Modeling microbial cross-feeding at intermediate scale portrays community dynamics and species coexistence

Chen Liao[1], Tong Wang[2,3], Sergei Maslov[3,4], Joao B. Xavier[1,*]

[1]Program for Computational and Systems Biology, Memorial Sloan-Kettering Cancer Center, New York, NY, United States of America, [2]Department of Physics, University of Illinois at Urbana-Champaign, IL 61801, USA, [3]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, IL 61801, USA, [4]Department of Bioengineering, University of Illinois at Urbana-Champaign, IL 61801, USA

[*]To whom correspondence may be addressed: J. Xavier (Email: XavierJ@mskcc.org)

Keywords:
cross-feeding; coarse-grained modeling; microbial community; ecological relationship; species coexistence; population dynamics

Abstract

Social interaction between microbes can be described at many levels of details: from the biochemistry of cell-cell interactions to the ecological dynamics of populations. Choosing an appropriate level to model microbial communities without losing generality remains a challenge. Here we show that modeling cross-feeding interactions at an intermediate level between genome-scale metabolic models of individual species and consumer-resource models of ecosystems is suitable to experimental data. We applied our modeling framework to three published examples of multi-strain *Escherichia coli* communities with increasing complexity: uni-, bi-, and multi-directional cross-feeding of either substitutable metabolic byproducts or essential nutrients. The intermediate-scale model accurately fit empirical data and quantified metabolic exchange rates that are hard to measure experimentally, even for a complex community of 14 amino acid auxotrophies. By studying the conditions of species coexistence, the ecological outcomes of cross-feeding interactions, and each community's robustness to perturbations, we extracted new quantitative insights from these three published experimental datasets. Our analysis provides a foundation to quantify cross-feeding interactions from experimental data, and highlights the importance of metabolic exchanges in the dynamics and stability of microbial communities.

Author summary

The behavior of microbial communities such as the human microbiome is hard to predict by its species composition alone. Our efforts to engineer microbiomes—for example to improve human health—would benefit from mathematical models that accurately describe how microbes exchange metabolites with each other and how their environment shapes these exchanges. But what is an appropriate level of details for those models? We propose an intermediate level to model metabolic exchanges between microbes. We show that these models can accurately describe population dynamics in three laboratory communities and predicts their stability in response to perturbations such as changes in the nutrients available in the medium that they grow on. Our work suggests that a highly detailed metabolic network model is unnecessary for extracting ecological insights from experimental data and improves mathematical models so that one day we may be able to predict the behavior of real-world communities such as the human microbiome.

Introduction

Most microorganisms that affect the environments we live in[1] and that impact our health[2] do not live in isolation: they live in complex communities where they interact with other strains and species. The past decade has seen a surge of scientific interest in microbial communities, such as the human microbiome, but most studies remain limited to cataloguing community composition[3]. Our mechanistic understanding of how biochemical processes occurring inside individual microbial cells command interaction between cells, and lead to the emergent properties of multi-species communities remains limited[4].

Microorganisms consume, transform and secrete many kinds of chemicals, including nutrients, metabolic wastes, extracellular enzymes, antibiotics and cell-cell signaling molecules such as quorum sensing autoinducers[5–8]. The chemicals produced by one microbe can impact the behaviors of others by promoting or inhibiting their growth[9], creating multi-directional feedbacks that can benefit or harm the partners involved[10,11].

If a community is well-characterized and given sufficient data on population dynamics, it should be possible to parameterize the processes involved in microbe-microbe interactions by fitting mathematical models[12]. Any model can potentially yield insights[13], but the complexity of most models so far has been either too high for parameterization[14], or too low to shed light on cellular mechanisms[15]. Microbial processes may be modelled across a range of details: At the low end of the spectrum we have population dynamic models such as generalized Lotka-Volterra (gLV)[16] and Consumer-Resource (C-R) models[17], which treat each organism as a 'black-box'. For example, C-R models assume a linear or Monod dependence of microbial growth on resource uptake kinetics. At the high end of the spectrum, we have detailed single-cell models such as dynamic flux balance analysis (dFBA)[18] and agent-based models[19] that have too many parameters

to be parameterizable by experimental data. For example, the linear equations for fluxes obtained from quasi-steady-state assumption of dFBA are underdetermined. What is an appropriate level of detail to model and constrain microbial processes using data, to produce accurate predictions as well as new mechanistic insights?

Here we propose a generalizable framework that couples classical ecological models of population and resource dynamics with coarse-grained intra-species metabolic networks. We show that modeling communities at this intermediate scale can accurately quantify metabolic processes from population dynamics data acquired in the laboratory. We demonstrate the approach on three evolved/engineered communities of *Escherichia coli* (*E. coli*) strains with increasing levels of complexity: (1) unilateral acetate-mediated cross-feeding[20], (2) bilateral amino-acid-mediated cross-feeding between leucine and lysine auxotrophies[21], and (3) multilateral amino-acid-mediated cross-feeding between 14 distinct amino acid autotrophies[22]. The parameterized models report inferred leakage fractions of metabolic byproducts that are difficult to measure directly by experiments, reveal how resource supply and partitioning alter the coexistence and ecological relationships between cross-feeders, and predict the limits of community robustness against external perturbations.

Results

Modeling microbial metabolic processes at an intermediate level. Inspired by the classical MacArthur's CR models[23] and many follow-ups[17,24–26], we propose to integrate CR models with a coarse-grained yet mechanistic description of cell metabolism. Metabolic reactions can be broadly classified as catabolic and anabolic, where catabolic reactions break down complex substrates from culture media into smaller metabolic intermediates that can be used to build up biomass

components by anabolic reactions. A minimal representation of cell metabolism is a three-layer network composed of growth substrates at the top, building block metabolites in the middle, and biomass at the bottom (Fig. 1). The growth substrates can be either substitutable (e.g., glucose and acetate) or non-substitutable (e.g., glucose and ammonium); however, in our model we consider only the non-substitutable building blocks for cell growth. In fact, substitutable metabolites can be mathematically lumped into complementary functional groups that together make a non-substitutable group when coarse-graining metabolic network. Despite its simplicity, this model is flexible enough to describe the transformation of resources into other resources, non-consumable chemicals and biomass, regardless of the specific reactions involved.

Based on these assumptions, we developed a dynamic modeling framework that contains eight kinds of biochemical reactions describing resource uptake, transformation, secretion, utilization, and degradation (Fig. 1, Supplementary Texts 1.1). Briefly, substrates available in the growth media can be imported into cells. A certain fraction of the imported substrates is then broken down into building block metabolites, which can be released back to the surrounding environment, used by cells for biomass production, consumed by other non-growth processes, and degraded. Secretable metabolites, when released, can be imported by cells in a way similar to externally supplied substrates, except that their uptake may be inhibited by other substitutable substrates that are assumed to be preferentially used (e.g., catabolite repression). The dynamics of population size change is affected by two elements: population growth and cell death, where the former may depend on both building blocks and substrates. Here the substrate dependency lumps the growth effects from metabolites that are not explicitly modeled, which can substantially reduce model size by defining and choosing model variables for only metabolites known to mediate

interpopulation interactions. To model the effects of toxic compounds[27] we allow the growth rate of any cell population can be inhibited by accumulation of toxic metabolites in the environment.

The eight types of reactions can be translated to differential equations. We assumed quasi-steady-state for intracellular substrates and metabolites, as metabolic reactions typically occur at faster time scales compared to ecological dynamics. The time-scale separation thus simplifies our model by excluding intracellular variables, leaving only three types of variables that describe the population density of active cells ($N_l,\ l = 1,2,,\cdots,n_c$), the extracellular concentrations of substrates ($[S_i],\ i = 1,2,\cdots,n_s$), and the concentrations of metabolic byproducts excreted by cells ($[M_j],\ j = 1,2,\cdots,n_m$). Assuming a chemostat environment with dilution rate D (which reduces to a batch culture when $D = 0$), the differential equations associated with the three state variables are given below (Supplementary Equations (9)-(11))

$$\frac{d[S_i]}{dt} = D\left(S_{0,i} - [S_i]\right) - \sum_{l=1}^{n_c} J_{l,i}^{upt,S} N_l \tag{1}$$

$$\frac{dN_l}{dt} = N_l\left(J_l^{grow} - J_l^{death} - D\right) \tag{2}$$

$$\frac{d[M_j]}{dt} = D\left(M_{0,j} - [M_j]\right) + \sum_{l=1}^{n_c}\left(J_{l,j}^{leak,M} - J_{l,j}^{upt,M}\right)N_l \tag{3}$$

where $S_{0,i}$ and $M_{0,j}$ are the feed medium concentrations of substrate S_i and metabolite M_j respectively. $J_{l,i}^{upt,S}$ and $J_{l,j}^{upt,M}$ represent uptake fluxes of substrates and metabolites respectively, $J_{l,j}^{leak,M}$ are metabolite secretion fluxes, and J_l^{grow} and J_l^{death} stand for per-capita growth and death rates respectively. We used Monod kinetics for resource uptake ($J_{l,i}^{upt,S}$ and $J_{l,j}^{upt,M}$; Supplementary Equation (16) and (17)), derived mathematical expressions for metabolite leakage ($J_{l,j}^{leak,M}$; Supplementary Equation (18) and (19)) and biomass production (J_l^{grow}; Supplementary Equation

(20)) using the Liebig's Law of the Minimum[28] (growth rate is proportional to the flux of the scarcest resource), and modelled cell death using first-order kinetics with constant specific mortality rate (J_l^{death}; Supplementary Equation (23)). The functional forms of these kinetic laws and other details of model derivation are described in Supplementary Texts 1.1.

Example 1: unilateral acetate-mediated cross-feeding. We first applied our modeling framework to a well-documented unilateral acetate-mediated cross-feeding polymorphism evolved from a single ancestral lineage of *E. coli* in laboratory conditions[20] (Supplementary Texts 1.2.1). The community contains two polymorphic subpopulations (*E. coli* subspecies) whose metabolism differs in their quantitative ability to uptake and efflux carbon sources: a glucose specialist strain (CV103) which has a faster glucose uptake rate but cannot grow on acetate, and an acetate specialist strain (CV101) which can grow on acetate but has a lower glucose uptake rate. CV103 secretes acetate—a major by-product of its aerobic metabolism—and this way creates a new ecological niche for CV101. For simplicity, we assumed that glucose and acetate are fully substitutable resources since *E. coli* cells can grow on either carbon source with similar yields (Supplementary Texts 1.2.2). Compared to its complete form (Supplementary Fig. 1), the simplified model diverts all glucose flux to acetate that acts as the only growth limiting factor (Fig. 2A). Using parameters estimated by manual fitting (Materials and methods, Supplementary Table 1), we show that the model accurately reproduced the observed changes in growth and acetate concentration in both monoculture and coculture experiments over time (Fig. 2B-E). Particularly, Fig. 2D shows that acetate is toxic to both strains and CV101 is more susceptible. Although Fig. 2E shows coexistence of CV101 and CV103 within 40 generations, our model predicts that CV103 would be eventually excluded from the community in the long run (Supplementary Fig. 2).

The simplified model has 11 parameters, including 8 free parameters, 2 parameters fixed at literature values, and 1 biological constant (Supplementary Table 1). To assess parameter uncertainty, we sampled posterior distribution of all free parameters using Markov-Chain-Monte-Carlo (MCMC) algorithm (Material and methods), finding that their medians coincide well with the default values obtained by manual fitting and used in the simulations (Supplementary Fig. 3, Supplementary Table 1). Compared to other free parameters, $C_{1,g}$ (half maximum inhibitory concentration of glucose for acetate uptake by CV101) and $I_{3,a}$ (half maximum inhibitory concentration of acetate for CV103 growth) have much wider distributions, suggesting the dataset (Fig. 2B-E) used to constrain the model is relatively insensitive to changes in their values. We did not find strong correlations among parameters, except for the maximum glucose uptake rate of CV101 and CV103 ($V_{1,g}$ and $V_{3,g}$ respectively), which has a Pearson correlation coefficient (PCC) 99.6%. Particularly, the distribution of the acetate leakage fraction has a median 36.7% with interquartile range from 29.8% to 44.6%, which is consistent with the manually optimized value 33.0%. This value suggests that both cell types have nearly equal carbon flux values between acetate secretion and glucose uptake, a quantitative relationship that has been observed in a different *E. coli* strain[29]. The high efflux of acetate may be a consequence of adaptive co-evolution and accumulation of degenerative mutations[20].

Our model indicated that the competition outcome depends on the acetate level in the feed medium (Fig. 2E): CV103 dominates the community without acetate supplementation while CV101 dominates when 1 mM acetate was supplemented. Fig. 2F outlines the region in the nutritional space when CV101 grows faster than (gray shading) and equal to (shading boundary) CV103. The region has a bell shape with the maximum at 0.81 mM glucose and is almost symmetric around 1 mM acetate. The dose-dependent growth effects can be explained by the

175 conflicting role of acetate which is both a source of carbon and a toxic waste. Acetate at low
176 concentration serves as nutrient for CV101 and increases its growth rate. However, too much
177 acetate is toxic and has stronger inhibitory effects on the growth of CV101 compared to CV103
178 (Fig. 2D). The growth advantage of CV101 conferred by an intermediate level of acetate can be
179 negated at high glucose level (> 0.81 mM) due to strong carbon catabolite repression resulting in
180 reduced assimilation of acetate by CV101.
181
182 **Coexistence of CV101 and CV103.** Coexistence of CV101 and CV103 requires that the growth
183 rate of both strains is equal to the dilution rate. The nutritional space has two solutions (Fig. 2F,
184 gray circles) that satisfy the criteria at dilution rate of 0.2 h^{-1} (the value used in the experiment[20]).
185 We then constructed a phase diagram (Fig. 2G) that spans a wide range of acetate leakage fraction
186 and the feed medium glucose concentration via simulations. Since acetate is not supplemented,
187 increasing glucose supplementation induces higher release of acetate to the environment. The
188 entire phase space is divided into five distinct regions with four outcomes, including population
189 collapse, extinction of CV103 (CV101 wins), extinction of CV101 (CV103 wins) and stable
190 coexistence. In general, CV103 wins when the supplementation level of glucose is either very low
191 (acetate level is too low to compensate for the growth disadvantage of CV101 due to slower
192 glucose uptake) or very high (acetate level is too high to be toxic and strongly inhibits CV101).
193 Stable coexistence can be maintained within a narrow range of acetate leakage fraction. We show
194 that the coexistence region is robust to changes in the two most uncertain parameters determined
195 by MCMC (Supplementary Fig. 4). Note that the narrow coexistence regime does not necessarily
196 conflict with the observed transient coexistence in Fig. 2E because the theoretical phase diagram
197 was constructed at steady state when time goes to infinity.

Using Chesson's coexistence theory[30], the boundaries of the coexistence region can be interpreted as the conditions when the fitness (growth rate) difference between CV101 and CV103 is exactly balanced by the stabilizing effects of their niche differences (differential use of carbon sources; in general, it is a collective name for all mechanisms that lower interspecific competition relative to intraspecific competition). When acetate is not leaked (i.e., the acetate leakage fraction is 0), there is no niche difference (the only available carbon source is glucose) and the fitness difference is determined by the basal growth advantage of CV103 due to faster glucose uptake rate. Increasing leakage fraction of acetate leads to higher niche difference since acetate accumulation in the culture allows CV101 to utilize acetate as alternative carbon source and effectively reduces inter-population competition with CV103 for glucose. Meanwhile, increased acetate leakage also causes CV101 to grow faster, first reducing the fitness difference between the two strains to 0 (by overcoming its basal growth disadvantage) and increasing the difference afterwards. As the acetate leakage fraction increases, the lines of niche and fitness difference can possibly have two intersection points (Supplementary Fig. 5), between which CV101 and CV103 coexist stably because their fitness difference is smaller than their niche difference.

Example 2: bilateral amino-acid-mediated cross-feeding. The second community is characterized by a synthetic cross-feeding mutualism between lysine and leucine auxotrophies of *E. coli*[21] (Supplementary Texts 1.3.1). The two mutants differ only by single gene deletions in the lysine (Δ*lysA*) and leucine (Δ*leuA*) biosynthesis pathways. Neither mutant can grow in monoculture, but their coculture can survive by exporting essential amino acids that are needed by their partners to the extracellular environment and developing a bidirectional, obligate relationship. For simplicity, we assumed (1) leucine or lysine does not limit growth of the strain that synthesizes

it *de novo* (i.e., its producing strain); (2) environment leucine or lysine is not assimilated by its producing auxotrophic strain (Supplementary Texts 1.3.2). Using MCMC algorithm to estimate parameter values of the model that does not take these assumptions (Supplementary Fig. 6), we justified the second assumption by showing that the amino acid uptake rates by their producing strains are 1-2 orders of magnitude lower than the rates by their non-producing strains (Supplementary Fig. 7), suggesting that the majority of amino acids in the environment are assimilated by their auxotrophies. However, it is important to note that the assumption is specific to nutrient auxotrophies and may not be generalized to non-auxotrophic, wild-type cells. For example, wild-type *E. coli* cells that are able to synthesize all amino acids *de novo* still grow faster when supplemented with additional amino acids. Using parameters obtained through manual fitting (Materials and methods, Supplementary Table 2), we show that the simplified model (Fig. 3A) remains effective for quantitatively reproducing the growth and nutrient dynamics in both monoculture and coculture conditions (Fig. 3B,C).

The simplified model has 15 parameters, including 9 free parameters, 4 parameters fixed at literature values, and 2 biological constants (Supplementary Table 2). MCMC simulations confirmed that the posterior median of the free parameters and their values obtained from manual fitting are close to each other (Supplementary Fig. 8, Supplementary Table 2), except that we underestimated the mortality rate constant of the leucine auxotroph ($\eta_{\Delta l}$). Relative to other free parameters, the distributions of K_g (half-maximal concentration for glucose uptake), $\eta_{\Delta k}$ (mortality rate constant of the lysine auxotroph), and $\eta_{\Delta l}$ are much wider and span orders of magnitudes, suggesting that they are loosely constrained by experimental data. In addition, strong negative correlations between the maximum uptake rate and yield of the two amino acids (PCC = -86.9% and -65.5% for leucine and lysine respectively) were found. Finally, the engineered

interaction between the lysine and leucine auxotrophies is much weaker with only 0.66% (interquartile range [0.52%, 0.85%]) leucine and 1.13% (interquartile range [0.91%, 1.41%]) lysine released back to the environment (their corresponding values obtained through manual fitting are 0.32% and 1.39% respectively), compared to the evolved acetate-mediated cross-feeding interaction (~30% acetate leakage) we studied in Example 1.

Coexistence of the lysine and leucine auxotrophies. We sought to explore when the two auxotrophic strains coexist in chemostat. Fig. 3D shows the phase diagram at different combinations of the lysine and leucine leakage fraction via simulations. We did not see competitive exclusion, which is expected because the interdependence between the two strains is mutually obligate. It is important to note from Fig. 3D that the minimum leakage fraction of leucine (5.50%) and lysine (9.50%) required by coexistence at dilution rate 0.1 h^{-1} are far larger than the actual secreted percentages that we fit from experimental data (posterior median 0.66% and 1.13% for leucine and lysine leakage respectively), suggesting that the two engineered strains may not be able to coexist in such condition (but they may coexist at lower dilution rate). Interestingly, the bottom left boundary of the coexistence region describes a negative interaction between the minimum of the two leakage fractions, suggesting that decreasing leakage of one amino acid must be compensated by increasing leakage of the other in order to satisfy the minimum requirement of coexistence.

Coexistence is possible in the majority of the phase space, suggesting that the community is insensitive to the changes in the leakage rates. A striking feature of the diagram is that, increasing the fraction of lysine leakage fraction may trigger a discontinuous, abrupt switch from a steady state dominated by the leucine auxotroph (regime R1) to a steady state dominated by the lysine

auxotroph (regime R2). Such abrupt, discontinuous regime shifts are a common feature of microbial communities limited by several essential nutrients[31]. What accompanies with the compositional shift is the qualitative change in the nutrient utilization strategies adopted by the two strains (Fig. 3E). Before the switch, growth of the lysine auxotroph is limited by shared glucose while that of the leucine auxotroph is limited by leucine secreted by the lysine auxotroph. When the lysine leakage fraction increases over the threshold of the shift, the lysine auxotroph is limited by lysine secreted by the leucine auxotroph while the leucine auxotroph is limited by shared glucose. Our results indicate that the cellular metabolic strategies that are needed to maintain stable coexistence of the two amino acid auxotrophies vary in a discontinuous manner with continuous changes in amino acid leakage fractions.

Supplementation of cross-fed metabolites can reverse the sign of microbial social interactions. Cross-feeding interactions within a microbial community may be described as social interactions with costs and benefits to the members involved[32,33]. Those costs and benefits can be altered by environmental perturbations that supply or remove the cross-fed metabolites from the environment. Using the bilateral amino-acid-mediated cross-feeding model, we investigated how the supplementation of amino acids affected ecological relationships between cross-feeders at the steady state (Material and methods).

The phase space that spans a wide range of the leucine and lysine concentrations in the feed medium suggest four possible ecological relationships, including competition, amensalism, mutualism and parasitism (Fig. 4A). Mutualism was maintained over a broad range of supplied amino acid concentrations, even though amino acid supplementation releases the dependence of one auxotroph on the other and is hence detrimental to the mutualistic relationship. In the

mutualism regime, glucose is in excess and both amino acid auxotrophies are limited by the essential amino acids they cannot produce (Fig. 4B, left column). Further addition of amino acids leads to compositional dominance of one auxotrophic strain, but not necessarily competitive exclusion. Supplementation of leucine destabilizes the community by excluding the lysine auxotroph (Fig. 4B, middle column), whereas adding lysine only reduced the relative abundance of the leucine auxotroph, rather than leading to the loss of its entire population (Fig. 4B, right column). These results suggest that adding cross-fed nutrients can induce competition between community members that previously interacted mutualistically, and shift positive interactions to negative interactions.

Why supplementation of lysine and leucine cause such asymmetrical long-term effects on the community's composition and stability? We found that the outcome may be dependent on whether one or both auxotrophies is limited by glucose. When glucose limits both auxotrophies (Fig. 4B, middle column), competitive exclusion occurs and the leucine auxotroph wins because it has the same glucose uptake kinetics as the lysine auxotroph but lower mortality rate (Supplementary Table 2). When only the lysine auxotroph is limited by glucose (Fig. 4B, right column), the leucine auxotroph can sustain its population by growing on leucine released by its competitor. Whether coexistence of the two auxotrophies remains stable with increased level of amino acids supplementation can also be understood from the conditions when the net growth rate (growth rate minus mortality rate) of both populations equal to the dilution rate in the nutritional space (Fig. 4C). Coexistence requires that the steady state leucine must be equal to 5.25×10^{-4} mM, suggesting that supplementing too much leucine would devastate the ability of the system to self-regulate and reach that level at steady state. By contrast, a solution with high lysine

312 concentration is always feasible, which explains why coexistence can be maintained at very high
313 level of lysine supplementation.
314
315 **Example 3: multilateral cross-feeding between 14 amino acid auxotrophies.** To further
316 demonstrate the utility of our modeling framework, we studied cross-feeding interactions within
317 communities of more than two members. We modeled a community of 14 amino acid auxotrophies
318 engineered from *E. coli* by genetic knockout[22] (Fig. 5A). The 14-auxotroph model was directly
319 extended from the 2-auxotroph model developed above by considering each auxotroph can
320 potentially release all other 13 amino acids to the shared environment (Supplementary Texts 1.4.1).
321 Although all feeding possibilities are known, the consumer feeding preferences are not. By fitting
322 experimental data on the population compositions we aimed to infer the unknown feeding
323 pattern—what amino acids and how much they are released by each auxotrophic strain to feed
324 each other.
325 The model has 269 parameters, including 219 free parameters, 36 parameters fixed at
326 literature values, and 14 biological constants. Parameter values were obtained through both
327 automatic (amino acid leakage fractions) and manual (the rest) data fitting (Material and methods,
328 Supplementary Table 3). We show that the fit gave an excellent match to the observed population
329 density fold changes in pairwise cocultures (Fig. 5B, PCC = 94%), except for cross-feeding pairs
330 whose fold change values are less than 1. The observed reduction in population density may be
331 caused by cell death in the absence of nutrients but it is difficult to know because the measurement
332 of optical density at low inoculation amount (10^7 cells/mL) is highly noisy. For simplicity, we
333 assumed no cell death and set mortality rates to zero in the simulation, which explains why the
334 simulated population density fold changes are always non-decreasing. To compare our model with

higher-level models that do not include explicit nutrients, we adopted a Lotka-Volterra (LV) type model used in the literature[22,33], which guarantees that cross-feeding is obligate for growth

$$\frac{dx_1}{dt} = C_{1,2}x_2\left(\frac{x_1}{x_1 + b}\right)\left(1 - \frac{x_1 + x_2}{k}\right) \quad (4)$$

$$\frac{dx_2}{dt} = C_{2,1}x_1\left(\frac{x_2}{x_2 + b}\right)\left(1 - \frac{x_1 + x_2}{k}\right) \quad (5)$$

x_1 and x_2 are cell densities of any two amino acid auxotrophies, $C_{1,2}$ and $C_{2,1}$ are their cooperative coefficients, b is a constant that tunes the saturation concentration of x_1 and x_2, and k is another constant that represents carrying capacity. We show that the LV-type model can at best achieve a PCC of 33%, using parameters optimized by MCMC algorithm (i.e., parameters from the MCMC sample with the highest PCC). Although this LV-type model has a smaller number of parameters than ours (198 vs. 269), the number of free parameters between the two models is of similar size and comparable (198 vs. 205; note that 14 mortality rates in our model were set to zero).

Fig. 5C reports the estimated leakage fractions of 14 amino acids by their amino acid auxotrophies in a matrix form. Although the 14 auxotrophies were derived from the same parent strain, they showed very different profiles of amino acid leakage: some auxotrophies such as the methionine auxotroph ΔM (36.41% total carbon loss) are highly cooperative whereas others such as the tryptophan auxotroph ΔW (1.37% total carbon loss) have very low cooperativity. These differences may be attributed to how metabolic network structure was disrupted to generate the auxotrophies and the concomitant changes in metabolic fluxes. One such example is the strong release (13.32%) of threonine by the methionine auxotroph. Since methionine and threonine biosynthesis pathways branch off from the same precursor homoserine, block of one pathway may lead to increased fluxes of another pathway and leakage of corresponding amino acids. However, the leakage fraction of methionine by the threonine auxotroph is very low (0.1%), suggesting that

355 network topology is not the only factor that affects leakage flux. Since methionine is the most
356 expensive amino acid to produce in terms of ATP consumption[34], its biosynthesis and leakage rates
357 may be tightly regulated and only loosely depend on the level of its precursors.
358
359 **The 14-member community converges to a stable coexisting subset at steady state.** Besides
360 the pairwise coculture data (Fig. 5B), our model also reproduced the population dynamics of
361 serially diluted cocultures of all 14 auxotrophies and four selected 13-auxotroph combinations (Fig.
362 6A). The fit is reasonably good at the log scale, except for the methionine-auxotroph-absent
363 community which seems to undergo non-ecological processes that rescue the threonine auxotroph
364 (ΔT) from the brink of extinction between day 2 and day 3. Quantitatively, the PCCs between
365 observed and predicted values on the log scale are 88.71% (all 14 auxotrophies), 75.30% (lysine-
366 auxotroph-absent), 78.34% (arginine-auxotroph-absent), 52.93% (threonine-auxotroph-absent),
367 and 8.90% (methionine-auxotroph-absent).
368 As shown in Fig. 6A, most amino acid auxotrophies were diluted away very quickly but
369 some, such as the isoleucine auxotroph (ΔI), exhibited transient recovery dynamics after the initial
370 decay. To understand the transient dynamics, we used the same model to infer the concentration
371 dynamics of glucose and all amino acids, which are hidden states (not yet observed) that are
372 relatively costly and inaccurate to measure in experiments. Supplementary Fig. 9 shows that the
373 population density of the isoleucine auxotroph had an initial drop because the isoleucine pool had
374 not been accumulated to a critical size that allows the actual growth to compensate for its mortality
375 and system dilution. As the pool size increases, its net growth rate (growth minus mortality)
376 surpasses the dilution rate and recovers its population density, which eventually levels off when
377 the positive and negative effects are balanced.

By simulating the 14-auxotroph community model to steady state, we further predicted that the initial 14-strain mixture converges to a stable coexisting subset that contains 4 amino acid auxotrophies that are deficient in biosynthesis of isoleucine (ΔI), lysine (ΔK), methionine (ΔM), and threonine (ΔT) (Fig. 6B). The predicted coexistence state was successfully validated by two independent observations over 50-day serial dilution[22], a much longer period of time than the duration of the training dataset (7-day serial dilution; Fig. 6A). The resource-consumer relationships of the 4-member community are shown in a bipartite network (Fig. 7A), where 3 amino acid secretion fluxes were identified as essential (solid arrows) as their deletions resulted in community member loss (Supplementary Fig. 10). These essential fluxes suggest that the primary feeders for ΔK, ΔM, ΔT are ΔT, ΔI, ΔM respectively; however, none of ΔK, ΔM, ΔT dominates the feeding of ΔI and their contributions to the isoleucine pool in the environment are substitutable.

Mutualistic cross-feeding network is prone to collapse after external perturbations. Using the model developed above, we computationally tested how external perturbations, including nutrient downshift, the addition of antibiotics, and invasion of cheating phenotypes (the same auxotrophic dependence but no amino acid leakage) affect the stability of coexistence among the 4 auxotrophic strains that would otherwise be stable (Material and methods). The 4-member community was able to cope with these disturbances to a certain extent and remained integrated. Beyond the thresholds, all three perturbation types resulted in community collapse as a result of domino effect (Fig. 7B-D), implying that tightly coupled cooperative communities are fragile and prone to collapse. Since antibiotics inhibit growth of individual strains (targeting consumer nodes in the bipartite network) while cheaters are amino acid sinks (targeting resource nodes in the bipartite network), we identified that ΔT and methionine as the weakest consumer node (Fig. 7C) and resource node (Fig.

401 7D) in the bipartite network respectively. Our results suggest that ΔT→K (secretion of lysine by
402 the threonine auxotroph) and M→ΔM (uptake of methionine by the methionine auxotroph)—the
403 outgoing links from the two weakest nodes that are also essential to maintain community
404 integrity—are the weakest metabolic fluxes that may set the resistance level of the community to
405 external perturbations[35].
406

407 ## Discussion

408 Predicting population dynamics of a microbial community from interactions between its
409 members is difficult because interaction happens across multiple scales of biological
410 organization[36]. Here we propose a mechanistic ecology model based on a coarse-grained
411 representation of cell metabolism that accurately describes the population dynamics of three
412 laboratory communities with well-defined metabolic exchanges. Previous studies have used
413 genome-scale models and metabolic flux analysis, but these studies require flux measurements by
414 isotope tracing and metabolomics to fit the adjustable flux parameters. Some success was also
415 achieved by fitting the time series data with coarser-grained ecological models[37–41] such as the
416 gLV equations; however, in gLV-type models, interspecific interactions are phenomenologically
417 defined based on density dependency, which gives little mechanistic understanding of the
418 underlying mechanism[15]. By contrast, our model has explicit formulations of context dependency
419 by representing the chemical flows within and between microbes and thus can explain the
420 metabolic part of microbe-microbe interactions.

421 When we have limited prior knowledge and data on a given community it becomes critical
422 to choose the right level of details. However, by applying our approach to well-defined laboratory
423 systems, we show that a highly detailed metabolic network is not necessary for developing useful

ecological models. In single-bacteria studies, coarse-grained metabolic models have been employed to understand the design principles of metabolic networks and their regulation[42], as well as to predict metabolic flux distributions useful for synthetic biology[43] and industrial[44] applications. Compared to genome-scale models, using coarse-grained models linking ecology and metabolism is simple and has recently become popular[26,45,46]. Depending on the research question, a coarse-grained metabolic network can be created at any level of granularity from a single reaction to the complete whole genome-scale reconstruction. The choice of granularity and how to derive a simpler model from the more complex one are usually empirical but can be facilitated by more systematic approaches to reduce dimensionality.

Our model could extract new insights from those previously published empirical data on well-defined laboratory systems. The analysis shows that unidirectional cross-feeding is equivalent to a commensalism and bidirectional cross-feeding is equivalent to a mutualism. As shown by our study (Fig. 2-4) and previous work[27,32], the actual relationship between cross-feeders, however, can be diverse in simple environments (e.g., glucose minimal medium) with constant resource supply due to a combination of positive effects of cross-feeding with negative effects of competition and toxicity of cross-fed metabolites, suggesting that the exact outcome cannot be precisely delineated by the cross-feeding type alone. For example, we predicted that, without supplementation of amino acids, coexistence of the leucine and lysine auxtrophies can only be achieved when one strain is limited in growth by glucose while the other strain is limited by the amino acid it is auxotrophic for (Fig. 3E). Although it is theoretically possible that growth of the two auxotrophies is simultaneously limited by the amino acids they are auxotrophic for (i.e., the lysine auxotroph limited by lysine and the leucine auxotroph limited by leucine), this interaction pattern does not occur in the phase diagram because glucose will always be sufficiently depleted

447 to a level that becomes growth limiting to at least one strain. The control of resource pool
448 availability via population dynamics has been demonstrated to be a key mechanism for microbial
449 community to optimize the metabolic strategy of its members to yield resistance to invasions and
450 to achieve maximum biomass[46].

451 Mechanistic models including explicit nutrients and other realistic features, such as the
452 models presented in this study, can help identify knowledge gaps[47]. For example, recent
453 experiments have demonstrated that the coexistence of two carbon source specialists in the
454 unilateral cross-feeding example is mutualistic in the sense that the consortium is fitter than the
455 individuals[48]. The syntropy can be explained by a null expectation from theoretical ecology
456 models[49]: the glucose specialist provides acetate in an exchange for a service provided by the
457 acetate specialist which scavenges the acetate down to a level at which growth inhibition is
458 insignificant. Although the mechanism of resource-service exchange has been considered in our
459 model, the coexistence regime in the phase diagram (Fig. 2G) is competitive, rather than
460 mutualistic. Since mutualism occurs when the reciprocal benefits associated with cross-feeding
461 outweigh competitive costs[50], our model may predict either or both of lower benefits and higher
462 costs than needed to achieve mutualistic coexistence. Overall, the cost-benefit nature of the cross-
463 feeding interaction between polymorphic *E. coli* strains is more complex than thought and warrants
464 further research.

465 Our modeling framework explains well the three published experiments but has noteworthy
466 limitations. For example, we assume that the leakage flux is proportional to the conversion rate
467 from substrate to metabolite (proportionality assumption), rather than proportional to the internal
468 metabolite concentration. When does this assumption remain valid and how does it break down?
469 By leveraging our previous experiences in modeling *E. coli* growth and resource allocation[43,51],

we developed a coarse-grained single-strain model that explicitly assumes a linear dependency of leakage rate on metabolite concentration (Supplementary Texts 1.5). We found that the proportionality assumption remains valid for an internal metabolite when its concentration was perturbed at the upstream, rather than the downstream of the metabolite (Supplementary Fig. 11). This makes sense because the proportionality assumption couples metabolite leakage with upstream biosynthesis but does not take feedback regulation from downstream reactions and metabolites into accounts. When a perturbation is imposed from the downstream side, the proportionality assumption can lead to undesired behavior such as high leakage flux at low metabolite concentration. Although the assumption remains valid in the context of the current study where resource availability is the only varying external condition, it may prevent us from generalizing our modeling framework to different types of perturbations. Future studies may correct this limitation by incorporating metabolite concentration and associated reaction kinetics.

So far, the current framework has been applied to well-characterized communities with known chemicals and associated interactions which provided a ground through to assess our model. Can the same approach be applied to infer community structure of complex microbiomes (e.g., human gut microbiome) where most of the metabolic exchanges involved in microbe-microbe interactions are still unknown? Our model has the potential if some technical challenges can be solved. First, direct modeling of a real-world microbiome with hundreds of species would be hurdled by too many unknown model parameters. One way to solve this problem is to simply ignore the rare species[38]. Another—arguably better—approach might be by grouping species composition into functional guilds using unsupervised methods that infer those groups from the data alone[52], or to use prior knowledge from genomics or taxonomy to create such functional groups. Second, inferring chemical mediators within a community of interacting populations is a

493 nontrivial task. It can be facilitated by prior knowledge such as searching the literature or
494 leveraging systems biology tools such as community-level metabolic network reconstruction[53].
495 Finally, our model is nonlinear, so that an efficient and robust nonlinear regression approach for
496 parameter estimation is essential. For a model with similar size to the 14-auxotroph community
497 we analyzed here, non-linear optimization algorithms may fail to converge to a realistic set of
498 parameters and manual parameter selection is often the only feasible approach. Although we
499 primarily chose the manual method to calibrate our models in this proof-of-concept study, manual
500 fitting is a subjective and time-consuming process, requires an expert operator with prior
501 knowledge to choose physically and biologically realistic values, and perhaps more importantly,
502 is unable to infer correlations among parameters. These downsides of manual parameter fitting has,
503 at least for now, precluded it from being applied to large-scale microbial communities. On the
504 positive side, the process of trial-and-error was greatly improved by the speed at which the
505 intermediate-scale model runs simulations on a regular desktop computer. Beyond these technical
506 issues, the model itself can be extended in multiple ways such as incorporating mechanisms of
507 resource allocation[46]. Despite any present limitations, we anticipate that network inference using
508 mechanism-explicit models can open new avenues for microbiome research towards more
509 quantitative, mechanistic, and predictive science.
510

511 ## Materials and methods

512 **Cross-feeding models.** The modeling framework presented in this study was developed by
513 integrating a classical ecology model for population and nutrient dynamics with a coarse-grained
514 description of cell metabolism. Custom MATLAB R2018a (The MathWorks, Inc., Natick, MA,
515 USA) codes were developed to perform computational simulations and analyses of all three cross-

516 feeding communities. Please refer to Supplementary Texts for a detailed description of the general
517 modeling framework and its applications to each cross-feeding community.
518
519 **Parameter estimation**. Our goal was to manually parameterize cross-feeding models directly
520 from experimental data, which are typically cell density and metabolite concentrations in the
521 culture. The manual process of parameter estimation began with initial values of parameters
522 selected to be either equal to their previously reported values or assumed to be of the same order
523 of magnitude based on the literature data. This was followed by the iterative evaluation of model
524 outputs and refinement until sufficient concordance between the model predictions and the
525 experimental data is achieved.
526 The only exception of parameters that were fit automatically are the amino acid leakage
527 fractions of the 14 amino acid auxotrophies. Under a few assumptions, our model can be simplified
528 and exactly solved for steady state population density in pairwise cocultures (Supplementary Texts
529 1.4.2). The values of these parameters were then estimated by minimizing the least square error
530 between observed and calculated fold changes of population density across all pairwise batch
531 cocultures. Once obtained, these values were fixed in the process of manually fitting the other
532 parameters of the model.
533
534 **Parameter sensitivity analysis.** To estimate parameter uncertainty and identify their potential
535 correlations, we used an adaptive MCMC (Markov-Chain-Monte-Carlo) method for sampling the
536 posterior distribution of parameters under constraints of experimental data. We obtained the
537 MATLAB code for this method from https://github.com/mjlaine/mcmcstat. Briefly, this method
538 constructs a sequence of random samples in the parameter space by the Metropolis-Hastings

539 algorithm: at each iteration, the algorithm randomly picks a candidate of the next sample (i.e.,
540 parameter set) based on the current sample value. The candidate is accepted with a probability
541 determined by the ratio of the likelihood of the new sample to that of the current sample and the
542 likelihood is given by a negative exponential function where the exponent is the prediction error
543 of our model using a given parameter set. Please refer to the original publication[54] for details of
544 the method.

545 We ran MCMC simulations for both 2-membebr communities with unilateral and bilateral
546 cross-feeding relationships. The posterior distribution of each parameter was estimated from
547 100,000 MCMC samples after a burn-in period of 10,000 samples. We assumed a Gaussian prior
548 with standard deviation 0.01. We used symmetric mean absolute percentage error as the cost
549 function that is minimized by the Metropolis-Hastings algorithm:

550 Unilateral cross-feeding: $\frac{1}{N_{data}}\left(0.1\sum_{i\in Fig.2\mathrm{B,C}}\frac{|y_{obs,i}-y_{sim,,i}|}{|y_{obs,i}|+|y_{sim,,i}|}+\sum_{i\in Fig.2D,E}\frac{|y_{obs,i}-y_{sim,,i}|}{|y_{obs,i}|+|y_{sim,,i}|}\right)$

551 Bilateral cross-feeding: $\frac{1}{N_{data}}\sum_{i\in Fig.3B,C}\frac{|y_{obs,i}-y_{sim,,i}|}{|y_{obs,i}|+|y_{sim,,i}|}$

552 where $y_{obs,i}$ is the observed datum i, $y_{sim,i}$ is its simulated value, and N_{data} is the total number of
553 data points. Note that the data from different experiments have unequal weights in the unilateral
554 cross-feeding example.

555

556 **Simulation of batch, continuous, and serially diluted culture.** Deterministic trajectories and
557 their steady states in batch and chemostat conditions were simulated by solving the differential
558 equations from the beginning to the end. Simulations of serial dilution transfer were slightly
559 different in the aspect that the equations were only integrated within each day. The initial condition
560 at the beginning of a day was obtained by dividing all population densities and nutrient

561 concentrations at the end of the previous day by the dilution factor and resetting the feed medium
562 nutrient concentrations to their initial values at day 0.
563
564 **Classification of interspecific ecological relationship.** We simulated chemostat cocultures of the
565 lysine and leucine auxotrophies at increasing levels of amino acid supplementation in the feed
566 medium, and computed the net effect (+,0,-) of one population on the other by comparing to
567 monoculture simulation. The pairwise ecological relationship between the two populations can
568 then be determined by the signs of their reciprocal impacts[55]: (+,+): mutualism; (-,-): competition;
569 (+,0) and (0,+): commensalism; (-,0) and (0,-): amensalism; (+,-) or (-,+): parasitism; (0,0): no
570 effect.
571
572 **Network perturbation.** External perturbations were exerted upon the steady state of the 4-
573 auxotroph community. Nutrient downshift was simulated by decreasing the feed medium
574 concentration of glucose at the beginning of simulations. The effects of an antibiotic that inhibit
575 growth of the amino acid auxotroph i was simulated by multiplying the growth rate of the
576 auxotroph by an inhibitory term, i.e., $J_l^{grow} \rightarrow J_l^{grow}/(1 + [A]/K_i)$, where $[A]$ is the antibiotic
577 concentration and K_i is the inhibition constant. We assumed antibiotic concentration remains
578 constant and chose $K_i = 1\ \mu M$. The cheaters of each amino acid auxotroph were simulated by
579 turning off all amino acid leakages of the auxotroph. They were mixed with the resident
580 community in varying ratios at the beginning of simulations. For all three perturbation types, the
581 feed medium glucose concentration is 0.2 wt.% in the unperturbed condition and serial dilution
582 was run to steady state at 60 days.
583

Acknowledgements

We thank Dr. Michael Mee, Dr. Harris Wang, and Dr. Jennifer Reed for provision and clarification of their experimental data. We also thank Dr. Jinyuan Yan for proofreading early drafts. This work was supported by NIH grants U01 AI124275 and R01 AI137269-01 to J.B.X. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data availability

The simulation data that support the conclusions of this study are available from the main text and Supplementary Information. The source codes for generating the figures of this study are available from https://github.com/liaochen1988/Source_code_for_cross_feeding_paper.

References

1. Kent, A. D. & Triplett, E. W. Microbial communities and their interactions in soil and rhizosphere ecosystems. *Annu. Rev. Microbiol.* **56,** 211–236 (2002).
2. Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat. Rev. Genet.* **13,** 260–270 (2012).
3. Li, H. Microbiome, Metagenomics, and High-Dimensional Compositional Data Analysis. *Annu. Rev. Stat. Appl.* **2,** 73–94 (2015).
4. Shank, E. A. Considering the lives of microbes in microbial communities. *mSystems* **3,** (2018).
5. Schmidt, R., Ulanova, D., Wick, L. Y., Bode, H. B. & Garbeva, P. Microbe-driven chemical ecology: past, present and future. *ISME J.* (2019).
6. Morris, J. J. Black Queen evolution: the role of leakiness in structuring microbial communities.

606 *Trends Genet.* **31,** 475–482 (2015).

607 7. Basan, M. *et al.* Overflow metabolism in *Escherichia coli* results from efficient proteome
608 allocation. *Nature* **528,** 99–104 (2015).

609 8. Paczia, N. *et al.* Extensive exometabolome analysis reveals extended overflow metabolism in
610 various microorganisms. *Microb. Cell Fact.* **11,** 122 (2012).

611 9. Hao, T. *et al.* The Genome-Scale Integrated Networks in Microorganisms. *Front. Microbiol.*
612 **9,** 296 (2018).

613 10. Morris, B. E. L., Henneberger, R., Huber, H. & Moissl-Eichinger, C. Microbial syntrophy:
614 interaction for the common good. *FEMS Microbiol. Rev.* **37,** 384–406 (2013).

615 11. Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving
616 and thriving in the microbial jungle. *Nat. Rev. Microbiol.* **8,** 15–25 (2010).

617 12. Hanemaaijer, M. *et al.* Systems modeling approaches for microbial community studies: from
618 metagenomics to inference of the community structure. *Front. Microbiol.* **6,** 213 (2015).

619 13. Zomorrodi, A. R. & Segrè, D. Synthetic ecology of microbes: mathematical models and
620 applications. *J. Mol. Biol.* **428,** 837–861 (2016).

621 14. Karr, J. R. *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell*
622 **150,** 389–401 (2012).

623 15. Momeni, B., Xie, L. & Shou, W. Lotka-Volterra pairwise modeling fails to capture diverse
624 pairwise microbial interactions. *Elife* **6,** (2017).

625 16. Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: Networks,
626 competition, and stability. *Science* **350,** 663–666 (2015).

627 17. Niehaus, L. *et al.* Microbial coexistence through chemical-mediated interactions. *Nat.*
628 *Commun.* **10,** 2052 (2019).

18. Mahadevan, R., Edwards, J. S. & Doyle, F. J. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* **83,** 1331–1340 (2002).
19. Nadell, C. D., Foster, K. R. & Xavier, J. B. Emergence of spatial structure in cell groups and the evolution of cooperation. *PLoS Comput. Biol.* **6,** e1000716 (2010).
20. Rosenzweig, R. F., Sharp, R. R., Treves, D. S. & Adams, J. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* **137,** 903–917 (1994).
21. Zhang, X. & Reed, J. L. Adaptive evolution of synthetic cooperating communities improves growth performance. *PLoS One* **9,** e108297 (2014).
22. Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci. USA* **111,** E2149-56 (2014).
23. MacArthur, R. Species packing and competitive equilibrium for many species. *Theor Popul Biol* **1,** 1–11 (1970).
24. Marsland, R. *et al.* Available energy fluxes drive a transition in the diversity, stability, and functional structure of microbial communities. *PLoS Comput. Biol.* **15,** e1006793 (2019).
25. Tilman, D. Resource competition and community structure. *Monogr Popul Biol* **17,** 1–296 (1982).
26. Brunner, J. D. & Chia, N. Metabolite-mediated modelling of microbial community dynamics captures emergent behaviour more effectively than species-species modelling. *J. R. Soc. Interface* **16,** 20190423 (2019).
27. LaSarre, B., McCully, A. L., Lennon, J. T. & McKinlay, J. B. Microbial mutualism dynamics governed by dose-dependent toxicity of cross-fed nutrients. *ISME J.* **11,** 337–348 (2017).
28. Liebig, J. & Playfair, L. Chemistry in its application to agriculture and physiology. (1842).

652 29. Enjalbert, B., Millard, P., Dinclaux, M., Portais, J.-C. & Létisse, F. Acetate fluxes in
653 *Escherichia coli* are determined by the thermodynamic control of the Pta-AckA pathway. *Sci.*
654 *Rep.* **7,** 42135 (2017).
655 30. Barabás, G., D'Andrea, R. & Stump, S. M. Chesson's coexistence theory. *Ecol. Monogr.* **88,**
656 277–303 (2018).
657 31. Dubinkina, V., Fridman, Y., Pandey, P. P. & Maslov, S. Multistability and regime shifts in
658 microbial communities explained by competition for essential nutrients. *Elife* **8,** (2019).
659 32. Hammarlund, S. P., Chacón, J. M. & Harcombe, W. R. A shared limiting resource leads to
660 competitive exclusion in a cross-feeding system. *Environ. Microbiol.* **21,** 759–771 (2019).
661 33. Hoek, T. A. *et al.* Resource Availability Modulates the Cooperative and Competitive Nature
662 of a Microbial Cross-Feeding Mutualism. *PLoS Biol.* **14,** e1002540 (2016).
663 34. Kaleta, C., Schäuble, S., Rinas, U. & Schuster, S. Metabolic costs of amino acid and protein
664 production in *Escherichia coli*. *Biotechnol. J.* **8,** 1105–1114 (2013).
665 35. Adamowicz, E. M., Flynn, J., Hunter, R. C. & Harcombe, W. R. Cross-feeding modulates
666 antibiotic tolerance in bacterial communities. *ISME J.* **12,** 2723–2735 (2018).
667 36. Widder, S. *et al.* Challenges in microbial ecology: building predictive understanding of
668 community function and dynamics. *ISME J.* **10,** 2557–2568 (2016).
669 37. Bucci, V. *et al.* MDSINE: Microbial Dynamical Systems INference Engine for microbiome
670 time-series analyses. *Genome Biol.* **17,** 121 (2016).
671 38. Stein, R. R. *et al.* Ecological modeling from time-series inference: insight into dynamics and
672 stability of intestinal microbiota. *PLoS Comput. Biol.* **9,** e1003388 (2013).
673 39. Buffie, C. G. *et al.* Precision microbiome reconstitution restores bile acid mediated resistance
674 to *Clostridium difficile*. *Nature* **517,** 205–208 (2015).

675 40. Venturelli, O. S. *et al.* Deciphering microbial interactions in synthetic human gut microbiome
676 communities. *Mol. Syst. Biol.* **14,** e8157 (2018).
677 41. Goldford, J. E. *et al.* Emergent simplicity in microbial community assembly. *Science* **361,**
678 469–474 (2018).
679 42. Kraft, B. *et al.* Nitrogen cycling. The environmental controls that govern the end product of
680 bacterial nitrate respiration. *Science* **345,** 676–679 (2014).
681 43. Liao, C., Blanchard, A. E. & Lu, T. An integrative circuit-host modelling framework for
682 predicting synthetic gene network behaviours. *Nat. Microbiol.* **2,** 1658–1666 (2017).
683 44. Liao, C. *et al.* Integrated, systems metabolic picture of acetone-butanol-ethanol fermentation
684 by *Clostridium acetobutylicum*. *Proc. Natl. Acad. Sci. USA* **112,** 8505–8510 (2015).
685 45. Muscarella, M. E. & O'Dwyer, J. P. Species dynamics and interactions via metabolically
686 informed consumer-resource models. *BioRxiv* (2019). doi:10.1101/518449
687 46. Taillefumier, T., Posfai, A., Meir, Y. & Wingreen, N. S. Microbial consortia at steady supply.
688 *Elife* **6,** (2017).
689 47. Hanemaaijer, M., Olivier, B. G., Röling, W. F. M., Bruggeman, F. J. & Teusink, B. Model-
690 based quantification of metabolic interactions from dynamic microbial-community data. *PLoS*
691 *One* **12,** e0173183 (2017).
692 48. Yang, D.-D. *et al.* Fitness and productivity increase with ecotypic diversity among *E. coli*
693 evolved in a simple, constant environment. *BioRxiv* (2019). doi:10.1101/679969
694 49. Estrela, S., Trisos, C. H. & Brown, S. P. From metabolism to ecology: cross-feeding
695 interactions shape the balance between polymicrobial conflict and mutualism. *Am. Nat.* **180,**
696 566–576 (2012).
697 50. Wu, F. *et al.* A unifying framework for interpreting and predicting mutualistic systems. *Nat.*

Commun. **10,** 242 (2019).

51. Liao, C. Integrated modeling of bacterial metabolism. PhD thesis, University of Illinois at Urbana-Champaign (2018).
52. Gibson, T. E. & Gerber, G. K. Robust and scalable models of microbiome dynamics. *arXiv preprint arXiv:1805.04591* (2018).
53. Levy, R. & Borenstein, E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc. Natl. Acad. Sci. USA* **110,** 12804–12809 (2013).
54. Haario, H., Laine, M., Mira, A. & Saksman, E. DRAM: Efficient adaptive MCMC. *Stat Comput* **16,** 339–354 (2006).
55. Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* **10,** 538–550 (2012).
56. Gudelj, I., Kinnersley, M., Rashkov, P., Schmidt, K. & Rosenzweig, F. Stability of Cross-Feeding Polymorphisms in Microbial Communities. *PLoS Comput. Biol.* **12,** e1005269 (2016).

Figure Legends:

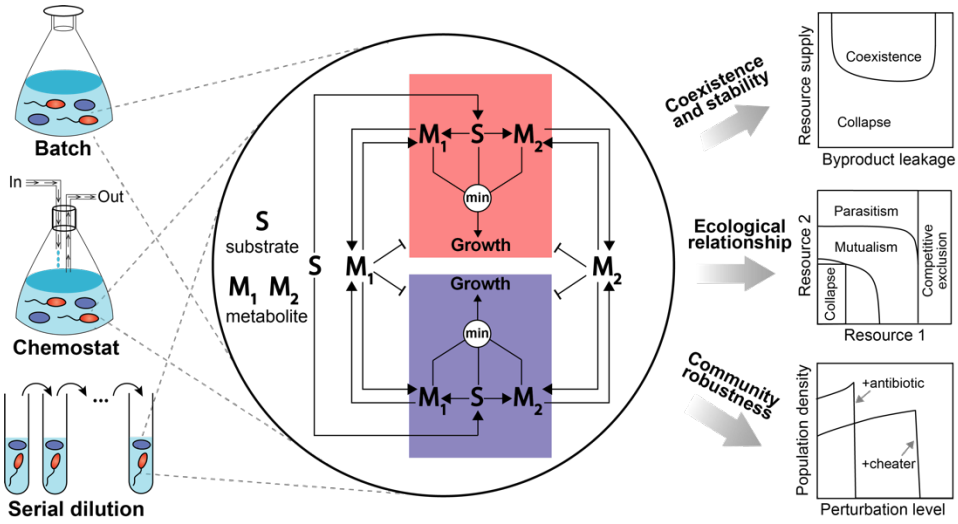
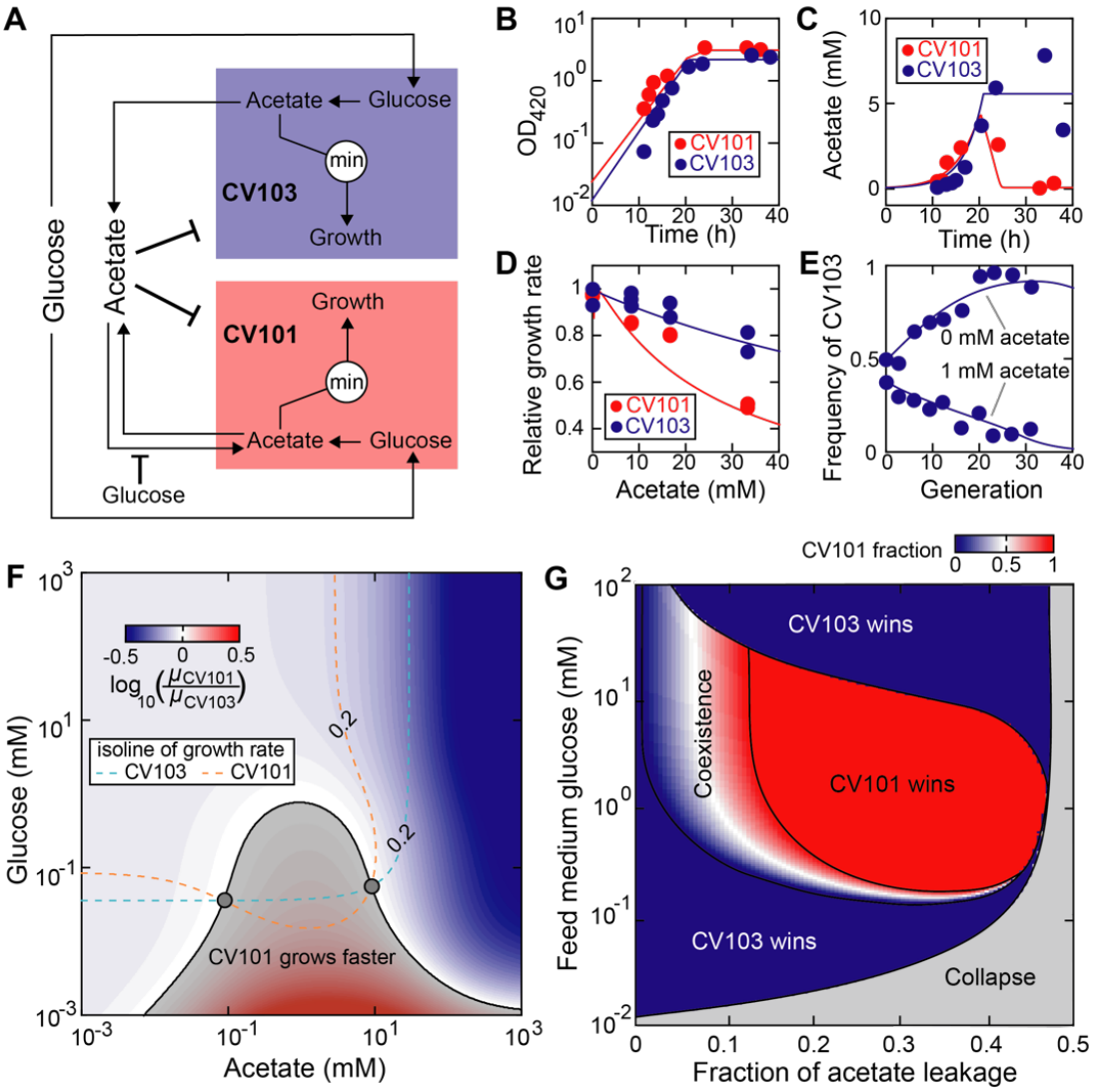


Figure 1 | Schematic diagram illustrating our model and its potential applications in microbial ecology research. A distinguishing feature of our microbial community model is that each community member harbors a coarse-grained metabolic network. Briefly, the metabolic network transforms growth substrates (**S**) to non-substitutable building block metabolites ($\mathbf{M_1}$, $\mathbf{M_2}$) and then to biomass whose production rate is set by the supply flux of the most limiting resource among all substrates and metabolites. The intracellularly synthesized metabolites can be secreted to the environment and then utilized by the community as public goods. For simplicity, the network is visually illustrated using one substrate and two metabolites but it can be extended to any number of nutrients. Enabled by the simplified metabolic network, different community members can interact through a variety of mechanisms, including exploitative competitions for shared substrates, cooperative exchanges of nutritional metabolites, and direct inhibition by secreting toxic compounds. Using training data from batch, chemostat or serial dilution cultures, our model can be parameterized to infer microbial processes underlying the data and then used to explore ecological questions and generate testable predictions. Pointed arrows denote the material flow and blunt-end arrows represent growth inhibition.



730

731 **Figure 2 | Unilateral acetate-mediated cross-feeding between evolved *E. coli* isolates.** (A)
732 Schematic diagram of the model. The glucose specialist (CV103) and acetate specialist (CV101)
733 are two *E. coli* mutants with different metabolic strategies[20]: the glucose specialist has improved
734 glucose uptake kinetics while the acetate specialist is able to use acetate as an additional carbon
735 source. At high concentration, acetate inhibits growth of both strains and its uptake by the acetate
736 specialist strain is weakly repressed by glucose. Since glucose and acetate are substitutable, all
737 glucose is converted to acetate which serves as the sole limiting factor for cell growth. (B-E)
738 Manual model calibration. Circles: experimental data; lines: simulations. (B,C) 0.1% glucose-

limited batch monoculture without supplementing acetate[20]. (D) 0.0125% glucose-limited batch monoculture supplemented with different concentrations of acetate[56]. (E) 0.00625% glucose-limited chemostat (dilution rate: D=0.2 h^{-1}) coculture with (1 mM) and without acetate supplementation[20]. The time for one generation is defined as log(2)/D. (F) Growth rate ratio of CV101 to CV103 in the nutritional space. The gray shading indicates when CV101 grows faster than CV103 and the gray circles mark when their growth rates are both equal to the dilution rate 0.2 h^{-1}. (G) The simulated steady-state phase diagram.

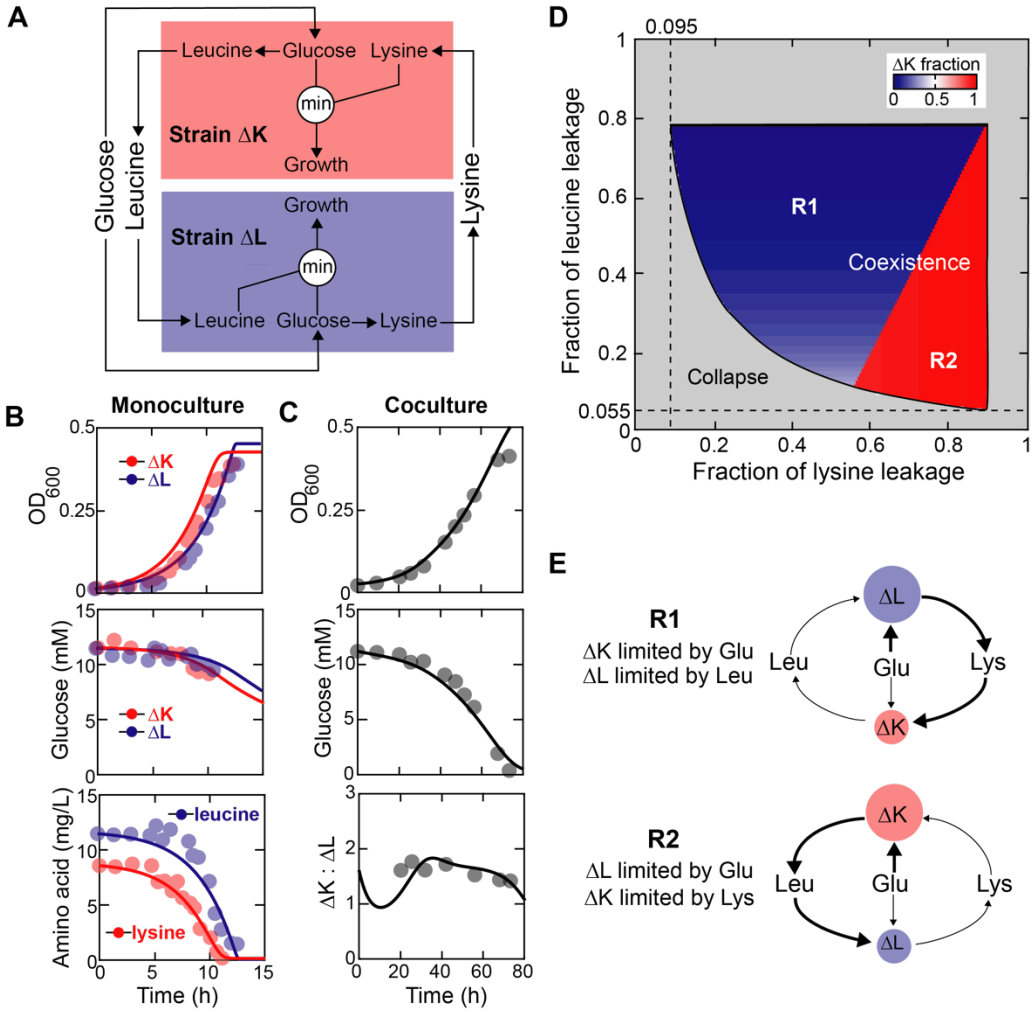


Figure 3 | Bilateral cross-feeding between engineered *E. coli* amino acid auxotrophies. (A) Schematic diagram of the model. The *E. coli* lysine auxotroph (ΔK) and leucine auxotroph (ΔL) compete for glucose while additionally acquiring essential amino acids from each other. Growth of each auxotroph is determined by the more limiting resource between glucose and the amino acid it needs to grow. (B,C) Manual model calibration. Circles: data; lines: simulation. (B) 2 g/L glucose-limited batch monoculture supplemented with 10 mg/L amino acids[21]. (C) 2 g/L glucose-limited batch coculture without amino acid supplementation[21]. (D) The simulated steady-state phase diagram. The feed medium glucose concentration is 10 mM. (E) The metabolic strategies

755 adopted by ΔK and ΔL in the coexistence regime. All chemostat simulations were run at dilution

756 rate of 0.1 h^{-1}.

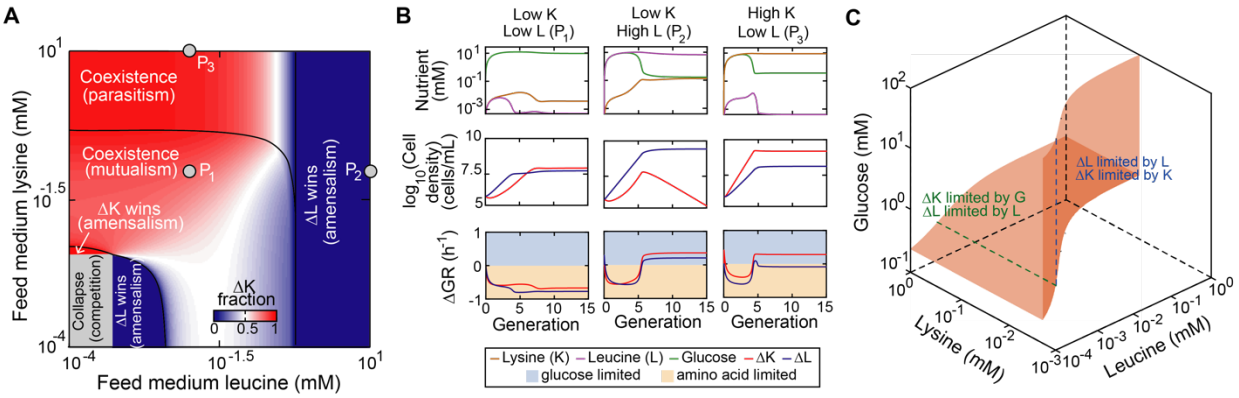


Figure 4 | Impacts of amino acid supplementation on ecological relationships between two amino acid cross-feeders. (A) The simulated steady-state phase diagram for different levels of amino acid supplementation. (B) Representative system dynamic trajectories of specific phases in (A). ΔGR: the difference between growth rate when set by amino acid as the sole limiting factor and when set by glucose as the sole limiting factor (the minimum of the two determines the actual growth rate). A positive or negative value of ΔGR indicates that cell growth is limited by glucose or amino acid respectively. (C) The isosurface of equal net growth rate (growth rate minus mortality rate) between the lysine and leucine auxotrophies. The dashed lines (blue and green) indicate when their net growth rates are equal to 0.1 h^{-1} (the dilution rate used throughout the figure). Abbreviations: glucose (G); lysine (K); leucine (L); lysine auxotroph (ΔK); leucine auxotroph (ΔL).

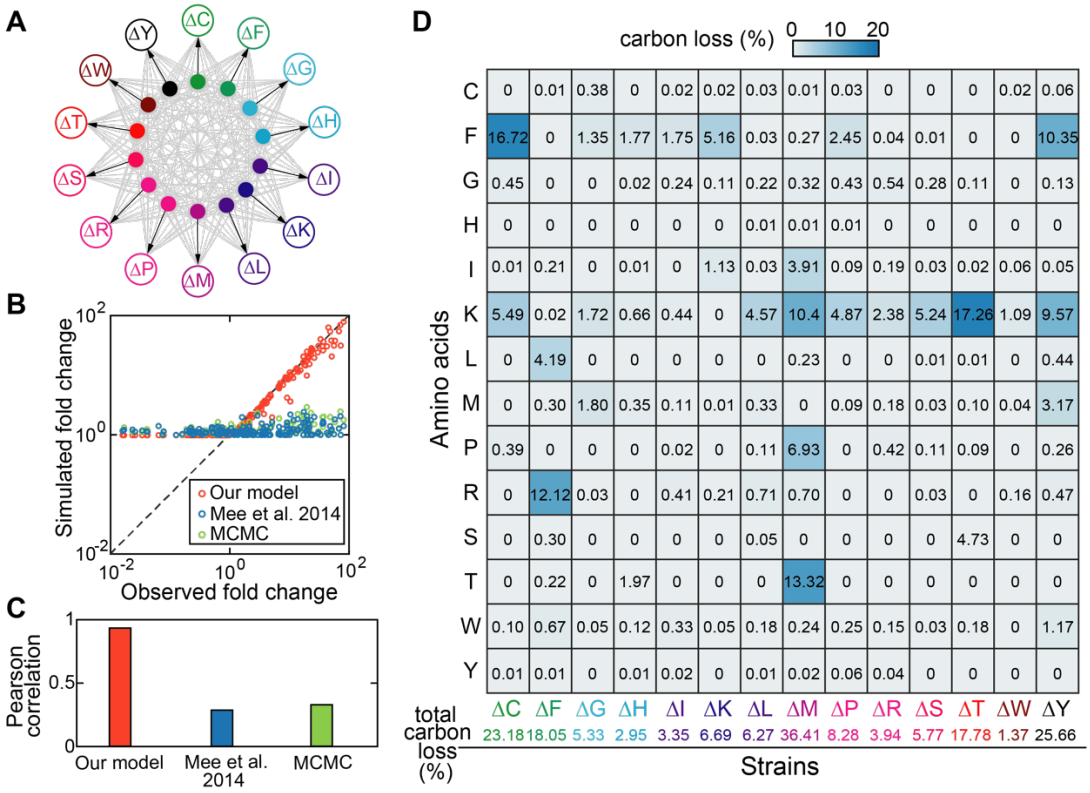
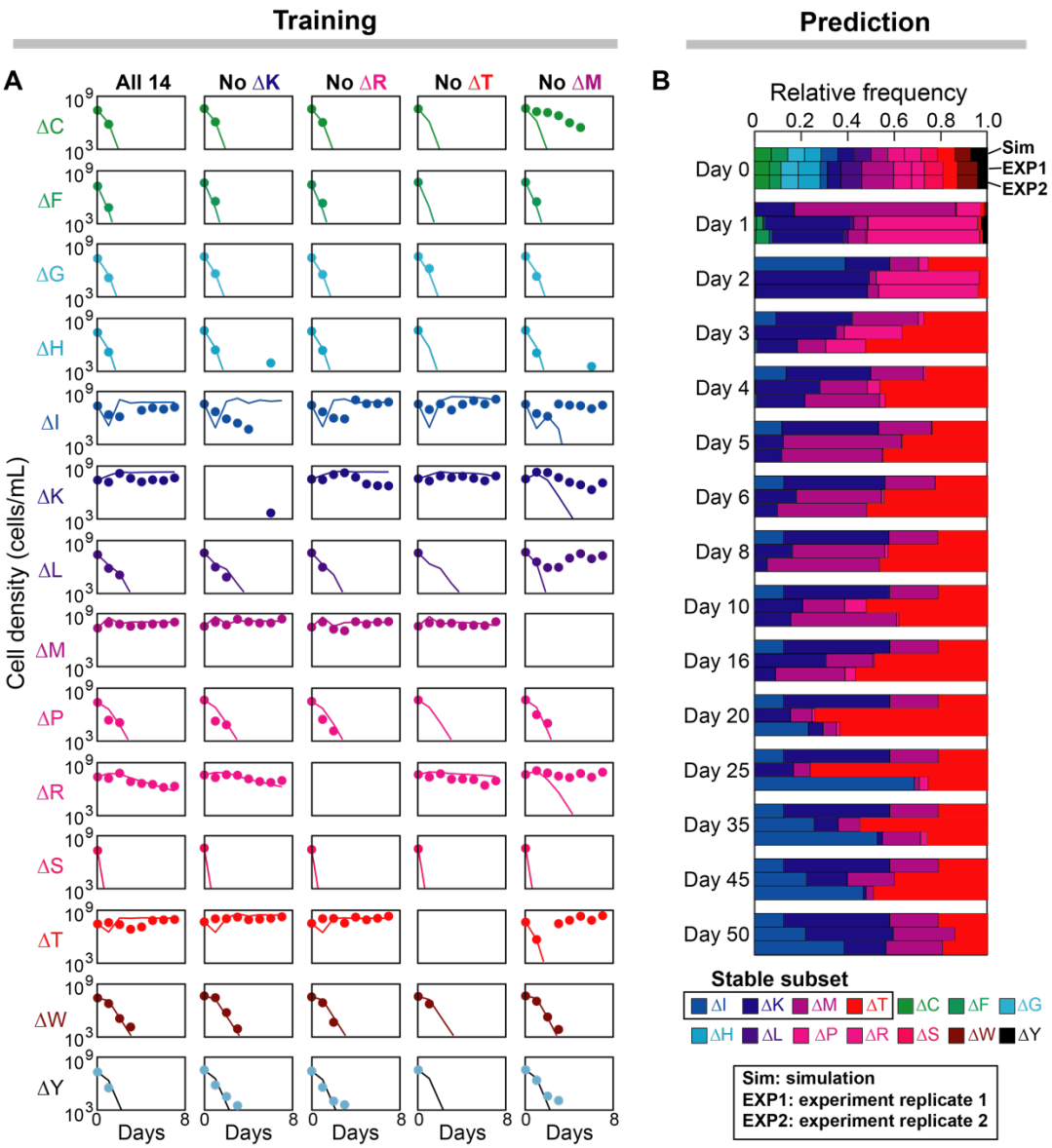


Figure 5 | Modeling a consortium of 14 amino acid auxotrophies. (A) Schematic diagram of the model. Each labeled empty circle represents one amino acid auxotroph and each filled circle with the same color corresponds to the amino acid that it is auxotrophic for. Gray arrows indicate production and release of amino acids to the environment and black arrows indicate the uptake of amino acids by their auxotrophies. (B) Scatter plot (upper panel) and Pearson correlation (bottom panel) between observed[22] and predicted cell density fold changes across all pairwise batch coculture of 14 *E. coli* amino acid auxotrophies. Orange circles: our model with manually curated parameters; Blue circles: a Lotka-Volterra-type model with parameters adopted from Mee *et al.*[22]; Green circles: the same Lotka-Volterra-type model with parameters optimized by Markov-Chain-Monte-Carlo (MCMC) algorithm. (C) Predicted amino acid leakage profiles (converted to percentage of carbon loss) for the 14 amino acid auxotrophies. Each value in the matrix describes the fraction of carbon loss due to release of the amino acid in the row by the auxotroph in the

783 column. Abbreviations: cysteine auxotroph (ΔC), phenylalanine auxotroph (ΔF), glycine
784 auxotroph (ΔG), histidine auxotroph (ΔH), isoleucine auxotroph (ΔI), lysine auxotroph (ΔK),
785 leucine auxotroph (ΔL), methionine auxotroph (ΔM), proline auxotroph (ΔP), arginine auxotroph
786 (ΔR), serine auxotroph (ΔS), threonine auxotroph (ΔT), tryptophan auxotroph (ΔW), and tyrosine
787 auxotroph (ΔY).



788

789 **Figure 6 | Prediction of the long-term steady state of the community of 14 amino acid**
790 **auxotrophies.** (A) Parameters other than the amino acid leakage fractions (obtained from fitting
791 pairwise coculture data in Fig. 5) were manually optimized from the observed population density
792 during a 7-day 100-fold serial dilution of one 14-auxotroph and four 13-auxotroph communities.
793 Filled circles: experiments[22]; Lines: simulation. (B) Simulation of the trained 14-member model
794 over 50 daily passages of the community into fresh medium. The observed long-term stable
795 coexistence of a four-auxotroph subset (ΔI, ΔK, ΔM, ΔT) was correctly predicted. The two

796 replicates of experimental observations were adopted from Mee *et al.*[22]. See Fig. 5 legend for

797 abbreviations of the names of amino acid auxotrophies.

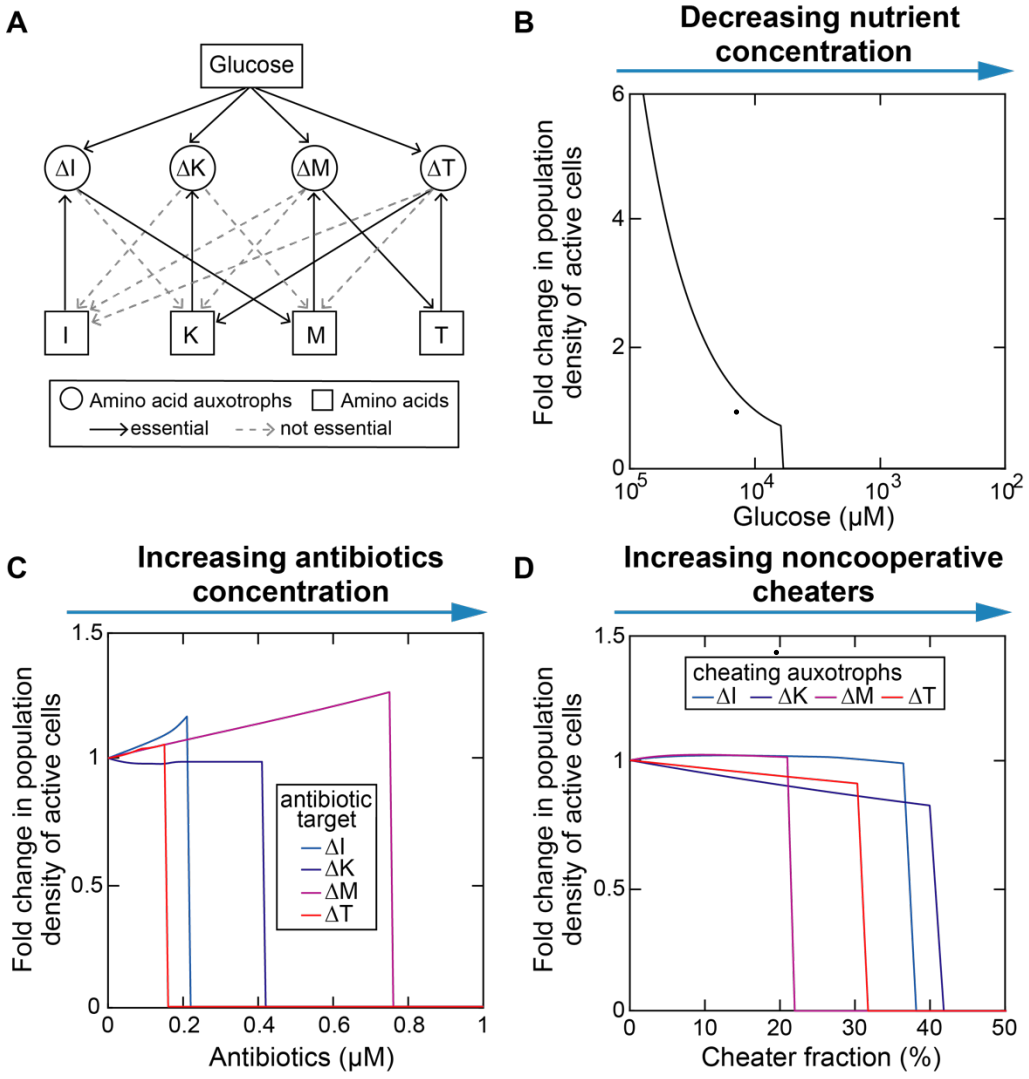


Figure 7 | Collapse of mutualistic cross-feeding network following external perturbations. (A) Bipartite interaction network of the subset of amino acid auxotrophies that stably coexist over long-term serial dilution (see also Fig. 6B). The network contains resource nodes (I, K, M, T for isoleucine, lysine, methionine, and threonine respectively) and consumer nodes (ΔI, ΔK, ΔM, ΔT are their corresponding auxotrophies). Each directed link indicates the presence of a resource-consumer relationship whose corresponding parameter value is not zero. An directed link is essential if its removal leads to loss of community members. (B-D) External perturbations,

807 including decreasing nutrient concentration (B), increasing antibiotic concentration (C), and
808 introducing noncooperative cheaters (D), result in an abrupt collapse of the community when the
809 perturbation level exceeds a certain threshold.