

3622 (errors:5) words

File: main.tex
Encoding: ascii
Sum count: 3622
Words in text: 3109
Words in headers: 52
Words outside text (captions, etc.): 396
Number of headers: 18
Number of floats/tables/figures: 8
Number of math inlines: 53
Number of math displayed: 12
Subcounts:
text+headers+captions (#headers/#floats/#inlines/#displayed)
26+8+0 (1/0/0/0) _top_
321+1+0 (1/0/0/0) Section: Abstract
724+1+81 (1/1/1/0) Section: Introduction
37+1+12 (1/1/0/0) Section: Methods} \label{methods
343+6+0 (4/0/30/8) Subsection: Constraints
72+1+0 (1/0/5/1) Subsection: Objective
331+2+0 (1/0/0/0) Subsection: Metabolic phenotypes
653+18+37 (4/1/17/3) Subsection: E. coli-Pseudomonas competitive community
0+3+0 (1/0/0/0) Section: Results and Discussion
155+8+163 (1/1/0/0) Subsection: New data fitting method for iterative experimental design
228+2+103 (1/4/0/0) Subsection: Simulation insights
219+1+0 (1/0/0/0) Section: Conclusion

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: 3622
Words in text: 3109
Words in headers: 52
Words outside text (captions, etc.): 396
Number of headers: 18
Number of floats/tables/figures: 8
Number of math inlines: 53
Number of math displayed: 12
Files: 2
Subcounts:
text+headers+captions (#headers/#floats/#inlines/#displayed)
3109+52+396 (18/8/53/12) 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 phenotypic investigations

Andrew P. Freiburger¹, Fatima Foflonker¹, Jeffrey A. Dewey², Gyorgy Babnigg², Dionysios A. Antonopoulos², and Christopher Henry^{1*}

¹Argonne National Laboratory, Computing, Environment, and Life Sciences Division, Lemont, IL 60148 USA

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

November 19, 2022

1 Abstract

Microbial communities are increasingly recognized as decisive agents in animal health, agricultural productivity, industrial operations, and ecological systems. The plethora of chemical interactions and dynamic systems in these complex communities, however, can complicate or evade experimental studies, which leaves insufficient metabolic resolution to fully understand their mechanisms and begin to explore rational design for applications in the aforementioned fields. Numerous computational approaches have been applied to these probe these metabolic nuances, including metabolic flux balance analysis (FBA) and systems of differential equations, yet, these methods a) do not natively parameterize experimental data, b) are biased with determinism, or c) do not capture phenotypic adjustments of community members to changing conditions.

We therefore developed a dynamic model (CommPhitting) that can resolve phenotype adjustments and metabolic nuances of a microbial community by coupling a precise FBA sequence with an original global fitting method. CommPhitting parameterizes high-dimensional growth and -omics data as variables and coefficients of a linear problem, coupled with constraints that embody biological processes and metabolic profiles of each pertinent member phenotype. The linear problem is then optimized to determine the requisite species and phenotype biomasses, kinetic constants, and metabolite concentrations at each time point that are necessary to recapitulate the parameterized data. We exemplify CommPhitting with an idealized yet unstudied two-member community of model organisms (*Escherichia coli* and *Pseudomonas fluorescens*) that further exhibits cross-feeding dependency in maltose media. CommPhitting simulations of this community with batch, metabolomics, and BIOLOG data resolved kinetics parameters and phenotype proportions that predict new knowledge about the community which were difficult to experimentally ascertain, and invites application to further engineer this community: e.g. for bioproduction. We believe that CommPhitting – which is generalized to accommodate a diversity of data types and formats, and is further available and amply documented as Python API in the ModelSEEDpy library – will augment basic understanding and accelerate the engineering of microbial communities in a range of fields including medicine, agriculture, industry, and ecology.

2 Introduction

Microbial communities are ubiquitous on Earth [1], as eukaryotic symbionts [2, 3, 4] and as agents in recycling vital nutrients throughout the biosphere [5, 6, 7]. These communities are therefore essential to understand fields as diverse as medicine and climatology [8, 9, 10]. The assemblage of microbes into a community, while inviting intimate competition [11], offers a few advantages that make it biologically favorable in most conditions. Firstly, diversification in a) genetic potential, b) biochemical vulnerabilities [12], and c) metabolic machinery [13, 14, 15] allows community members to grow in otherwise inhospitable conditions by relying upon other members. Secondly, specialization among the members enables the community to achieve greater productivity than the sum of its isolated members. These positive member interactions, which justify community assemblage, are both predicated upon a metabolic economy of cross-feeding, where thriving members support struggling members and metabolic specialists exchange their nutrients to mutually satisfy nutritional requirements [16].

Metabolic cross-feeding (or syntrophy), despite being the crux of microbial communities, is difficult to experimentally resolve. One reason is that some metabolites do not detectably accumulate in the media, which obviates all but a few experimental methods, such as fluxomics isotope labeling [17]. A second difficulty, which proves to be more formidable, is that the exponentially increasing quantity of cross-feeding exchanges with increasing community size creates an untenable multitude of experiments to resolve all possible interactions. These challenges are further compounded by phenotype acclimation to dynamic community conditions, which adds a dimension of time-dependent phenotype distribution to an already experimentally challenging analysis.

Computational biology offers a few tools to that may supplement experimental methods in resolving community interactions. Flux Balance Analysis (FBA) [18], for example, is a prominent framework for simulating cellular metabolism and has been applied in numerous software tools that seek to resolve metabolic interactions within microbial communities: e.g. BacArena [19], μ bialSim [20], dOptCom [21], SMETANA [22], CASINO [2], and MICOM [23]). FBA offers explanatory power for predicting interactions from established knowledge, but it is difficult to directly parameterize these models with experimental data and discern new interactions or phenotype abundances that may be implicit in the data. Machine-learning algorithms [24, 25] offer more exploratory flexibility to predict interactions and phenotype abundances from experimental data, however, these methods often lack mechanistic explanations that cultivate understanding. Differential equation models have the converse problem, where they are insufficiently broad and unbiased to flexibly explore new knowledge but they offer precise mechanistic explanations for predictions.

A fitting model, by contrast, could be flexible enough to acquiesce new information from experimental data [26, 27, 28] while maintaining mechanistic resolution that would be necessary for understanding and rationally designing a community. These methods, despite an attractive blend of attributes, have not yet been applied to study community dynamics. We therefore developed a global fitting model (CommPhitting) that deduces time-resolved biomasses of each metabolic phenotype, concentrations of pertinent cross-feeding agents, and chemical parameters – including kinetic growth rate constants for each phenotype and conversion factors from each experimental signal to biomass – by parsing high-dimensional growth and -omics data of a community. CommPhitting captures numerous experimental dimensions – 1) community species, 2) species phenotypes, 3) time, 4) experimental signal, and 5) media concentrations – and simultaneously fits all data points, as a linear problem [29], to ensure that a global optimum is determined [30]. The flux profiles from each metabolic phenotype derive from FBA simulations of genome-scale metabolic models that are tailored with experimentally-specified sources and excreta. We exemplify this dynamic model with an idealized 2-member community of (*Escherichia coli* and *Pseudomonas fluorescens*), whose members have been extensively studied yet never as a coculture. This community demonstrated pivotal cross-feeding with acetate excreta from *E. coli* that was able to cultivate growth of *P. fluorescens* in a non-viable media of maltose (Figure 1). The predicted total biomasses of the community and members fit remarkably well to the fluorescence growth curves, which encourages the accuracy of the predicted phenotype biomasses, the kinetic growth rates of each phenotype, and the concentrations of each pertinent metabolite. We further simulated this community from BIOLOG data, which revealed chemical parameters of this community for each condition. We anticipate using this information to guide the engineering of *E. coli* for community bioproduction and for predicting community behaviors in a diverse range of conditions. CommPhitting is available as an open-source Python module in the ModelSEEDpy and is operable with the KBase [31] ecosystem. We believe that this light-weight yet robust fitting model will illuminate nuances of community interactions and cultivate rational design of microbial communities for myriad fundamental and industrial applications.

3 Methods

CommPhitting defines numerous parameters and variables in Table 1, linear constraints, and an objective function to capture a microbial community over time into a linear problem. These aspects of the linear problem are elaborated in the following sub-sections.

3.1 Constraints

The following linear constraints of CommPhitting categorically represent aspects of community biology.

3.1.1 Biomass abundances

The experimental biomass abundance ($EB_{s,t,j}$), over all experimental trials $j \in J$ and time points $t \in T$, is determined by converting the experimental signal ($E_{s,t,j}$) via a unique coefficient (EC_s) for each member $s \in S$

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

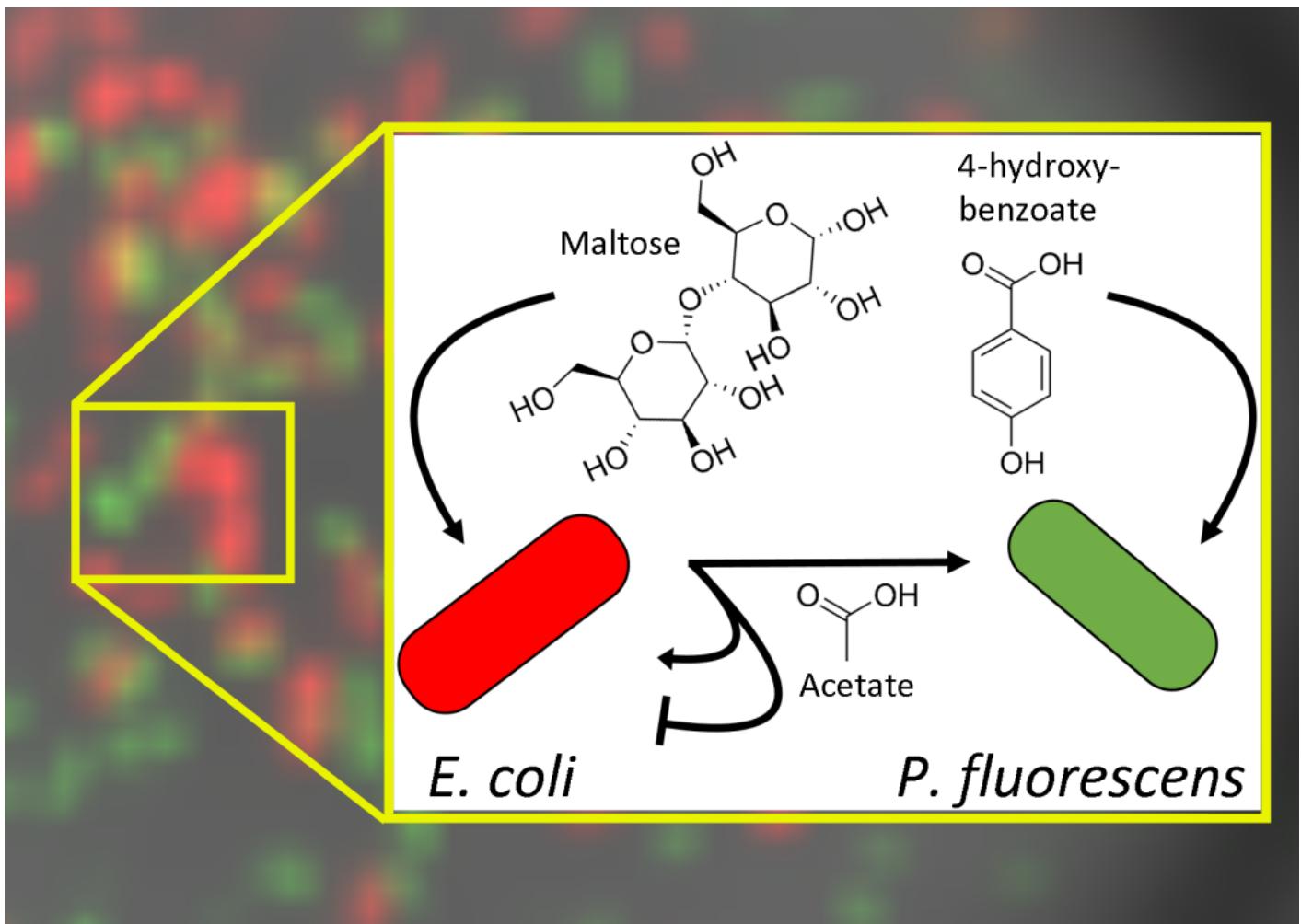


Figure 1: The primary metabolic exchanges that were experimentally elucidated and computationally modeled. The acetate byproduct of *E. coli* is the pivotal exchange of metabolic interest, where this source subsists *P. fluorescens*'s existence in maltose and exhibits interesting dual effects upon *E. 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*.

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

| Term | Description |
|-------------------------------|---|
| DIMENSIONS | |
| s | A species in the examined community. j . |
| k | A growth phenotype of species s . |
| t | An experimental time point. |
| j | An experimental trial. |
| i | Extracellular metabolite. |
| PARAMETERS | |
| $E_{s,t,j}$ | The experimental growth signal for a species at instant t in trial j . |
| $es_{s,k}$ | A boolean description of $k \in s$ |
| Δt | The seconds per timestep, which determines the amount of biomass growth per timestep. |
| $n_{k,i}$ | The exchange flux of each metabolite i in each strain k . |
| $v_{k,t,j}$ | The rate constant for growth of strain k at instant t in trial j . This parameter may be either a global value or a Michaelis-Menten flux such as $\frac{v_{max, k}}{k_m, k + c_{t,j,i}}$ that considers the concentration c of i . |
| $cvcf$ & $cvcf$ | Conversion coefficients of phenotype biomass to and from the stationary phase, respectively. |
| bcv_k | The greatest fraction of biomass ($0 < bcv < 1$) of strain k that can transition phenotypes in a timestep. |
| $cvmn$ | The minimal value of variable $cvt_{k,t,j}$. |
| VARIABLES | |
| EC_k | The conversion coefficient ($0 < EC < 1000$) from parameter $E_{s,t,j}$ into biomass, which is unique for each strain k . |
| $EB_{s,t,j}$ | The computed biomass from each experimental datum, as the product of EC_k & $E_{s,t,j}$. |
| $b_{k,t,j}$ | The predicted biomass from the fitting model. |
| $EV_{s,t,j}$ | The variance between the computed experimental biomass $EB_{s,t,j}$ and the predicted biomass $b_{k,t,j}$. |
| $c_{t,j,i}$ | The concentration of metabolite i at an experimental datum. |
| $g_{k,t,j}$ | The predicted growth rate for each strain at each datum. |
| $cvt_{k,t,j}$ & $cvf_{k,t,j}$ | The quantity of strain k biomass that transitions to and from the stationary phase, respectively, at an experimental datum. |

The variance ($EV_{s,t,j}$) between the converted experimental biomass and the predicted biomass abundance $b_{k,t,j}$ is determined for each phenotype $k \in K$

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

where only the phenotypes of each respective species are considered with the binary es_s variable.

3.1.2 Phenotype transitions

The biomass change over time is calculated for non-stationary (growing) phenotypes

$$b_{t,j,k} + \frac{\Delta t}{2} (g_{t,j,k} + g_{t+1,j,k}) + cvf_{t,j,k} - cvt_{t,j,k} = b_{t+1,j,k}. \quad (3)$$

The constraint consists of terms for the current biomass ($b_{k,t,j}$), the biomass growth rate $g_{k,t,j}$ as the biomass derivative ($\frac{\Delta \text{biomass}}{\Delta t}$), the biomass in the next timestep ($b_{k,t+1,j}$), and the net transition of biomass to non-stationary metabolic phenotypes $cvf_{k,t,j} - cvt_{k,t,j}$. The growth rate is constrained

$$b_{k,t,j} * v_{k,t,j} = g_{k,t,j}, \quad (4)$$

as the product of the current biomass abundance $b_{k,t,j}$ and the growth rate constant $v_{k,t,j}$, which reflects 1st-order kinetics and can be tailored for the examined system. The stationary phenotype (not growing) mediates phenotypic transitions, which subtly delays phenotype transitions as a reflection of cellular delays (e.g. diffusion). Future biomass for the stationary phenotype is analogous to eq. (3),

$$b_{k,t,j} - \sum_s^S (es_s * (cvf_{k,t,j} - cvt_{k,t,j})) = b_{k,t+1,j} \quad (5)$$

except that a) $g_{k,t,j} = 0$ by definition; b) the net transition to non-stationary phenotypes is negative; and c) all of the non-stationary strain phenotypes are summed for each species, since there are numerous non-stationary phenotypes yet only one stationary phenotype. The fraction of biomass that can transition is constrained

$$cvt_{k,t,j} \leq bcv * b_{k,t,j} + cvmin \quad (6)$$

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

This constraint is an application of Heun's integration method [32, 33], which – for an arbitrary function y , its derivative y' , and a timestep Δt – is defined as

$$y_{t+1} = y_t + \frac{1}{2} * \Delta t * (y'_t + y'_{t+1}). \quad (7)$$

This method is a 2nd-order Runge-Kutta formulation [34] that captures dynamic changes with high numerical accuracy [35] while maintaining a linear formulation that is amenable with our LP method.

3.1.3 Concentrations

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

$$c_{t,j,i} + \frac{\Delta t}{2} \sum_k^K (n_{i,k}(g_{t,j,k} + g_{t+1,j,k})) = c_{t+1,j,i} \quad (8)$$

through Heun's integration method from eq. (7). This implementation includes the current concentration ($c_{t,j,i}$) and concentration changes that are the product of Δt and the inner product of uptake fluxes ($n_{i,k}$) – which are determined by simulating metabolic models [36] of each member – and the biomass growth rate for a member.

3.2 Objective

The objective function of the model

$$\sum_{s,t,j}^{S,T,J} (EV_{s,t,j}^2) - \sum_{k,t,j}^{k,t,j} (cvct * cvt_{k,t,j}) - \sum_{k,t,j}^{T,J,K} (cvcf * cvf_{k,t,j}) \quad (9)$$

minimizes both the sum of variance (EV) from eq. (2) – between the predicted (b) and experimentally calculated (EB) biomass from eq. (10) – and the biomass transitions to $cvc\ell * cvt$ and from $cvcf * cvf$ the stationary metabolic phase. This objective therefore optimally fits the data with the least phenotypic transitions, which prevents overfitting of the models with an unrealistic excess of phenotypic transitions, which presumes that metabolic transitions have a non-trivial cost to undertake.

3.3 Metabolic phenotypes

The phenotypes for each member, except for the stationary phase where no fluxes occur by definition, were designed from genome-scale metabolic models (GEM's) of the members. CommPhitting-compatible GEM's are constructed from experimental genomes through the ModelSEED pipeline []. The phenotypes for each species then leverage the corresponding species GEM through the following precise sequence of optimizations, constraints, and FBA simulations.

1. A minimal biomass growth and the environmental media are defined and constrained in the GEM, which includes prohibiting hydrogen consumption and limits oxygen consumption to twice the total consumption of the defining phenotype sources.
2. The total influx of carbonaceous compounds excluding the phenotype sources and compounds for which a formula is not defined was minimized. This optimally isolates the solution fluxes to only those that correspond with metabolism from the phenotype sources. The resultant fluxes from this minimization, excluding the phenotype sources, are fixed to the model and are unchanged in subsequent optimizations.
3. The total flux of the phenotype sources is minimized, which optimizes biomass yield from the phenotype sources. This presumably finds the most biologically desirable metabolic profile for this phenotype to which evolution has presumably steered the species' metabolome. The phenotype source fluxes from this minimization are fixed to the model.
4. Optionally, the excretion flux for assigned excreta of a phenotype (where experimental evidence is available) is maximized. The excreta flux from this maximization is fixed to the model.
5. The final step applies parsimonious FBA from COBRApy [36] – a method that minimizes the total flux and thereby finds the most efficient means of achieving defined metabolic goals – to the model while maintaining all of the aforementioned constraints and constants. This parsimonious simulation presumably emulates the efficiencies that evolution found for the simulated organism. The fluxes from this optimization are the phenotype fluxes that are employed by CommPhitting.

The above five-step optimization sequence is repeated for each phenotype of all species in the community. This robust method creates an irreducible metabolic profile for each phenotype that fosters pure phenotype predictions by CommPhitting.

3.4 *E. coli*-*Pseudomonas* competitive community

3.4.1 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 encompass a wide range of disciplines, such as a) the human microbiome, b) bioproduction, c) synthetic biology, d) the rhizosphere, e) agriculture, and f) ecology. Secondly, despite these members being robustly studied individually, have not been studied as a coculture, which both offers new experimental information and mitigates biases and preconceptions in hypothesis development and results interpretation.

The coculture was created through the following protocol. The *P. fluorescens* and *E. coli* strains were purchased from ATC, were stored at -80degC before preparation for electrocompetancy, and were transformed with a plasmid to constitutively express either mNeongreen or mRuby2 fluorescent proteins (GFP and RFP), respectively [37]. Transformed cells were freshly streaked on a LB agar plate with appropriate antibiotics from -80degC glycerol stocks, and were incubated overnight at 30degC. A single colony from the plate was picked, placed into liquid LB (Lennox) broth with the antibiotics, and shaken @ 250 RPM overnight at 30degC. 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 μ 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

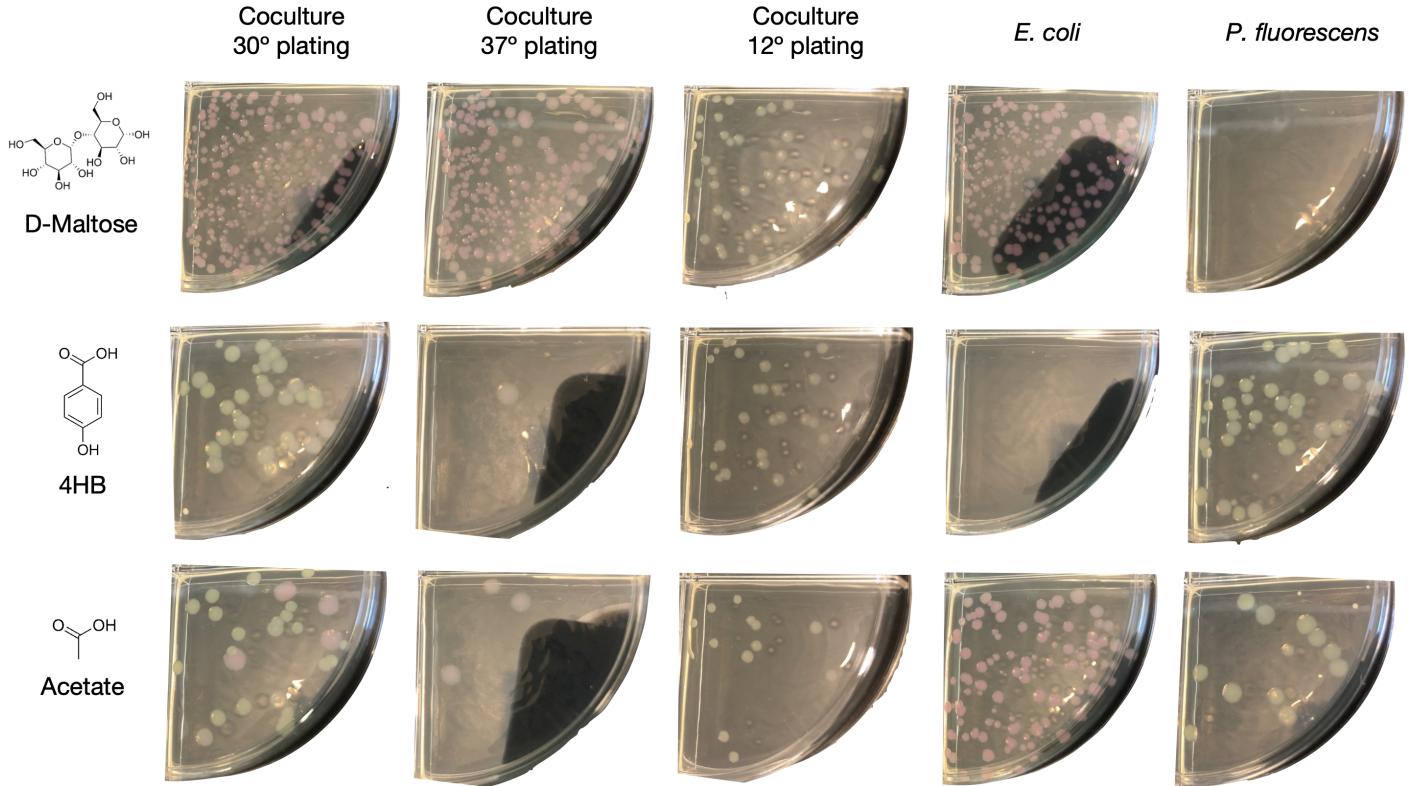


Figure 2: The agar growth from mono- and co-cultural experiments of *E. coli* and *P. fluorescens*. The similar growth profiles, and member exclusive conditions, between temperature and media compounds corroborates the application of fluorescence reporters to monitor member growth.

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 μ 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 a Hidex ... plate reader and was prewarmed to 30degC. The cells were shaken with (... settings) 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 deg C and 37 deg 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 if depicted through the agar plates of Figure 2, which supports their application in our study as a resolving mechanism for the community members.

3.4.2 Adapting the formulation for growth data

The general formulation in Section 3 was tailored to this particular 2-member community community in numerous aspects. First, the signal to biomass conversion constraint in eq. (10) was adapted to this community system

$$\begin{aligned}
 RFP_{t,j} * RFPC &= RFPB_{t,j} \\
 GFP_{t,j} * GFPC &= GFPB_{t,j} \\
 OD_{t,j} * ODC &= ODB_{t,j}
 \end{aligned} \tag{10}$$

where the RFP_C , GFP_C , and OD_C conversion factors were defined for each member and for each $RFP_{t,j}$, $GFP_{t,j}$, and $OD_{t,j}$ experimental signal. Second, the variance constraint in eq. (2) is adapted for each experimental signal

$$\begin{aligned} RFPB_{t,j} - \sum_k^K (pf_k * b_{t,j,k}) &= RFPV_{t,j} \\ GFPB_{t,j} - \sum_k^K (ec_k * b_{t,j,k}) &= GFPV_{t,j} \\ OD_{t,j} - \sum_k^K (b_{t,j,k}) &= ODV_{t,j} \end{aligned} \quad (11)$$

where pf_k and ec_k are binary variables that filter biomass abundances for each member. Third, the objective function of eq. (9) is defined in terms of the experimental signals

$$\sum_{t,j}^{T,J} (EV_{ecoli,t,j}^2) + \sum_{t,j}^{T,J} (EV_{pseudo,t,j}^2) + \sum_{t,j}^{T,J} (EV_{OD,t,j}^2) - \sum_{k,t,j}^{T,J,K} (cvct * cvt_{k,t,j}) - \sum_{k,t,j}^{T,J,K} (cvcf * cvf_{k,t,j}) . \quad (12)$$

Finally, the data was processed to conform with the model framework, primarily by vetting and omitting trials and times that were determined by ~~-----~~ that were inconsistent.

The *P. fluorescens* and *E. coli* genome-scale metabolic models were constructed in KBase Narrative 93465 through leveraging RAST to annotate the genomes [38] in the ModelSEED pipeline [39]. The *E. coli* model derived from the ASM584v2 experimental genome assembly and was gapfilled with acetate and maltose carbon sources that were studied. The *P. fluorescens* model derived from the ASM161270v1 experimental genome assembly and was gapfilled with acetate and 4-hydroxybenzoate carbon sources that were studied.

3.4.3 Adapting the formulation to BIOLOG data

4 Results and Discussion

4.1 New data fitting method for iterative experimental design

The CommPhitting model was developed, through the aforementioned methods, to resolve community interactions and phenotype distributions in communities while maintaining flexibility to diverse structures of community growth, and complementary -omics, data. The logical workflow of the model is illustrated in Figure 3, where 1) experimental data is parsed and parameterized into a dynamic optimization model of the community; 2) the model is simulated to the global optimum; and 3) the simulation predictions – namely phenotype abundances and chemical parameters – are interpreted, graphically or in spreadsheets, to either inform engineering designs or suggest additional measurements that are necessary to improve the fit and thus predictive precision. We leveraged open-source packages and initiatives in developing the model – e.g. the Optlang linear programming module [40] and the DOE Biological KnowledgeBase (KBase) ecosystem [31] – and have released our model in the ModelSEEDpy repository as a generalized tool that can be adapted communities and for which ample documentation is available in the ModelSEEDpy ReadTheDocs.

4.2 Simulation insights

The CommPhitting simulations of our 2-member community generated falsifiable predictions of metabolic concentrations and cross-feeding and irreplaceable time-resolution of phenotype biomasses within the community. The predicted abundances in Figure 6, for example, can be applied to The cross-feeding predictions additionally have actionable value for rational designing communities.

The results of CommPhitting can be validated through a few approaches. Firstly, the concentration profiles can be validated with metabolomics data. Secondly, the phenotype abundances can be validated with transcriptomics data. These validation approaches should mutually consistent in simulation that possess a high fit between the predicted and simulated biomasses. We demonstrate each of these approaches with our sample community ...[TODO]

We envision that this fitting model can be extended from the elucidation stage to the exploration stage that is encapsulated by Step 4 in Figure 3 as it guides engineering designs. This stage will apply the fitted parameters and variable values, from simulations of adequate mechanistic resolution, to either investigate community behaviors in various conditions (such as different media or the presence of toxins) or to guide the rational design of the community through

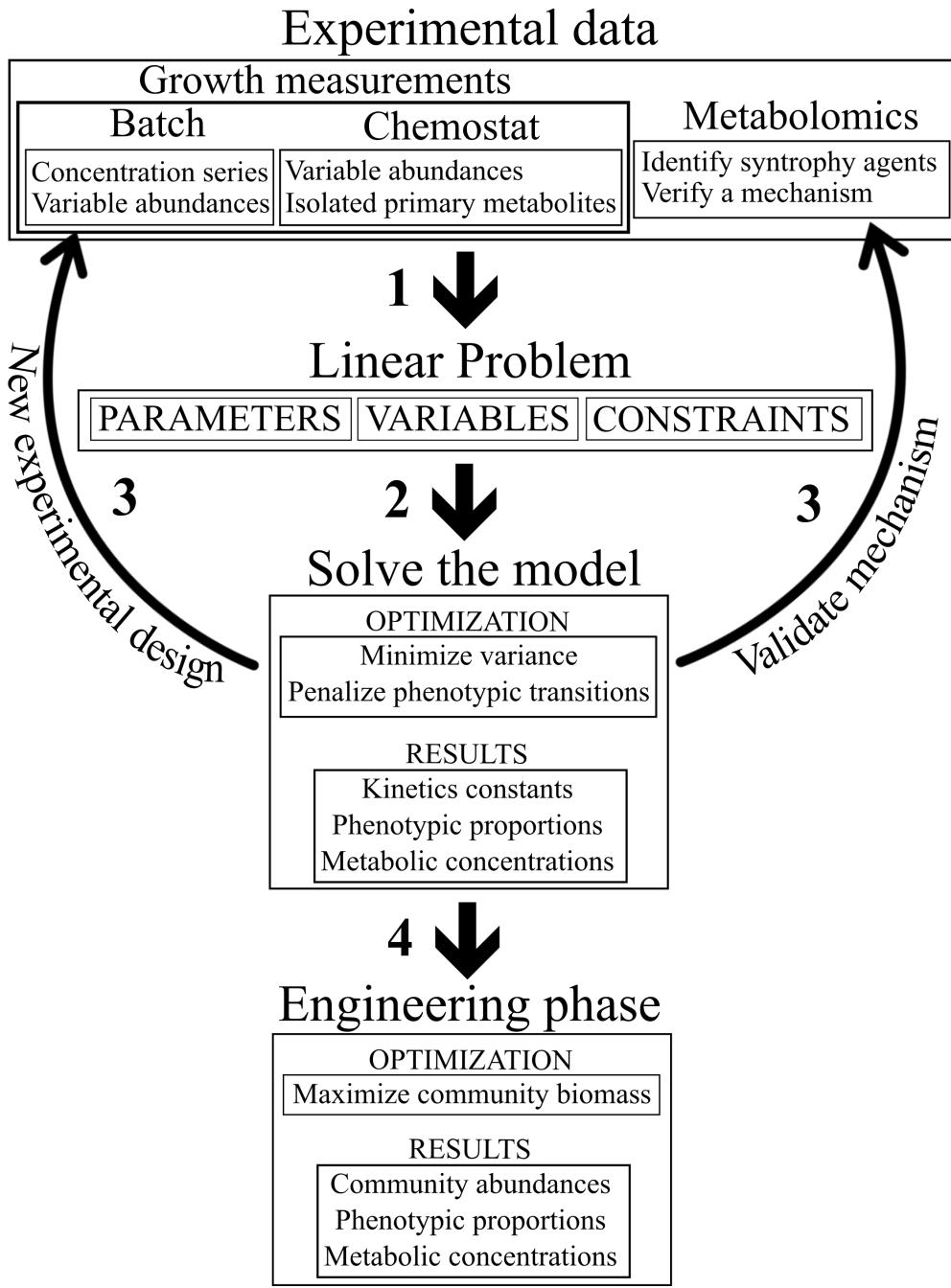


Figure 3: A workflow of the fitting model. **Step 1:** experimental data – from growth and possibly metabolomics measurements – is parsed into a linear problem that consists of the parameters and variables that are detailed in Table 1 and the constraints that are explained in Section 3. **Step 2:** the linear problem is executed with the objective function of eq. (9). **Step 3:** the simulation results are interpreted to either identify experimental changes that will improve modeling fit or to propose select metabolomic measurements that can crystallize 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 optimizes for community growth. The system of steps 1-4 create an intricate method of gleaning mechanistic insights of microbial communities and then immediately using these insights to rationally design a community with desirable activity.

Table 2: Table of final kinetic parameters predicted by model

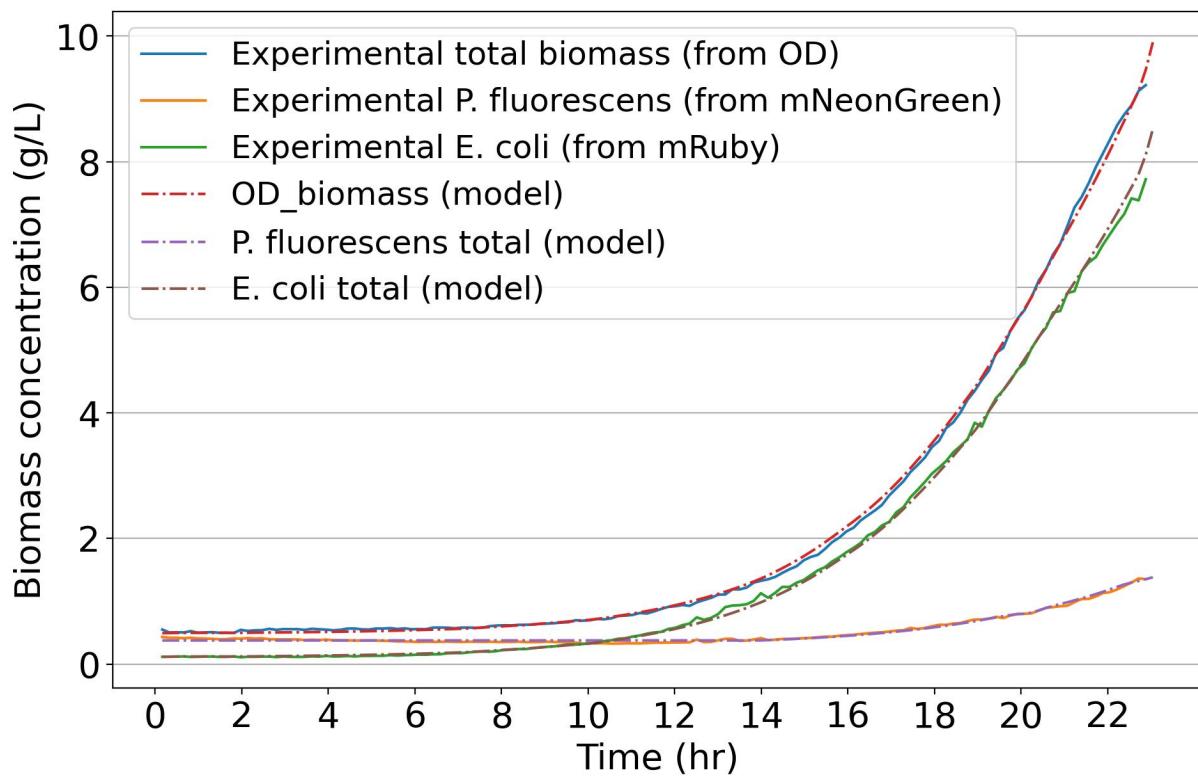


Figure 4: The converted experimental and predicted biomasses of the total community (OD), *E. coli* (RFP), and *P. fluorescens* (GFP) for the coculture experiment on maltose media (Table ??). Tight agreement between the experimental and predicted biomass values improves confidence in the predicted community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

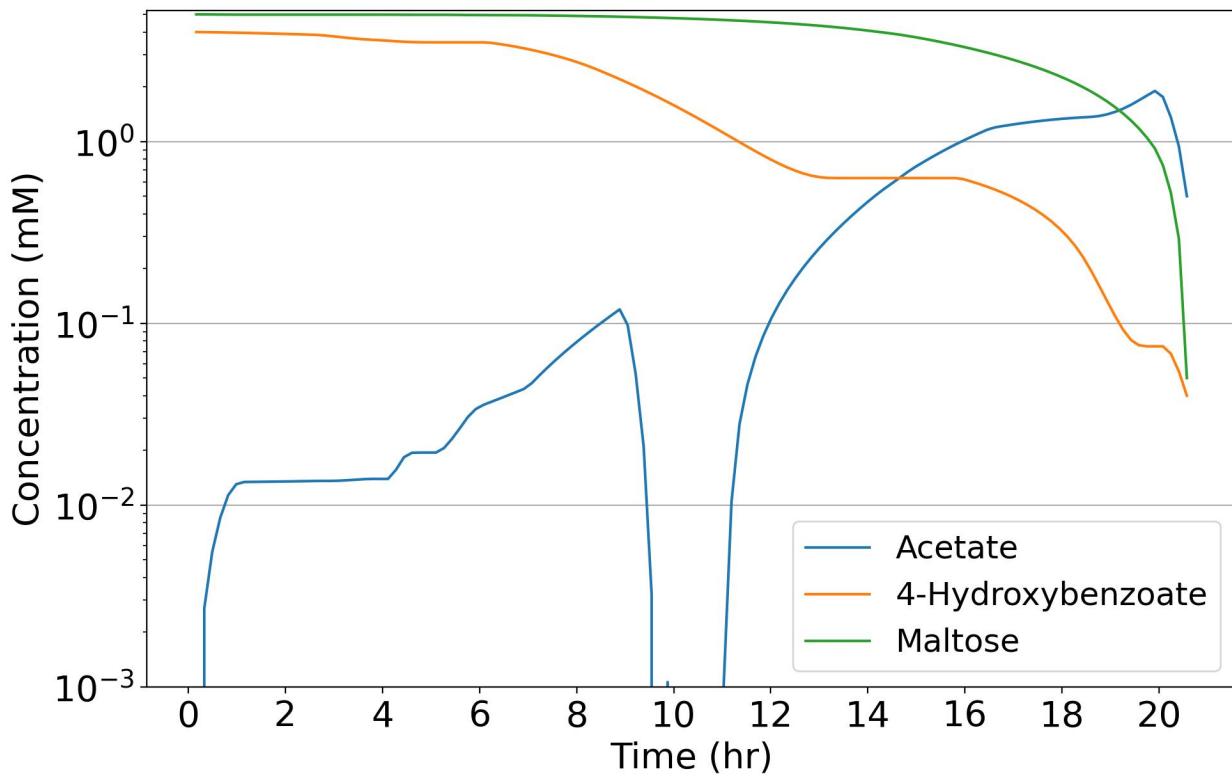


Figure 5: The concentrations of primary and secondary carbon sources, corresponding with the coculture experiments of a combined maltose and 4HB media. The figure illustrates dynamic acetate production from maltose consumption and the pronounced effect of phenotype transitions in concentration perturbations.

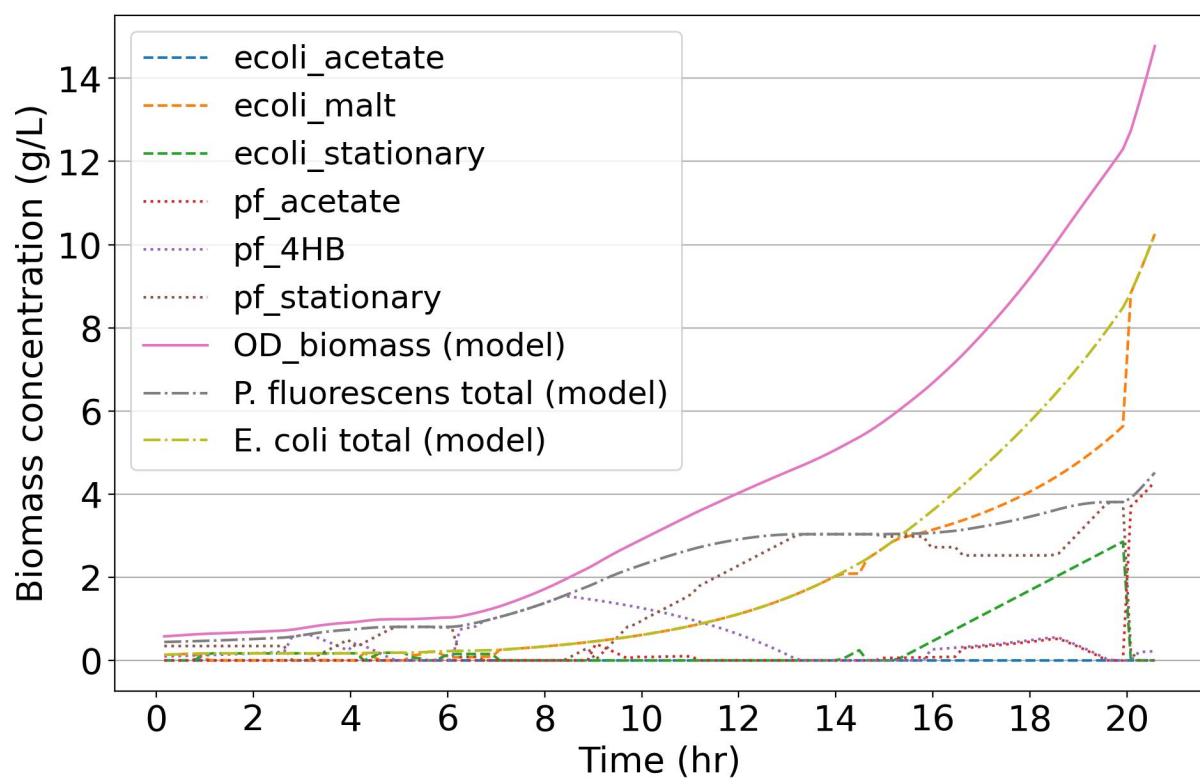


Figure 6: Phenotype abundances in the coculture experiment of maltose and 4HB.

engineering the member phenotypes. This second stage may ultimately loop into the first stage again when the engineered community is itself experimentally measured for growth and fit via our model to discern its interaction mechanisms and phenotype behaviors over time, thus creating a build/design/test/learn cycle.

5 Conclusion

A robust understanding of member interactions is necessary to precisely engineer microbial communities; yet, these interactions remain a formidable challenge for both experimental methods as the combinatoric possibilities becomes exponentially untenable as community size grows, which is compounded by dynamic metabolic phenotype adjustments to varying conditions, and existing computational methods suffer from either inadequate mechanistic resolution or accommodation for parameterized data. We proposed and derived herein a fitting method to extract knowledge and information from member genomes and integrated growth and multi-layered -omics data that best captures community dynamics while clearly elucidating community dynamics. The time-resolved predictions of metabolic concentrations in the media, kinetic growth constants for each phenotype, and phenotype distributions within the community allows researchers to judiciously direct limited resources towards metabolic hypotheses, both qualitative and quantitative, for which there is the most mechanistic support and likelihood of success, thereby improving experimental efficiency and the rate of advancement towards synthetic communities. Dissonance between predicted and experimental concentrations can further suggest cross-feeding interactions or other contributing member phenotypes that are not accounted by the parameterized model, which can be experimentally actionable hypotheses for future study. We suspect that the basic knowledge which is acquired from CommPhitting will accelerate RD in fields as diverse as bioproduction, medical therapies, and national security through the rational design of microbial communities.

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 Bertagnolli, 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 Wootlet, 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. Kyrpides, 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] Saeed Shoaei, Pouyan Ghaffari, Petia Kovatcheva-Datchary, Adil Mardinoglu, Partho Sen, Estelle Pujos-Guillot, Tomas De Wouters, Catherine Juste, Salwa Rizkalla, Julien Chilloux, Lesley Hoyles, Jeremy K. Nicholson, Joel Dore, Marc E. Dumas, Karine Clement, Fredrik Bäckhed, and Jens Nielsen. Quantifying Diet-Induced Metabolic Changes of the Human Gut Microbiome. *Cell Metabolism*, 22(2):320–331, 2015. ISSN 19327420. doi: 10.1016/j.cmet.2015.07.001.
- [3] 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>.
- [4] Saeed Shoaei, 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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>.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.

- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] Nikodimos A. Gebreselassie and Maciek R. Antoniewicz. ^{13}C -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>.
- [18] 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.
- [19] Eugen Bauer, Johannes Zimmermann, Federico Baldini, Ines Thiele, and Christoph Kaleta. BacArena: Individual-based metabolic modeling of heterogeneous microbes in complex communities. *PLoS Computational Biology*, 13(5): 1–22, 2017. ISSN 15537358. doi: 10.1371/journal.pcbi.1005544.
- [20] Denny Popp and Florian Centler. μ BialSim: Constraint-Based Dynamic Simulation of Complex Microbiomes. *Frontiers in Bioengineering and Biotechnology*, 8(June), 2020. ISSN 22964185. doi: 10.3389/fbioe.2020.00574.
- [21] Ali R. Zomorodi, Mohammad Mazharul Islam, and Costas D. Maranas. D-OptCom: Dynamic Multi-level and Multi-objective Metabolic Modeling of Microbial Communities. *ACS Synthetic Biology*, 3(4):247–257, 2014. ISSN 21615063. doi: 10.1021/sb4001307.
- [22] Aleksej Zelezniak, Sergej Andrejev, Olga Ponomarova, Daniel R. Mende, Peer Bork, and Kiran Raosaheb Patil. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 112(20):6449–6454, 2015. ISSN 10916490. doi: 10.1073/pnas.1421834112.
- [23] Christian Diener, Sean M. Gibbons, and Osbaldo Resendis-Antonio. MICOM: Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems*, 5(1), 2020. ISSN 2379-5077. doi: 10.1128/msystems.00606-19.
- [24] 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.
- [25] 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>.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] 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. Report SAND2013-8847. (October):62, 2013. URL <http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online>.

- [30] 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.
- [31] 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 Devold, 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. Novichkov, 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.
- [32] 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.
- [33] 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>.
- [34] J. C. Butcher. A history of Runge-Kutta methods. *Applied Numerical Mathematics*, 20:247–260, 1996.
- [35] 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.
- [36] Ali Ebrahim, Joshua A. Lerman, Bernhard O. Palsson, and Daniel R. Hyduke. COBRAPy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7, 2013. ISSN 17520509. doi: 10.1186/1752-0509-7-74.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] 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.