PNAS
www.pnas.org

Supplementary Information for

Identifying determinants of bacterial fitness in a model of human gut microbial succession

Lihui Feng, Arjun S. Raman, Matthew C. Hibberd, Jiye Cheng, Nicholas W. Griffin, Yangqing Peng, Semen A. Leyn, Dmitry A. Rodionov, Andrei L. Osterman, and Jeffrey I. Gordon

Email: jgordon@wustl.edu

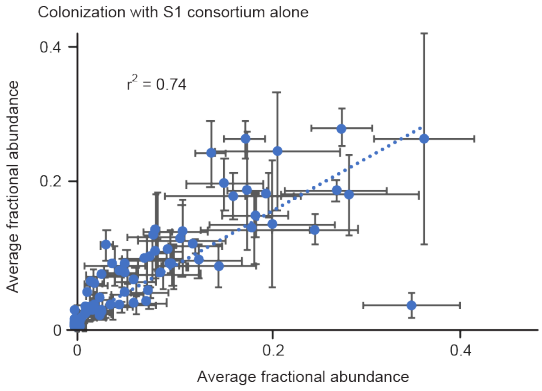
This PDF file includes:

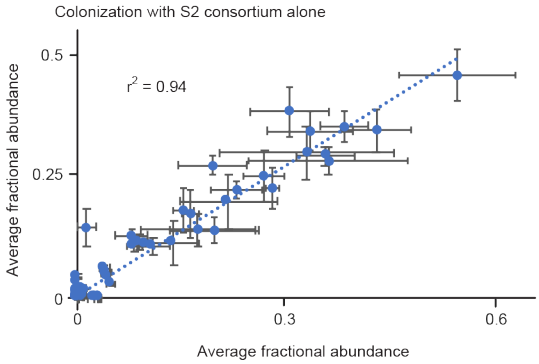
Figs. S1 to S6
Captions for databases S1 to S13

Other supplementary materials for this manuscript include the following:

Datasets S1 to S13

Supplemental Figures


Colonization with S1 consortium alone
r² = 0.74
Average fractional abundance
Average fractional abundance
0
0.2
0.4


Colonization with S2 consortium alone
r² = 0.94
Average fractional abundance
Average fractional abundance
0
0.25
0.5
0.3
0.6

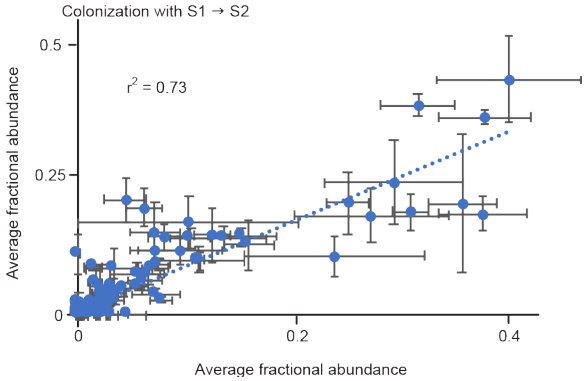

Colonization with S1 → S2
r² = 0.73
Average fractional abundance
Average fractional abundance
0
0.25
0.5
0.2
0.4

Fig. S1 – Average fractional abundance of organisms is reproducible over two independent colonization experiments where animals were colonized with S1 alone, S2 alone and S1→S2. The x-axis plots data from the experiments shown in **Fig. 1** while the y-axis plots data obtained from an independent series of experiments whose results are presented in **Dataset S3B**. The closed circle and associated horizontal and vertical lines represent the mean fractional abundance ± SD of a given organism in a given colonization condition. Pearson correlation values are shown.

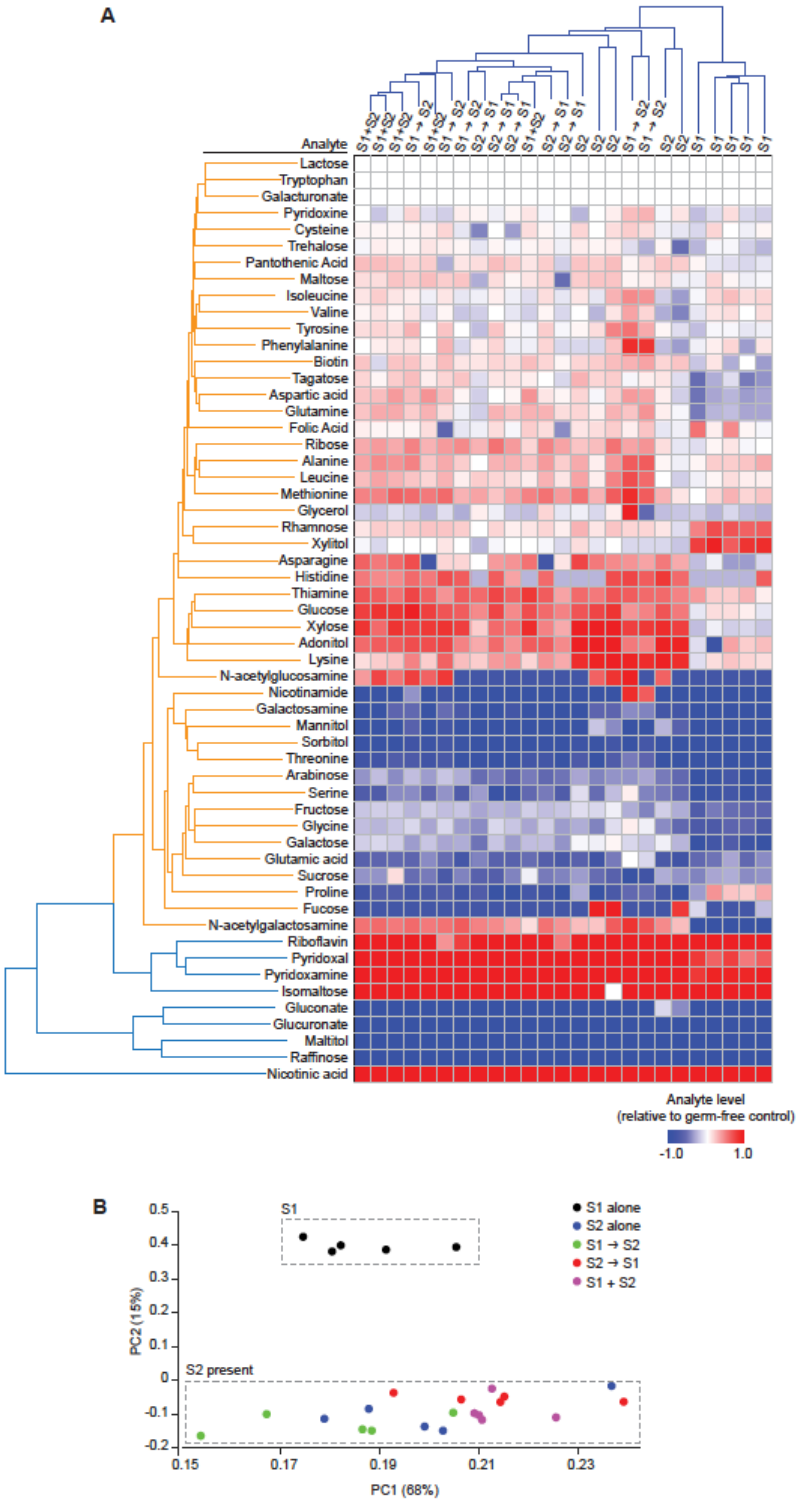

A
Analyte
Lactose
Tryptophan
Galacturonate
Pyridoxine
Cysteine
Trehalose
Pantothenic Acid
Maltose
Isoleucine
Valine
Tyrosine
Phenylalanine
Biotin
Tagatose
Aspartic acid
Glutamine
Folic Acid
Ribose
Alanine
Leucine
Methionine
Glycerol
Rhamnose
Xylitol
Asparagine
Histidine
Thiamine
Glucose
Xylose
Adonitol
Lysine
N-acetylglucosamine
Nicotinamide
Galactosamine
Mannitol
Sorbitol
Threonine
Arabinose
Serine
Fructose
Glycine
Galactose
Glutamic acid
Sucrose
Proline
Fucose
N-acetylgalactosamine
Riboflavin
Pyridoxal
Pyridoxamine
Isomaltose
Gluconate
Glucuronate
Maltitol
Raffinose
Nicotinic acid
Analyte level
(relative to germ-free control)
-1.0
1.0
B
S1
S2 present
S1 alone
S2 alone
S1 → S2
S2 → S1
S1 + S2
PC2 (15%)
PC1 (68%)
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
0.15
0.17
0.19
0.21
0.23

Fig. S2 – Cecal analytes measured in mice subjected to different colonization histories. (A) Mass spectrometry of the concentrations of 27 carbohydrates, 19 amino acids and 10 vitamins/cofactors in cecal contents recovered from mice subjected to each of the colonization sequences tested, as well as from germ-free controls (see **Dataset S4A,B**). The extent to which an analyte was increased or decreased in the cecum for a particular colonization condition relative to germ-free controls was calculated for all analytes and the results are shown in the form of a heatmap (average values displayed, see **Dataset S4C** for data from individual animals). Hierarchical clustering reveals nine analytes whose relative levels are not dependent upon colonization conditions (blue components of the dendrogram representing cluster 1). (**B**) PCA was performed on the 47 analytes whose levels are dependent on colonization sequence (orange color components of dendrogram in panel A comprising cluster 2). Plotting the pattern of relative levels of analytes in cluster 2 in a given mouse subjected to a given colonization sequence shows that the metabolite landscape of animals colonized with S2 consortium members is distinct from that produced by colonization with the S1 consortium alone (see **Dataset S4D)**.

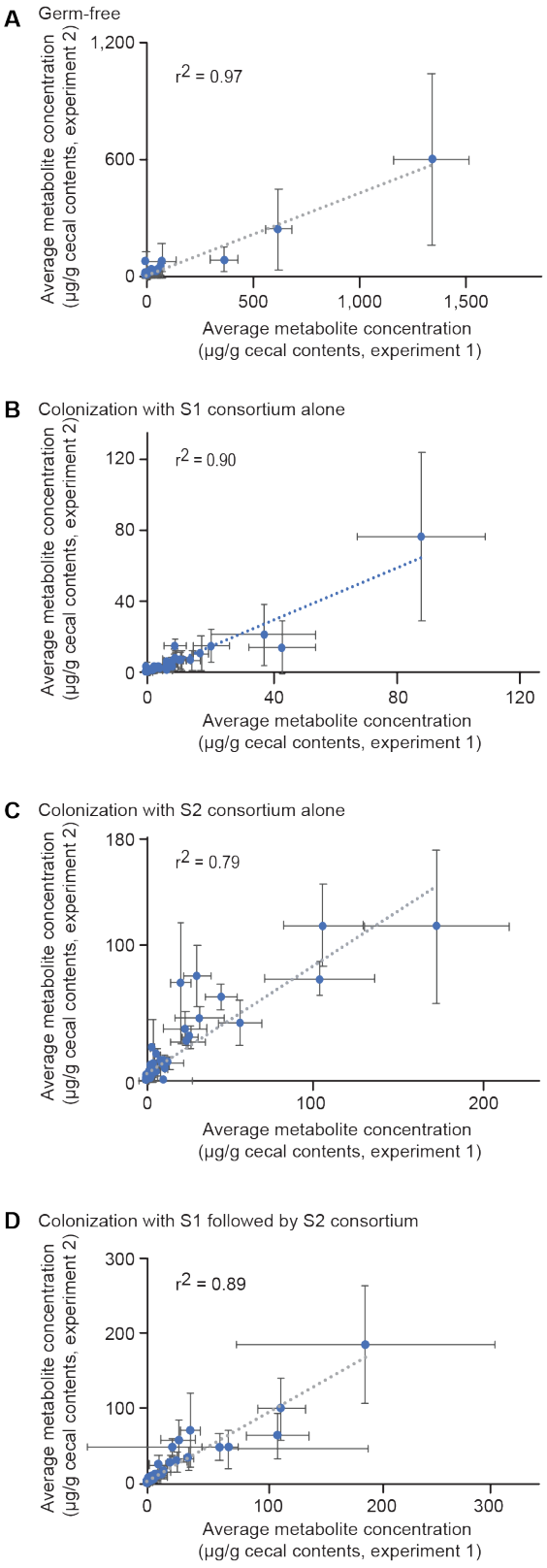

A
Germ-free
Average metabolite concentration (µg/g cecal contents, experiment 2)
1,200
600
0
r2 = 0.97
0
500
1,000
1,500
Average metabolite concentration (µg/g cecal contents, experiment 1)
B
Colonization with S1 consortium alone
Average metabolite concentration (µg/g cecal contents, experiment 2)
120
80
40
0
r2 = 0.90
0
40
80
120
Average metabolite concentration (µg/g cecal contents, experiment 1)
C
Colonization with S2 consortium alone
Average metabolite concentration (µg/g cecal contents, experiment 2)
180
100
0
r2 = 0.79
0
100
200
Average metabolite concentration (µg/g cecal contents, experiment 1)
D
Colonization with S1 followed by S2 consortium
Average metabolite concentration (µg/g cecal contents, experiment 2)
300
200
100
0
r2 = 0.89
0
100
200
300
Average metabolite concentration (µg/g cecal contents, experiment 1)

Fig. S3 - Cecal analyte concentrations in gnotobiotic mice with different colonization histories. (**A-D**) Average analyte concentrations, measured by targeted mass spectrometry in cecal contents harvested from germ-free mice (panel A), and mice colonized with the S1 consortium alone (panel B), the S2 consortium alone (panel C), or the S1 followed by the S2 consortium (panel D), are reproducible over two separate experiments. The closed circles and associated horizontal and vertical lines represent mean concentration ± SD of a given analyte in a given colonization condition. Pearson correlation values are shown.


mcSEED metabolic pathway/module
Bifidobacterium dentium
Citrobacter amalonaticus
Clostridium bolteae
Clostridium butyricum
Clostridium innocuum
Clostridium paraputrificum
Clostridium ramosum
Clostridium scindens
Enterobacter massiliensis
Enterococcus casseliflavus
Enterococcus faecium
Escherichia coli
Klebsiella oxytoca
Lactobacillus rhamnosus
Staphylococcus aureus
Staphylococcus simulans
Streptococcus lutetiensis
Streptococcus oralis
Streptococcus parasanguinis
Streptococcus salivarius
Veillonella atypica
Veillonella dispar
Actinomyces odontolyticus
Akkermansia muciniphila
Bifidobacterium adolescentis
Bifidobacterium breve
Bifidobacterium longum
Collinsella intestinalis
Eubacterium limosum
Flavonifractor plautii
Parabacteroides distasonis
Peptoclostridium difficile
Peptoniphilus gorbachii
Ruminococcus gnavus
Ruminococcus gnavus
Amino Acids
Arginine
Lysine
Histidine
Tryptophan
tyrosine
Phenylalanine
Chorismate
Isoleucine
Valine
Leucine
Serine
Cysteine
Threonine
Methionine
Proline
Glycine
Glutamate
Glutamine
Aspartate
Asparagine
Vitamins/ cofactors
B1
B2
B3
B5
B6
B7
B9
B12
Queuosine
Lipoate
Menaquinone
Short chain fatty acids
Butyrate
Propionate
Acetate
Formate
Lactate
Carbohydrates
Glucose
Galactose
Fructose
Mannose
Tagatose
Glucuronate
Galacturonate
N-acetylgalactosamine
Galactosamine
N-acetylglucosamine
N-acetylmannosamine
N-acetylneuraminate
Unsaturated glucuronate
Xylose
Arabinose
Fucose
Rhamnose
Ribose
Gluconate
Maltose
Sucrose
Trehalose
Chitobiose
Lacto-N-biose
lactose
Mannosides
Fructooligosaccharides
Alpha-galactosides
Melibiose
Raffinose
Beta-galactosides
Beta-glucosides
Oligogalacturonate
Glucuronides
Maltodextrin
Beta-xylosides
Alpha-xylosides
Beta-araninosides
Alpha-araninosides
Fucosides
Rhamnoglucuronides
Hyaluronate
Mannitol
Arabinitol
Xylitol
Sorbitol
Galactitol
Ribitol
Inositol
Fructose-Asparagine
Fructoselysine
Glucoselysine
Psicoselysine
Prototrophy/utilizer
Auxotrophy/non-utilizer

Fig. S4 – Binary Phenotype Matrix (BPM) encompassing the 34 organisms studied. Red cells indicate the presence of a complete biosynthetic pathway for an amino acid or a vitamin/cofactor/essential nutrient (prototrophy), or a carbohydrate utilization pathway (utilizer), or a SCFA production pathway (producer), whereas white cells indicate the absence of biosynthetic pathway (auxotrophy), carbohydrate utilization (non-utilizer), or SCFA fermentation pathway (non-producer).

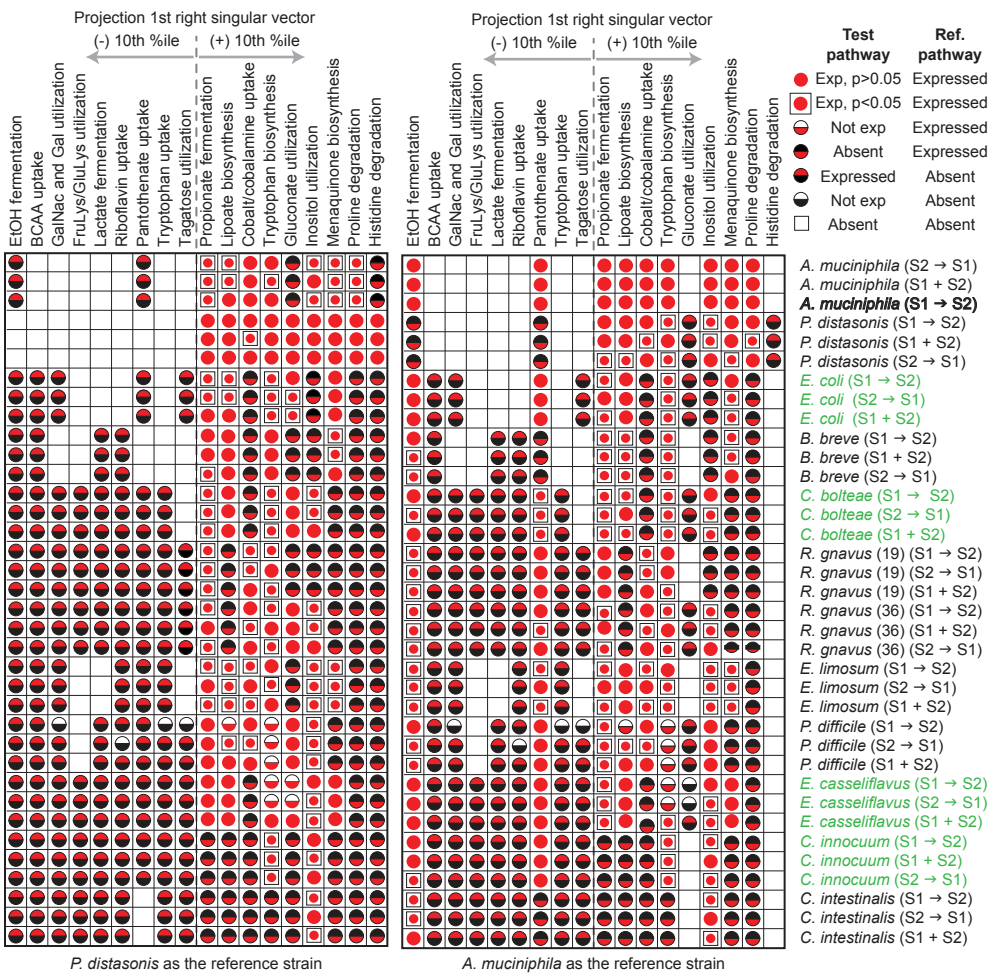

Projection 1st right singular vector
(-) 10th %ile
(+) 10th %ile
EtOH fermentation
BCAA uptake
GalNac and Gal utilization
FruLys/GluLys utilization
Lactate fermentation
Riboflavin uptake
Pantothenate uptake
Tryptophan uptake
Tagatose utilization
Propionate fermentation
Lipoate biosynthesis
Cobalt/cobalamine uptake
Tryptophan biosynthesis
Gluconate utilization
Inositol utilization
Menaquinone biosynthesis
Proline degradation
Histidine degradation
Test pathway
Ref. pathway
Exp, p>0.05 Expressed
Exp, p<0.05 Expressed
Not exp Expressed
Absent Expressed
Expressed Absent
Not exp Absent
Absent Absent
A. muciniphila (S2 → S1)
A. muciniphila (S1 + S2)
A. muciniphila (S1 → S2)
P. distasonis (S1 → S2)
P. distasonis (S1 + S2)
P. distasonis (S2 → S1)
E. coli (S1 → S2)
E. coli (S2 → S1)
E. coli (S1 + S2)
B. breve (S1 → S2)
B. breve (S1 + S2)
B. breve (S2 → S1)
C. bolteae (S1 → S2)
C. bolteae (S2 → S1)
C. bolteae (S1 + S2)
R. gnavus (19) (S1 → S2)
R. gnavus (19) (S2 → S1)
R. gnavus (19) (S1 + S2)
R. gnavus (36) (S1 → S2)
R. gnavus (36) (S1 + S2)
R. gnavus (36) (S2 → S1)
E. limosum (S1 → S2)
E. limosum (S2 → S1)
E. limosum (S1 + S2)
P. difficile (S1 → S2)
P. difficile (S2 → S1)
P. difficile (S1 + S2)
E. casseliflavus (S1 → S2)
E. casseliflavus (S2 → S1)
E. casseliflavus (S1 + S2)
C. innocuum (S1 → S2)
C. innocuum (S1 + S2)
C. innocuum (S2 → S1)
C. intestinalis (S1 → S2)
C. intestinalis (S2 → S1)
C. intestinalis (S1 + S2)
P. distasonis as the reference strain
A. muciniphila as the reference strain

Fig. S5 – Comparing the expression patterns of metabolic pathways within S1 and S2 organisms using *Akkermansia muciniphila* as the reference S1 strain instead of *Parabacteroides distasonis*. See legend to **Fig. 3D** for definition of symbols.

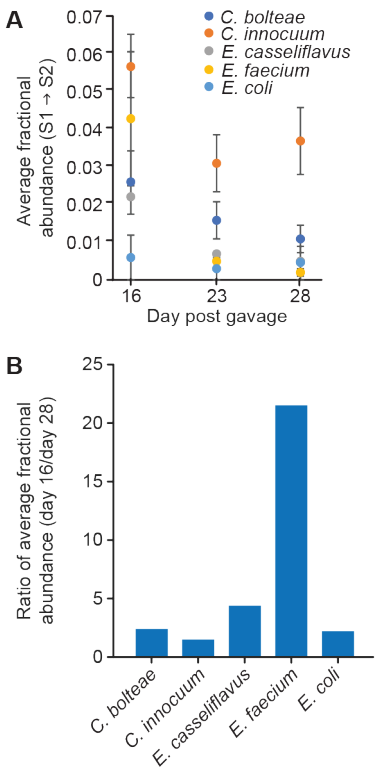

A
C. bolteae
C. innocuum
E. casseliflavus
E. faecium
E. coli
Average fractional abundance (S1 → S2)
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
16
23
28
Day post gavage
B
Ratio of average fractional abundance (day 16/day 28)
25
20
15
10
5
0
C. bolteae
C. innocuum
E. casseliflavus
E. faecium
E. coli

Fig. S6 – Fitness of the S1 consortium members *C. bolteae*, *C. innocuum*, *E. casseliflavus*, *E. coli* and *E. faecium* from dpg 16 to 28 in the S1→S2 colonization sequence. (A) Average fractional representation. Closed circles and vertical bars represent mean values ± SD. **(B)** Ratio of average fractional representation at dpg 16 compared to dpg 28 (see **Dataset S3A** for the complete time-course).

Supplemental Datasets

Dataset S1 - Information about the infant donor from whom the culture collection was generated.

Dataset S2 - Strains in S1 and S2 consortia and media used for their growth *in vitro*.

Dataset S3 - COPRO-Seq data of fractional abundances of each organism in each mouse used to generate the results shown in Fig. 1 and fig. S1.

Dataset S4 - Cecal analyte concentrations in mice belonging to the various treatment groups depicted in fig. S2 and fig.S3.

Dataset S5 - Genome annotations of S1 and S2 bacterial strains introduced into gnotobiotic mice, corresponding TPM-normalized transcript data generated under each of the indicated colonization conditions, and correlation between transcriptional and organism abundance.

Dataset S6 – Data related to mcSEED metabolic pathway/module relative expression matrix and corresponding eigendecomposition for data shown in Fig. 2B.

Dataset S7 – Relationship between position along PC1 of Fig. 2B,C and bacterial fitness on dpg 28 as a function of changing the reference colonization condition for computing the mcSEED relative expression matrix.

Dataset S8 - SVD of the mcSEED relative expression matrix presented in Dataset S6 and Fig. 3C.

Dataset S9 – Results of statistical analysis of differences in mcSEED pathway expression related to Fig. 3D.

Dataset S10 – Results of statistical analysis of differences in mcSEED pathway expression related to Fig. 3D, compared to *Akkermansia muciniphila*.

Dataset S11 – Data analysis related to Fig. 4.

Dataset S12 – Microbial RNA-Seq count statistics for *E. faecium*.

Dataset S13 – Results of statistical analysis of differences in mcSEED pathway expression related to Fig. 4E.