid a,c-diamide
CbiP 6.3.5.10
Adenosylcobyrinic acid
CbiB 6.3.1.10
Adenosylcobinamide
CobU 2.7.1.156 2.7.7.62
Adenosylcobinamide-GDP
CobS 2.7.8.26
Adenosylcobalamin
BtuR 2.5.1.17
Cob(I)yrinic acid a,c-diamide
BtuR 2.5.1.17
Cobalamin
BtuFCD
B12
Aspartate
NadB 1.4.3.16
Iminoaspartate
NadA 2.5.1.72
Quinolinate
NadC 2.4.2.19
Nicotinate ribonucleotide (NaMN)
NadD 2.7.7.18
Deamido-NAD
NadE 6.3.5.1
NAD
NadF 2.7.1.23
NADP
PncB 2.4.2.11
Nicotinate
NiaY
B3
PncA 3.5.1.19
Nicotinamide
NiaY
Nicotinamide
3-Methyl-2-oxobutanoate
PanB 2.1.2.11
2-Dehydropantoate
Aspartate
PanD 4.1.1.11
PanE or PanG 1.1.1.169
β-alanine
Pantoate
PanC 6.3.2.1
B5
PanT
Pantothenate
CoaX 2.7.1.33
(R)-4'-Phosphopantothenate
CoaB 6.3.2.5
(R)-4'-Phosphopantothenoyl-L-cysteine
CoaC 4.1.1.36
Pantetheine 4'-phosphate
CoaD 2.7.7.3
Dephospho-CoA
CoaE 2.7.1.24
Coenzyme A (CoA)
Pantetheine
PanF*
Pantetheine
2.7.1.34 CoaX
Malonyl-ACP
BioCG
Pimeloyl-ACP
BioF 2.3.1.47
8-Amino-7-oxononanoate
BioY2*
BioA 2.6.1.62
7,8-Diaminononanoate
BioD 6.3.3.3
Dethiobiotin
BioB 2.8.1.6
?
Biotinylated proteins
Biotinyl-5'-AMP
BirA 6.2.1.11
Biotin
BioY1
B7
GTP
FolE 3.5.4.16
7,8-Dihydroneopterin 3'-triphosphate
phosphatase
7,8-Dihydroneopterin
FolB 4.1.2.25
2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine
FolK 2.7.6.3
2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine diphosphate
FolP 2.5.1.15
7,8-Dihydropteroate
FolC 6.3.2.12
Dihydrofolate
FolA 1.5.1.3
Tetrahydrofolate
FolC 6.3.2.17
THF-polyglutamate
Chorismate
PabB 2.6.1.85
4-amino-4-deoxychorismate
PabA 4.1.3.38
p-aminobenzoate
B9
FolT
Folate
FolA 1.5.1.3
glyceraldehyde 3P, ribulose-5P
PdxTS 2.4.2.-
Pyridoxine phosphate
PdxH 1.4.3.5
Pyridoxal phosphate
PdxK 2.7.1.35
PdxK 2.7.1.35
B6
PdxU
Pyridoxine
PdxH
Pyridoxal
Transcriptional regulation by:
THF riboswitch
TPPriboswitch
FMN riboswitch
B12 riboswitch
BirA repressor
NaiR repressor
Pdxr repressor
Genomes:
Ruminococcus gnavus TS8243C
Clostridium symbiosum TS8243C
Function/Localization:
Biosynthetic enzymes
Degradation enzymes
Transporters

Reconstructed *Ruminococcus gnavus* TS8243C and *Clostridium symbiosum* TS8243C metabolic pathways involved in carbohydrate and central carbon metabolism, amino acid biosynthesis and degradation, and vitamin and cofactor biosynthesis. Enzymatic reactions and transporters are depicted by solid and dashed lines, respectively. Red and purple lines indicate the presence of the enzyme or transporter in the cultured strains of *R. gnavus and C. symbiosum*, respectively, while gray lines indicate absence of the enzyme or transporter in either strain. Common names of enzymes and transporters are shown in black and blue font, respectively. Enzymes are also identified by their enzyme commission (EC) numbers. Major families of carbohydrate transporters indicated in parentheses are ATP-Binding Cassette (ABC) transporters, Major Facilitator Superfamily (MFS) permeases, TRipartite ATP-independent Periplasmic (TRAP) transporters, and phosphoenolpyruvate-dependent PhosphoTransferase System (PTS) transporters. Cognate transcriptional regulators of sugar utilization pathways are highlighted in yellow ovals. Central carbon metabolism genes controlled by the global redox-responsive regulator Rex in both genomes are marked by yellow vertical stripes. Genes involved in amino acid and vitamin metabolism that are controlled by local regulators and metabolite responsive riboswitches are denoted by colored circles. Gene locus tags and functional annotations are listed in **Table S13.** Genes with novel predicted functions are marked with an asterisk. Putative pathways and transporters for which no gene annotation could be identified are indicated with a question mark.

(A) Carbon and energy: central carbon metabolism, carbohydrate utilization, and short chain fatty acid production

Central carbon metabolism – Both the *R. gnavus* and *C. symbiosum* strains possess the complete glycolysis/gluconeogenesis and pentose phosphate (PP) pathways, as well as the key enzymes involved in conversion of pyruvate to acetyl-CoA or oxaloacetate, and reverse conversion of oxaloacetate to phosphoenolpyruvate (PEP). The PP pathway in *C. symbiosum* includes multiple paralogs of genes encoding transketolase (TktAB) and transaldolase (Tal). In addition, *C. symbiosum* (but not *R. gnavus*) encodes the oxidative pathway for utilization of glucose-6 phosphate via the PP pathway (Zwf, Pgl, Gnd). The tricarboxylic acid (TCA) cycle is incomplete in both strains. In *R. gnavus*, it involves two parts: (i) interconversion of pyruvate to malate and fumarate, and (ii) fusion of oxaloacetate and acetyl-CoA to produce 2-oxoglutarate, a key precursor for glutamate biosynthesis. *C. symbiosum* possesses only the former pathway for conversion of pyruvate to malate and fumarate.

Carbohydrate utilization (CU) pathways – The reconstructed carbohydrate utilization (CU) potential of the two Malawian TS8243C strains differ substantially and include novel carbohydrate-specific regulators, transporters, and enzymes that are not orthologous to previously characterized proteins. *R. gnavus* has 28 predicted novel regulators, most of which belong to the AraC family, and 35 transporter genes encoding 16 transporter complexes (including 7 ABC transporters). *C. symbiosum* contains 5 novel regulators from various structural families and 17 transporter genes (5 ABC transporter complexes). Notable novel monosaccharide catabolizing enzymes include xylose isomerase (XylA), tagatose-6P isomerase (LacC), galactitol kinase (GatK), and bifunctional fuculose-P/rhamnulose-P aldolase (FucA) in *R. gnavus*, as well as arabinose isomerase (AraI) and mannitol dehydrogenase (MtlD) in *C. symbiosum*.

R. gnavus has predicted pathways for utilization of glucose, fructose and its oligosaccharides, sucrose, lactose, galactosides, fucose and fucosides, N-acetylglucosamine, galacto-N-biose, N-acetylneuraminate, sialic acids, beta-xylosides, rhamnogalacturonides, glucuronides, polygalacturonate, maltodextrins, beta-glucosides and cellobiose, meliobiose, fucosyllactose, hyaluronate and galactitol. *C. symbiosum* is predicted to utilize glucose, fructose, beta-galactosides, N-acetylneuraminate, chitobiose, galacto-N-biose, arabinose, mannitol and gluconate.

Utilization of other carbon sources – R. gnavus and *C. symbiosum* have shared and unique pathways for utilization of other carbon sources. Both strains share the glycerol and D-lactate utilization pathways. *C. symbiosum* has unique pathways for utilization of ethanolamine and citrate. *R. gnavus* has a unique pathway for utilization of D-glycerate.

Short-chain fatty acid (SCFA) production – R. gnavus and *C. symbiosum* possess both shared and strain-specific pathways for fermentation of various carbon sources and production of SCFAs. Both strains have complete pathways for production of formate, lactate, acetate and ethanol from pyruvate. *R. gnavus* has a unique pathway for production of propionate/propanol by fermentation of fucose and rhamnose. This pathway involves lactaldehyde reductase (FucO) and B_{12}-dependent propanediol dehydratase (PduABCD), plus several other downstream enzymes for conversion of propionaldehyde to propionate and propanol. *C. symbiosum* has the complete pathway for conversion of acetyl-CoA to butyrate and butanol via crotonoyl-CoA and butyryl-CoA intermediates. Most of the butyrate synthesis pathway genes are organized into a single locus controlled by the Rex repressor. *C. symbiosum* also has the complete pathway for further conversion of lactate to propionate, which involves propionate CoA-transferase (Pct), lactoyl-CoA dehydratase (Lcd), acryloyl-CoA reductase (Acr) and is equipped with predicted propionate permease (PrpP*) for end product excretion. Thus, *C. symbiosum* and *R. gnavus* possess distinct pathways for production of propionate.

(B) Amino acids

Both strains are capable of synthesizing most amino acids *de novo*. The histidine biosynthesis pathway in *R. gnavus* is composed of 11 committed enzymes, most of which are organized into a single operon controlled by a HisR repressor. The HisR regulon in *R. gnavus* also includes the predicted ABC-type His transporter HisXYZ* suggesting it can also salvage exogenous His. In contrast, the reconstructed HisR regulon in *C. symbiosum* includes the *hutH-yuiF* operon involved in His uptake and degradation; other His degradation genes were also identified in its genome. Thus, *C. symbiosum* is a predicted His auxotroph able to salvage exogenous His and use it both for protein synthesis and as a source of carbon, energy and nitrogen. Glutamate is synthesized from 2-oxoglutarate using the committed enzyme glutamate synthase GltBD, which is present in *R. gnavus* but not in *C. symbiosum*, suggesting the latter is a Glu auxotroph. In support of this prediction, there are two copies of the glutamate uptake transporter gene *gltS* in *C. symbiosum* but not in *R. gnavus*. Furthermore, *C. symbiosum* but not *R. gnavus* has a unique pathway for degradation of exogenous glutamate to produce butyrate.

In addition to the His and Glu uptake transporters described above, specific amino acid transporters were predicted for arginine, tyrosine, tryptophan, glutamine, branched chain amino acids (isoleucine/leucine/valine), alanine and methionine. These transporters are present in both *R. gnavus* and *C. symbiosum* suggesting that they can salvage these

amino acids from the environment, thus providing an alternative way to supply their biosynthetic needs. Tryptophan is a precursor of the neurotransmitter tryptamine. In contrast to *C. symbiosum, R. gnavus* contains a tryptophan decarboxylase ortholog (Trd) suggesting that is able to produce tryptamine.

(C) B-vitamins

Thiamin (vitamin B1) is a common precursor of thiamin pyrophosphate (TPP), an essential cofactor of several ubiquitous enzymes in central carbon metabolism, most notably the pyruvate dehydrogenase complex. Neither *R. gnavus* nor *C. symbiosum* have a complete set of enzymes for *de novo* synthesis of TPP, and thus should be considered B1 auxotrophs. Both strains contain the thiamin salvage enzyme ThiN (thiamin pyrophosphokinase). The B1-specific transporter ThiT was identified in *R. gnavus*, but not in *C. symbiosum*. Alternatively, or in addition to thiamin salvage, both strains are capable of synthesizing TPP using exogenous thiazole via the dedicated salvage enzyme ThiM (hydroxyethylthiazole kinase). However, the respective thiazole transporter, ThiW was confidently identified only in *C. symbiosum.* Both strains have the complete machinery for production of the second biosynthetic intermediate (HMP-PP), joining it with thiazole-P, and downstream phosphorylation to yield the mature TPP cofactor. *C. symbiosum*, but not *R. gnavus*, also has the dedicated transporter ThiXYZ for the opportunistic salvage of the HMP intermediate.

Riboflavin (vitamin B2) is a precursor and biosynthetic intermediate of the essential redox cofactors flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD). The complete canonical *de novo* riboflavin biosynthetic pathway, from GTP and D-ribulose-5-phosphate, is present in *C. symbiosum* but absent in *R. gnavus*. This is one of the two primary distinctions between these two strains, which are otherwise quite similar in terms of B-vitamin proto- and auxotrophies. *R. gnavus*, unlike *C. symbiosum*, lacks fumarate reductase, one of the primary flavin-dependent enzymes in central carbon metabolism. Both strains have a committed riboflavin transporter RibU*,* plus an essential bi-functional enzyme RibF for converting riboflavin to FMN and FAD. A FMN-sensitive riboswitch is present to regulate RibU expression in both strains and *de novo* synthesis in *C. symbiosum*.

Niacin (vitamin B3), or nicotinic acid, as well as nicotinamide are converted to the most abundant redox cofactors, NAD and NADP; these cofactors are collectively involved in ~10% of all cellular biochemical transformations. Reconstruction of this subsystem in both *R. gnavus* and *C. symbiosum* includes both *de novo* synthesis of NAD from aspartate (prototrophy), and potential (optional) salvage of B3 and its amidated precursor. However, the B3 transporter NiaY is only present in *C. symbiosum*. In the absence of additional data, it is impossible to distinguish between two possible interpretations: (i) *R. gnavus* has a B3 transporter of another, yet unknown, type or (ii) the enzymes PncA and PncB are predominantly involved in recycling of intracellular nicotinamide, a well-known product of NAD degradation.

Pantothenate (vitamin B5) is an intermediate in the biosynthesis of coenzyme A (CoA), an indispensable cofactor in a variety of acyl transfer reactions, most importantly in fatty acid biosynthesis, degradation and the TCA cycle. Both strains lack components for *de novo* synthesis of pantothenate. However, auxotrophy is fully applicable only to *R. gnavus*, which has a complete set of downstream enzymes for the 5-step conversion of pantothenate to CoA. An interesting distinction of *C. symbiosum* is the absence of two

enzymes CoaB and CoaC. Pantetheine, a product of CoA degradation, can be phosphorylated (apparently via the second activity of pantothenate kinase), bypassing CoaBC-mediated steps, yielding a substrate for the next step catalyzed by CoaD. Therefore, *C. symbiosum* should be considered a pantetheine rather than a pantothenate auxotroph.

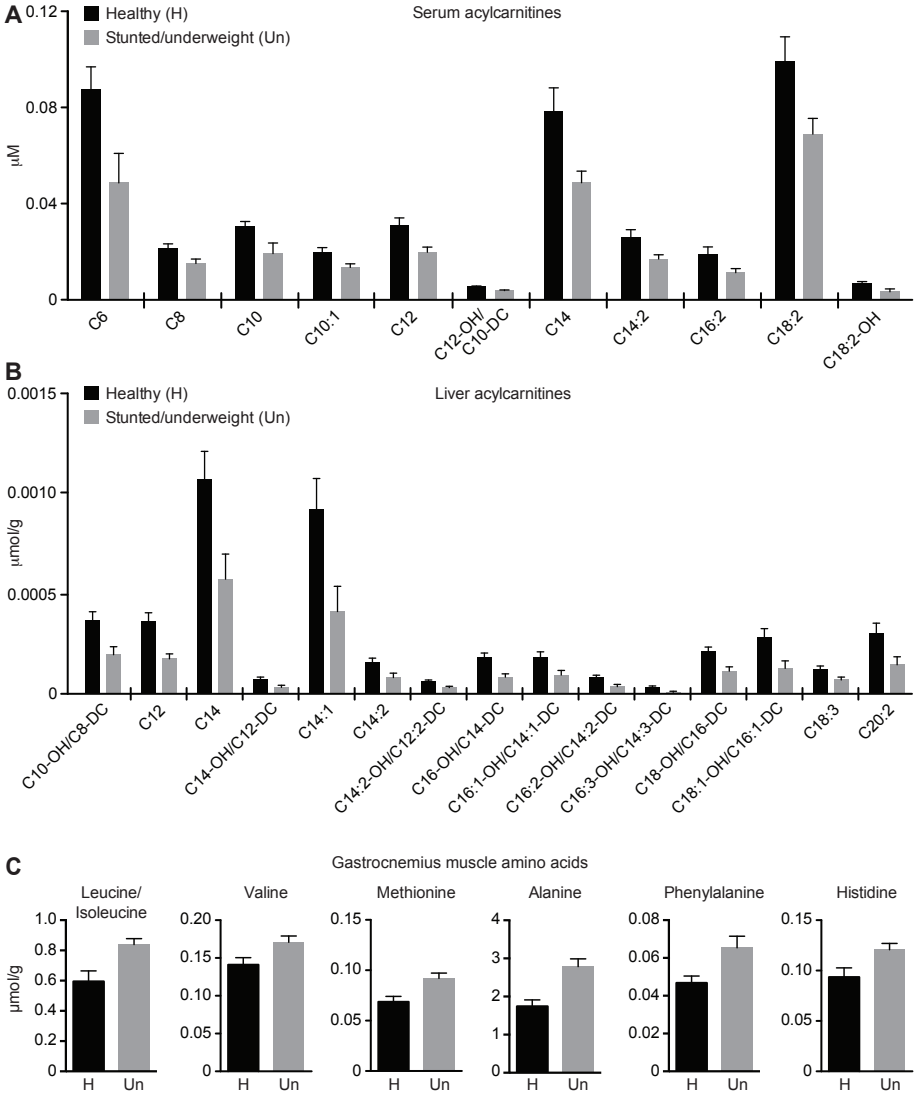
Pyridoxine (B6) is a precursor of the pyridoxal phosphate (PLP) cofactor that drives numerous transamination, decarboxylation and epimerization reactions of various amino acids. This pathway reveals a second major phenotypic distinction between *R. gnavus* and *C. symbiosum*. In contrast to the pattern with vitamin B2, *C. symbiosum* is a B6 auxotroph, whereas *R. gnavus* is a B6 prototroph. Consistent with this conclusion, only *C. symbiosum* contains a B6 transporter (PdxU).

Biotin (B7) is the critical cofactor in transcarboxylation reactions (including the initiation of fatty acid synthesis); it performs its function after covalent attachment to carrier protein(s) by the universally conserved and essential enzyme BirA (biotin ligase). Neither *R. gnavus* nor *C. symbiosum* has recognizable enzymes involved in the upstream steps of B7 biosynthesis, including the critical enzyme BioF (8-amino-7-oxononanoate synthase). Both strains contain the biotin transporter BioY, predicted to be under control of the same transcriptional repressor (the second function of the bifunctional protein BirA). However, both species also contain BioB, pointing to dethiobiotin as a potential alternative vitamin. Moreover, the presence of BioA and BioD, upstream of BioB, in *C. symbiosum* (but not in *R. gnavus*) points to two other possibilities; (i) salvage of 8-amino-7-oxononanoate as yet another undocumented vitamin, and (ii) an alternative upstream pathway merging with the canonical route at the stage of 8-amino-7-oxononanoate. The first possibility appears more likely, and is indirectly supported by the presence of the second BioY paralog in *C. symbiosum* (but not in *R. gnavus*) in the chromosomal neighborhood of the *bioA-bioB* operon. The second possibility would make *C. symbiosum* a B7 prototroph.

Folate (B9) is a precursor of folate-derived cofactors DHF, THF and its polyglutamate derivatives that play a critical role in single carbon metabolism, including numerous methylation reactions. Both strains are obvious B9 auxotrophs containing only downstream enzymes and the folate transporter FolT. In *C. symbiosum*, this transporter and FolC are under regulation of the THF-responding riboswitch.

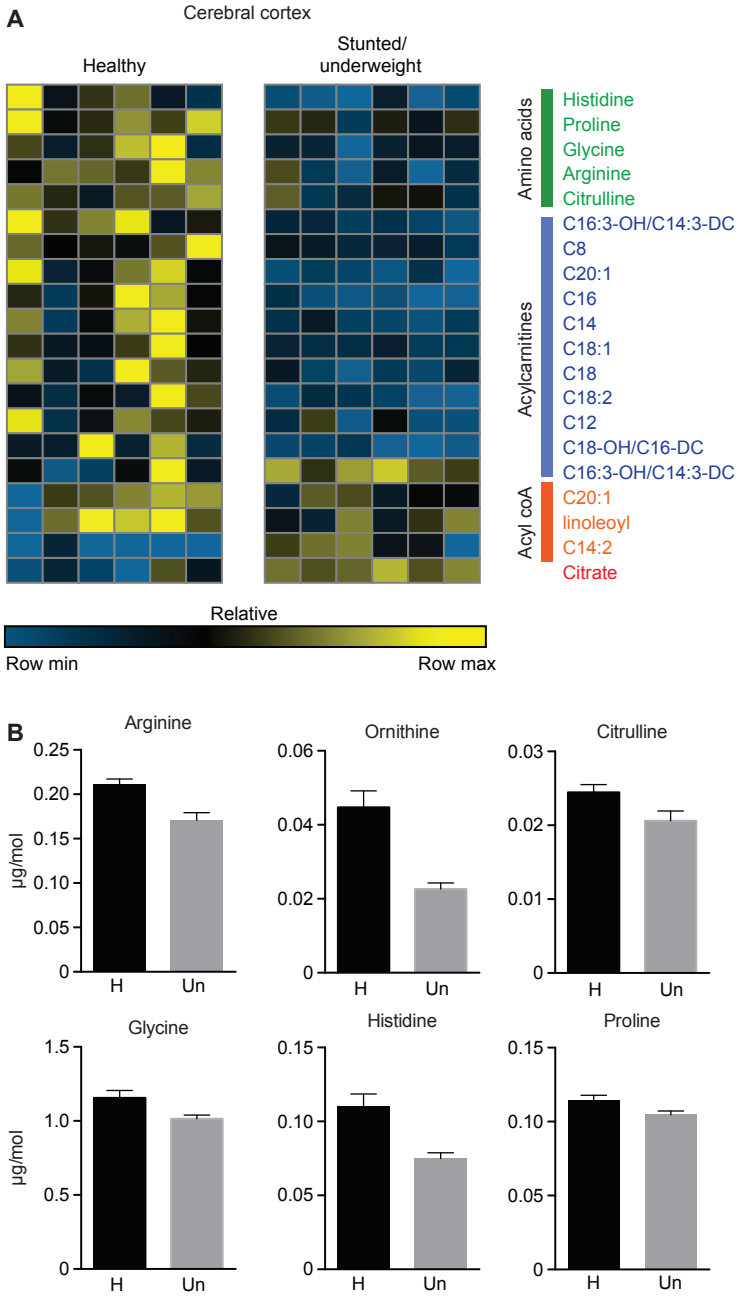
Cobalamin (B12) is involved in a limited number of very important reactions (e.g., methionine biosynthesis). Both analyzed strains have a complete B12 biosynthesis pathway. In addition, *C. symbiosum* (but not *R. gnavus*) contains the B12-specific transporter BtuFCD under regulation of a B12 riboswitch.

Fig. S7.



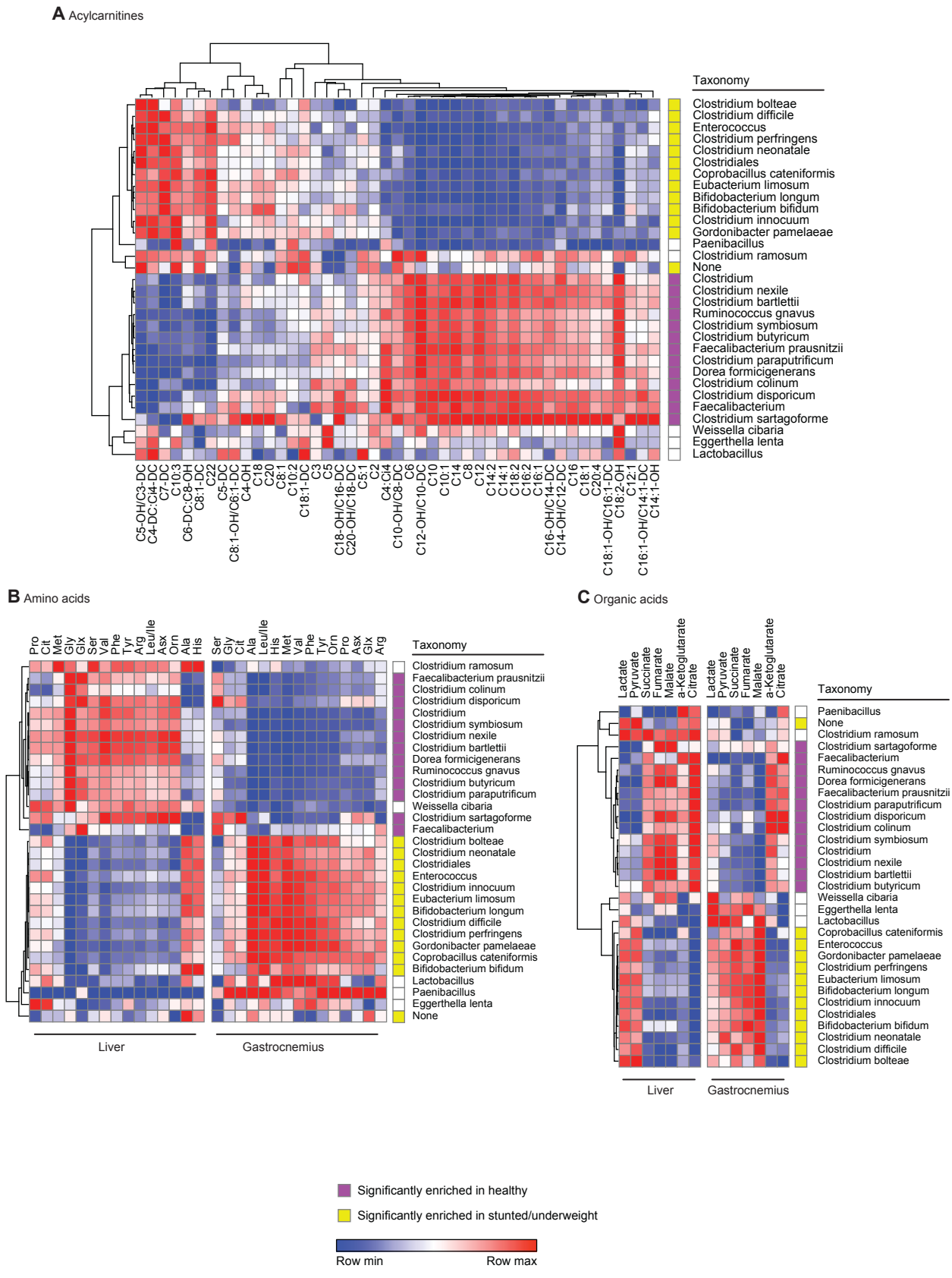
Metabolic features that distinguish gnotobiotic mice colonized with a microbiota from a 6-month-old healthy infant or an immature microbiota from a 6-month-old severely stunted/underweight infant. Metabolic profiles generated by targeted mass spectrometry of **(A)** acylcarnitines in serum **(B)** acylcarnitines in liver, and **(C)** amino acids in gastrocnemius muscle. Each metabolite's concentration differs significantly between the two recipient groups of mice ($P < 0.05$; Student's t-test).

Fig. S8.



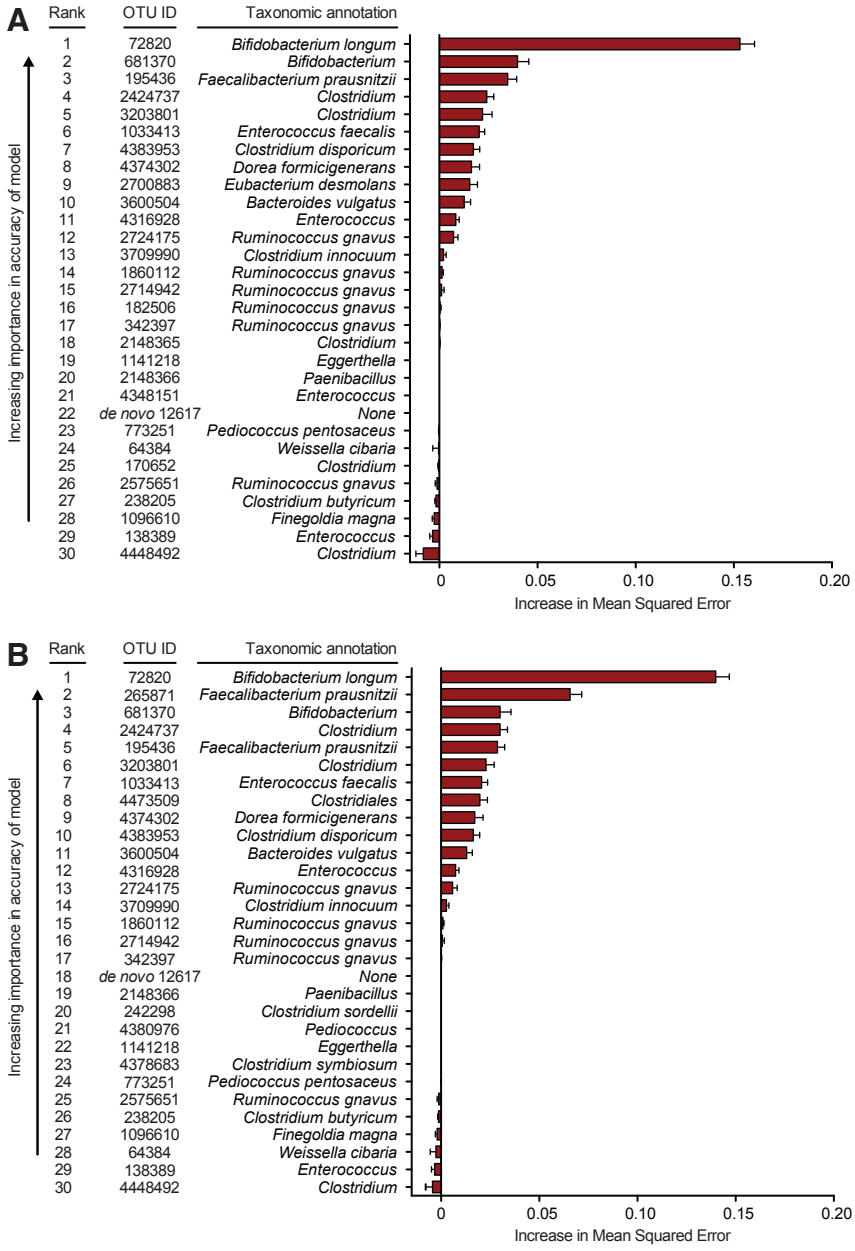
Differences in brain metabolism in gnotobiotic recipients of 6-month-old healthy compared to severely stunted/underweight donor microbiota. (A) Metabolic profile of the cerebral cortex recovered at sacrifice. Columns represent individual animals. Rows represent metabolites whose concentrations are significantly different between the two treatment groups ($P < 0.05$ for all metabolites shown, Student's t-test). **(B)** Concentrations of amino acids in the cerebral cortex of mice colonized with the healthy (H) or stunted/underweight (Un) donor microbiota (mean±SEM values plotted; $P < 0.05$ for all metabolites shown; Student's t-test).

Fig. S9.


A Acylcarnitines
C5-OH/C3-DC
C4-DC:C4-DC
C7-DC
C10:3
C6-DC:C8-OH
C8:1-DC
C22
C5-DC
C8:1-OH/C6:1-DC
C4-OH
C18
C20
C8:1
C10:2
C18:1-DC
C3
C5
C18-OH/C16-DC
C20-OH/C18-DC
C5:1
C2
C4:Ci4
C10-OH/C8-DC
C6
C12-OH/C10-DC
C10
C10:1
C14
C8
C12
C14:2
C14:1
C16:2
C16:1
C16-OH/C14-DC
C14-OH/C12-DC
C16
C18:1
C20:4
C18:1-OH/C16:1-DC
C18:2-OH
C12:1
C16:1-OH/C14:1-DC
C14:1-OH
Taxonomy
Clostridium bolteae
Clostridium difficile
Enterococcus
Clostridium perfringens
Clostridium neonatale
Clostridiales
Coprobacillus cateniformis
Eubacterium limosum
Bifidobacterium longum
Bifidobacterium bifidum
Clostridium innocuum
Gordonibacter pamelaeae
Paenibacillus
Clostridium ramosum
None
Clostridium
Clostridium nexile
Clostridium bartlettii
Ruminococcus gnavus
Clostridium symbiosum
Clostridium butyricum
Faecalibacterium prausnitzii
Clostridium paraputrificum
Dorea formicigenerans
Clostridium colinum
Clostridium disporicum
Faecalibacterium
Clostridium sartagoforme
Weissella cibaria
Eggerthella lenta
Lactobacillus
B Amino acids
Pro
Cit
Met
Gly
Glx
Ser
Val
Phe
Tyr
Arg
Leu/Ile
Asx
Orn
Ala
His
Ser
Gly
Cit
Ala
Leu/Ile
His
Met
Val
Phe
Tyr
Orn
Pro
Asx
Glx
Arg
Taxonomy
Clostridium ramosum
Faecalibacterium prausnitzii
Clostridium colinum
Clostridium disporicum
Clostridium
Clostridium symbiosum
Clostridium nexile
Clostridium bartlettii
Dorea formicigenerans
Ruminococcus gnavus
Clostridium butyricum
Clostridium paraputrificum
Weissella cibaria
Clostridium sartagoforme
Faecalibacterium
Clostridium bolteae
Clostridium neonatale
Clostridiales
Enterococcus
Clostridium innocuum
Eubacterium limosum
Bifidobacterium longum
Clostridium difficile
Clostridium perfringens
Gordonibacter pamelaeae
Coprobacillus cateniformis
Bifidobacterium bifidum
Lactobacillus
Paenibacillus
Eggerthella lenta
None
Liver
Gastrocnemius
C Organic acids
Lactate
Pyruvate
Succinate
Fumarate
Malate
a-Ketoglutarate
Citrate
Lactate
Pyruvate
Succinate
Fumarate
Malate
a-Ketoglutarate
Citrate
Taxonomy
Paenibacillus
None
Clostridium ramosum
Clostridium sartagoforme
Faecalibacterium
Ruminococcus gnavus
Dorea formicigenerans
Faecalibacterium prausnitzii
Clostridium paraputrificum
Clostridium disporicum
Clostridium colinum
Clostridium symbiosum
Clostridium
Clostridium nexile
Clostridium bartlettii
Clostridium butyricum
Weissella cibaria
Eggerthella lenta
Lactobacillus
Coprobacillus cateniformis
Enterococcus
Gordonibacter pamelaeae
Clostridium perfringens
Eubacterium limosum
Bifidobacterium longum
Clostridium innocuum
Clostridiales
Bifidobacterium bifidum
Clostridium neonatale
Clostridium difficile
Clostridium bolteae
Liver
Gastrocnemius
Significantly enriched in healthy
Significantly enriched in stunted/underweight
Row min
Row max

Spearman's rank correlation coefficients for bacterial species relative abundances in the fecal microbiota versus concentrations of metabolites in gnotobiotic mouse recipients of healthy and severely stunted/underweight infant donor gut communities. Spearman correlations involving **(A)** acylcarnitines in serum, **(B)** amino acids in liver and gastrocnemius muscle, and **(C)** organic acids in liver and gastrocnemius muscle. Rows are hierarchically clustered based on pairwise distances calculated from Pearson correlation values. Bars to the right of each heatmap indicate significant species enrichment ($P < 0.05$; Student's t-test) in the fecal microbiota of mice colonized with the healthy (purple) or stunted/underweight (yellow) donor's gut community.

Fig. S10.



Feature importance scores for OTUs, identified as weight gain and lean body mass gain discriminatory in mice, applied to members of the Malawian twin birth cohort. (A) The 30 most weight gain discriminatory OTUs (as shown in Fig. 2D), and **(B)** the 30 most lean body mass gain discriminatory OTUs (as shown in Fig. S2), ranked by increase in mean squared error (MSE) when these taxa are regressed against the WHZ scores of infants and children in the healthy Malawian cohort. Bars show mean±SD of the MSE.

Supplementary Data Tables

Table S1.
Summary of the ages, anthropometry, and time of fecal collection from Malawian cohorts. (A) Concordant healthy Malawian twins pairs/triplet sets. **(B)** Members of the iLiNS-DYAD-M birth cohort with fewer than 3 reported days of antibiotic consumption.

Table S2.
Summary of bacterial V4-16S rRNA datasets from human donors. (A) Malawi Twin cohort. **(B)** iLiNS-DYAD-M cohort.

Table S3.
Random Forests-derived sparse model of microbiota maturation in healthy Malawian infants/children. (A) Rank order of OTU feature importance. **(B)** Comparison of OTUs comprising the sparse Malawian and sparse Bangladeshi models.

Table S4.
V4-16S rRNA-based characterization of fecal microbiota sampled from transplant recipients. (A) Maturity indices (microbiota age and MAZ scores) for microbiota selected for transplantation into gnotobiotic mice. **(B)** Summary of 16S rRNA datasets generated from the fecal microbiota of recipient mice in the initial screen of 19 donor microbiota. **(C)** Summary of 16S rRNA datasets generated from the follow-up transplant studies using microbiota from a 6-month-old healthy and a 6-month-old severely stunted/ underweight donor.

Table S5.
Composition of the Malawi-8 (M8) diet. **(A)** Summary of ingredients. **(B)** Nutritional analysis.

Table S6.
Summary of growth phenotypes in recipients of the 19 microbiota from 6- and 18-month-old healthy and undernourished Malawian donors. (A) Weight gain. **(B)** Lean body mass gain and fat body mass gain. Spearman's rank correlation of Shannon Index to lean body mass gain at day 28 following gavage of the donors' microbiota is non-significant ($P = 0.3275$).

Table S7.
Consumption of M8 chow in each donor group of gnotobiotic recipient mice from the screen of 19 donor microbiota.

Table S8.
Rank order list of feature importance scores of OTUs in Random Forest models based on the growth phenotypes of gnotobiotic mice with 19 different transplanted microbiota from 6- and 18-month-old Malawian donors. (A) Weight gain. **(B)** Lean body mass gain.

Table S9.
Spearman's rank correlations of OTUs to mouse weight and lean body mass gain as well as to the chronological ages of members of concordant healthy Malawian twin pairs/triplet sets and the of iLiNS-DYAD-M cohort. (A) Weight gain. **(B)** Lean body mass gain.

Table S10.
Rank order list of OTU feature importance scores in Random Forest models based on the femoral bone metrics of groups of recipient mice with 19 different transplanted microbiota from 6- and 18-month-old Malawian donors.

Table S11.
List of species and their relative abundances used to construct the bargraphs shown in Fig. S4B.

Table S12.
Direction and success of invasion in cohousing experiments involving H^{CH} and Un^{CH} mice.

Table S13.

Reconstructed metabolic pathways for *Ruminococcus gnavus* TS8243C and *Clostridium symbiosum* TS8243C. (A) Overall number of genes in the reconstructed metabolic pathways of the two strains. The three major functional categories of studied genes are those involved in (i) carbohydrate and central carbon metabolism, (ii) amino acid metabolism and (iii) vitamin and cofactor biosynthesis. **(B,C)** Detailed lists of genes in *R. gnavus* TS8243C and *C. symbiosum* TS8243C, and their orthologs in the type strains, *R. gnavus* ATCC 29149 and *C. symbiosum* WAL-14163. Enzymes are identified by their enzyme commission (EC) numbers. Functional descriptions for carbohydrate active enzymes (CAZymes) include their assigned glycoside hydrolase (GH) families and their predicted localization (intracellular versus secreted as defined by SignalP). Transcriptional regulators and riboswitches that are predicted to control expression of individual genes from the reconstructed pathways (according to conservative propagation in the RegPrecise database) are listed in the last column. Central carbon metabolism genes controlled by the global redox-responsive regulator Rex are also indicated. Genes with novel predicted functions are marked with an asterisk. **(D)** Summarized metabolic capabilities. Predicted ability (+) or inability (-) to: (i) synthesize amino acids, (ii) synthesize B-vitamin related cofactors, (iii) utilize sugars, (iv) utilize other sources of carbon and energy, and (v) generate fermentation products.

Table S14.
List of taxa and their relative abundances (%) used to construct the bargraphs shown in Fig. 4C.

Table S15.
Metabolic phenotypes, as defined by targeted mass spectrometry, of gnotobiotic mice colonized with microbiota from the severely stunted/underweight infant donor or with the donor's microbiota together with *R. gnavus* TS8243C and *C. symbiosum* TS8243C.

Table S16.
Metabolic phenotypes of gnotobiotic mice colonized with a normally mature microbiota from a 6-month-old healthy (H) infant and a significantly immature microbiota from a 6-month-old severely stunted/underweight (Un) donor.

Table S17.
Spearman's rank correlations between levels of metabolites and the abundances of bacterial species in the fecal microbiota of mice colonized with the healthy versus the severely stunted/underweight infant microbiota. (A) Serum acylcarnitines. **(B)** Amino acids in liver and gastrocnemius muscle. **(C)** Organic acids assayed in liver and gastrocnemius muscle.