MICOM: metagenome-scale modeling to infer metabolic interactions in the gut microbiota.

Christian Diener[1,2], Sean M. Gibbons[2,3,*] and Osbaldo Resendis-Antonio[1,4,*].

[1] Instituto Nacional de Medicina Genómica (INMEGEN), Mexico City 14610, México

[2] Institute for Systems Biology, Seattle, WA, USA

[3] eScience Institute, University of Washington, Seattle, WA, USA

[4] Human Systems Biology Laboratory. Coordinación de la Investigación Científica - Red de Apoyo a la Investigación, UNAM.

[*] Corresponding authors: oresendis@inmegen.gob.mx, sgibbons@isbscience.org

Abstract

Compositional changes in the gut microbiota have been associated with a variety of medical conditions such as obesity, Crohn's disease and diabetes. However, connecting microbial community composition to ecosystem function remains a challenge. Here, we introduce MICOM - a customizable metabolic model of the human gut microbiome. By using a heuristic optimization approach based on L2 regularization we were able to obtain a unique set of realistic growth rates that corresponded well with observed replication rates. We integrated adjustable dietary and taxon abundance constraints to generate personalized metabolic models for individual metagenomic samples. We applied MICOM to a balanced cohort of metagenomes from 186 people, including a metabolically healthy population and individuals with type 1 and type 2 diabetes. Model results showed that individual bacterial genera maintained conserved niche structures across humans, while the community-level production of short chain fatty acids (SCFAs) was heterogeneous and highly individual-specific. Model output revealed complex cross-feeding interactions that would be difficult to measure *in vivo*. Metabolic interaction networks differed somewhat consistently between healthy and diabetic subjects. In particular MICOM predicted reduced butyrate and propionate production in a diabetic cohort, with restoration of SCFA production profiles found in healthy subjects following metformin treatment. Overall, we found that changes in diet or taxon abundances have highly personalized effects. We believe MICOM can serve as a useful tool for generating mechanistic hypotheses for how diet

and microbiome composition influence community function. All methods are implemented in the open source Python package, which is available at https://github.com/micom-dev/micom.

Introduction

The composition of the gut microbiome can influence host metabolism (1) and has been associated with a variety of health conditions such as obesity, Crohn's Disease, diabetes and colorectal cancer (2–6). However, the causal roles played by the gut microbiota in host physiology and disease remain unclear. Several studies have mapped individual gut microbial genes to functions (7–9). However, these mappings are largely qualitative, as the presence of a particular gene does not guarantee expression of a functional enzyme. An alternative strategy to quantify the metabolic capacity of a microbial community is to use computational models for inferring fluxes in biochemical networks (10, 11). While direct experimental measurement of fluxes by carbon or nitrogen labeling is costly, one can readily estimate the metabolic fluxes of a model organism using genome-scale metabolic models. For individual bacteria, metabolic modeling using flux balance analysis (FBA) has been a valuable tool for exploring metabolic capacities under varying conditions and has been used extensively in basic research, biochemical strain engineering, and *in vitro* models of bacterial interactions (12–15). In FBA, fluxes are usually approximated from a genome-scale model containing all known biochemical reactions by maximizing the production of biomass under constraints mirroring enzymatic, thermodynamic and environmental conditions (13). For instance, one can restrict metabolic import fluxes to those whose substrates are present in the growth medium (12, 14, 16) in order to simulate a particular nutrient environment. Extending FBA to microbial communities is challenging due to the necessity of modeling metabolic exchanges between many taxa and selecting an appropriate objective function to account for potential tradeoffs between species and community growth rates.

Maximizing the community growth rate is at odds with maximizing individual species growth rates. Multi-objective methods, like OptCom, attempt to find the joint maximum of individual and community growth rates (17). However, these multi-objective methods are limited to smaller-sized communities. The human gut microbiome, on the other hand, may contain up to several hundred distinct species (18). An additional challenge is the integration of relative abundances obtained from 16S amplicon or metagenomic shotgun sequencing into a community FBA model. This is particularly important for accurately inferring the metabolic exchanges taking place between different species within the community. A very abundant species should import and export much greater absolute quantities of metabolites than a very rare species, which in turn

impacts the resulting community-level biochemical fluxes. Despite the challenges, genome-scale metabolic modeling of microbial communities holds great promise as a tool for estimating the metabolic potential of an individual's gut microbiome. In particular, this approach could yield valuable insights into possible metabolic mechanisms underlying host disease states.

Here, we present a computational approach that efficiently extends metabolic modeling to entire microbial communities. Using a two-step optimization procedure, we were able to simulate growth and metabolic exchange fluxes for metagenome-scale metabolic models of ecologically diverse bacterial systems. Additionally, we explicitly included microbial abundances from metagenomic shotgun sequencing and realistic dietary inputs in order to make quantitative, personalized, metabolic predictions. This entire strategy is implemented in an open-access Python software package called "MICOM" (**MI**crobial **COM**munity).

We tested our approach by applying MICOM to a balanced data set of 186 Danish and Swedish individuals, including healthy controls, patients with type 1 diabetes, and patients with type 2 diabetes with and without metformin treatment. We show that individual bacterial growth rates vary greatly across samples and are correlated with independently measured replication rates. We quantified exchanges between the gut microbiota and gut lumen and studied the effect of the microbiota composition on the production of short chain fatty acids (SCFAs) across samples from healthy and diabetic individuals. Overall, we found that MICOM predicted a bimodal usage pattern of dietary metabolites, ecological interactions between microbes tended to be community-specific and largely competitive, key gut genera associated with health participated in the largest number of ecological interactions, inferred SCFA production was lower in diabetic patients, and targeted dietary or probiotic interventions had unique functional consequences for each individual.

Results

A regularization strategy for microbial community models.

Metabolic modeling is commonly applied to model a single strain of bacteria in log-phase, where the growth rate is approximately constant and the log of the bacterial abundance increases linearly with time. Modeling bacterial growth in natural environments is often more complex than this, but some information on environmental context can be extracted from the relative abundances of bacterial taxa. Within a single individual, and in the absence of persistent dietary

changes, gut microbial relative abundances tend to fluctuate around a fixed median value over month-to-year timescales (19–21). This is consistent with a steady state model where bacterial growth is in equilibrium with a dilution process that continuously removes biomass from the system (22). Under this approximation bacterial growth rates are constants μ_i (in 1/h), which is compatible with the assumption of FBA. All bacteria in the microbial community then contribute to the production of total biomass, with an overall growth rate constant μ_c. The community growth rate μ_c is obtained from the individual growth rates μ_i by a weighted mean, with the relative contribution of species i (a_i) to the total biomass serving as the weight (17, 22).

$$\mu_C = \sum_i \; a_i \mu_i \qquad (1)$$

Even though FBA can be used to obtain the maximum community growth rate, one can see from Equation 1 that there is an infinite combination of different individual growth rates μ_i for any given community growth rate μ_c (see Fig. 1A for an example). Various strategies have been employed in order to deal with this limitation, where the simplest strategy is to report any one of the possible growth rates distributions for μ_i. Other approaches attempt to find the set of growth rates that maximize community growth and individual growth at the same time (17), but this is computationally intensive and may not scale well to the species-diverse gut microbiome (18, 23). Thus, we formulated a strategy that allows us to identify a realistic set of individual growth rates μ_i and scales to large communities. The simplest case of a microbial community is a community composed of two identical clonal strains, each present in the same abundance. Assuming that the maximum community and individual growth rates are equal to 1.0 there are now many alternative solutions giving maximal community growth (Fig. 1A). However, the two populations are identical and present in the same abundance, so one would expect that both grow at the same rate. In order to enforce a particular distribution of individual growth rates one can try to optimize an additional function over the individual growth rates μ_i. This is known as regularization and a feasible regularization function should enrich for biologically relevant growth rate distributions. As a heuristic, our minimal requirement for a feasible regularization function was consistency with the observed metagenomic abundances. This means that a taxon that is observed in the data should be able to grow. Thus the growth rate of a taxon should be non-zero if its abundance is non-zero. We show in Supplemental Text S1 that no linear regularization function can comply with that requirement whereas a simple quadratic regularization, also known as L2 regularization, does fulfill that requirement (24, 25). L2 regularization is known to distribute magnitude over all

variables, which is also consistent with a maximization of individual growth rates and thus forms a heuristic for the simultaneous maximization of individual and community growth rates.

L2 regularization can be readily integrated into FBA as a quadratic optimization problem, which is not necessarily true for any generic function. In the previous example of two identical strains only the L2 norm correctly identifies the solution where both strains grow at the same rate as optimal. Additionally, the L2 norm has a unique minimum. Thus, there is only one configuration of individual growth rates μ_i that minimizes the L2 norm for a given community growth rate μ_c. In practice, maximal community growth might only be achievable if many taxa are excluded from growth, for instance by giving all resources to a fast-growing subpopulation. Again, this is inconsistent with reality if one has prior knowledge that the other taxa are present in the gut and should be able to grow. Instead of enforcing the maximal community growth rate one can limit community growth to only a fraction of its maximum rate, thus creating a tradeoff between optimal community growth and individual growth rate maximization. Community growth maximization requires full cooperativity, whereas the L2 norm minimization represents selfish individual growth maximization. Thus we call our two-step strategy of first fixing community growth rate to a fraction of its optimum and then minimizing the L2 norm of individual growth rates a “cooperative trade-off”. Even though it is difficult to formulate a closed form solution for this two-step optimization, one can obtain a solution for the second optimization (minimization of regularization term) when dropping additional constraints for growth rates (see Supplemental Text S1 for derivation). In that case, growth rates are given by:

$$\mu_i = \frac{\alpha \mu_c}{a^T a} a_i \ . \qquad (2)$$

Thus, optimal growth rates will be approximately correlated with abundance where the slope depends on the abundance distribution and the maximum community growth rate.

We found that computation time generally scaled well with the community size (with most individual optimizations taking less than 5 minutes) when using interior point methods, which are known to provide better performance for larger models (26). However, we found that it was difficult to maintain numerical stability with large community models. None of the tested solvers were able to converge to optimality when solving the quadratic programming problem posed by the L2 norm minimization (see Methods). Thus, we used a crossover strategy to identify an optimal solution to the L2 minimization (see Methods).

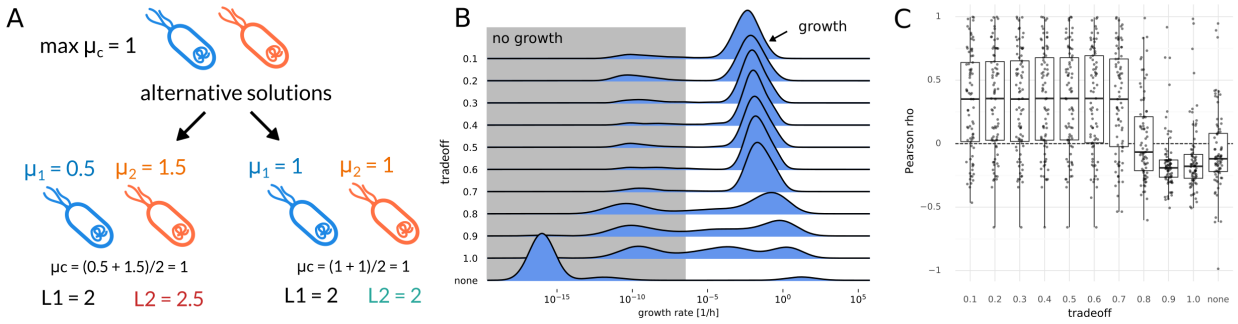


Figure 1: Regularization of growth rates. (A) Regularization values for a toy model of two identical *E. coli* populations. Two alternative solutions are shown with different individual growth rates and respective values of regularization optima. Here L1 denotes minimizing the sum of growth rates whereas L2 denotes minimization of the sum of squared growth rates. Only L2 regularization favors one over the other and identifies the expected solution where both populations grow with the same rate. (B) Effect of different trade-off values (fraction of maximum community growth rate) on the distribution of individual genus growth rates. Zero growth rates were assigned a value of 10^{-16} which was smaller than the observed non-zero minimum. Growth rates smaller than 10^{-6} were considered to not represent growth (shaded area). (C) Pearson correlation between replication rates and inferred growth rates under varying trade-off values. "None" indicates a model without regularization returning arbitrary alternative solutions (see Methods). The dashed line indicates a correlation coefficient of zero.

Regularization by cooperative trade-off yields realistic growth rate estimates.

In order to test whether cooperative trade-off yields realistic growth rates, we implemented and applied it to a set of 186 metagenome samples from Swedish and Danish individuals (27), consisting of healthy individuals, individuals with type 1 diabetes and individuals with type 2 diabetes stratified by metformin treatment (a known modulator of the gut microbiome) (28). Relative abundances and cleaned coverage profiles for a total of 239 bacterial genera and 637 species were obtained with SLIMM (29) from previously published metagenomic reads (27, 29) as described in the Methods section. We used ratios in coverage between replication initiator and terminus as a measure for replication rates, which have been reported to be good proxies for bacterial growth rates *in vivo* (30). This provided a set of 1571 strain level replication rate measurements across the 186 samples that were used for validation of the inferred growth rates (1062 and 1113 on genus and species levels respectively, see Methods). Abundance profiles for

all identified genera across all samples were connected with the AGORA models, a set of manually curated metabolic models which currently comprises 818 bacterial species (31). 93 gut-associated genera within the AGORA reconstructions represented more than 96% of metagenomic reads across the 186 samples (91% vs 94%, see Table 1 “genus” row). Even though the cooperative tradeoff strategy is applicable to species or even strain-level data, the AGORA reconstructions accounted for only 52% of all bacterial species in the data set. Thus, we decided to perform community model construction separately on species level as well as the genus level, which covered a larger fraction of the observed microbiome. To accomplish this, individual strain models from AGORA were pooled into the higher phylogenetic ranks (see Methods). After removing low abundance taxa (<0.1% for genera and <0.01% for species), the resulting communities contained between 12 and 30 taxa at the genus level and between 23 and 81 taxa at the species level. Each taxon was represented by a full genome-scale metabolic model and connected by exchange reactions with the gut lumen, thus yielding two sets of 186 complete metagenome-scale metabolic models (one set for the species level and one for the genus level). We used the relative read abundances as a proxy for the relative biomass of each taxa in each sample (see Methods). Even though relative abundances from shotgun metagenomes are not an exact representation of bacterial mass (in grams dry weight), we argue that the discrepancy between the two is probably much smaller than the variation in taxon abundances, which spans several orders of magnitude (18).

taxa	unique taxa	assigned reads	with model
kingdom	1	100% ± 0%	100% ± 0%
phylum	22	100% ± 0%	99% ± 0%
class	32	100% ± 0%	99% ± 0%
family	102	100% ± 0%	91% ± 0%
genus	239	94% ± 5%	91% ± 5%
species	637	79% ± 9%	52% ± 9%

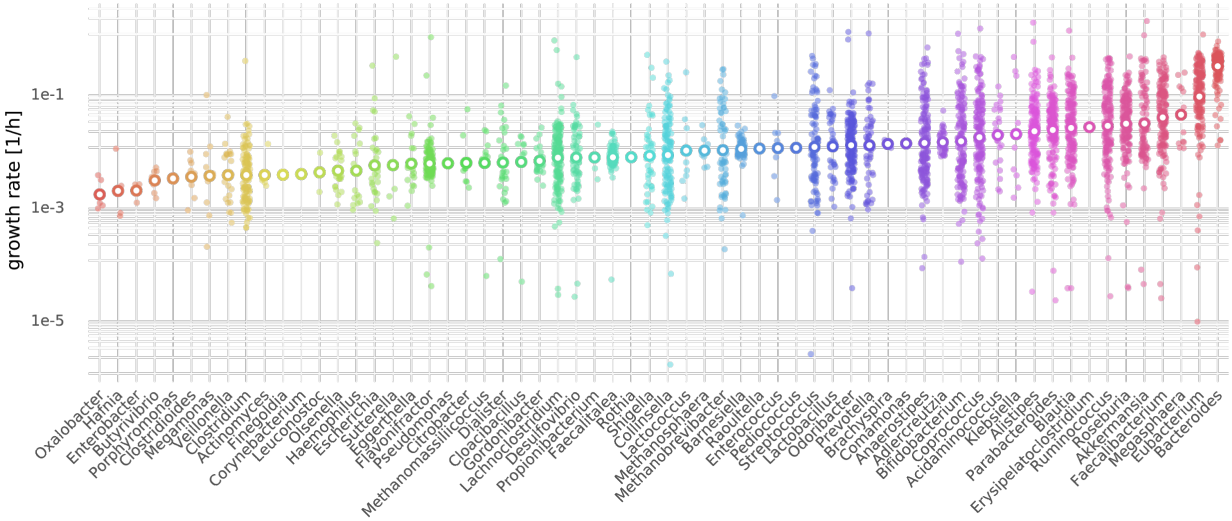
Table 1: Distribution of taxa assignments across ranks. Only reads classified as bacteria were considered. Shown are the number of unique taxa for each rank together with the percentage of mapped reads that could be uniquely assigned to taxa within the rank, as well as the percentage of reads whose taxa had at

least one representative in the AGORA genome-scale metabolic models. Percentages are shown as mean ± standard deviation across the 186 samples.

The data on 186 individuals used in this analysis did not include diet, metabolomics, or data on total microbial load. Thus, we were limited to study metabolic effects that are driven by microbiota composition alone and not by additional factors such as diet or total bacterial biomass. To use a moderately realistic set of import constraints for the community models, we modeled all individuals as consuming an average Western diet (32). Import fluxes for external metabolites were based on a reported set of fluxes for an average Western diet (31, 33). To account for uptake in the small intestine, we reduced all import fluxes for metabolites commonly absorbed in the small intestine by a factor of 10.

To evaluate the performance of the cooperative tradeoff we compared the inferred growth rates with the replication rates obtained directly from sequencing data. First, to establish a baseline we ran an optimization that only maximized the community growth rate and used the distribution of growth rates returned by the solver when applying no regularization. This was followed by applying the cooperative trade-off strategy with varying levels of suboptimality ranging from 10% to 100% of the maximum community growth rate. As stated above, we observed that simply optimizing the community growth rate with no regularization of the individual growth rates led to solutions where only a few taxa grew with unreasonably high growth rates (doubling times shorter than 5 minutes), whereas the rest of the microbial community had growth rates near zero (compare Fig. 1B with strategy marked by “none”). Consequently, the resulting model growth rates were uncorrelated with replication rates (mean Pearson rho=-0.02). Adding the L2 norm minimization while maintaining maximum community growth allowed more genera to grow (see Fig. S1) but yielded growth rates that were anticorrelated with replication rates (mean r = -0.11). Lowering the community growth rate to suboptimal levels strongly increased the growing fraction of the population (Fig. S1) and led to a much better agreement with replication rates for tradeoff values smaller than 70% (mean Pearson rho ≅ 0.4). Calculating correlations across all samples rather than within samples showed a similar tendency, with no regularization showing no correlation with replication rates (r = -0.05, p = 0.07) and increased agreement up to a tradeoff of

242 50% (r = 0.21, p = 2e-12). The lower magnitude correlations in the across samples setting is likely
243 due to differences in diet or bacterial load across people that were not taken into account. Overall,
244 the best agreement with the observed replication rates across and within samples was observed
245 at 50% sub-optimal community growth. We observed similar performance with the species level
246 models (Fig. S2). However, the best agreement with in vivo replication rates was observed for a
247 tradeoff parameter of 0.7 (Fig. S2C). Because genus level model performed equally well as
248 species level models but represented a higher percentage of observed reads (compare Table 1)
249 we decided to continue all further analysis with the genus models and a tradeoff parameter of 0.5.
250



251
252 Figure 2: Non-zero growth rates (> 10^{-6}) across genera obtained by cooperative trade-off (50% maximum
253 community growth rate). Each small filled point denotes a growth rate in one of the 186 samples and
254 larger points with white fill denote the mean growth rate for the genus (see Methods). Genera are sorted
255 by mean growth rate from left (lowest) to right (largest).
256
257

Growth rates are heterogeneous and depend on the community composition

A tradeoff of 50% maximal community growth led to good agreement with replication rates. Bacterial communities showed an average doubling time of about 6 hours, where individual genera had an average doubling time of 11 hours. Community growth did not vary substantially across samples (0.246 +- 0.002 1/h) indicating that each individuals' microbiota were almost equally efficient at converting dietary metabolites into biomass at the community level. However, we found that individual genus-level growth rates often varied over five orders of magnitude (Fig. 2). *Bacteroides* was predicted to be the fastest growing genus overall and was closely followed by *Eubacterium*, which is consistent with the ubiquitous presence of these abundant taxa in microbiome samples (34, 35).

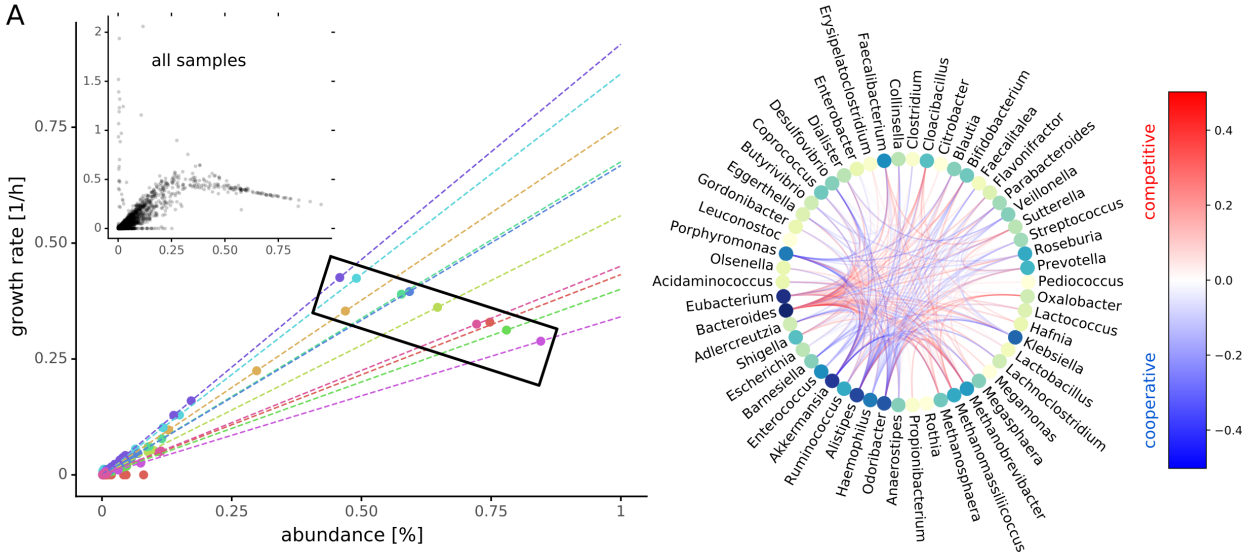


Figure 3: Co-dependencies of growth rates. (A) Relationship between abundance and growth rate across samples. Shown in the larger scatter plot are growth rates and abundances for the first 10 samples. Each dot denotes one genus in one sample and is colored by sample provenance. Dashed lines denote the linear relationship between growth rates and abundances predicted by Equation 2 for each sample. The black box demonstrates how varying slopes (i.e. as community evenness declines, so does the within-sample slope) can result in negative correlation between abundance and growth rate across samples. Smaller inset scatter plot shows data from all samples (Pearson rho=0.69, n=39,815). (B) Growth rate interactions between genera as estimated by genus knockouts. Shown are only interaction that induce a mean growth rate change of 0.1 across all samples (i.e. ubiquitous interactions). Color of edges indicates change of growth rate and type of interaction. Red edges denote competition where removal of one genus increases

the growth rate of the other and blue edges denote cooperation or syntropy where the removal of one genus lowers the growth rate of the other. Nodes are colored by the degree (number of total connections) from lime (few) to dark blue (many).

In the absence of additional constraints, L2 regularization will result in growth rates that are linearly dependent on the taxa abundances (see Equation 2 and Supplemental Text S1). However, this requires some simplifying assumptions that may not be met in the particular constraints of the full metabolic community models. We compared growth rates estimates obtained from numerical optimization with the approximation from Equation 2. We found that growth rates obtained with the cooperative tradeoff usually followed the derived linear relationship, albeit with a large variation (mean R^2 = 0.94, sd = 0.34, Fig. 3A). Deviations from that relationship were mostly observed for small growth rates (see Fig. 3A) which could not reach the suggested growth rate due to additional constraints on growth. Thus, the linear relationship between growth rates and abundance holds for most growth rates but is likely inaccurate for very small growth rates. It is important to note that even though abundances are positively correlated with growth rates within a single individual this is not true across samples where one can observe a negative correlation for abundant taxa (see Fig. 3A). This is a consequence of the coefficient in Equation 2, which depends on the actual abundance distribution as well. In particular, the slope of the linear relationship between abundance and growth rate will be the greatest if all taxa have equal abundances and take its lowest value when one taxon dominates.

We observed a wide variation in individual taxon growth rates across samples. Because all of the community models were constrained under the same diet this phenomenon was due to microbiota composition only. To explain this variation in individual growth rates, we hypothesized that different genera might influence each other's growth rate, either by competition or by cooperation. In order to quantify growth rate interdependencies we performed *in silico* knockouts for each genus in each sample and tracked the change in growth rates for all remaining genera in the sample (see Methods). Here we found that the growth rate of each genus was influenced by another genus in at least one of the 186 samples. As would be expected for bacterial species competing for the same resources, most interactions were competitive (red edges in Fig. 3B). However, we observed a distinct subset of bacteria that were interconnected by a network of cooperative interactions, including *Akkermansia* and *Faecalibacterium* (blue edges in Fig. 3B, also Fig. S3). Strikingly, genera participating in many interactions across all samples, such as

316 *Bacteroides*, *Eubacterium*, *Akkermansia*, *Alistipes* and *Faecalibacterium*, are known to be
317 ubiquitous members of the gut microbiome and are often associated with health (6, 36–40). We
318 found that the prevalence and strength of interactions were highly dependent on the composition
319 of the microbiome. The vast majority of strong growth interactions were present in only 1-5
320 samples, whereas all other samples showed very few strong interactions (Fig. S3). This result is
321 perhaps unsurprising, as many strong species-species interactions are thought to be destabilizing
322 to ecological communities (41, 42).
323

324 Analysis of exchange fluxes reveals differential use of diet components and
325 niche partitioning in the microbiota

326
327 One of the major modes of interaction between the gut microbiota and the host is by means of
328 consumption or production of the metabolite pool in the gut. In our simulations all individuals were
329 under the same average Western diet (see Methods), which imposes an upper bound for the flux
330 of metabolites into the gut lumen. However this does not determine *a priori* which components of
331 the diet are consumed at what rate in each sample, because individual microbiota may consume
332 less than what is imposed by that maximum diet flux. We quantified this effect by obtaining all
333 import and export fluxes for each individual genus across all samples (1,613 exchange reactions
334 in each of 62 genera) as well as metabolite exchanges between the microbiota and the gut lumen
335 (152 metabolites). This was done in the absence of a metabolic model for enterocytes,
336 colonocytes or goblet cells due to the lack of a curated metabolic reconstruction and validated
337 objective function for those cells. A unique set of exchange fluxes was obtained by considering
338 the set of exchange fluxes with smallest total import flux for the growth rates obtained by the
339 cooperative trade-off (see Methods). This assumes that the microbiota competes for resources
340 with the host gut and will thus favor an efficient import that yields the maximum growth rate. This
341 also corresponds to the particular distribution of import fluxes an individual microbiota is most
342 adapted to.
343
344 Even though the minimization of total import fluxes favors simpler media compositions, most
345 samples showed a diverse consumption of metabolites from the gut, particularly in the wide array
346 of carbon and nitrogen sources (Fig. 4A). There was a large set of metabolites that were
347 consumed across all samples but we also observed many metabolites with differential import

fluxes across individuals. In particular, we observed a bimodal distribution where microbiota either consumed fibers and starches or branched chain amino acids (see indicated metabolites in Fig. 4A). This bimodal pattern did not correlate with health or disease states. As expected, all communities showed net anaerobic growth.

Given the observed heterogeneity in taxon growth rates and the large number of interactions we were interested in looking at the uptake rates of metabolites from the gut lumen by each genus. There was overlap of metabolite usage across genera. On average 32% of the metabolites were shared between any two genera in any sample (standard deviation of 13%, see Methods).

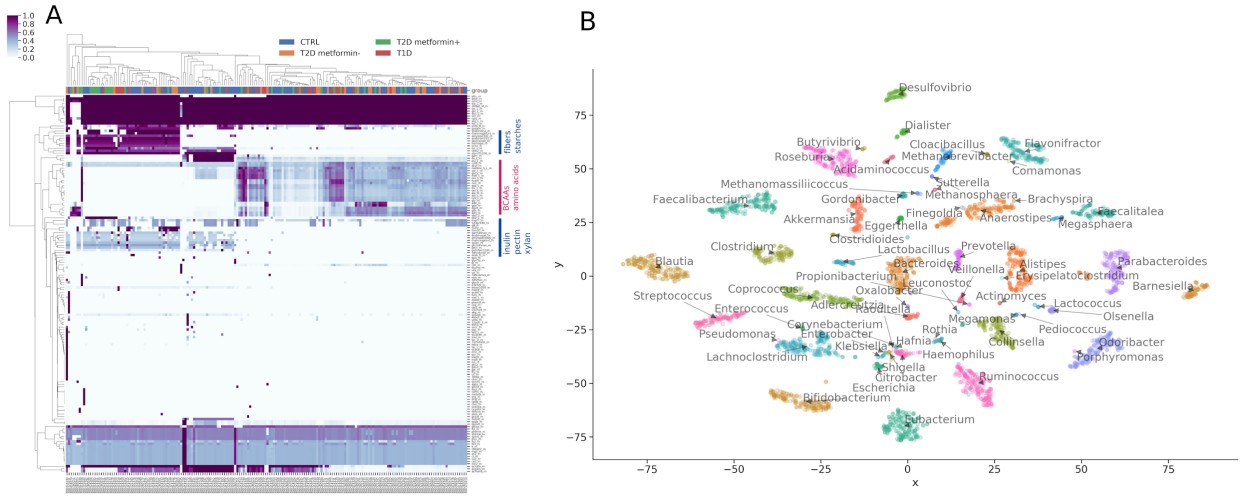


Figure 4: Microbiota import fluxes across samples. Exchange fluxes were calculated as the smallest set of import fluxes that could maintain the genera growth rates obtained by the cooperative trade-off. (A) Import fluxes across samples. Rows were normalized to their absolute maximum and colors denote the import rate ranging from 0% to 100% maximum import. Metabolite groups of interest are marked by blue and red lines (BCAA = branched chain amino acids). Column headers are colored by host metabolic health state (CTRL = metabolically healthy, T1D = type 1 diabetes, T2D = type 2 diabetes, metformin +/- = with/without metformin treatment). (B) Growth niche map for gut genera. Import fluxes for each genus in each sample were reduced to two dimensions using t-SNE. Each point denotes a genus in one sample and is colored and named by its genus.

Consequently, more than two thirds of metabolites were used differentially between pairs of genera. To visualize the structure of metabolite consumption by individual bacterial genera in the gut we used t-SNE dimensionality reductions on the individual genus-specific import fluxes (43). This revealed clear genus-specific niche structure across samples, where individual genera could

be uniquely identified by their particular set of import fluxes (Fig. 4B). Here, taxa closer to one another overlap more in consumed metabolites (Fig. 4B). For instance, *Bacteroides* and *Prevotella* were relatively close to each other in the center of the map, which may help explain the observed tradeoff between *Bacteroides* and *Prevotella* abundances across humans (44). *Blautia*, *Desulfovibrio*, *Bifidobacterium* and *Eubacterium* had the most unique growth niches overall. Genus identity alone explained 61% of the variance in import fluxes (Euclidean PERMANOVA p=0.001). Thus, there was extensive growth niche partitioning between bacterial genera.

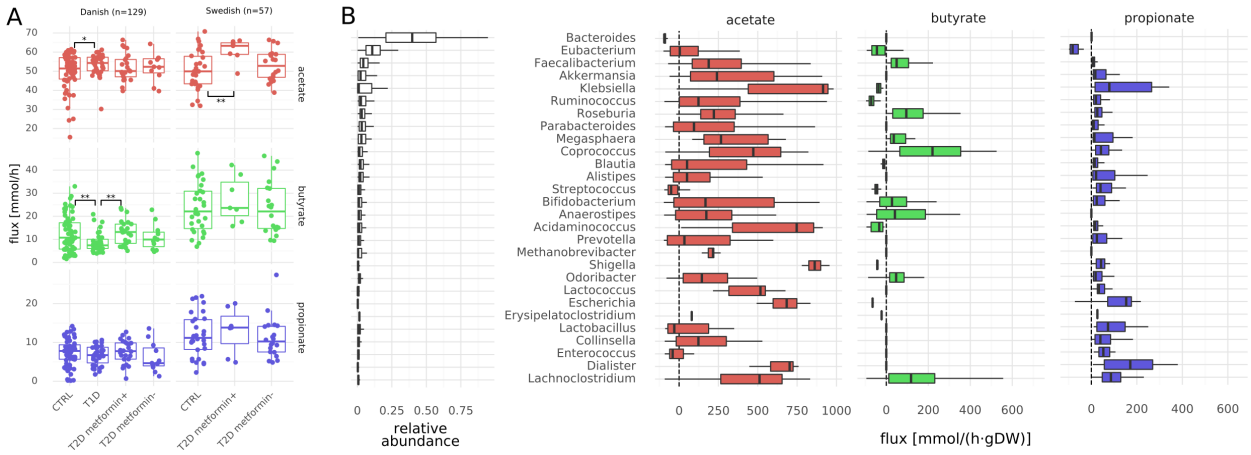


Figure 5: SCFA fluxes. (A) Production capacities of the major SCFAs stratified by population. Fluxes denote total amount of SCFA produced by 1g of bacterial biomass in the gut. Stars denote significance under Welch's t-test (* p<0.05, ** p<0.01). (B) Genus-specific fluxes for the three major SCFAs. Shown are only genera with a relative abundance >1%. Fluxes denote total production/consumption for each genus (see Methods) and are directed towards exports. Thus, positive fluxes denote production of the metabolite and negative fluxes consumption. Genera are ordered by average relative abundance (relative abundances shown in the first column) from top to bottom.

SCFA production is driven by extensive cross-feeding within the microbiota and can be modulated by personalized interventions

Given the association between SCFAs and disease phenotypes we investigated the degree of SCFA production by the model microbiota (2, 45, 46). Intestinal cells have access to the full pool of SCFAs in the gut lumen and would probably take up a significant fraction of those extracellular SCFAs. Thus, the total export flux of any SCFA into the gut lumen by all taxa in a specific model is a measure of host-available SCFA production by the microbiota (see Methods for details on

399 computation). Overall SCFA production for the major SCFAs showed large variations even in
400 healthy individuals, which indicates a large impact of gut microbiota composition on SCFA
401 availability. In particular we observed that Swedish individuals showed higher SCFA production
402 rates than the Danish individuals in the study. Butyrate production was diminished by about 2-fold
403 in Danish individuals with type 1 diabetes (Welch's t-test, p=0.004) but not in Swedish individuals.
404 Danish and Swedish individuals with type 1 diabetes had microbiota that produced more acetate
405 than healthy individuals (Welch's t-test, p=0.02 and 0.003, respectively; Fig. 5A). Metformin
406 treatment had a moderate effect in increasing butyrate productions in both cohorts, however this
407 effect was strongest when comparing Danish individuals with T1D and metformin-treated danish
408 individuals with T2D (Welch's t-test, p=0.003). Higher production of SCFAs was usually
409 accompanied by an increased consumption of SCFAs within the gut microbiota (Fig. S4). This is
410 consistent with prior findings in Danish and Chinese populations (4, 27, 47).
411
412 Decreases in butyrate production were usually accompanied by increases in acetate production.
413 This appeared to indicate SCFA cross-feeding within the microbiota, which we confirmed by
414 comparing the total production and consumption fluxes for each bacterial genus across all
415 samples (Fig. 5B). We observed that butyrate was almost exclusively formed in an acetate-
416 dependent manner from acetyl-CoA, which is the most prevalent butyrate-production pathway in
417 bacteria (48). In particular, production of butyrate depended on acetate production in the
418 community. This was enabled by an extensive cross-feeding between the genera. All SCFAs were
419 produced by a heterogeneous set of taxa, with acetate and propionate production being spread
420 out across most taxa in the system and butyrate produciton being somewhat more restricted to a
421 smaller set of taxa (Fig. 5B). Across samples, the most efficient butyrate producers were
422 *Faecalibacterium*, *Coprococcus*, *Roseburia*, *Anaerostipes* and *Lachnoclostridium*, all of which are
423 known butyrate producers (48, 49). However, the models also predicted consumption of acetate
424 by *Bacteroides* and consumption of butyrate by *Eubacterium*, which is not commonly observed *in*
425 *vivo.* Production of SCFAs was complemented by several other genera, generating a network of
426 SCFA cycling within the microbiota. SCFA production by any genus showed high variation across
427 samples and in some cases would even switch between consumption and production of a
428 particular SCFA, which shows how specific SCFA production is to a particular microbial
429 community (compare Fig 5B). Net production of SCFAs was low compared to overall production
430 (Fig. 5A and S4B) which indicates that most SCFAs in our models were cycled within the bacterial
431 community.
432

Finally, as a proof of concept for the utility of MICOM, we aimed to quantify the impact of targeted interventions on the net consumption or production of SCFAs by the microbiota. For this we chose three Swedish samples (normal, T2D metformin-, T2D metformin+). The impacts of a particular univariate interventions were then quantified by using elasticity coefficients (50, 51), which are dimensionless measures of how strongly a parameter affects a given flux (see Methods). Univariate interventions included increasing the availability of a single metabolite in the diet or increasing the abundance of a single bacterial genus. We observed that the effect of these single interventions were very heterogeneous across all 3 samples (see Fig. 6). The strongest and most common observed effects were to diminish overall SCFA production. However, we observed a few interventions that were able to weakly increase SCFA production. There was a distinct set of metabolites that would increase butyrate production in the T2D individuals but not in the healthy individual. This was also dependent on metformin status. For instance, Arabinan increased butyrate production in the metformin positive individual and D-Xylose increased butyrate production in the metformin negative individual, but not *vice versa*. Thus, MICOM is able to explore the potential functional consequences of targeted dietary or probiotic interventions, which can differ greatly depending on the context of the microbiota in which the interventions are made.

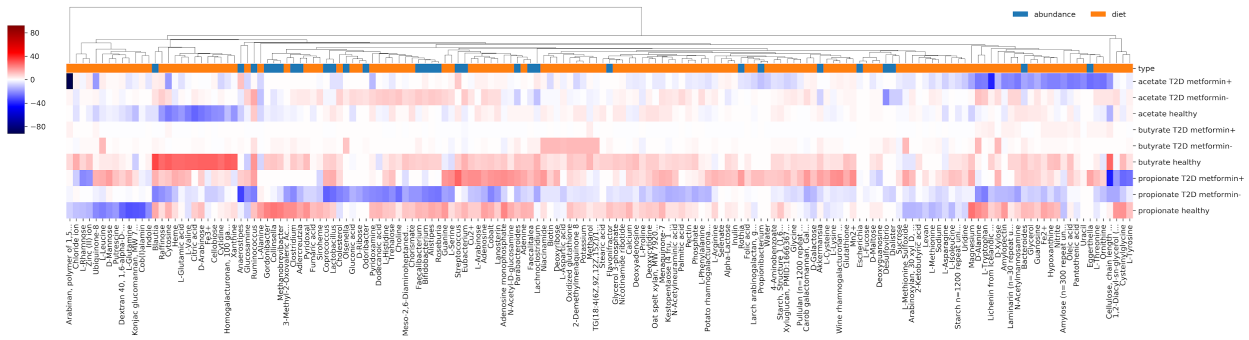


Figure 6: Effect of interventions on SCFA production in 3 samples. Each row denotes a SCFA in a specific individual and each column either denotes a diet component or bacterial genus. Colors denote the elasticity (i.e. the percent change in SCFA production given a percent increase in the specific effector). Red denotes interventions that would increase SCFA production and blue interventions that would decrease production. Shown are only interventions with non-zero elasticities in at least one sample.

Discussion

There is a large amount of sequencing data on microbial communities available today. This is mostly due to the falling cost of 16S rRNA sequencing or shallow shotgun sequencing (52). There

is wide interest in extracting information from sequencing data that goes beyond bacterial proportions (53). Metabolic modeling incorporates a rich knowledge base from genomics and biochemistry and is a valuable resource for adding value to existing datasets. Specifically, MICOM allows for the integration of genome-scale metabolic models, dietary information in the form of import flux bounds, and abundance data from metagenomic shotgun or marker gene sequencing. This framework enables *in silico* mechanistic predictions concerning ecological interactions within microbial communities, inferred exchanges between microbial communities and their environment, and mechanistic hypotheses for how metabolic interactions can be modulated by changes in the environment. The design of a reasonable metagenome-scale metabolic models is challenging due to apparent tradeoffs between individual and community growth rates and issues with computational tractability. Here, we provided a viable strategy that allows for complex analysis of the metabolic consequences of variation in microbial community composition. Our regularization strategy allows for fast identification of unique sets of individual growth rates, which operate in biologically realistic ranges. Our assumption that there is a tradeoff between community growth rate and individual taxon growth rates is supported by the observation that most microbial communities are composed of a large number of species with non-negligible abundances. Individual growth rates for bacterial genera varied greatly across samples (Fig. 2) and were tightly coupled to genus abundances within a sample (Fig. 3A). However, there may be other regularization strategies that provide better agreement with the underlying biology. Our validation strategy using replication rates obtained directly from metagenomic data provides a simple framework to test new regularization functions in the future. It seems that the large variation of growth rates can be explained by a dependency of the growth rate on the presence of other bacteria in the sample (Fig. 3B). Thus, bacterial growth in the gut microbiota is not only dictated by abundance but also by taxon-taxon interactions.

Our predictions are somewhat limited by a variety of factors. For instance, the lack of metabolic models for the major cell types of the gut epithelium (especially goblet cells, enterocytes and colonocytes) and sample-specific metabolite availability in the gut lumen limits the accuracy of MICOM's predictions. Additionally, the use of representative models for bacterial genera may also have caveats. Available metabolic reconstructions are often based on laboratory strains that may not represent the exact metabolic capacity of strains in the human gut. Thus, reconstructions may lack certain metabolic pathways present in the sample and yield inaccurate results, especially when not applying appropriate bounds for the underlying diet (33, 54). Model-predictions may become more quantitative as better personalized data becomes available. The incorporation of

494 personalized data on diet and a better grasp on what fraction of metabolites are absorbed in the
495 small intestine should help to improve model-based predictions. Personalized reconstruction of
496 microbial community metabolic models directly from metagenomic sequencing data may provide
497 more accurate predictions as well, but this approach is currently limited by insufficient sequencing
498 coverage for low abundance taxa.
499
500 MICOM provided valuable ecological insights into the gut microbiota. For instance, the
501 cooperative tradeoff in Equation 2 indicates that a more diverse microbiome (i.e. higher evenness)
502 results in higher individual growth rates (on average) due to the magnitude scaling of the
503 abundance vector (denominator in Equation 2). We also found strong niche partitioning in the
504 model, where taxa showed minimal overlap with each other in resource utilization space. This
505 minimal overlap implies that there is likely an upper bound on alpha diversity in the gut, due to
506 the fact that growth niches eventually saturate and limit the number of taxa that can engraft. Even
507 though only about a third of metabolites were consumed by any pair of taxa in the models, this
508 small amount of niche overlap still resulted in resource competition between taxa. This was
509 particularly true for dominant taxa (e.g. *Bacteroides*), which tended to show competitive
510 interactions with many other genera, likely due to the comparatively higher resource requirements
511 of these abundant taxa for maintaining growth. This community-wide resource competition fits
512 well with the observed growth dependency on amino acid import fluxes across all taxa (Fig. 3A),
513 which is consistent with prior work that suggest that nitrogen may be the global limiting factor for
514 microbial growth in the gut (55). Finally, the methods here extend to any ecosystem containing
515 many microbial taxa and is applicable to abundance data summarized at various phylogenetic
516 ranks. As such MICOM can be employed to perform functional analyses on a wide range of
517 microbial ecosystems.
518
519 It has been difficult and time-consuming to obtain empirical evidence for the mechanistic basis of
520 gut-microbiota interactions. MICOM provides a high-throughput platform for generating
521 mechanistic hypotheses and running *in silico* experiments that would be impossible to perform *in*
522 *vivo*. Thus, we feel that the major application for MICOM is to provide detailed functional
523 hypotheses that can serve as targets for experimental validation. For example, MICOM reveals
524 widespread SCFA cross-feeding in the gut microbiota. The mere presence of butyrate producers
525 was not enough for stable SCFA production - acetate production was also required. Furthermore,
526 MICOM generated personalized predictions for how dietary and probiotic interventions influenced
527 SCFA production capacity. This basic approach could be extended to any number of clinically-

528 relevant metabolites. Thus, we hope that the method presented here will aid researchers in
529 leveraging existing gut microbiome data to design and test personalized intervention strategies.
530

531 # Methods

532 ## Data availability and reproducibility

533
534 All data to reproduce the manuscript, intermediate results as well as Python scripts to reproduce
535 the figures in this manuscript are available in a data repository at
536 https://github.com/resendislab/micom_study. Metagenomic reads for the 186 individuals were
537 obtained from Forslund et. al. (27) and can be downloaded from the Sequence Read Archive
538 (https://www.ncbi.nlm.nih.gov/sra) with the SRA toolkit
539 (https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/). A full list of run accession IDs for the
540 individual samples is provided in the data repository ("recent.csv"). All algorithms and methods
541 used here were implemented in a Python package and can be easily applied to different data
542 sets. The Python package "MICOM" (**MI**crobial **COM**munities) along with documentation and
543 installation instructions are available at https://github.com/resendislab/micom. MICOM is based
544 on the popular COBRApy Python package for constraint-based modeling of biological networks
545 and is compatible with its API (56). The cooperative trade-off strategy as described here was
546 introduced to MICOM in version 0.9.0. The AGORA reference reconstructions with an already
547 applied average Western diet can be downloaded from https://vmh.uni.lu/#downloadview. Several
548 methods used in MICOM require an interior point solver with capabilities for quadratic
549 programming problems (QPs) for which there is currently only commercial software available.
550 MICOM supports CPLEX (https://cplex.org) and Gurobi (https://gurobi.org) both of which have
551 free licenses for academic use. Intermediate results that required those solvers are also provided
552 in the data repository to permit reproduction of our major conclusions.

553 ## Metagenomic shotgun data analysis

554 All metagenomic analyses were performed in R using an in-house pipeline which is available as
555 an open source package along with documentation at https://github.com/resendislab/mbtools.
556 Sample FASTQ files were downloaded using the SRA toolkit and trimmed and filtered using the
557 DADA2 "filter_and_trim" function (57) with a left trimming of 10 bp, no right trimming, a quality

cutoff of 10 and a maximum number of 2 expected errors under the Illumina model. Abundances across different taxa levels were then obtained using SLIMM (29) which was chosen because it supported one of the largest references (almost 5,000 reference bacterial genomes). In brief, all sample FASTQ files were first aligned to the SLIMM reference using Bowtie2 saving the 100 best matches for each read. Taxa abundance profiles were then obtained using SLIMM with a window size of 100bps and assembled into a single abundance file. SLIMM coverage profiles resolved to single strains where then used to infer replication rates using the iRep method (58). In brief, coverage profiles were first smoothed with a rolling mean over 5kbp windows and only genomes with at least a mean coverage of 2 and with at least 60% of total length covered were considered. Coverage values were log-transformed, sorted, and the lowest and highest 10% of the data points were removed to obtain the linear part of the curve. Replication rates were then inferred from the slope of a regression on that linear part. An estimate for the minimum coverage was then obtained from the intercept of the regression and only replication rates for strains with a minimum coverage >2 were kept. No correction for GC content was performed. Before model construction, genus-level and species-level quantifications for each sample were matched separately to the AGORA models by name. The final quantification and mapping is provided in the data repository ("genera.csv" and "species.csv" at https://github.com/micom-dev/paper).

Strategies used in MICOM

Flux balance analysis obtains approximate fluxes for a given organism by assuming a steady state for all fluxes in the biological system and optimizing an organism-specific biomass reaction. Using the stoichiometric matrix S which contains reaction in its columns and metabolites in its rows this can be formulated as a constrained linear programming problem for the fluxes v_i (in mmol/[gDW h]):

$$maximize\ v_{bm}$$
$$s.t. Sv = 0$$
$$lb_i \geq v_i \geq ub_i$$

The biomass reaction v_{bm} is usually normalized such that it will produce 1g of biomass which results in a unit 1/h corresponding to the growth rate μ of the organism. The upper and lower bounds (lb_i and ub_i, respectively) impose additional thermodynamic constraints on the fluxes or restrict exchanges with the environment (in the case of exchange fluxes). In order to describe a community model containing several organisms each with a particular abundance a_i (in gDW) one

usually embeds each organism in an external compartment which represents the community environment (for instance the gut lumen for models of the gut microbiota). Adding exchanges for the environment compartment and exchanges between a particular organism and the environment one obtains a community model with the following constraints:

$$maximize\ \mu_c = \sum_i a_i \cdot \mu_i$$

$$s.t. \forall i: Sv = 0$$

$$\mu_i = v_i^{bm} \geq \mu_i^{min}$$

$$lb_i \geq v_i \geq ub_i$$

$$lb_i^{ex} \geq a_i \cdot v_i^{ex} \geq ub_i^{ex}$$

$$lb_i^{m} \geq v_i^{m} \geq ub_i^{m}$$

Here, a_i denotes the relative abundance of genus i, μ_i its growth rate, v_i^{bm} its biomass flux, μ_i^{min} a user specified minimum growth rate, v_i^{ex} the exchange fluxes with the external environment, and lb and ub the respective lower and upper bounds. Additionally, μ_c denotes the community growth rate and v_i^{m} the exchanges between the entire community and the gut lumen. The described constraints are identical to the ones employed in SteadyCom (22, 29). We assigned an upper bound of 100 mmol/[gDW h] for the internal exchange fluxes v_i^{ex}. Assuming a total microbiota biomass of 200 g and a representative bacterial cell dry weight of 2 pg (59), this corresponds to a maximum import or export of more than 100,000 molecules/[cell s]. Diet derived lower bounds with values smaller 10^{-6} mmol/[gDW h] were set to zero as they would have been lower than the numerical tolerance of the solver. Taxa with relative abundances a_i smaller 10^{-3} for the genus models or 10^{-4} for the species models were discarded since they would not be able to affect the external metabolite levels in a significant way but do increase computation time. Internal fluxes v_i received respective bounds of 1000.0 (or 0 if irreversible) making them essentially unbounded. The described constraints are applied to all optimization problems in MICOM and will be further called the “community constraints”.

The cooperative trade-off method consists of two sequential problems. First, maximize the community growth rate μ_c to obtain μ_c^{max}. Using a user specified trade-off α now solve the following quadratic minimization problem:

$$minimize \sum_i {\mu_i}^2$$

$$s.t. \mu_c \geq \alpha \cdot {\mu_c}^{max}$$

$$and\ community\ constraints$$

The knockout for a genus i was performed by setting all fluxes belonging to this genus along with its exchanges with the external environment to zero (lb=0 and ub=0). This is followed by running cooperative trade-off on the knockout model and comparing the growth rates after the knockout with the ones without the knockout.

Solvers and Numerical stabilization

Most genome-scale metabolic models usually do not treat more than 10,000 variables in the corresponding linear or quadratic programming problems. However, in microbial community models we usually treat 10s to 100s of distinct genome-scale models, which makes the corresponding problem much larger. Unfortunately, many open and commercial solvers have difficulties solving problems of that scale, so we also implemented strategies to increase the success rate of those optimizations. All linear and quadratic programming problems were solved using interior point methods as those were much faster than simplex methods for problems with more than 100,000 variables. Here, we used CPLEX (https://www.ibm.com/analytics/cplex-optimizer) but also tested all methods with Gurobi (https://www.gurobi.com). Since growth rates tend to be small we also multiplied the objectives used in the cooperative tradeoff (maximization of community growth rate and minimization of regularization term) with a scaling factor in order to avoid near-zero objective coefficients. A scaling factor in the order of the largest constraint (1000.0) seemed to work well. Nevertheless, the default interior point methods for quadratic problems in CPLEX or Gurobi were usually not capable of solving the minimization of the regularization term to optimality and usually failed due to numerical instability. The solutions reported by the aborted optimization run were usually close to the optimum, but tended to violate some numerically ill-conditioned constraints. To alleviate this problem, we implemented a crossover strategy where we took the solution of the numerically ill-conditioned quadratic interior point method as a candidate solution set μ_i^{ca}. Based on that we then optimized the following linear programming problem in order to restore feasibility:

$$maximize\ \mu_c = \sum_i a_i \cdot \mu_i$$

$$s.t.\ \mu_i \leq {\mu_i}^{ca}$$

$$and\ community\ constraints$$

Linear interior point methods are usually numerically stable so this linear programming problem can usually be solved to optimality. The maximization together with the new constraints will push the individual growth rates towards the candidate solution as long as it is numerically feasible.

Minimal media and exchange fluxes

By convention MICOM formulates all exchange fluxes in the export direction so that all import fluxes are positive and export fluxes are negative. Based on this, the minimal medium for a community was obtained by minimizing the total import flux:

$$minimize\ v_{tot} = \sum_i \{|{v_i}^m|, {v_i}^m < 0\}$$

$$s.t. \forall i: \mu_i \geq {\mu_i}^{ct}$$

$$\mu_c \geq \alpha \cdot {\mu_c}^{max}$$

$$and\ community\ constraints$$

Here μ_i^{ct} denotes the optimal genera growth rates obtained by cooperative trade-off. The community exchanges were then obtained by extracting all v_i^m, whereas genus-specific exchanges were given by all v_i^{ex} as defined earlier.

Overall production fluxes were calculated as

$${v_{tot}}^m = \sum_{i,v_i^m>0} a_i \cdot {v_i}^m,$$

where v_i^m denotes an exchange flux for the metabolite m in taxon i. Overall consumption rates were calculated in a similar manner but restricting fluxes to ones with $v_i^m<0$ (imports).

Single target intervention studies

We used elasticity coefficients (50, 51) to evaluate the sensitivity of exchange fluxes to changes in exchange flux bounds (*ergo* diet changes) or changes in genus abundances. The logarithmic formulation of elasticity coefficients is given by

$${\varepsilon_p}^v = \frac{\partial\ ln\ |v|}{\partial\ ln\ |p|}$$

where v denotes the exchange flux of interest and p the changed parameter. Since the absolute value removes information about the directionality of the flux this was logged separately to

maintain this information. We used a value of 0.1 as differentiation step size in log space, which corresponds to a bound or abundance increase of about 10.5% in the native scale. To enable efficient computation, elasticity coefficients were grouped by the p parameter, then the cooperative trade-off was run once without modification, the p parameter was increased, the cooperative trade-off was run again, and differentiation was performed for all exchange fluxes at once.

Acknowledgements

ORA and CD were supported by an internal grant from the National Institute of Genomic Medicine (INMEGEN/México). SMG and CD were supported by a Washington Research Foundation Distinguished Investigator Award and startup funds from the Institute for Systems Biology.

Author contributions

CD developed/implemented the methods and performed the analysis. SMG helped design the metagenomic analyses and growth rate validations. ORA developed the methods and designed the meta-analysis. All authors wrote the manuscript.

Supplementary Material

Supplementary Text S1. Closed form solutions for regularization.

Additional materials and data are available at https://github.com/micom-dev/micom_study.

Supplemental Figures

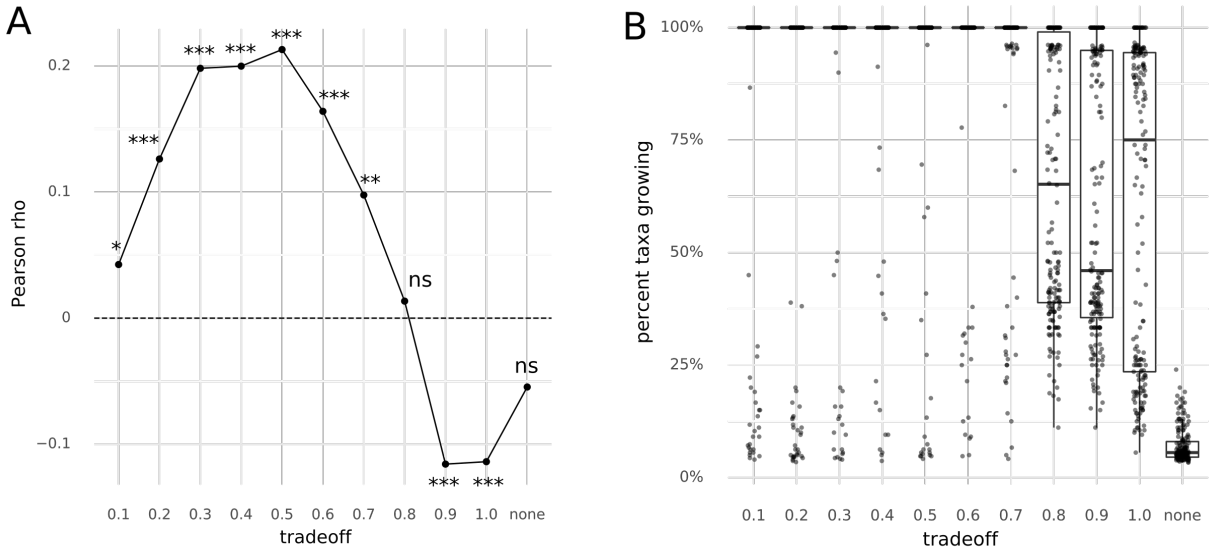


Figure S1. (A) Pearson correlations between inferred growth rates and replication rates across all samples (n=1,062). Stars denote significance of a Pearson test (ns - p>0.05, * - p<0.05, ** - p<0.01, *** - p<0.001). The dashed line denotes a zero correlation coefficient. (B) Fraction of observed genera growing in each sample. Each dot denotes a single sample. A trade-off of “none” means optimization without L2 regularization and only maximizing the community growth rate.

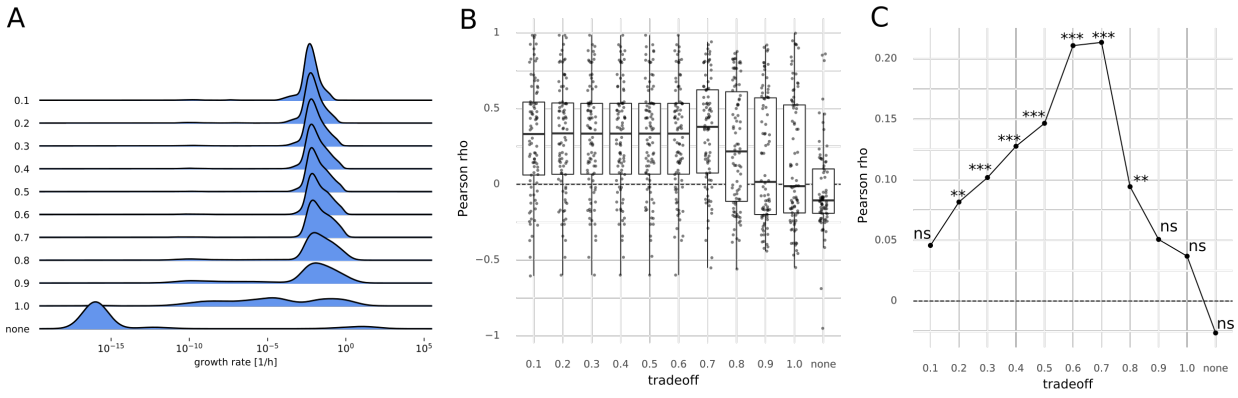
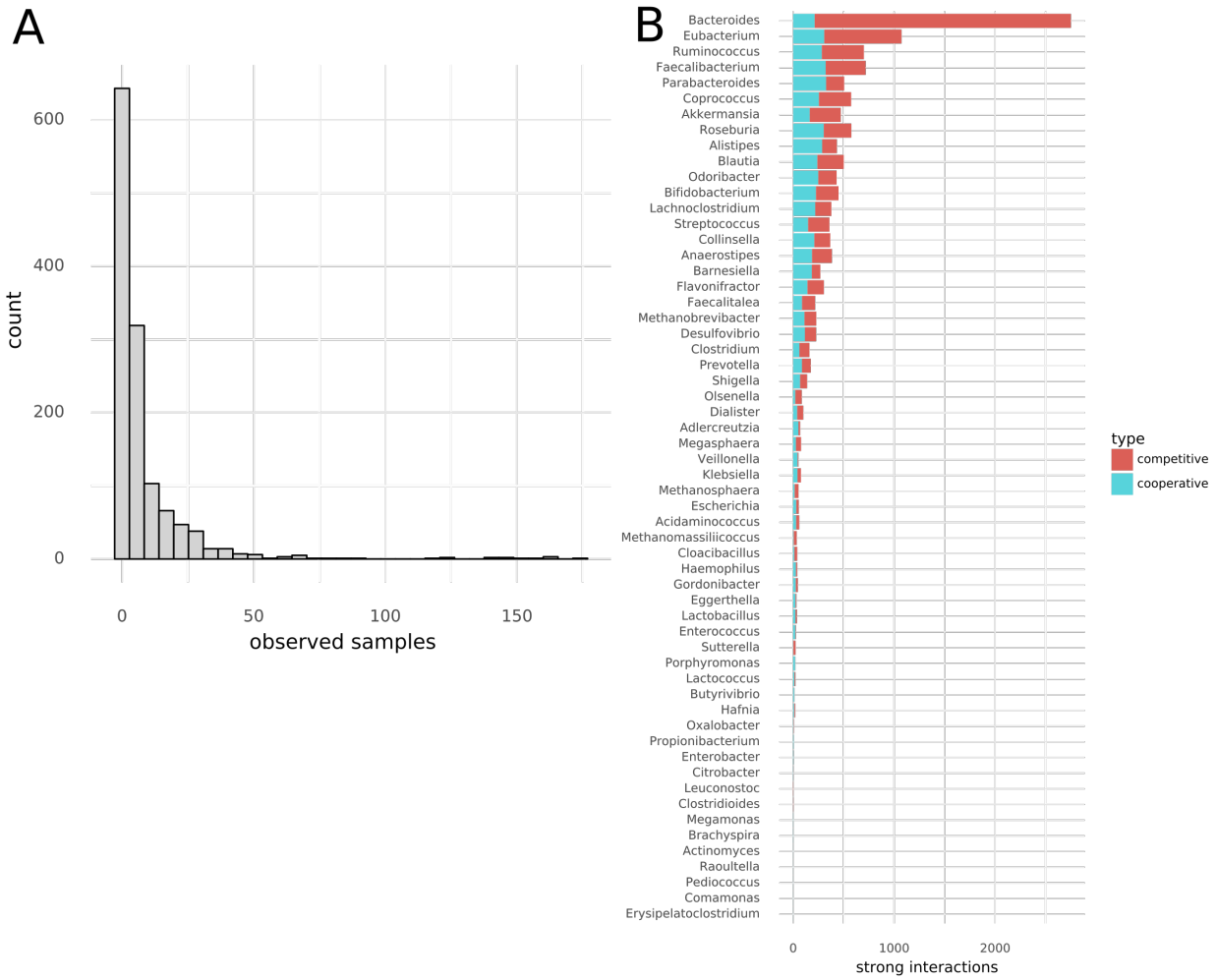


Figure S2. Metrics on the species level. (A) Effect of different trade-off values (fraction of maximum community growth rate) on the distribution of individual species growth rates. Zero growth rates were assigned a value of 10^{-16} which was smaller than the observed non-zero minimum. Growth rates smaller than 10^{-6} were considered to not represent growth. (B) Pearson correlation between replication rates and inferred species growth rates under varying trade-off values. “None” indicates a model without regularization returning arbitrary alternative solutions (see Methods). The dashed line indicates a correlation coefficient of zero. (C) Pearson correlations between inferred species growth rates and

720 replication rates across all samples (n=1,113). Stars denote significance of a Pearson test (ns - p>0.05, *
721 - p<0.05, ** - p<0.01, *** - p<0.001). The dashed lines denotes a zero correlation coefficient.

722

723



724

725

726 Figure S3. Strong interactions are heterogeneously distributed across samples. (A) Histogram of the
727 number of samples a particular interaction occurs in. Lower numbers on the x-axis denote sample-specific
728 interactions and higher numbers denote ubiquitous interactions. (B) Interactions observed across all
729 samples stratified by genus and type.

730

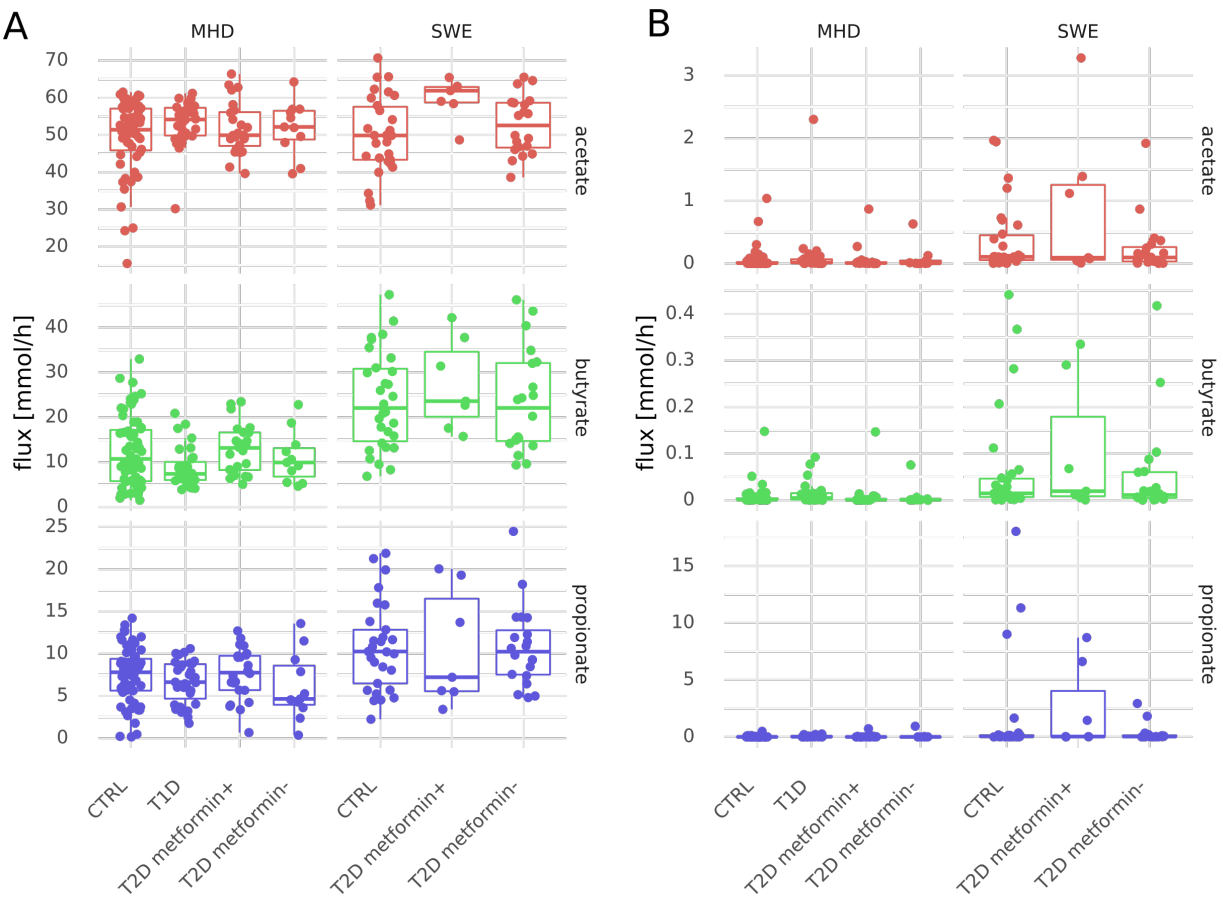


Figure S4. Total consumption and net flux of SCFAs by the microbiota. (A) Total consumption flux of SCFAs. (B) Net production of SCFAs (i.e. difference of total production and total consumption).

References

1. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A, Ben-Yacov O, Lador D, Avnit-Sagi T, Lotan-Pompan M, Suez J, Mahdi JA, Matot E, Malka G, Kosower N, Rein M, Zilberman-Schapira G, Dohnalová L, Pevsner-Fischer M, Bikovsky R, Halpern Z, Elinav E, Segal E. 2015. Personalized Nutrition by Prediction of Glycemic Responses. Cell 163:1079–1094.

2. Cho I, Blaser MJ. 2012. The human microbiome: at the interface of health and disease. Nat Rev Genet 13:260–270.

3. Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, Bittinger K, Bailey A, Friedman ES, Hoffmann C, Albenberg L, Sinha R, Compher C, Gilroy E, Nessel L, Grant A, Chehoud C, Li H, Wu GD, Bushman FD. 2015. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host Microbe 18:489–500.

4. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng Y, Zhang D, Jie Z, Wu W, Qin Y, Xue W, Li J, Han L, Lu D, Wu P, Dai Y, Sun X, Li Z, Tang A, Zhong S, Li X, Chen W, Xu R, Wang M, Feng Q, Gong M, Yu J, Zhang Y, Zhang M, Hansen T, Sanchez G, Raes J, Falony G, Okuda S, Almeida M, LeChatelier E, Renault P, Pons N, Batto J-M, Zhang Z, Chen H, Yang R, Zheng W, Li S, Yang H, Wang J, Ehrlich SD, Nielsen R, Pedersen O, Kristiansen K, Wang J. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490:55–60.

5. Livanos AE, Greiner TU, Vangay P, Pathmasiri W, Stewart D, McRitchie S, Li H, Chung J, Sohn J, Kim S, Gao Z, Barber C, Kim J, Ng S, Rogers AB, Sumner S, Zhang X-S, Cadwell K, Knights D, Alekseyenko A, Bäckhed F, Blaser MJ. 2016. Antibiotic-mediated gut microbiome perturbation accelerates development of type 1 diabetes in mice. Nat Microbiol 1:16140.

6. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. 2017. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun 8:1784.

7. Xu Z, Malmer D, Langille MGI, Way SF, Knight R. 2014. Which is more important for classifying microbial communities: who's there or what they can do? ISME J 8:2357–2359.

8. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol 31:814–821.

9. Aßhauer KP, Wemheuer B, Daniel R, Meinicke P. 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics 31:2882–2884.

10. Bauer E, Thiele I. 2018. From Network Analysis to Functional Metabolic Modeling of the Human Gut Microbiota. mSystems 3.

11. Heinken A, Sahoo S, Fleming RMT, Thiele I. 2013. Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut. Gut Microbes 4:28–40.

12. Resendis-Antonio O, Reed JL, Encarnación S, Collado-Vides J, Palsson BØ. 2007. Metabolic reconstruction and modeling of nitrogen fixation in Rhizobium etli. PLoS Comput Biol 3:1887–1895.

13. Orth JD, Thiele I, Palsson BØ. 2010. What is flux balance analysis? Nat Biotechnol 28:245–248.

14. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, König R, Smith RD, Palsson BØ. 2010. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol Syst Biol 6:390.

15. Medlock GL, Carey MA, McDuffie DG, Mundy MB, Giallourou N, Swann JR, Kolling GL, Papin JA. 2018. Inferring Metabolic Mechanisms of Interaction within a Defined Gut Microbiota. Cell Systems 7:245–257.e7.

16. Long MR, Ong WK, Reed JL. 2015. Computational methods in metabolic engineering for strain design. Curr Opin Biotechnol 34:135–141.

17. Zomorrodi AR, Maranas CD. 2012. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput Biol 8:e1002363.

18. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y,

González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R, Earth Microbiome Project Consortium. 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551:457–463.

19. Gibbons SM, Kearney SM, Smillie CS, Alm EJ. 2017. Two dynamic regimes in the human gut microbiome. PLoS Comput Biol 13:e1005364.

20. Johnson AJ, Vangay P, Al-Ghalith GA, Hillmann BM, Ward TL, Shields-Cutler RR, Kim AD, Shmagel AK, Syed AN, Personalized Microbiome Class Students, Walter J, Menon R, Koecher K, Knights D. 2019. Daily Sampling Reveals Personalized Diet-Microbiome Associations in Humans. Cell Host Microbe 25:789–802.e5.

21. Poyet M, Groussin M, Gibbons SM, Avila-Pacheco J, Jiang X, Kearney SM, Perrotta AR, Berdy B, Zhao S, Lieberman TD, Swanson PK, Smith M, Roesemann S, Alexander JE, Rich SA, Livny J, Vlamakis H, Clish C, Bullock K, Deik A, Scott J, Pierce KA, Xavier RJ, Alm EJ. 2019. A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. Nat Med 25:1442–1452.

22. Chan SHJ, Simons MN, Maranas CD. 2017. SteadyCom: Predicting microbial abundances while ensuring community stability. PLoS Comput Biol 13:e1005539.

23. The Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214.

24. Engl HW, Hanke M, Neubauer A. 2000. Regularization of Inverse Problems. Springer Science & Business Media.

25. Hoerl AE, Kennard RW. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12:55–67.

26. Potra FA, Wright SJ. 2000. Interior-point methods. J Comput Appl Math 124:281–302.

27. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, Prifti E, Vieira-Silva S, Gudmundsdottir V, Pedersen HK, Arumugam M, Kristiansen K, Voigt AY, Vestergaard H, Hercog R, Costea PI, Kultima JR, Li J, Jørgensen T, Levenez F, Dore J,

MetaHIT consortium, Nielsen HB, Brunak S, Raes J, Hansen T, Wang J, Ehrlich SD, Bork P, Pedersen O. 2015. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature 528:262–266.

28. Wu H, Esteve E, Tremaroli V, Khan MT, Caesar R, Mannerås-Holm L, Ståhlman M, Olsson LM, Serino M, Planas-Fèlix M, Xifra G, Mercader JM, Torrents D, Burcelin R, Ricart W, Perkins R, Fernàndez-Real JM, Bäckhed F. 2017. Metformin alters the gut microbiome of individuals with treatment-naive type 2 diabetes, contributing to the therapeutic effects of the drug. Nat Med 23:850–858.

29. Dadi TH, Renard BY, Wieler LH, Semmler T, Reinert K. 2017. SLIMM: species level identification of microorganisms from metagenomes. PeerJ 5:e3138.

30. Korem T, Zeevi D, Suez J, Weinberger A, Avnit-Sagi T, Pompan-Lotan M, Matot E, Jona G, Harmelin A, Cohen N, Sirota-Madi A, Thaiss CA, Pevsner-Fischer M, Sorek R, Xavier R, Elinav E, Segal E. 2015. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. Science 349:1101–1106.

31. Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K, Jäger C, Baginska J, Wilmes P, Fleming RMT, Thiele I. 2017. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol 35:81–89.

32. Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G, Daníelsdóttir AD, Krecke M, Merten D, Haraldsdóttir HS, Heinken A, Heirendt L, Magnúsdóttir S, Ravcheev DA, Sahoo S, Gawron P, Friscioni L, Garcia B, Prendergast M, Puente A, Rodrigues M, Roy A, Rouquaya M, Wiltgen L, Žagare A, John E, Krueger M, Kuperstein I, Zinovyev A, Schneider R, Fleming RMT, Thiele I. 2019. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Res 47:D614–D624.

33. Magnúsdóttir S, Heinken A, Fleming RMT, Thiele I. 2018. Reply to “Challenges in modeling the human gut microbiome.” Nat Biotechnol.

34. Lagier J-C, Million M, Hugon P, Armougom F, Raoult D. 2012. Human gut microbiota: repertoire and variations. Front Cell Infect Microbiol 2:136.

35. Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Bäckhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R, Bäckhed F, Isolauri E, Salminen S, Ley RE. 2012. Host remodeling of the gut microbiome and metabolic changes during pregnancy. Cell 150:470–480.

36. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, Reyes JA, Shah SA, LeLeiko N, Snapper SB, Bousvaros A, Korzenik J, Sands BE, Xavier RJ, Huttenhower C. 2012. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol 13:R79.

37. Brown K, DeCoffe D, Molcan E, Gibson DL. 2012. Diet-induced dysbiosis of the intestinal microbiota and the effects on immunity and disease. Nutrients 4:1095–1119.

38. Bajaj JS, Hylemon PB, Ridlon JM, Heuman DM, Daita K, White MB, Monteith P, Noble NA, Sikaroodi M, Gillevet PM. 2012. Colonic mucosal microbiome differs from stool microbiome in cirrhosis and hepatic encephalopathy and is linked to cognition and inflammation. Am J Physiol Gastrointest Liver Physiol 303:G675–85.

39. Chen W, Liu F, Ling Z, Tong X, Xiang C. 2012. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS One 7:e39743.

40. Murri M, Leiva I, Gomez-Zumaquero JM, Tinahones FJ, Cardona F, Soriguer F, Queipo-Ortuño MI. 2013. Gut microbiota in children with type 1 diabetes differs from that in healthy children: a case-control study. BMC Med 11:46.

41. McCann KS. 2000. The diversity–stability debate. Nature.

42. Coyte KZ, Schluter J, Foster KR. 2015. The ecology of the microbiome: Networks, competition, and stability. Science 350:663–666.

43. Mahfouz A, van de Giessen M, van der Maaten L, Huisman S, Reinders M, Hawrylycz MJ, Lelieveldt BPF. 2015. Visualizing the spatial gene expression organization in the brain through non-linear similarity embeddings. Methods 73:79–89.

44. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S,

879 Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak
880 S, Doré J, MetaHIT Consortium, Antolín M, Artiguenave F, Blottiere HM, Almeida M,
881 Brechot C, Cara C, Chervaux C, Cultrone A, Delorme C, Denariaz G, Dervyn R, Foerstner
882 KU, Friss C, van de Guchte M, Guedon E, Haimet F, Huber W, van Hylckama-Vlieg J,
883 Jamet A, Juste C, Kaci G, Knol J, Lakhdari O, Layec S, Le Roux K, Maguin E, Mérieux A,
884 Melo Minardi R, M'rini C, Muller J, Oozeer R, Parkhill J, Renault P, Rescigno M, Sanchez
885 N, Sunagawa S, Torrejon A, Turner K, Vandemeulebrouck G, Varela E, Winogradsky Y,
886 Zeller G, Weissenbach J, Ehrlich SD, Bork P. 2011. Enterotypes of the human gut
887 microbiome. Nature 473:174–180.

888 45. Kinross JM, Darzi AW, Nicholson JK. 2011. Gut microbiome-host interactions in health and
889 disease. Genome Med 3:14.

890 46. Tan J, McKenzie C, Potamitis M, Thorburn AN, Mackay CR, Macia L. 2014. The role of
891 short-chain fatty acids in health and disease. Adv Immunol 121:91–119.

892 47. Tremaroli V, Bäckhed F. 2012. Functional interactions between the gut microbiota and host
893 metabolism. Nature 489:242–249.

894 48. Vital M, Karch A, Pieper DH. 2017. Colonic Butyrate-Producing Communities in Humans:
895 an Overview Using Omics Data. mSystems 2.

896 49. Baxter NT, Schmidt AW, Venkataraman A, Kim KS, Waldron C, Schmidt TM. 2019.
897 Dynamics of Human Gut Microbiota and Short-Chain Fatty Acids in Response to Dietary
898 Interventions with Three Fermentable Fibers. MBio 10:e02566–18.

899 50. Savageau MA. 2010. Biochemical Systems Analysis: A Study of Function and Design in
900 Molecular Biology. CreateSpace.

901 51. Heinrich R, Rapoport TA. 1974. A Linear Steady-State Treatment of Enzymatic Chains.
902 General Properties, Control and Effector Strength. Eur J Biochem 42:89–95.

903 52. Hillmann B, Al-Ghalith GA, Shields-Cutler R, Zhu Q, Gohl D, Beckman KB, Knight R,
904 Knights D. 2018. Evaluating the information content of shallow shotgun metagenomics.

905 53. Gilbert JA, Blaser MJ, Gregory Caporaso J, Jansson JK, Lynch SV, Knight R. 2018.
906 Current understanding of the human microbiome. Nat Med 24:392–400.

907 54. Babaei P, Shoaie S, Ji B, Nielsen J. 2018. Challenges in modeling the human gut
908 microbiome. Nat Biotechnol 36:682–686.

909 55. Reese AT, Pereira FC, Schintlmeister A, Berry D, Wagner M, Hale LP, Wu A, Jiang S,
910 Durand HK, Zhou X, Premont RT, Diehl AM, O'Connell TM, Alberts SC, Kartzinel TR,
911 Pringle RM, Dunn RR, Wright JP, David LA. 2018. Microbial nitrogen limitation in the
912 mammalian large intestine. Nat Microbiol 3:1441–1450.

913 56. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. 2013. COBRApy: COnstraints-Based
914 Reconstruction and Analysis for Python. BMC Syst Biol 7:74.

915 57. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2:
916 High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583.

917 58. Brown CT, Olm MR, Thomas BC, Banfield JF. 2016. Measurement of bacterial replication
918 rates in microbial communities. Nat Biotechnol 34:1256–1263.

919 59. Sender R, Fuchs S, Milo R. 2016. Revised Estimates for the Number of Human and
920 Bacteria Cells in the Body. PLoS Biol 14:e1002533.

921