

5831 (errors:5) words

File: main.tex  
 Encoding: utf8  
 Sum count: 5831  
 Words in text: 5102  
 Words in headers: 54  
 Words outside text (captions, etc.): 572  
 Number of headers: 22  
 Number of floats/tables/figures: 8  
 Number of math inlines: 86  
 Number of math displayed: 17  
 Subcounts:  
   text+headers+captions (#headers/#floats/#inlines/#displayed)  
   25+9+0 (1/0/0/0) \_top\_  
   224+1+0 (1/0/0/0) Section: Abstract  
   910+1+0 (1/0/1/0) Section: Introduction  
   0+1+0 (1/0/0/0) Section: Methods} \label{methods}  
   778+3+0 (2/0/16/0) Subsection: Experimental Methods} \label{experimental\_methods  
   1237+16+12 (8/1/61/17) Subsection: CommPhitting  
   0+3+0 (1/0/0/0) Section: Results and Discussion  
   129+2+43 (1/1/0/0) Subsection: Original experiments  
   153+8+208 (1/1/0/0) Subsection: New data fitting method for iterative experimental design  
   276+2+96 (1/2/3/0) Subsection: Simulation insights  
   126+1+0 (1/0/2/0) Subsection: Metabolomics  
   538+5+213 (1/3/3/0) Subsection: \textit{E. coli} Knockout and simulation  
   687+1+0 (1/0/0/0) Section: Conclusion  
   19+1+0 (1/0/0/0) Section: Acknowledgements

File: output.bbl  
 Encoding: utf8  
 Sum count: 0  
 Words in text: 0  
 Words in headers: 0  
 Words outside text (captions, etc.): 0  
 Number of headers: 0  
 Number of floats/tables/figures: 0  
 Number of math inlines: 0  
 Number of math displayed: 0

Total  
 Sum count: 5831  
 Words in text: 5102  
 Words in headers: 54  
 Words outside text (captions, etc.): 572  
 Number of headers: 22  
 Number of floats/tables/figures: 8  
 Number of math inlines: 86  
 Number of math displayed: 17  
 Files: 2  
 Subcounts:  
   text+headers+captions (#headers/#floats/#inlines/#displayed)  
   5102+54+572 (22/8/86/17) File(s) total: main.tex  
   0+0+0 (0/0/0/0) File(s) total: output.bbl

(errors:5)

# A microbial community growth model for dynamic phenotype predictions

Andrew P. Freiburger<sup>ID</sup><sup>1</sup>, Jeffrey A. Dewey<sup>ID</sup><sup>2</sup>, Fatima Foflonker<sup>ID</sup><sup>1</sup>, Gyorgy Babnigg<sup>ID</sup><sup>2</sup>, Dionysios A. Antonopoulos<sup>ID</sup><sup>2</sup>, and Christopher Henry<sup>ID</sup><sup>1\*</sup>

<sup>1</sup>Argonne National Laboratory, Computing, Environment, and Life Sciences Division, Lemont, IL 60148 USA

<sup>2</sup>Argonne National Laboratory, Biosciences Division, Lemont, IL 60148 USA  
\*e-mail: cshenry@anl.gov

May 29, 2023

## 1 Abstract

Microbial communities are increasingly recognized as critical to animal health, agricultural productivity, industrial operations, and ecological systems. The abundance of chemical interactions in these complex communities, however, can evade experimentation and computational approaches – notably flux balance analysis (FBA) and systems of ordinary differential equations (ODEs) – that fail to capture the dynamic phenotype expression of community members. These shortcomings hinder basic understanding of microbial ecology and limits research efforts to rationally design community systems.

We therefore developed a dynamic model (CommPhitting) introduces the appropriate abstractions to fit or explain diverse experimental omics data. CommPhitting captures these data as variables and coefficients within a mixed integer linear optimization problem (MILP) that is globally optimized to most accurately acquire time-resolved predictions of (1) metabolite concentrations, and (2) abundances and (3) linearized growth kinetics for each phenotype of each member. We exemplify CommPhitting by applying it to an idealized two-member community of the model organisms (*Escherichia coli* and *Pseudomonas fluorescens*) that exhibits acetate cross-feeding. Simple batch growth of fluorescently-tagged organisms was sufficient to acquire growth kinetics and biomass abundances that would be difficult to ascertain experimentally, yet are important to understand community behavior. CommPhitting is a pioneering tool that is generalized for diverse growth data, and as an open-source Python API, will foster basic understanding of microbial communities towards synthetic communities engineering in medicine, agriculture, industry, and ecology.

## 2 Introduction

Microbial communities are ubiquitous on Earth [1], and serve fundamental roles as eukaryotic symbionts [2, 3] and pathogens [4, 5], industrial assets [6] and antagonists [7], and ecological agents of biogeochemical cycling [8, 9, 10]. Microbial communities are therefore essential to understand in fields as diverse as medicine [11, 12] and climatology [13, 14, 15], but communities remain understudied [16] despite their potential to test ecological theories: such as how the competition from intimate co-habitation [17] is balanced by synergistic benefits. One benefit of community formation is diversity in a) genetics, b) biochemical vulnerabilities [18], and c) metabolic machinery [19, 20, 21], which allow community members to grow in nutrient-poor or toxic environments by relying on well-adapted members. Another ecological feature of communities is that members differentiate into ecological niches, to minimize competition and augment production efficiency [22] and biomass growth rates [23], which gradually increases synergistic interactions over time. These competitive and synergistic dynamics manifest from chemical exchanges, primarily metabolic cross-feeding (syntrophy) [24, 25], where auxotrophs can be satisfied by the excreted byproducts of other members. Understanding these cross-feeding dynamics has allowed communities to be engineered for desired attributes [26], such as greater bioproduction than isolated monocultures [6] or photosynthetic powered bioproduction [27].

A defining challenge in microbial community research is predictably controlling a particular organism's prevalence or ecological role by adjusting syntrophic dynamics. *P. fluorescens*, for example, has emerged as a bioproduction chassis and plant growth promoting rhizobacteria but it struggles to consume many complex sugars that are present in lignocellulosic or rhizosphere nutrient streams. Rationally designing a community, where secondary organisms predigest complex sugars and other intractable carbon sources into amenable nutrients for *Pseudomonas fluorescens*, could therefore advantageously

improve bioproduction efficiency and crop productivity. Knowledge of these dynamics could further enable engineering of the overflow metabolism of secondary organisms to excrete high-energy carbon sources and thereby maximize the growth of a desired organism, such as *P. fluorescens*.

Experimental methods, however, have several notable problems measuring syntrophy and community dynamics. One problem is that cross-feeding can sometimes be sufficiently rapid to prevent accumulation which disables their experimental detection, except for fluxomics isotope labeling [28]. Another more formidable problem is that syntrophic exchanges scale combinatorially with the size of a community, so detecting and tracing these exchanges quickly becomes untenable for even communities with several members. A third problem is that member phenotypes dynamically change with environmental conditions, even with monocultural isolates [29], which introduces considerable variability that can evade experimental controls.

Computational biology offers a few tools that can compensate for experimental limitations in resolving syntrophic exchanges. Flux Balance Analysis (FBA) [30], for example, is a prominent framework for simulating cellular metabolism without data [31] or can optionally be parameterized with metabolomics [32, 33], proteomics [34], transcriptomics [35, 36], or multi-omics [37] data. FBA has several limitations, however, that hinder its ability to study community systems: 1) FBA, and even Dynamic FBA (dFBA) [38] and other variants of the FBA algorithm, cannot capture phenotypic variability; and 2) the under-determined problem of more flux variables than chemical equations in FBA optimizations compromises precision and reproducibility. Machine-learning (ML) biological models [39, 40] in contrast can lack mechanistic information that is necessary to advance understanding and design principles in basic research. Differential equation models, conversely, are too stiff [41] to accommodate variability in phenotype abundances or diverse experimental datasets. Optimized data fitting attractively balances the strengths and weaknesses of the aforementioned methods by flexibly accommodating diverse experimental data [42, 43, 44] and requiring minimal parameterization while providing sufficient mechanistic resolution to improve understanding and advance engineering principles. Data fitting methods have been applied to deduce kinetics coefficients [45] and phenotype expression [46] from experimental -omics data, but these fitting methods neglect dynamic phenotype abundances and variable growth yields, which are essential insights to precisely engineer microbial communities.

We therefore selected a model benchtop community of *Escherichia coli* and *P. fluorescens* [47] to explore design principles that allow us to understand the community dynamics and to ultimately engineer *E. coli* for maximal growth of the aforementioned bioproduction and PGPR organism *P. fluorescens* in media of complex sugars. These members furthermore exhibit complementary carbon preferences [48] and have not before been investigated as a coculture, which encouraged their coupling for study. This experimental effort was accelerated and enabled by the development of a globally-optimized data fitting model of community phenotypes (CommPhitting) that elucidates extracellular concentrations, conversion factors from experimental signals to species  $\frac{g}{L}$  biomass, and abundances, uptake kinetics [49], and growth yields for each phenotype of each community member over the experimental time-course. CommPhitting additionally leverages FBA to define metabolite flux profiles for each phenotype of each member, which the optimization selectively expresses to best recapitulate the experimental data. CommPhitting solves all data points simultaneously in a linear problem [50] to ascertain the global optimum [51], which circumvents deceptive local minima and uniquely allows the optimization to "plan" for future timepoints. The simulation results revealed that acetate cross-feeding from *E. coli* nourishes *P. fluorescens* in media maltose (an intractable complex sugar), which was validated by numerous metabolomic measurements, an knockout experiment, and subsequent simulations. Simple monocultural and cocultural growth data of our 2-member community was sufficient for CommPhitting to offer numerous testable hypotheses of community behaviors, to suggest mechanisms for engineering community dynamics, and to quantify uptake kinetics, growth yields, and phenotype abundances for each member. CommPhitting is available as an open-source Python module in the ModelSEEDpy, which will further augment the ability of this light-weight yet robust fitting model to illuminate nuances of community interactions and ultimately foster the rational design of microbial communities for myriad fundamental and industrial applications. Our study reveals fundamental principles of community dynamics and presents a simple, generalizable, method for engineering community dynamics towards a prescribed goal.

## 3 Methods

### 3.1 Experimental Methods

The exemplary *P. fluorescens* SBW25 and *E. coli* (minimally modified) MG1655 community was selected for several reasons. Firstly, these members are model organisms that are relevant to diverse disciplines: human microbiome, industrial bioproduction, synthetic biology, and ecological rhizospheres. Secondly, despite that these members are robustly studied as isolates or in other cocultures, they have not been studied together as a coculture.

The coculture was created through the following protocol. The *P. fluorescens* and *E. coli* strains were purchased

from ATC, were stored at  $-80^{\circ}\text{C}$  before preparation for electrocompetency, and were transformed with a plasmid to constitutively express either mNeongreen or mRuby2 fluorescent proteins (GFP and RFP), respectively [52]. Transformed cells were freshly streaked on a LB agar plate with appropriate antibiotics from  $-80^{\circ}\text{C}$  glycerol stocks, and were incubated overnight at  $30^{\circ}\text{C}$ . A single colony from the plate was picked, placed into liquid LB (Lennox) broth with the antibiotics, and shaken @ 250 RPM overnight at  $30^{\circ}\text{C}$ . The 2 mL overnight culture was pelleted (4000x g for 10 min), the supernatant was removed, and the cells were resuspended in 1 mL of M9 media that contains no carbon source. This washing sequence was repeated twice. A 20  $\mu\text{L}$  aliquot of the washed cells was combined with 2 mL of M9 media that contained the appropriate carbon source for each strain – 10 mM D-maltose for *E. coli* and 6 mM 4-hydroxybenzoate for *P. fluorescens* – and was shaken @ 250 RPM for at least 16 hours. These overnight cultures were washed twice, following the same procedure as the overnight culture. These cells were finally analytically examined via optical density (OD 590 nm) and fluorescence using a plate reader (Hidex).

The aforementioned M9 cultured cells were mixed in fresh M9 media to achieve the desired initial cell ratio at OD 0.1 (590 nm) and carbon source concentration/ratios. A 200  $\mu\text{L}$  aliquot of the cell mixture was then added to wells of a sterile, black wall & clear bottom, 96-well imaging plate (Costar). The 96-well plate was then added to the Hidex plate reader and was prewarmed to  $30^{\circ}\text{C}$ . The cells were shaken orbitally at 900 RPM in the plate reader for, at least, 24h while being measured for optical density (600 nm), red fluorescence (544 excitation, 590 emission), and green fluorescence (485 excitation, 535 emission) every 10 minutes.

We utilized two methods to accurately disentangle the composition of a liquid coculture:  $12^{\circ}\text{C}$  and  $37^{\circ}\text{C}$  temperature variability and fluorescence reporter with the green- and red-fluorescence proteins, for *P. fluorescens* and *E. coli* respectively. The precision and mutual validation of these methods is depicted through a combined liquid culture and agar plating experiment (Figure 1), which supports their application in our study as a resolving mechanism for the community members. In this qualitative experiment, 2 mL of M9 media containing either maltose, 4-hydroxybenzoate, or acetate are seeded with mono- or cocultures of *E. coli* and *P. fluorescens* to 0.1 OD. After 24h growth with shaking at  $30^{\circ}\text{C}$ , the liquid cultures are diluted into M9 media and streaked out on LB agar plates. The plates are placed at three different temperatures to enable selective growth of specific organisms. Specifically, *P. fluorescens* grows perceptively after 60h at  $12^{\circ}\text{C}$  but does not grow at  $37^{\circ}\text{C}$ , *E. coli* shows the opposite growth trend, and both organisms grow robustly overnight at  $30^{\circ}\text{C}$ . The results and significance of this experiment are discussed below, but the temperature dependent plating method confirms that fluorescence changes observed in coculture accurately represent changes in each organism's growth/abundance.

Subsequent knockout experimentation were conducted on *E. coli*. We purchased *E. coli* mutants from the *E. coli* Genetic Stock Center [53] for knockouts of the four proteins – PoxB, ACS, Pta, and AckA – that directly contribute to Acetate metabolism in MG1655. Each knockout mutant (KO) was simulated via the same experimental growth protocols that are previously enumerated.

### 3.1.1 Metabolomics

Various metabolomics measurements were acquired for our community system. The cells were prepared and analyzed as described above in the 'Plate-based Coculture Growth', 3.4.1.x section. Subsequently, the cell supernatant was separated via centrifugation at 13000xg for 10 minutes prior to metabolomics analysis.

The University of Chicago metabolomics lab incubated the samples at  $-80^{\circ}\text{C}$  for at least one hour, or up to overnight. Extraction solvent (4 volumes of 100% methanol spiked with internal standards and stored at  $-80^{\circ}\text{C}$ ) was added to the liquid sample (1 volume) in a microcentrifuge tube. Tubes were then centrifuged at  $-10^{\circ}\text{C}$ , 20,000 x g, for 15 min, and the supernatant was used for metabolomic analysis. The targeted data were acquired via an Agilent 8890/5977B and 7890B/5977B GCMS with chemical ionization (CI) and a Positive electron impact - gas chromatography - mass spectrometry (+EI-GC-MS, Agilent, 7890B).

## 3.2 CommPhitting

Our CommPhitting method essentially transforms the ordinary differential equations that define microbial growth and chemical activity into a mixed integer linear optimization problem (MILP) with the variables, constraints, and parameters that are described in Table 1. The MILP is globally optimized to minimize error between the predicted and experimental abundances and to minimize the stationary phenotype for each member, which creates the optimal recapitulation of the data with the least overfitting from phenotype transitions, respectively. The following sections detail each MILP constraint and the objective expression.

Table 1: A glossary of dimensions, parameters, and variables that comprise the fitting model.

Term	Type	Description
<b>DIMENSIONS</b>		
$s$	String	A species in the examined community.
$k$	String	A growth phenotype of species $s$ .
$t$	Float	An experimental time point.
$i$	String	Extracellular metabolite.
$z$	Integer	A biomass partition.
<b>PARAMETERS</b>		
$E_{s,t}$	Float	The experimental growth signal for a species at instant $t$ .
$es_{s,k}$	Boolean	A designation of truth for $k \in s$
$\Delta t$	Float	The seconds per timestep, which determines the amount of biomass growth per timestep.
$n_{k,i}$	Float	The exchange flux of each metabolite $i$ in each strain $k$ .
$cvcf$	Float	Conversion coefficients of phenotype biomass to and from the stationary phase, respectively.
$bcv_k$	Float	The greatest fraction of biomass ( $0 < bcv < 1$ ) of strain $k$ that can transition phenotypes in a timestep.
$cvmn$	Float	The minimal value of variable $cvt_{k,t}$ .
$stat$	Float	The optimization penalty for the stationary phenotype of each species.
$kcat_z$	Float	The $kcat$ growth rate constant for the biomass partition $z$ .
$kcat_k$	Float	The $kcat$ growth rate constant for the phenotype $k$ .
<b>VARIABLES</b>		
$EC_k$	Continuous	The conversion coefficient ( $0 < EC < 1000$ ) from parameter $E_{s,t}$ into biomass, which is unique for each strain $k$ .
$EB_{s,t}$	Continuous	The computed biomass from each experimental datum, as the product of $EC_k$ & $E_{s,t}$ .
$bin_k^z$	Binary	A binary switch that determines whether a given biomass partition of a phenotype is active and contributes to the total $kcat$ of the phenotype.
$b_{k,t}^z$	Continuous	The phenotype biomass partitions, which each exhibit a distinct $kcat$ .
$b_{k,t}$	Continuous	The total predicted phenotype biomass from the fitting model.
$EV_{s,t}$	Continuous	The variance between the computed experimental biomass $EB_{s,t}$ and the predicted biomass $b_{k,t}$ .
$g_{k,t}$	Continuous	The predicted growth rate for each strain at each datum.
$c_{t,i}$	Continuous	The concentration of metabolite $i$ at an experimental datum.
$cvt_{k,t}$ & $cvf_{k,t}$	Continuous	The quantity of strain $k$ biomass that transitions to and from the stationary phase, respectively, at an experimental datum.

### 3.2.1 Metabolic phenotypes

CommPhitting first defines the metabolic phenotypes for each member as the profile excretion and consumption fluxes through the following sequence of constraints, objectives, and optimizations that manipulate ModelSEED [54] genome-scale metabolic models (GEMs) for each specified phenotype of each member.

1. The member GEM is constrained by several cellular and environmental conditions. First, hydrogen consumption is prohibited to force utilization of the phenotype's carbon source(s). Second, oxygen consumption is limited to stoichiometric equivalency with the total consumption of phenotype carbon source(s), although, this threshold can be tailored to better reflect the member's biology and overflow metabolism. Finally, a minimal biomass growth is defined to prevent the subsequent steps from optimizing to extermination.
2. The total influx of all carbon compounds and compounds with undefined formula are minimized, with a 1000x smaller coefficient on the defined phenotype carbon source(s) to incentivize its utilization relative to other carbon compounds. This minimization chisels the metabolic network around the consumption of the phenotype carbon source(s) while also maximizing their biomass yield (although, this biomass yield is later tuned by the optimization through the adjustment of ATP stoichiometry in the biomass reaction). The simulated exchanges are constrained to the fluxes from this minimization and remain unchanged in subsequent optimizations.
3. Optionally, flux of a specified excreta for the phenotype is maximized, which aligns the metabolic profile with experimental observations. The exchange fluxes for these specified excreta are constrained after this maximization.
4. The model is simulated, either with standard FBA or with parsimonious FBA [55] that selects the most efficient pathway with the same objective value and presumably embodies the evolutionarily favorable metabolism.
5. The non-zero fluxes are assembled into the metabolic profile for the given member phenotype.

The above five-step sequence is repeated for each phenotype of each community member. These pre-optimized representations of metabolic phenotypes remain unchanged throughout the simulation except for the noted adjustment of the biomass yield by editing ATP stoichiometry in the biomass reaction. This approach of static metabolic profiles improves simulation efficiency relative to ODE and FBA methods that re-compute mostly the same fluxes each timestep.

### 3.2.2 Constraints

The following several constraints of cellular chemistry and growth dynamics capture community biology in CommPhitting.

**Biomass abundances** The experimental biomass ( $EB_{s,t}$ ), for each member  $s \in S$  over all time points  $t \in T$ , is acquired by converting the experimental growth signal ( $E_{s,t}$ ) via a coefficient ( $EC_s$ )

$$E_{s,t} * EC_s = EB_{s,t}. \quad (1)$$

The final  $EB_{s,t}$  can be explicitly constrained to within 30% of the requisite biomass to consume 90% of the carbon source(s), given a fixed growth yield for the member phenotypes, which can improve accuracy of the  $EC_s$  conversion coefficient. Variance ( $EV_{s,t}$ ) between the converted experimental biomass  $EB_{s,t}$  and the predicted species biomass, as the sum of all  $b_{k,t}$  biomasses for the  $k \in K$  phenotypes of the member, is defined

$$EB_{s,t} - \sum_{s,k}^{S,K} (es_s * b_{k,t}) = EV_{s,t}, \quad (2)$$

where the binary  $es_s$  variable filters for only the phenotypes of species  $s$ . Each predicted biomass  $b_{k,t}$  is further partitioned into five partitions  $\forall z \in Z$

$$b_{k,t}^z \leq kcat_z * b_{k,t} \quad \forall z \in Z \quad (3)$$

that each exhibit a distinct  $kcat_z$  growth uptake rate. The expression of each biomass partition  $b_{k,t}^z$  is controlled by its binary variable  $bin_k^z$

$$b_{k,t}^z \leq 1000 - 1000 * bin_k^z \quad \forall z \in Z \quad (4)$$

and

$$kcat_z * b_{k,t} - b_{k,t}^z - 1000 * bin_k^z \leq 0 \quad \forall z \in Z, \quad (5)$$

where  $bin_k^z = 0$  activates and  $bin_k^z = 1$  deactivates the given partition. One of the biomass partitions must be expressed in each simulation

$$0 <= \sum_{z,k}^{Z,K} (bin_k^z) <= 4 \quad (6)$$

to maintain mass balance continuity through the simulation. The  $kcat_k$  value of each phenotype is calculated as a linear combination of the biomass partitions weighted by their  $kcat_z$  values over three iterations that successively hone towards an increasingly precise  $kcat_k$  prediction

$$kcat_{k,new} = \sum_z^Z (kcat_z * bin_k^z) * kcat_{k,old} \quad (7)$$

where  $kcat_{k,new}$  becomes  $kcat_{k,old}$  in the next loop and the first  $kcat_{k,old}$  is an arbitrary guess. The  $kcat_z$  values for each iteration of the  $5 z \in Z$  partitions [10, 1, 0.1, 0.01, 0.001], then [4, 2, 1, 0.5, 0.25], and finally [1, 0.33, 0.11, 0.037, 0.0123]. This approach critically maintains linearity in eq. (10).

**Phenotype transitions** The biomass change over time is calculated for non-stationary (growing) phenotypes via Heun's integration method [56, 57], which – for an arbitrary function  $y$ , its derivative  $y'$ , and a timestep  $\Delta t$  – is defined as

$$y_{t+1} = y_t + \frac{1}{2} * \Delta t * (y'_t + y'_{t+1}) . \quad (8)$$

Heun's method is a 2<sup>nd</sup>-order Runge-Kutta formulation [58] that captures dynamic changes with high numerical accuracy [59] while maintaining a linear formulation that is amenable with our MILP. Our application of Heun's method

$$b_{t,k} + \frac{\Delta t}{2} (g_{t,k} + g_{t+1,k}) + cvf_{t,k} - cvt_{t,k} = b_{t+1,k} . \quad (9)$$

equates biomass in the next timestep ( $b_{k,t+1}$ ) to the current biomass ( $b_{k,t}$ ) plus the midpoint of biomass growth over the timestep (where the biomass growth rate  $g_{k,t}$  is the biomass derivative  $\frac{\Delta \text{biomass}}{\Delta t}$ ) and plus the net transition of biomass to growth phenotypes  $cvf_{k,t} - cvt_{k,t}$ . The growth rate for phenotype  $k$  is constrained

$$kcat_k * \sum_z^Z (b_{k,t}^z) = g_{k,t} , \quad (10)$$

as the product of the current biomass  $b_{k,t}$  and the growth rate constant  $kcat_k$ , which reflects 1<sup>st</sup>-order kinetics with respect to biomass abundance and zero-order kinetics with respect to substrate concentration. Every species is defined with a stationary (not growing) phenotype that mediates, and subtly delays, phenotype transitions to reflect lagging expression changes. Further, the stationary phenotype allows each species to cease growth once carbon source(s) are exhausted. Future biomass for the stationary phenotype is analogous to eq. (9),

$$b_{k,t} - \sum_s^S (es_s * (cvf_{k,t} - cvt_{k,t})) = b_{k,t+1} \quad (11)$$

except that a)  $g_{k,t} = 0$  by definition; b) the net transition to growth phenotypes is negative, to reflect to opposite direction of biomass transition; and c) the many possible transitions to/from the stationary phenotype are summed. The fraction of biomass that can transition is constrained

$$cvt_{k,t} \leq bcv * b_{k,t} + cvmin \quad (12)$$

to greater than a minimum limit ( $cvmmin$ ) and lesser than a fraction of biomass ( $bcv$ ) above this minimum.

**Concentrations** Future concentrations of substrates ( $c_{t+1,i}$ ) for all  $i \in I$  substrates are constrained

$$c_{t,i} + \frac{\Delta t}{2} \sum_k^K (n_{i,k}(g_{t,k} + g_{t+1,k})) = c_{t+1,i} \quad (13)$$

as another application of Heun's method in eq. (8), where the future concentration  $c_{t+1,i}$  equates the current concentration  $c_{t,i}$  plus the inner product of the exchange flux  $n_{i,k}$  for metabolite  $i$  from the phenotype profile and the biomass growth midpoint from eq. (9).

### 3.2.3 Objective

The objective expression

$$\sum_{s,t}^{S,T} (EV_{s,t}^2 + stat * b_{s,k,t}) , \quad k = stationary \quad (14)$$

minimizes the sum of variance ( $EV$ ) from eq. (2) and the sum total amount of biomass in the stationary phenotypes to best fit the data while avoiding overfitting of excessive phenotype transitions or stationary phenotype expression, where the stationary phenotype mediates all phenotype transitions in eq. (9).

### 3.2.4 Adapting the formulation for growth data

The *P. fluorescens* and *E. coli* GEMs were constructed in KBase Narrative 93465 by leveraging RAST [60] and the ModelSEED pipeline [54]. The *E. coli* model derived from the ASM584v2 experimental genome assembly and was gapfilled with acetate and maltose carbon sources. The *P. fluorescens* model derived from the ASM161270v1 experimental genome assembly and was gapfilled with acetate and 4-hydroxybenzoate carbon sources. We attempted to use published metabolic models for *E. coli* [61] and *P. fluorescens* [62], however, these models were a) not sufficiently responsive to our phenotyping methods (3.2.1) and b) contained incompatible metabolites labels that hinders community cross-feeding.

The general CommPhitting formulation in Sections 3.2.1-3.2.3 was tailored to our particular 2-member community. First, each member's growth phenotypes that we hypothesized were active in the simulations were defined: maltose and acetate for *E. coli*, and 4HB and acetate for *P. fluorescens*. Second, signal-to-biomass conversion factors ( $RFPC$ ,  $GFPC$ , and  $ODC$ ) were defined from eq. (1) to convert each experimental fluorescence growth signal ( $RFP_t$ ,  $GFP_t$ , and  $OD_t$ ) into a corresponding biomass

$$\begin{aligned} RFP_t * RFPC &= RFPB_t \\ GFP_t * GFPC &= GFPB_t \\ OD_t * ODC &= ODB_t . \end{aligned} \quad (15)$$

Third, the variance constraint in eq. (2) was adapted for each experimental signal

$$\begin{aligned} RFPB_t - \sum_k^K (pf_k * b_{t,k}) &= RFPV_t \\ GFPB_t - \sum_k^K (ec_k * b_{t,k}) &= GFPV_t \\ OD_t - \sum_k^K (b_{t,k}) &= ODV_t \end{aligned} \quad (16)$$

where  $pf_k$  and  $ec_k$  are binary variables that filter for *Pseudomonas* and *E. coli*, respectively. Fourth, the objective function of eq. (14) is applied to these members

$$\sum_t^T (EV_{ecoli,t}^2 + EV_{pseudo,t}^2 + EV_{OD,t}^2 + stat * (b_{ecoli,k,t} + b_{pseudo,k,t} + b_{OD,k,t})) , \quad k = stationary . \quad (17)$$

Finally, data at and beyond plateau of the OD signal are removed, which prevents over-production anomalies of fluorescence protein during the stationary phase of growth from skewing the optimization.

## 4 Results and Discussion

### 4.1 Original experiments

The *E. coli* and *P. fluorescens* co-culture was grown in Maltose, 4HB, and Maltose+4HB media and at various temperatures (Figure fig. 1). We specifically sought to understand how environmental substrates – an intractable glucose dimer, maltose, and a common lignin degradation product, 4-hydroxybenzoate – are consumed by *E. coli* and *P. fluorescens*, and especially how to adjust the community such that *Pseudomonas* can grow in otherwise inaccessible environments. We observed that *Pseudomonas* grew monoculturally on 4-HB but was only able to grow in maltose media while in a coculture, which we

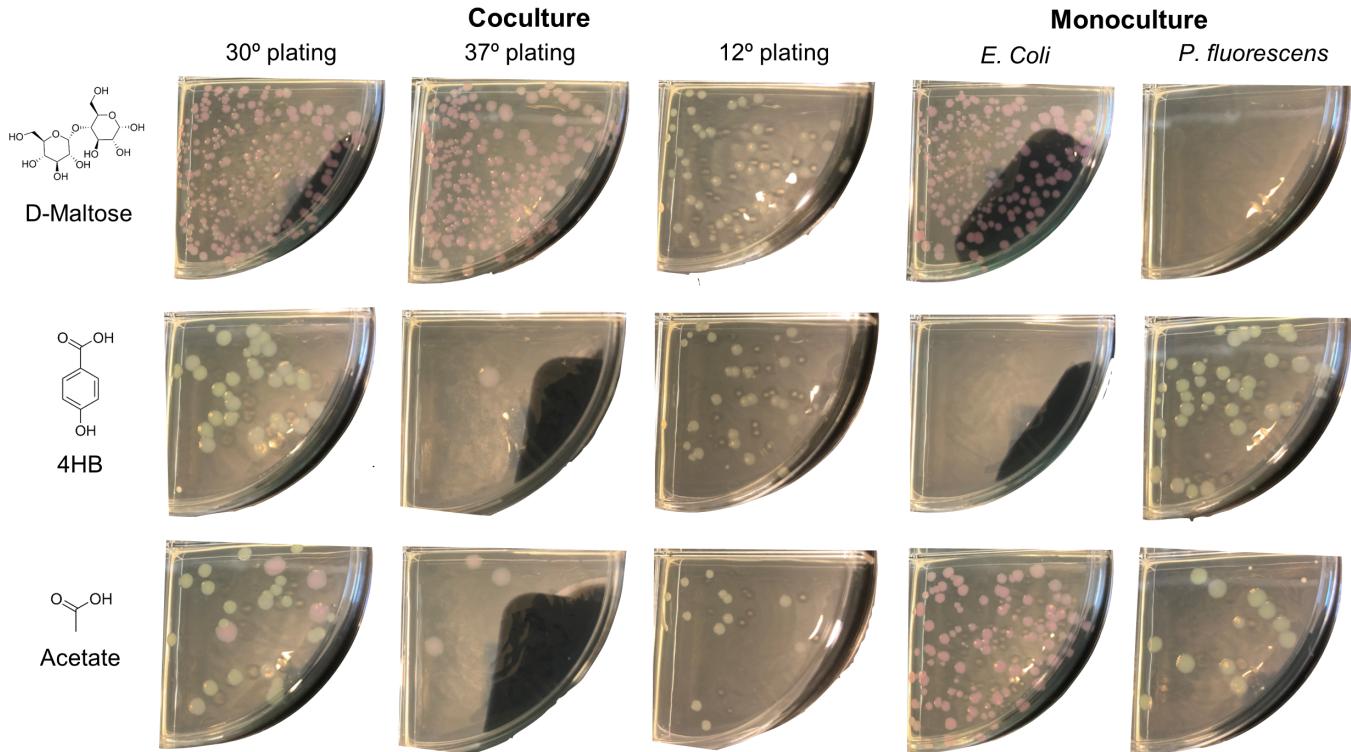


Figure 1: Organism abundance after coculture and monoculture experiments in M9 media containing different carbon sources is visualized via plating at select temperatures. Coculture growth patterns replicate those observed in 96-well plate reader assays, confirming that fluorescence reporters accurately represent each organism’s growth/abundance.

hypothesized was attributed to cross-feeding from an overflow metabolite from *E. coli*’s metabolism of maltose: possibly acetate. We promptly parameterized CommPhitting for this community (Section 3.2.4) to test whether this hypothesis of acetate syntrophy can recapitulate the growth data.

## 4.2 New data fitting method for iterative experimental design

CommPhitting is an original method to resolve extracellular concentrations, and biomass abundances, uptake kinetics, and growth yields for all member phenotypes in a microbial community from simple growth data. The use cases and workflow are illustrated in Figure 2. First, experimental data is parsed into a standardized format and the specified growth phenotypes for all members are defined. Second, the CommPhitting problem is globally optimized to determine the extracellular concentrations, and the biomass abundances, uptake kinetics, and growth yields for each phenotype that best recapitulate the experimental data while minimizing overfitting. Finally, these predicted values are processed through robust, built-in, functions that consolidate the high-dimensional simulation results into exportable figures, HDF5 files, and spreadsheets, which streamlines biological discovery and improves accessibility to investigators who are not data scientists. CommPhitting is available as an open-source Python package in the ModelSEEDpy library ModelSEEDpy ReadTheDocs, and leverages the Optlang optimization module [63] to construct and execute the method.

## 4.3 Simulation insights

The acetate phenotypes of *E. coli* and *P. fluorescens* were both activated while recapitulating the fluorescence growth data, which supports our hypothesis of acetate syntrophy via *E. coli*’s overflow metabolism of maltose. The remarkable data fit (see Figure 4), notwithstanding data anomalies like those in the maltose + 4HB trials of Figure ??, demonstrates the capacity and flexibility of CommPhitting to capture the experimental systems. The simulations of maltose media 5 further demonstrate that the simulation only expresses biologically plausible phenotypes, since *Pseudomonas*’s 4-HB phenotype is never expressed when 4-HB substrate is omitted.

There are several noteworthy phenotypic observations from *E. coli*’s monocultural growth versus its cocultural growth in maltose media. First, the transition from expressing *E. coli*’s maltose phenotype versus its acetate phenotype in Figure

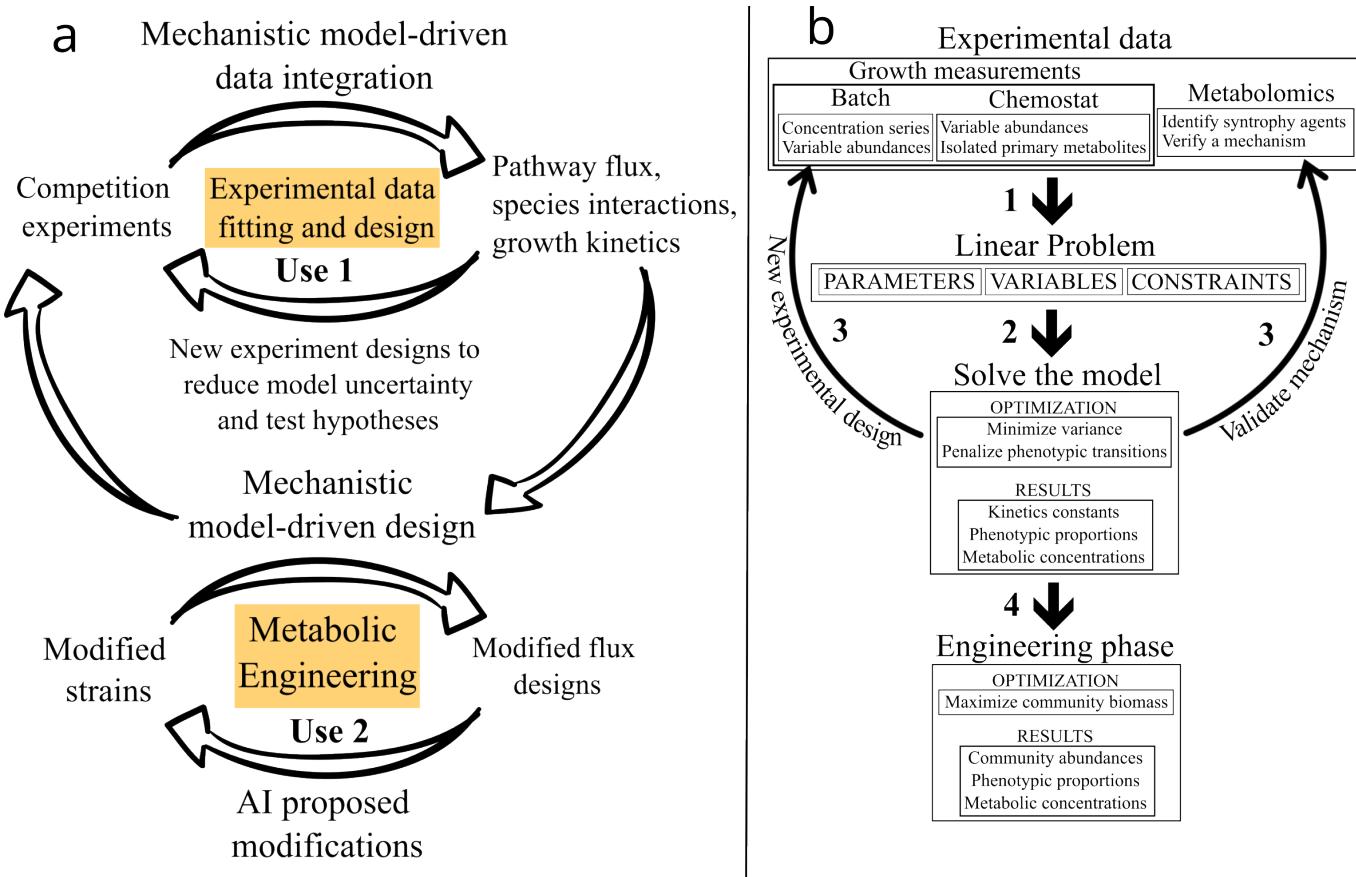


Figure 2: A workflow of the fitting model. Panel a) illustrates logically how CommPhitting can iteratively feedback with experimental efforts and ultimately integrate with engineering efforts and potential AI methods to further interpret experimental data. Panel b) details the technical steps of each CommPhitting simulation. **Step 1:** experimental data – from growth and possibly metabolomics measurements – is parsed into a MILP problem that consists of the parameters and variables that are detailed in Table 1 and the constraints that are explained in Section 3. **Step 2:** the linear problem is executed with the objective function of eq. (14). **Step 3:** the simulation results are interpreted to either identify experimental changes that will improve modeling fit or to propose select additional omics measurements that can further solidify mechanistic insights from the fit. Steps 1-3 repeat until a satisfactory fit and mechanistic resolution is achieved. **Step 4:** the fitted model can be used in a forward design-phase, instead of a purely retrospective fitting-phase, by replacing the objective function of the fitted model with one that maximizes community growth or targets other potential ecological objective functions. The system of steps 1-4 create an integrated method for gleaning mechanistic insights of microbial communities and then immediately using these insights to rationally design a community with desirable activity.

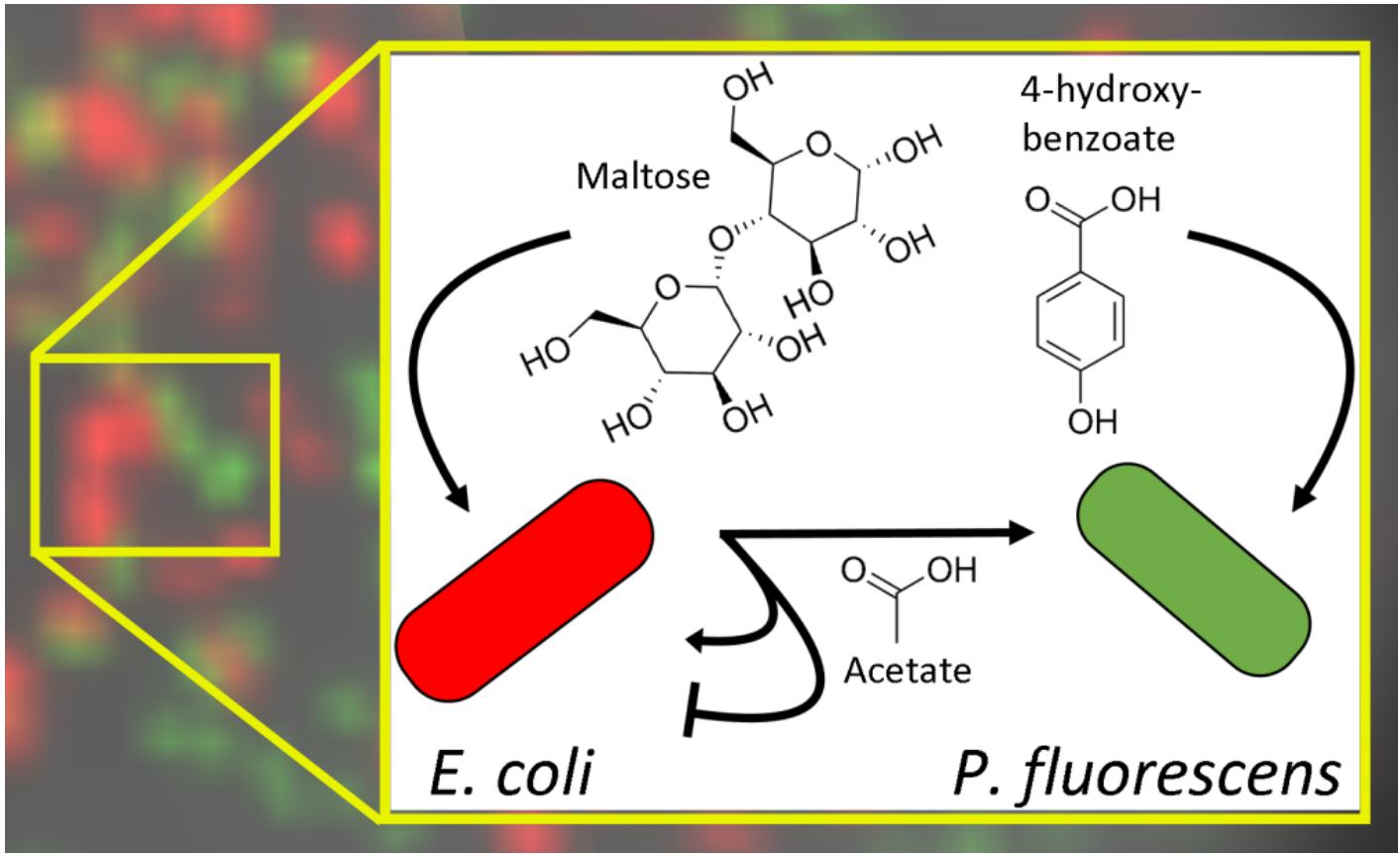


Figure 3: The primary metabolic exchanges that were experimentally elucidated and computationally modeled. The acetate byproduct of *E. coli* is the pivotal exchange of metabolic interest, where this source subsists *P. fluorescens*'s existence in maltose and exhibits interesting dual effects upon *E. coli* as both a secondary carbon source and a growth inhibitor. Additional experiments revealed that *E. coli* is reluctant to grow in pure acetate and most often fails to grow at all, particularly when in cocultures with *P. fluorescens*.

5 occurs sooner than in the coculture, where the transition coincides with when maltose is exhausted. This reflects the need in monoculture to prevent the accumulation of acetate to toxic concentrations while the consumption of acetate in coculture allows *E. coli* to delay consumption of acetate until its preferred food source of maltose becomes sufficiently scarce. Second, the total growth period is much shorter in monoculture than in coculture, perhaps because acetate accumulates to more inhibitory concentrations without *Pseudomonas* aggressively consume it. Third, following the previous observation, *E. coli* more readily expresses its acetate phenotype to presumably detoxify its environment, although it ultimately consumes less acetate as a monoculture as its growth is substantially impeded. Finally, the predicted  $k_{cat,acetate}$  is far greater ( $8.89 \frac{1}{hour}$  versus  $0.23 \frac{1}{hour}$ ) in Table 2, which may be another reflection of growth inhibition from the higher acetate concentrations in the monoculture. Figure 3 embodies the interpreted interactions of our community from these insights.

#### 4.4 Metabolomics

The hypothesis of acetate cross-feeding was further validated with metabolomics experiments. Our targeted data revealed that none of the hypothesized alternative cross-feeding agents, such as formate, were meaningfully detected above the

Simulation	<i>E. coli</i> Acetate	<i>P. fluor.</i> Acetate	<i>E. coli</i> Maltose	<i>P. fluor.</i> 4-HB
Maltose	6e-3	0.28	0.19	266
Maltose+4HB	5e-5	0.19	1	501
Acetate	5e-5	0.22	129	0.86

Table 2: Predicted uptake kinetic coefficients for the various community phenotypes in each of the simulation media.

media background. Our untargeted data mitigated potential bias in selecting cross-feeding agents by examining hundreds of metabolites, in numerous biological categories; yet, acetate is evidently still the prominent cross-feeding agent. No other metabolites accumulated to sufficiently high concentrations to support the observed *Pseudomonas* biomass growth. The role of acetate as an syntrophic exchange is observed in other community systems [1]. Only 10% of the targeted metabolites were detectable, and only 10% were detected above the base media concentrations and varied with experimental growth, which further supports the unique role of acetate as the syntrophic agent in this community.

#### 4.5 *E. coli* Knockout and simulation

We examined the coculture and both isolates in acetate media to further understand acetate utilization in this community. An evident and predictable trend is that *P. fluorescens* more rapidly consumes acetate than *E. coli* (excel file “PF-EC Acetate and Carbon 6-24-22” also “PF-EC-PFEC on Acetate 6-17-22” for relative acetate consumption rates and efficiency aka yields). We then added 4-HB to the acetate media to observe how 4-HB consumption compares with the syntrophic agent, where we observed its preferential consumption over acetate. This suggest that the syntropy is additive to *P. fluorescens* growth in the presence of 4-HB and may effectively augment *P. fluorescens* growth in complex carbon environments.

A final experimental confirmation of acetate as the central cross-feeding agent was the exploration of *E. coli* mutants whose acetate overflow metabolism has been eliminated. Monocultural growth of various KO candidates revealed that only the ΔPta mutant completely eliminated acetate production (Excel file “Acetate KO Ecoli only 10-21-22”) in addition to its consumption. Assessing each mutant in cocultures confirmed that this Pta mutant eliminated acetate overflow metabolism since it is the only KO that significantly altered *E. coli*’s growth (presumably since lactate is not inhibitory like acetate (comparing F7 to F10 in Excel “Acetate KO Ecoli only 10-21-22”)) and perturbed the growth of *P. fluorescens*, where its growth was augmented (comparing well G7 to G10, and column 7 to column 10 more broadly in Excel sheet “PF-ECKOs Ace maltose”). The Pta KO is known to excrete lactate instead of acetate during sugar metabolism [64], which explains the augmentation of *Pseudomonas*’ growth since lactate is a more energetic than acetate. The minimal perturbation of *E. coli*’s growth from the Pta KO further demonstrates the insignificance of its acetate phenotype, which is predicted by CommPhitting simulations where maltose was present.

A time course metabolomics experiment with wt and Pta-KO *E. coli* cocultures on either maltose or maltose + 4HB validated that lactate is indeed produced x( 7-10x-fold I think) more in Pta-KO cells than wt. These experiments further confirmed the preferred consumption of 4-HB to small organic acids where lactate consumption was delayed in the presence of 4HB (based on earlier metabolomics data from 08-2022). Inhibition of 4-HB on *E. coli*’s lactate consumption is also possible (Excel sheet “TMS metabolomics with Graphs 3-28-23”), although the monocultural growth in the presence of 4-HB was not observed. The metabolites whose concentrations significantly exceeded the media background and deviated from the wildtype metabolomics were presumed to represent unique chemistry in the Pta-KO phenotype.

The Pta mutant coculture was simulated by CommPhitting to elucidate dynamics between the phenotypes that were inferred from the metabolomics data. The Pta phenotype as inferred from the metabolomics data was mostly captured; however, the co-excretion of 2-hydroxybutyric acid with lactate (1:10 ratio) could not be considered since neither GEM model contained this metabolite. The chemical and functional similarity of this compound to butyric acid [1], however, suggests that it may be a peripheral cross-feeding agent in the Pta knockout, albeit with an order-of-magnitude smaller concentration according to the metabolomics data. The simulation results revealed similar behavior of the lactate phenotype with the acetate phenotype of the wildtype in coculture, which supports that this excreta results from the same overflow metabolism.

## 5 Conclusion

Microbial communities offer untold potential as biotechnologies and to expand basic biological and ecological understanding; however, progress is hindered by a) knowledge gaps of interspecies exchanges and their influence on community and phenotype dynamics, and b) an absence of predictive methods for rational design. Our study of a model coculture of *E. coli* and *Pseudomonas* revealed cross-feeding from complex sugar metabolism. We suspectec acetate as the syntrophic agent, but we wanted to mechanistically understand phenotype dynamics with this hypothesized syntropy to better understand the system and how to engineer it. Extant computational models were unequipped to revolve this information from our data, so we developed a Community Phenotype Fitting method (CommPhitting) that uniquely requires only two inputs – member metabolic models and growth data (OD and fluorescence) – to predict time-resolved metabolite concentrations, and abundances, growth yields, and uptake kinetics for each specified phenotype of each member. CommPhitting specifically can accelerate progress by a) indicating knowledge gaps where dissonance between predictions and experimental validation, or where poor data fitting, is observed; and b) generating molecular hypotheses that can be evaluated and

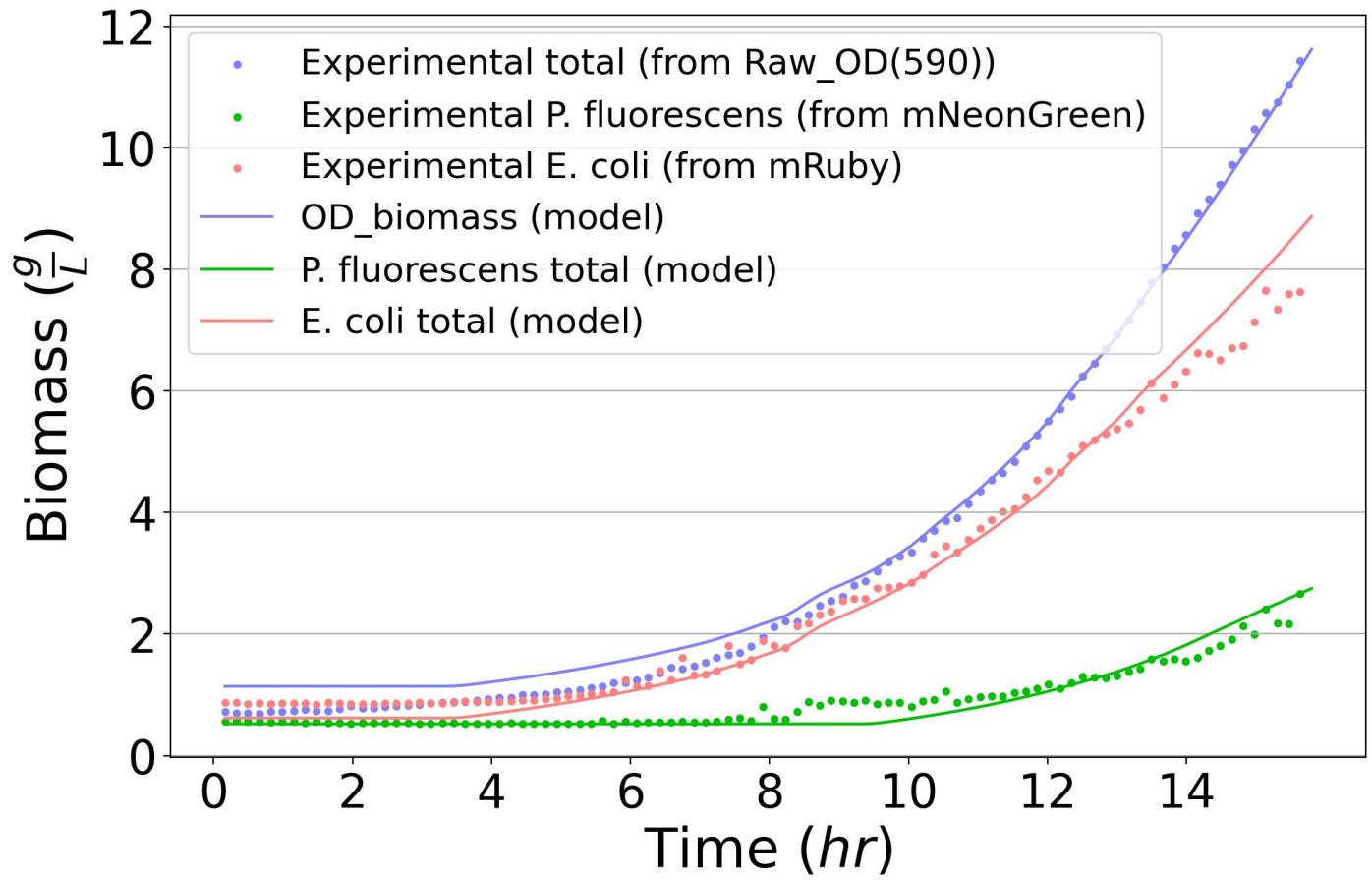


Figure 4: The converted experimental and predicted biomasses of the total community (OD), *E. coli* (RFP), and *P. fluorescens* (GFP) for the coculture experiment on maltose media (Table S1). Tight agreement between the experimental and predicted biomass values improves confidence in the predicted community behaviors and underlying chemical parameters.

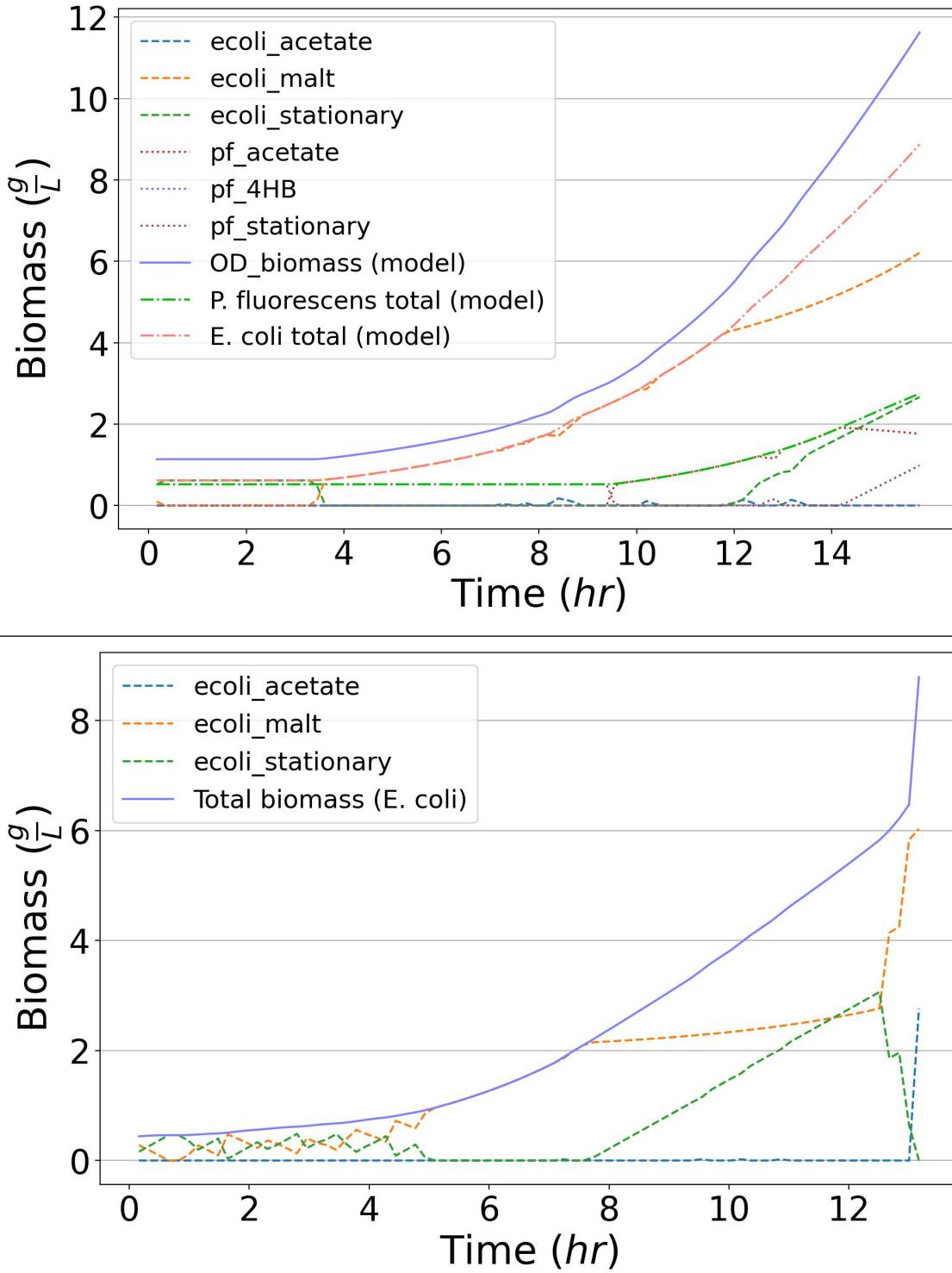


Figure 5: The phenotype abundances for an experiment with 5mM of maltose as the sole carbon sources. The top figure depicts the 1:1 coculture, where *E. coli* outcompeted *P. fluorescens*; however, *P. fluorescens* managed to subsist on the acetate excreta from *E. coli*. The bottom figure depicts a *E. coli* monoculture in maltose, where it reaches the same final biomass but consumes 50% more acetate, albeit very reluctantly. The monoculture interestingly also exhibited much more stationary phenotype during the lag phase than *E. coli* in the same conditions as a coculture. A *P. fluorescens* monoculture figure is not depicted since *P. fluorescens* exhibited no growth.

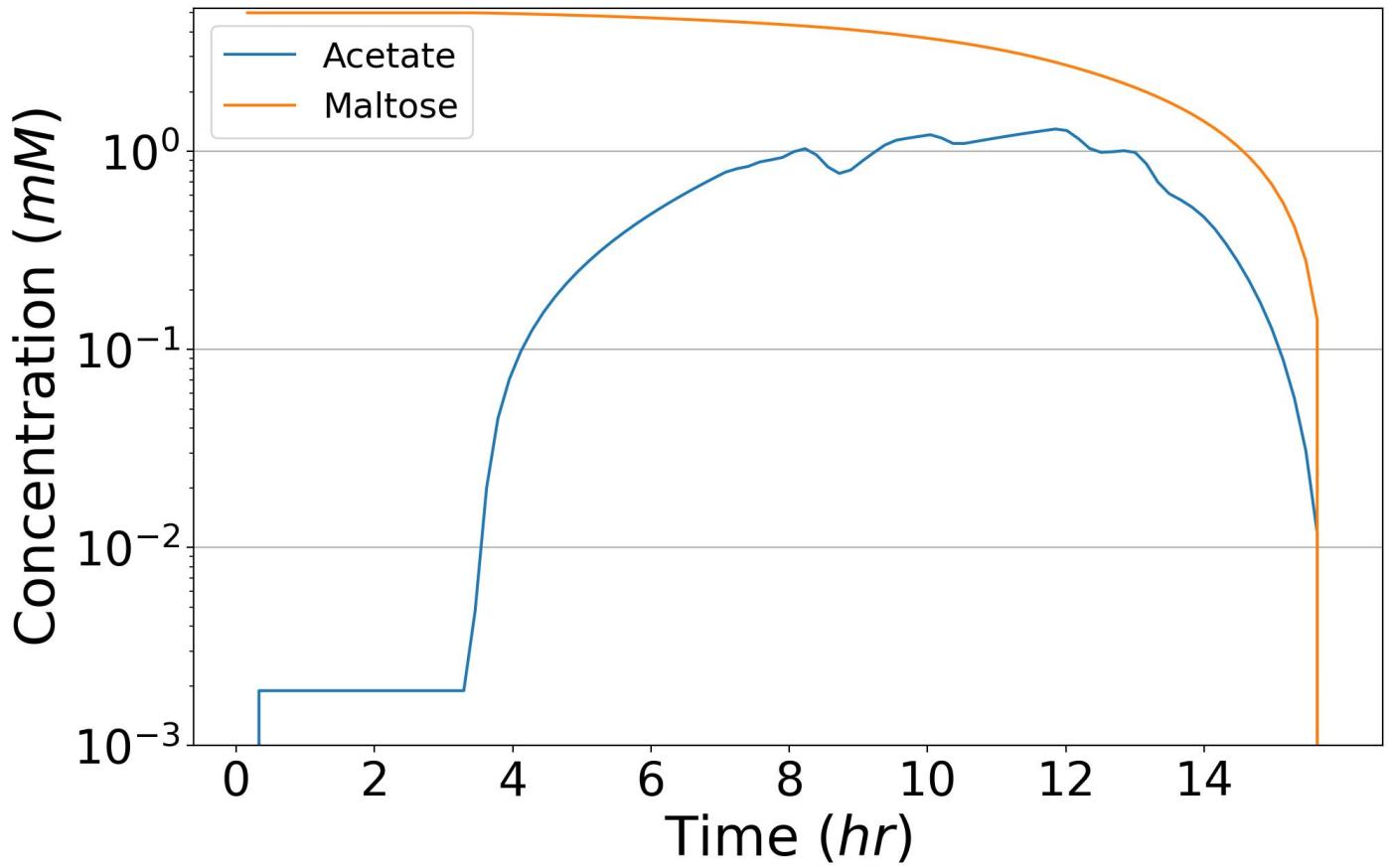


Figure 6: The concentrations of primary carbon sources and the predominate cross-feeding agent acetate from the coculture trial of Figure 5. The dynamic production and subsequent consumption of acetate elicits the unique information that can be derived from CommPhitting simulations. The peak acetate concentration at  $2.4 \text{ mM}$  corroborates with the peak magnitude from the metabolomics data at the same OD value, which validates the concentration mechanisms of CommPhitting.

re-simulated with simple growth experiments, thereby expanding basic knowledge of community dynamics and ultimately engineering these communities for diverse applications. The outputs from our simulations concisely revealed acetate dynamics in *E. coli* and *P. fluorescens* and illustrated the delay of *E. coli*'s consumption of acetate in coculture much longer than in monoculture, presumably since *Pseudomonas* lessens inhibitory pressure in the coculture.

The CommPhitting results emboldened us to allocate resources towards *E. coli* KOs that selectively perturb acetate overflow metabolism. We observed that only the Pta KO of *E. coli* (Pta-KO) eliminated acetate excretion, where lactate was excreted instead. The Pta-KO slightly improved *E. coli*'s growth, presumably because lactate is less inhibitory than acetate, and substantially augmented *Pseudomonas*' growth, since lactate is a much higher energy carbon source than acetate. This perturbation from a precise KO of acetate's overflow metabolism concretely affirms that acetate excretion is the primary syntrophic exchange with the wildtype system and that sufficiently accurate knowledge of the community dynamics were gleaned from the experiments and CommPhitting simulations to rationally engineer community that selectively augments the growth of a soil microbe. Simulations of the Pta KO coculture further revealed similar expression behavior of the lactate phenotype to the wildtype acetate phenotype, which supports that the Pta-KO simply replaces the overflow metabolite in the same underlying pathway. We also observed 4-HB inhibition on lactate and acetate consumption for both *Pseudomonas* and *E. coli*, which is unexpected and may result from one of several interactions. First, *Pseudomonas* may prefer over acetate/lactate and poison *E. coli*'s ability to metabolize them while it consumes 4-HB. Second, 4-HB, as another organic acid, could directly inhibit *E. coli* metabolism of acetate and lactate unlike maltose metabolism (column 11 + rows B, D, F of PF-EC-ACS-4-HB 5-10-22). Third, the decreasing concentration of 4-HB stimulates acetate/lactate production in *E. coli* without stimulating acetate/lactate consumption in *Pseudomonas*, which leads to the observed accumulation of acetate/lactate.

Our study, in summary, rapidly acquired molecular- and phenotype-level understanding of syntrophic dynamics of a model community through the unique interplay of experimentation – batch growth, knockout engineering, and metabolomics – and a novel modeling framework – CommPhitting. Our experiments and simulations suggest several testable hypotheses for future study. First, we plan to simulate community BIOLOG data, which would augment basic understanding of community dynamics, growth yields, and uptake kinetics in diverse conditions and provide more guidance for a range of community engineering efforts. The ecological role of *Pseudomonas* preferring lignin derivatives or overflow metabolites may also be clarified by BIOLOG experiments. We additionally envision that CommPhitting can be extended from the investigation phase that was applied herein to an exploration phase (Step 4 of Figure 2) where engineering hypotheses of community behaviors perturbations in arbitrary environments can be predicted from data in given environments, e.g. metabolic mutants or the presence of toxins/antibiotics. We finally plan to wrap CommPhitting into a point-and-click DOE KnowledgeBase (KBase) application [65], where non-technical users will be able to utilize this new tool for diverse investigations of community biology and synthetic community design.

## 6 Acknowledgements

We are immensely grateful to Maciek R. Antoniewicz of the University of Michigan for his insightful comments and perspectives.

## References

- [1] Stephen Nayfach, Simon Roux, Rekha Seshadri, Daniel Udwary, Neha Varghese, Frederik Schulz, Dongying Wu, David Paez-Espino, I. Min Chen, Marcel Huntemann, Krishna Palaniappan, Joshua Ladau, Supratim Mukherjee, T. B.K. Reddy, Torben Nielsen, Edward Kirton, José P. Faria, Janaka N. Edirisinghe, Christopher S. Henry, Sean P. Jungbluth, Dylan Chivian, Paramvir Dehal, Elisha M. Wood-Charlson, Adam P. Arkin, Susannah G. Tringe, Axel Visel, Helena Abreu, Silvia G. Acinas, Eric Allen, Michelle A. Allen, Lauren V. Alteio, Gary Andersen, Alexandre M. Anesio, Graeme Attwood, Viridiana Avila-Magaña, Yacine Badis, Jake Bailey, Brett Baker, Petr Baldrian, Hazel A. Barton, David A.C. Beck, Eric D. Becraft, Harry R. Beller, J. Michael Beman, Rizlan Bernier-Latmani, Timothy D. Berry, Anthony Bertagnoli, Stefan Bertilsson, Jennifer M. Bhatnagar, Jordan T. Bird, Jeffrey L. Blanchard, Sara E. Blumer-Schuette, Brendan Bohannan, Mikayla A. Borton, Allyson Brady, Susan H. Brawley, Juliet Brodie, Steven Brown, Jennifer R. Brum, Andreas Brune, Donald A. Bryant, Alison Buchan, Daniel H. Buckley, Joy Buongiorno, Hinsby Cadillo-Quiroz, Sean M. Caffrey, Ashley N. Campbell, Barbara Campbell, Stephanie Carr, Jo Lynn Carroll, S. Craig Cary, Anna M. Cates, Rose Ann Cattolico, Ricardo Cavicchioli, Ludmila Chistoserdova, Maureen L. Coleman, Philippe Constant, Jonathan M. Conway, Walter P. Mac Cormack, Sean Crowe, Byron Crump, Cameron Currie, Rebecca Daly, Kristen M. DeAngelis, Vincent Denef, Stuart E. Denman, Adey Desta, Hebe Dionisi, Jeremy Dodsworth, Nina Dombrowski, Timothy Donohue, Mark Dopson, Timothy Driscoll, Peter Dunfield, Christopher L.

Dupont, Katherine A. Dynarski, Virginia Edgcomb, Elizabeth A. Edwards, Mostafa S. Elshahed, Israel Figueroa, Beverly Flood, Nathaniel Fortney, Caroline S. Fortunato, Christopher Francis, Claire M.M. Gachon, Sarahi L. Garcia, Maria C. Gazitua, Terry Gentry, Lena Gerwick, Javad Gharechahi, Peter Girguis, John Gladden, Mary Gradoville, Stephen E. Grasby, Kelly Gravuer, Christen L. Grettenberger, Robert J. Gruninger, Jiarong Guo, Mussie Y. Habteselassie, Steven J. Hallam, Roland Hatzenpichler, Bela Hausmann, Terry C. Hazen, Brian Hedlund, Cynthia Henny, Lydie Herfort, Maria Hernandez, Olivia S. Hershey, Matthias Hess, Emily B. Hollister, Laura A. Hug, Dana Hunt, Janet Jansson, Jessica Jarett, Vitaly V. Kadnikov, Charlene Kelly, Robert Kelly, William Kelly, Cheryl A. Kerfeld, Jeff Kimbrel, Jonathan L. Klassen, Konstantinos T. Konstantinidis, Laura L. Lee, Wen Jun Li, Andrew J. Loder, Alexander Loy, Mariana Lozada, Barbara MacGregor, Cara Magnabosco, Aline Maria da Silva, R. Michael McKay, Katherine McMahon, Chris S. McSweeney, Mónica Medina, Laura Meredith, Jessica Mizzi, Thomas Mock, Lily Momper, Mary Ann Moran, Connor Morgan-Lang, Duane Moser, Gerard Muyzer, David Myrold, Maisie Nash, Camilla L. Nesbø, Anthony P. Neumann, Rebecca B. Neumann, Daniel Noguera, Trent Northen, Jeanette Norton, Brent Nowinski, Klaus Nüsslein, Michelle A. O’Malley, Rafael S. Oliveira, Valeria Maia de Oliveira, Tullis Onstott, Jay Osvatic, Yang Ouyang, Maria Pachiadaki, Jacob Parnell, Laila P. Partida-Martinez, Kabir G. Peay, Dale Pelletier, Xuefeng Peng, Michael Pester, Jennifer Pett-Ridge, Sari Peura, Petra Pjevac, Alvaro M. Plominsky, Anja Poehlein, Phillip B. Pope, Nikolai Ravin, Molly C. Redmond, Rebecca Reiss, Virginia Rich, Christian Rinke, Jorge L. Mazza Rodrigues, William Rodriguez-Reillo, Karen Rossmassler, Joshua Sackett, Ghasem Hosseini Salekdeh, Scott Saleska, Matthew Scarborough, Daniel Schachtman, Christopher W. Schadt, Matthew Schrenk, Alexander Sczyrba, Aditi Sengupta, Joao C. Setubal, Ashley Shade, Christine Sharp, David H. Sherman, Olga V. Shubenkova, Isabel Natalia Sierra-Garcia, Rachel Simister, Holly Simon, Sara Sjöling, Joan Slonczewski, Rafael Soares Correa de Souza, John R. Spear, James C. Stegen, Ramunas Stepanauskas, Frank Stewart, Garret Suen, Matthew Sullivan, Dawn Sumner, Brandon K. Swan, Wesley Swingley, Jonathan Tarn, Gordon T. Taylor, Hanno Teeling, Memory Tekere, Andreas Teske, Torsten Thomas, Cameron Thrash, James Tiedje, Claire S. Ting, Benjamin Tully, Gene Tyson, Osvaldo Ulloa, David L. Valentine, Marc W. Van Goethem, Jean VanderGheynst, Tobin J. Verbeke, John Vollmers, Aurèle Vuillemin, Nicholas B. Waldo, David A. Walsh, Bart C. Weimer, Thea Whitman, Paul van der Wielen, Michael Wilkins, Timothy J. Williams, Ben Woodcroft, Jamie Wootton, Kelly Wrighton, Jun Ye, Erica B. Young, Noha H. Youssef, Feiqiao Brian Yu, Tamara I. Zemskaya, Ryan Ziels, Tanja Woyke, Nigel J. Mouncey, Natalia N. Ivanova, Nikos C. Kyripides, and Emiley A. Eloe-Fadrosh. A genomic catalog of Earth’s microbiomes. *Nature Biotechnology*, 39(4):499–509, 2021. ISSN 15461696. doi: 10.1038/s41587-020-0718-6.

- [2] Manish Kumar, Boyang Ji, Karsten Zengler, and Jens Nielsen. Modelling approaches for studying the microbiome. *Nature Microbiology*, 4(8):1253–1267, 2019. ISSN 20585276. doi: 10.1038/s41564-019-0491-9. URL <http://dx.doi.org/10.1038/s41564-019-0491-9>.
- [3] Saeed Shoae, Fredrik Karlsson, Adil Mardinoglu, Intawat Nookaew, Sergio Bordel, and Jens Nielsen. Understanding the interactions between bacteria in the human gut through metabolic modeling. *Scientific Reports*, 3:1–10, 2013. ISSN 20452322. doi: 10.1038/srep02532.
- [4] Luke Mcguigan and Máire Callaghan. The evolving dynamics of the microbial community in the cystic fibrosis lung. *Environmental Microbiology*, 17(1):16–28, 2015. ISSN 14622920. doi: 10.1111/1462-2920.12504.
- [5] Howard F. Jenkinson and Richard J. Lamont. Oral microbial communities in sickness and in health. *Trends in Microbiology*, 13(12):589–595, 2005. ISSN 0966842X. doi: 10.1016/j.tim.2005.09.006.
- [6] Robin K. Pettit. Mixed fermentation for natural product drug discovery. *Applied Microbiology and Biotechnology*, 83(1):19–25, 2009. ISSN 01757598. doi: 10.1007/s00253-009-1916-9.
- [7] Ethan T. Hillman, Louis Edwards Caceres-Martinez, Gozdem Kilaz, and Kevin V. Solomon. Top-down enrichment of oil field microbiomes to limit souring and control oil composition during extraction operations. *AIChE Journal*, (September):1–13, 2022. ISSN 15475905. doi: 10.1002/aic.17927.
- [8] Bryan S. Griffiths and Laurent Philippot. Insights into the resistance and resilience of the soil microbial community. *FEMS Microbiology Reviews*, 37(2):112–129, 2013. ISSN 01686445. doi: 10.1111/j.1574-6976.2012.00343.x.
- [9] William B. Whitman, David C. Coleman, and William J. Wiebe. Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, 95(12):6578–6583, 1998. ISSN 00278424. doi: 10.1073/pnas.95.12.6578.
- [10] Anne E. Dukas, Rachel S. Poretsky, and Victoria J. Orphan. Deep-Sea Archaea Fix and Share Nitrogen in Methane-Consuming Microbial Consortia. *Science*, 326(October):422–427, 2009.

- [11] Xuesong He, Jeffrey S. McLean, Lihong Guo, Renate Lux, and Wenyuan Shi. The social structure of microbial community involved in colonization resistance. *ISME Journal*, 8(3):564–574, 2014. ISSN 17517370. doi: 10.1038/ismej.2013.172.
- [12] Daniel van der Lelie, Akihiko Oka, Safiyh Taghavi, Junji Umeno, Ting Jia Fan, Katherine E. Merrell, Sarah D. Watson, Lisa Ouellette, Bo Liu, Muyiwa Awoniyi, Yunjia Lai, Liang Chi, Kun Lu, Christopher S. Henry, and R. Balfour Sartor. Rationally designed bacterial consortia to treat chronic immune-mediated colitis and restore intestinal homeostasis. *Nature Communications*, 12(1):1–17, 2021. ISSN 20411723. doi: 10.1038/s41467-021-23460-x.
- [13] Ricardo Cavicchioli, William J. Ripple, Kenneth N. Timmis, Farooq Azam, Lars R. Bakken, Matthew Baylis, Michael J. Behrenfeld, Antje Boetius, Philip W. Boyd, Aimée T. Classen, Thomas W. Crowther, Roberto Danovaro, Christine M. Foreman, Jef Huisman, David A. Hutchins, Janet K. Jansson, David M. Karl, Britt Koskella, David B. Mark Welch, Jennifer B.H. Martiny, Mary Ann Moran, Victoria J. Orphan, David S. Reay, Justin V. Remais, Virginia I. Rich, Brajesh K. Singh, Lisa Y. Stein, Frank J. Stewart, Matthew B. Sullivan, Madeleine J.H. van Oppen, Scott C. Weaver, Eric A. Webb, and Nicole S. Webster. Scientists’ warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology*, 17(9):569–586, 2019. ISSN 17401534. doi: 10.1038/s41579-019-0222-5. URL <http://dx.doi.org/10.1038/s41579-019-0222-5>.
- [14] V. J. Orphan, C. H. House, K. U. Hinrichs, K. D. McKeegan, and E. F. DeLong. Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science*, 293(5529):484–487, 2001. ISSN 00368075. doi: 10.1126/science.1061338.
- [15] Jennifer B. Glass and Victoria J. Orphan. Trace metal requirements for microbial enzymes involved in the production and consumption of methane and nitrous oxide. *Frontiers in Microbiology*, 3(FEB):1–20, 2012. ISSN 1664302X. doi: 10.3389/fmicb.2012.00061.
- [16] A F White, S A Billings, A M Rossi, K R Hubbert, S M Mudd, E T A Mitchard, W E Dietrich, K Singha, B J Minsley, W S Holbrook, A F Sheehan, M Leopold, S P Anderson, J Taylor Perron, S J Martel, K Singha, R S Anderson, S P Anderson, S P Anderson, G E Tucker, and S L Brantley. A unified initiative to harness Earth’s microbiomes. *Science*, 315(6260):2015–2017, 2015.
- [17] Jacob D. Palmer and Kevin R. Foster. Bacterial species rarely work together. *Science*, 376(6593):581–582, 2022. ISSN 10959203. doi: 10.1126/science.abn5093.
- [18] Cristiane Aparecida Pereira, Rogério Lima Romeiro, Anna Carolina Borges Pereira Costa, Ana Karina Silva MacHado, Juliana Campos Junqueira, and Antonio Olavo Cardoso Jorge. Susceptibility of *Candida albicans*, *Staphylococcus aureus*, and *Streptococcus mutans* biofilms to photodynamic inactivation: An in vitro study. *Lasers in Medical Science*, 26(3):341–348, 2011. ISSN 02688921. doi: 10.1007/s10103-010-0852-3.
- [19] Thien Fah C. Mah and George A. O’Toole. Mechanisms of biofilm resistance to antimicrobial agents. *Trends in Microbiology*, 9(1):34–39, 2001. ISSN 0966842X. doi: 10.1016/S0966-842X(00)01913-2.
- [20] K. Sauer, M. C. Cullen, A. H. Rickard, L. A.H. Zeef, D. G. Davies, and P. Gilbert. Characterization of nutrient-induced dispersion in *Pseudomonas aeruginosa* PAO1 biofilm. *Journal of Bacteriology*, 186(21):7312–7326, 2004. ISSN 00219193. doi: 10.1128/JB.186.21.7312-7326.2004.
- [21] Prashanth Suntharalingam and Dennis G. Cvitkovitch. Quorum sensing in streptococcal biofilm formation. *Trends in Microbiology*, 13(1):3–6, 2005. ISSN 0966842X. doi: 10.1016/j.tim.2004.11.009.
- [22] Michael T. Mee, James J. Collins, George M. Church, and Harris H. Wang. Syntrophic exchange in synthetic microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20), 2014. ISSN 10916490. doi: 10.1073/pnas.1405641111.
- [23] Samay Pande, Holger Merker, Katrin Bohl, Michael Reichelt, Stefan Schuster, Luís F. De Figueiredo, Christoph Kaleta, and Christian Kost. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME Journal*, 8(5):953–962, 2014. ISSN 17517370. doi: 10.1038/ismej.2013.211.
- [24] Erica C. Seth and Michiko E. Taga. Nutrient cross-feeding in the microbial world. *Frontiers in Microbiology*, 5 (JULY):1–6, 2014. ISSN 1664302X. doi: 10.3389/fmicb.2014.00350.

- [25] Edwin H. Wintermute and Pamela A. Silver. Emergent cooperation in microbial metabolism. *Molecular Systems Biology*, 6(407):1–7, 2010. ISSN 17444292. doi: 10.1038/msb.2010.66. URL <http://dx.doi.org/10.1038/msb.2010.66>.
- [26] Jack A. Connolly, William R. Harcombe, Michael J. Smanski, Linda L. Kinkel, Eriko Takano, and Rainer Breitling. Harnessing intercellular signals to engineer the soil microbiome. *Natural Product Reports*, 39(2):311–324, 2022. ISSN 14604752. doi: 10.1039/d1np00034a.
- [27] Cristal Zuñiga, Tingting Li, Michael T. Guarnieri, Jackson P. Jenkins, Chien Ting Li, Kerem Bingol, Young Mo Kim, Michael J. Betenbaugh, and Karsten Zengler. Synthetic microbial communities of heterotrophs and phototrophs facilitate sustainable growth. *Nature Communications*, 11(1), 2020. ISSN 20411723. doi: 10.1038/s41467-020-17612-8. URL <http://dx.doi.org/10.1038/s41467-020-17612-8>.
- [28] Nikodimos A. Gebreselassie and Maciek R. Antoniewicz. 13C-metabolic flux analysis of co-cultures: A novel approach. *Metabolic Engineering*, 31:132–139, 2015. ISSN 10967184. doi: 10.1016/j.ymben.2015.07.005. URL <http://dx.doi.org/10.1016/j.ymben.2015.07.005>.
- [29] R Frank Rosenzweig, R R Sharp, David S Treves, and Julian Adams. Microbial evolution in a simple unstructured environment: Genetic differentiation in *Escherichia coli*. *Genetics Society of America*, 137:903–917, 1994.
- [30] Jeffrey D. Orth, Ines Thiele, and Bernhard O. Palsson. What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248, 2010. ISSN 10870156. doi: 10.1038/nbt.1614.
- [31] Daniel R. Hyduke, Nathan E. Lewis, and Bernhard O. Palsson. Analysis of omics data with genome-scale models of metabolism. *Molecular BioSystems*, 9(2):167–174, 2013. ISSN 1742206X. doi: 10.1039/c2mb25453k.
- [32] Robert A. Dromms and Mark P. Styczynski. Systematic applications of metabolomics in metabolic engineering. *Metabolites*, 2(4):1090–1122, 2012. ISSN 22181989. doi: 10.3390/metabo2041090.
- [33] Xuewen Chen, Ana P. Alonso, Doug K. Allen, Jennifer L. Reed, and Yair Shachar-Hill. Synergy between 13C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in *E. coli*. *Metabolic Engineering*, 13(1):38–48, 2011. ISSN 10967176. doi: 10.1016/j.ymben.2010.11.004. URL <http://dx.doi.org/10.1016/j.ymben.2010.11.004>.
- [34] Keren Yizhak, Tomer Benyaminini, Wolfram Liebermeister, Eytan Ruppin, and Tomer Shlomi. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, 26(12):255–260, 2010. ISSN 13674803. doi: 10.1093/bioinformatics/btq183.
- [35] Rogier J.P. Van Berlo, Dick De Ridder, Jean Marc Daran, Pascale A.S. Daran-Lapujade, Bas Teusink, and Marcel J.T. Reinders. Predicting metabolic fluxes using gene expression differences as constraints. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(1):206–216, 2011. ISSN 15455963. doi: 10.1109/TCBB.2009.55.
- [36] Min Kyung Kim and Desmond S. Lun. Methods for integration of transcriptomic data in genome-scale metabolic models. *Computational and Structural Biotechnology Journal*, 11(18):59–65, 2014. ISSN 20010370. doi: 10.1016/j.csbj.2014.08.009. URL <http://dx.doi.org/10.1016/j.csbj.2014.08.009>.
- [37] Vikash Pandey, Daniel Hernandez Gardiol, Anush Chiappino-pepe, and Vassily Hatzimanikatis. TEX-FBA : A constraint-based method for integrating gene expression , thermodynamics , and metabolomics data into genome-scale metabolic models. *bioRxiv*, pages 1–30, 2019.
- [38] Radhakrishnan Mahadevan, Jeremy S. Edwards, and Francis J. Doyle. Dynamic Flux Balance Analysis of diauxic growth in *Escherichia coli*. *Biophysical Journal*, 83(3):1331–1340, 2002. ISSN 00063495. doi: 10.1016/S0006-3495(02)73903-9. URL [http://dx.doi.org/10.1016/S0006-3495\(02\)73903-9](http://dx.doi.org/10.1016/S0006-3495(02)73903-9).
- [39] Davide Chicco. Ten quick tips for machine learning in computational biology. *BioData Mining*, 10(1):1–17, 2017. ISSN 17560381. doi: 10.1186/s13040-017-0155-3.
- [40] David T. Jones. Setting the standards for machine learning in biology. *Nature Reviews Molecular Cell Biology*, 20(11):659–660, 2019. ISSN 14710080. doi: 10.1038/s41580-019-0176-5. URL <http://dx.doi.org/10.1038/s41580-019-0176-5>.

- [41] Eva Maria Kapfer, Paul Stapor, and Jan Hasenauer. Challenges in the calibration of large-scale ordinary differential equation models. *IFAC-PapersOnLine*, 52(26):58–64, 2019. ISSN 24058963. doi: 10.1016/j.ifacol.2019.12.236. URL <https://doi.org/10.1016/j.ifacol.2019.12.236>.
- [42] Laura Lee, David Atkinson, Andrew G. Hirst, and Stephen J. Cornell. A new framework for growth curve fitting based on the von Bertalanffy Growth Function. *Scientific Reports*, 10(1):1–12, 2020. ISSN 20452322. doi: 10.1038/s41598-020-64839-y.
- [43] Jon E. Lindstrom, Ronald P. Barry, and Joan F. Braddock. Microbial community analysis: A kinetic approach to constructing potential C source utilization patterns. *Soil Biology and Biochemistry*, 30(2):231–239, 1997. ISSN 00380717. doi: 10.1016/S0038-0717(97)00113-2.
- [44] J L Garland. Potential and limitations of BIOLOG for microbial community analysis. *Microbial Biosystems: New Frontiers, Proceedings of the 8th International Symposium on Microbial Ecology*, pages 1–7, 1999. ISSN 1642395X.
- [45] Saratram Gopalakrishnan, Satyakam Dash, and Costas Maranas. K-FIT: An accelerated kinetic parameterization algorithm using steady-state fluxomic data. *Metabolic Engineering*, 61(January):197–205, 2020. ISSN 10967184. doi: 10.1016/j.mben.2020.03.001.
- [46] Weihua Guo and Xueyang Feng. OM-FBA: Integrate transcriptomics data with flux balance analysis to decipher the cell metabolism. *PLoS ONE*, 11(4):1–20, 2016. ISSN 19326203. doi: 10.1371/journal.pone.0154188.
- [47] Hans C. Bernstein. Reconciling Ecological and Engineering Design Principles for Building Microbiomes. *mSystems*, 4(3):1–5, 2019. ISSN 23795077. doi: 10.1128/msystems.00106-19.
- [48] Fernando Rojo. Carbon catabolite repression in *Pseudomonas*: Optimizing metabolic versatility and interactions with the environment. *FEMS Microbiology Reviews*, 34(5):658–684, 2010. ISSN 01686445. doi: 10.1111/j.1574-6976.2010.00218.x.
- [49] Raphy Zarecki, Matthew A. Oberhardt, Keren Yizhak, Allon Wagner, Ella Shtifman Segal, Shiri Freilich, Christopher S. Henry, Uri Gophna, and Eytan Ruppin. Maximal sum of metabolic exchange fluxes outperforms biomass yield as a predictor of growth rate of microorganisms. *PLoS ONE*, 9(5), 5 2014. ISSN 19326203. doi: 10.1371/journal.pone.0098372.
- [50] Jared L Gearhart, Kristin L Adair, Richard J Detry, Justin D Durfee, Katherine A Jones, and Nathaniel Martin. Comparison of Open-Source Linear Programming Solvers. Technical Report October, Sandia National Laboratories, 2013. URL <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>.
- [51] A. Cassioli, D. Di Lorenzo, M. Locatelli, F. Schoen, and M. Sciandrone. Machine learning for global optimization. *Computational Optimization and Applications*, 51(1):279–303, 2012. ISSN 09266003. doi: 10.1007/s10589-010-9330-x.
- [52] Rosemarie Wilton, Angela J. Ahrendt, Shalaka Shinde, Deirdre J. Sholto-Douglas, Jessica L. Johnson, Melissa B. Brennan, and Kenneth M. Kemner. A New Suite of Plasmid Vectors for Fluorescence-Based Imaging of Root Colonizing *Pseudomonads*. *Frontiers in Plant Science*, 8(February):1–15, 2018. ISSN 1664462X. doi: 10.3389/fpls.2017.02242.
- [53] Tomoya Baba, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A. Datsenko, Masaru Tomita, Barry L. Wanner, and Hirotada Mori. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Molecular Systems Biology*, 2, 2006. ISSN 17444292. doi: 10.1038/msb.4100050.
- [54] Christopher S. Henry, Matthew Dejongh, Aaron A. Best, Paul M. Frybarger, Ben Lindsay, and Rick L. Stevens. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, 28(9):977–982, 2010. ISSN 10870156. doi: 10.1038/nbt.1672.
- [55] Nathan E. Lewis, Kim K. Hixson, Tom M. Conrad, Joshua A. Lerman, Pep Charusanti, Ashoka D. Polpitiya, Joshua N. Adkins, Gunnar Schramm, Samuel O. Purvine, Daniel Lopez-Ferrer, Karl K. Weitz, Roland Eils, Rainer König, Richard D. Smith, and Bernhard Palsson. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, 6, 2010. ISSN 17444292. doi: 10.1038/msb.2010.47.
- [56] Mark Lotkin. A Note on the Midpoint Method of Integration. *Journal of the ACM (JACM)*, 3(3):208–211, 1956. ISSN 1557735X. doi: 10.1145/320831.320840.
- [57] L. F. Shampine. Stability of the leapfrog/midpoint method. *Applied Mathematics and Computation*, 208(1):293–298, 2009. ISSN 00963003. doi: 10.1016/j.amc.2008.11.029. URL <http://dx.doi.org/10.1016/j.amc.2008.11.029>.

- [58] J. C. Butcher. A history of Runge-Kutta methods. *Applied Numerical Mathematics*, 20:247–260, 1996.
- [59] W. H. Witty. A New Method of Numerical Integration of Differential Equations. *Mathematics of Computation*, 18(87):497, 1964. ISSN 00255718. doi: 10.2307/2003774.
- [60] Ramy K. Aziz, Daniela Bartels, Aaron Best, Matthew DeJongh, Terrence Disz, Robert A. Edwards, Kevin Formsma, Svetlana Gerdes, Elizabeth M. Glass, Michael Kubal, Folker Meyer, Gary J. Olsen, Robert Olson, Andrei L. Osterman, Ross A. Overbeek, Leslie K. McNeil, Daniel Paarmann, Tobias Paczian, Bruce Parrello, Gordon D. Pusch, Claudia Reich, Rick Stevens, Olga Vassieva, Veronika Vonstein, Andreas Wilke, and Olga Zagnitko. The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics*, 9:1–15, 2008. ISSN 14712164. doi: 10.1186/1471-2164-9-75.
- [61] Sven E.F. Borgos, Sergio Bordel, Håvard Sletta, Helga Ertesvåg, Øyvind Jakobsen, Per Bruheim, Trond E. Ellingsen, Jens Nielsen, and Svein Valla. Mapping global effects of the anti-sigma factor MucA in *Pseudomonas fluorescens* SBW25 through genome-scale metabolic modeling. *BMC Systems Biology*, 7:1–15, 2013. ISSN 17520509. doi: 10.1186/1752-0509-7-19.
- [62] Jonathan M Monk, Colton J Lloyd, Elizabeth Brunk, Nathan Mih, Anand Sastry, Zachary King, Rikiya Takeuchi, Wataru Nomura, Zhen Zhang, Hirotada Mori, Adam M Feist, and Bernhard O Palsson. iML1515 , a knowledgebase that computes *E. coli* traits. *Nature Biotechnology*, 35(10):8–12, 2017.
- [63] Kristian Jensen, Joao G.R. Cardoso, and Nikolaus Sonnenschein. Optlang: An algebraic modeling language for mathematical optimization. *The Journal of Open Source Software*, 2(9):139, 2017. doi: 10.21105/joss.00139.
- [64] Sara Castaño-Cerezo, José M. Pastor, Sergio Renilla, Vicente Bernal, José L. Iborra, and Manuel Cánovas. An insight into the role of phosphotransacetylase (pta) and the acetate/acetyl-CoA node in *Escherichia coli*. *Microbial Cell Factories*, 8:1–19, 2009. ISSN 14752859. doi: 10.1186/1475-2859-8-54.
- [65] Adam P. Arkin, Robert W. Cottingham, Christopher S. Henry, Nomi L. Harris, Rick L. Stevens, Sergei Maslov, Paramvir Dehal, Doreen Ware, Fernando Perez, Shane Canon, Michael W. Sneddon, Matthew L. Henderson, William J. Riehl, Dan Murphy-Olson, Stephen Y. Chan, Roy T. Kamimura, Sunita Kumari, Meghan M. Drake, Thomas S. Brettin, Elizabeth M. Glass, Dylan Chivian, Dan Gunter, David J. Weston, Benjamin H. Allen, Jason Baumohl, Aaron A. Best, Ben Bowen, Steven E. Brenner, Christopher C. Bun, John Marc Chandonia, Jer Ming Chia, Ric Colasanti, Neal Conrad, James J. Davis, Brian H. Davison, Matthew Dejongh, Scott Devoid, Emily Dietrich, Inna Dubchak, Janaka N. Edirisinghe, Gang Fang, José P. Faria, Paul M. Frybarger, Wolfgang Gerlach, Mark Gerstein, Annette Greiner, James Gurtowski, Holly L. Haun, Fei He, Rashmi Jain, Marcin P. Joachimiak, Kevin P. Keegan, Shinnosuke Kondo, Vivek Kumar, Miriam L. Land, Folker Meyer, Marissa Mills, Pavel S. Novickov, Taeyun Oh, Gary J. Olsen, Robert Olson, Bruce Parrello, Shiran Pasternak, Erik Pearson, Sarah S. Poon, Gavin A. Price, Srividya Ramakrishnan, Priya Ranjan, Pamela C. Ronald, Michael C. Schatz, Samuel M.D. Seaver, Maulik Shukla, Roman A. Sutormin, Mustafa H. Syed, James Thomason, Nathan L. Tintle, Daifeng Wang, Fangfang Xia, Hyunseung Yoo, Shinjae Yoo, and Dantong Yu. KBase: The United States department of energy systems biology knowledgebase. *Nature Biotechnology*, 36(7):566–569, 2018. ISSN 15461696. doi: 10.1038/nbt.4163.