



DOE Systems Biology Knowledgebase

Resolving chemical interactions of microbial
communities via a comprehensive API ecosystem

Andrew Freiburger
Chris Henry



Acknowledgements



Argonne - Biosciences
Dion Antonopoulos
Jeff Dewey
Gyorgy Bibnigg
²



ModelSEED

Colorado State
Kelly Wrighton

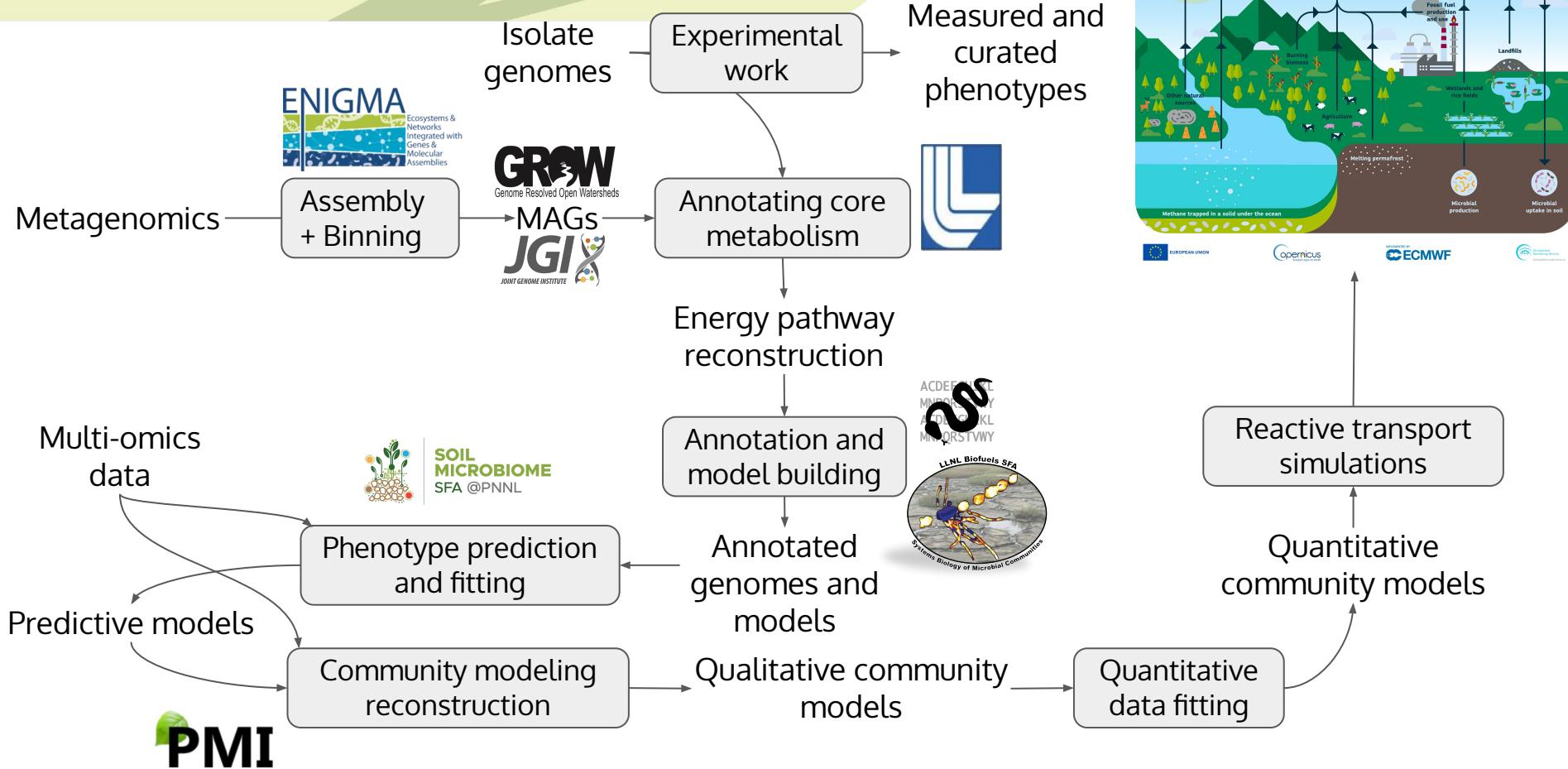
PMI

U Nebraska Collaborators
Hyun Seob Song

**U.S. DEPARTMENT OF
ENERGY**
Office of Science

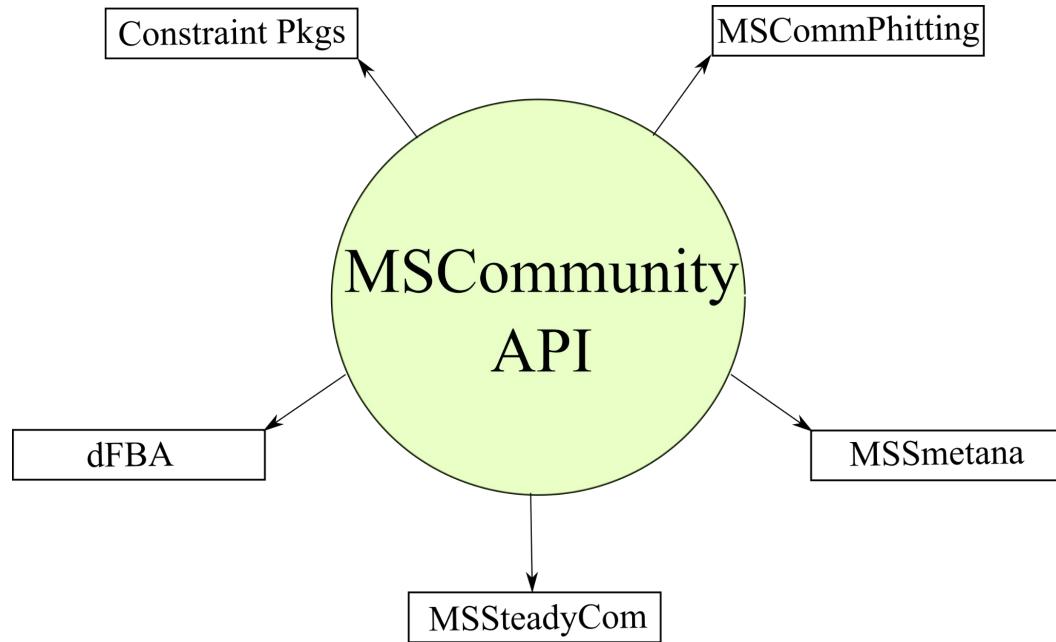
 **KBase**
DOE Systems Biology Knowledgebase

Enabling mechanistic understanding of environmental ecology



MSCommunity

- Isolate members
- A central object for various packages
 - MSCompatibility
 - MSSMETANA
 - MSMinimal media
 - MSSteadyCom
 - Community Kinetics
 - Steady-State flux constraints
 - SimpleThermo
 - Element Uptake
 - MSdFBA
- API backend for envisioned **KBase Community FBA Applications**



MSCompatibility

Converts all exchanges and/or cytoplasmic reactions to align with the ModelSEED Database ID's (MSID's)

- Compounds with non-MSID's are mapped by name to MSID's
- Compounds without matching names are left unchanged

```
from modelseedpy.community import MSCompatibility

new_models = MSCompatibility.standardize(models, conflicts_file_name="orig_conflicts.json")
```

Standardize exchange reactions in iML1515
=====

{'original': {'id': 'cpd03191_e0', 'name': 'D-Glucuronate 1-phosphate_e0'},
 'new': {'id': 'cpd00880_e0', 'name': 'D-Glucuronate 1-phosphate_e0'},
 'justification': 'The cpd03191_e0 and cpd00880_e0 distinction in iML1515 is '
 'incompatible; hence, the cpd00880_e0 ID and D-Glucuronate '
 '1-phosphate_e0 are used. The cpd03191_e0 and cpd00880_e0 '
 'metabolites were matched via their name. The ID match was '
 'verified with the ['BiGG'] cross-reference(s.)'}

{'original': {'reaction': 'cpd10516_e0 + cpd15411_e0 --> cpd03294_e0'},
 'new': {'reaction': 'cpd10516_e0 --> '},
 'justification': 'The new cpd03294_e0 ID for cpd15411_e0 already exists in '
 'model iML1515, so each reaction (here rxn08144_e0) must be '
 'replaced. The ID match was verified with the ['BiGG', '
 "BiGG'] cross-reference(s.)'}

{'original': {'reaction': 'cpd00067_p0 + cpd15411_p0 --> cpd00067_c0 + '
 'cpd15411_e0'},
 'new': {'reaction': 'cpd00067_p0 + cpd15411_p0 --> cpd00067_c0 + cpd03294_e0'},
 'justification': 'The new cpd03294_e0 ID for cpd15411_e0 already exists in '
 'model iML1515, so each reaction (here ARB1NTex_p0) must be '
 'replaced. The ID match was verified with the ['BiGG', '
 "BiGG'] cross-reference(s.)'}

{'original': {'reaction': 'cpd15411_e0 <=> '},
 'new': {'reaction': 'cpd03294_e0 <=> '},
 'justification': 'The new cpd03294_e0 ID for cpd15411_e0 already exists in '
 'model iML1515, so each reaction (here EX_cpd15411_e0) must '
 'be replaced. The ID match was verified with the ['BiGG', '
 "BiGG'] cross-reference(s.)'}

5 reactions were substituted and 14 metabolite IDs were redefined in iML1515,iSB1139 by standardize().

Standardize exchange reactions in iSB1139
=====

{'original': {'id': 'r1423_e0', 'name': 'alpha-ketoglutarate'},
 'new': {'id': 'cpd00024_e0', 'name': 'alpha-ketoglutarate_e0'},
 'justification': 'The r1423_e0 and cpd00024_e0 distinction in iSB1139 is '
 'incompatible; hence, the cpd00024_e0 ID and 'alpha-ketoglutarate_e0' are used. The r1423_e0 ID is not a 'ModelSEED Database ID. The r1423_e0 and cpd00024_e0 'metabolites were matched via their name.'}

{'original': {'id': 'r293_e0', 'name': 'Fe2+'},
 'new': {'id': 'cpd10515_e0', 'name': 'Fe2+_e0'},
 'justification': 'The r293_e0 and cpd10515_e0 distinction in iSB1139 is '
 'incompatible; hence, the cpd10515_e0 ID and Fe2+_e0 are 'used. The r293_e0 ID is not a ModelSEED Database ID. The 'r293_e0 and cpd10515_e0 metabolites were matched via their 'name.'}

{'original': {'id': 'r1262_e0', 'name': 'Mo2+'},
 'new': {'id': 'cpd00131_e0', 'name': 'Mo2+_e0'},
 'justification': 'The r1262_e0 and cpd00131_e0 distinction in iSB1139 is '
 'incompatible; hence, the cpd00131_e0 ID and Mo2+_e0 are 'used. The r1262_e0 ID is not a ModelSEED Database ID. The 'r1262_e0 and cpd00131_e0 metabolites were matched via their 'name.'}

{'original': {'id': 'r307_e0', 'name': 'Mn2+'},
 'new': {'id': 'cpd00030_e0', 'name': 'Mn2+_e0'},
 'justification': 'The r307_e0 and cpd00030_e0 distinction in iSB1139 is '
 'incompatible; hence, the cpd00030_e0 ID and Mn2+_e0 are 'used. The r307_e0 ID is not a ModelSEED Database ID. The 'r307_e0 and cpd00030_e0 metabolites were matched via their 'name.'

MSSteadyCom

- Steady member abundance during FBA
 - Member growth rate is a fraction of the community growth rate
- Members in a growing community must grow themselves to have flux

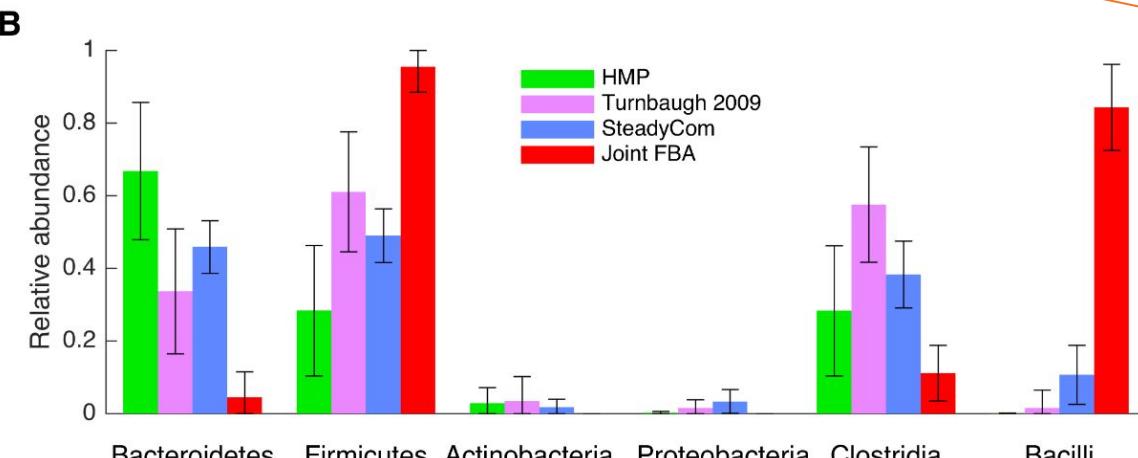


Fig 5. Distribution of the gut microbiota abundances simulated using 1000 sets of randomly assigned carbon uptake bounds for each species given the estimated average American diet. (A) Species abundances simulated using SteadyCom (blue) and joint FBA (red) respectively are displayed. (B) Comparison to the two sets of American gut microbiota data respectively from the Human Microbiome Project [70] (green) and Turnbaugh *et al.*, 2009 [65] (purple) at the phylum level. Two known important classes in Firmicutes, the Clostridia and Bacilli are also included.

$$\max \mu$$

subject to

$$\left[\begin{array}{l} \sum_{j \in \mathbf{J}^k} S_{ij}^k V_j^k = 0, \quad \forall i \in \mathbf{I}^k \\ LB_j^k X^k \leq V_j^k \leq UB_j^k X^k, \quad \forall j \in \mathbf{J}^k \\ V_{biomass}^k = X^k \mu \\ X^k \geq 0 \\ u_i^c - e_i^c + \sum_{k \in \mathbf{K}} V_{ex(i)}^k = 0, \quad \forall i \in \mathbf{I}^{com} \\ \sum_{k \in \mathbf{K}} X^k > 0 \\ \mu, \quad e_i^c \geq 0, \quad \forall i \in \mathbf{I}^{com} \end{array} \right] \quad \forall k \in \mathbf{K}$$

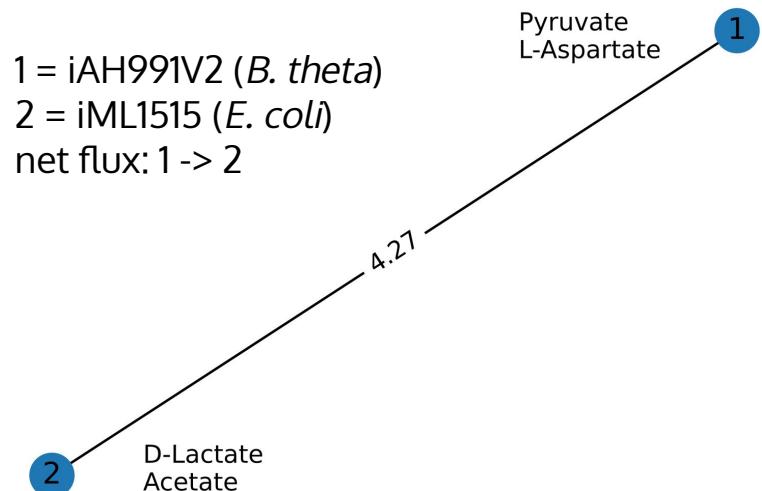
MSSteadyCom

https://link.springer.com/protocol/10.1007/978-1-0716-1585-0_13
(narrative 40576)

Environment	Species1	Species2
Metabolite/Donor ID		
12ethd_e0	-1.641505e-03	0.000000e+00
cpd00001_e0	-2.799308e+00	-3.746002e+00
cpd00009_e0	2.374552e-01	-8.380026e-02
cpd00011_e0	-3.820114e+00	6.134485e+00
cpd00013_e0	1.933339e+00	-3.109064e-01
cpd00020_e0	0.000000e+00	-5.399554e+00
cpd00027_e0	5.000000e+00	-1.629168e+00
cpd00028_e0	4.055331e-04	0.000000e+00
cpd00029_e0	-4.822227e+00	5.507701e+00
cpd00030_e0	4.655638e-04	-6.003071e-05
cpd00033_e0	0.000000e+00	6.599208e-03
cpd00034_e0	4.351575e-04	-2.962442e-05
cpd00036_e0	-1.896507e+00	5.775493e-02
cpd00041_e0	0.000000e+00	-4.927872e-01
cpd00047_e0	-3.649798e-03	0.000000e+00
cpd00048_e0	3.768643e-04	-3.768643e-04
cpd00051_e0	0.000000e+00	2.504436e-02
cpd00053_e0	0.000000e+00	-9.965315e-02
7	9.965315e-02	9.965315e-02

Tabular and graphical resolution of cross-feeding

- Fluxes
- Names

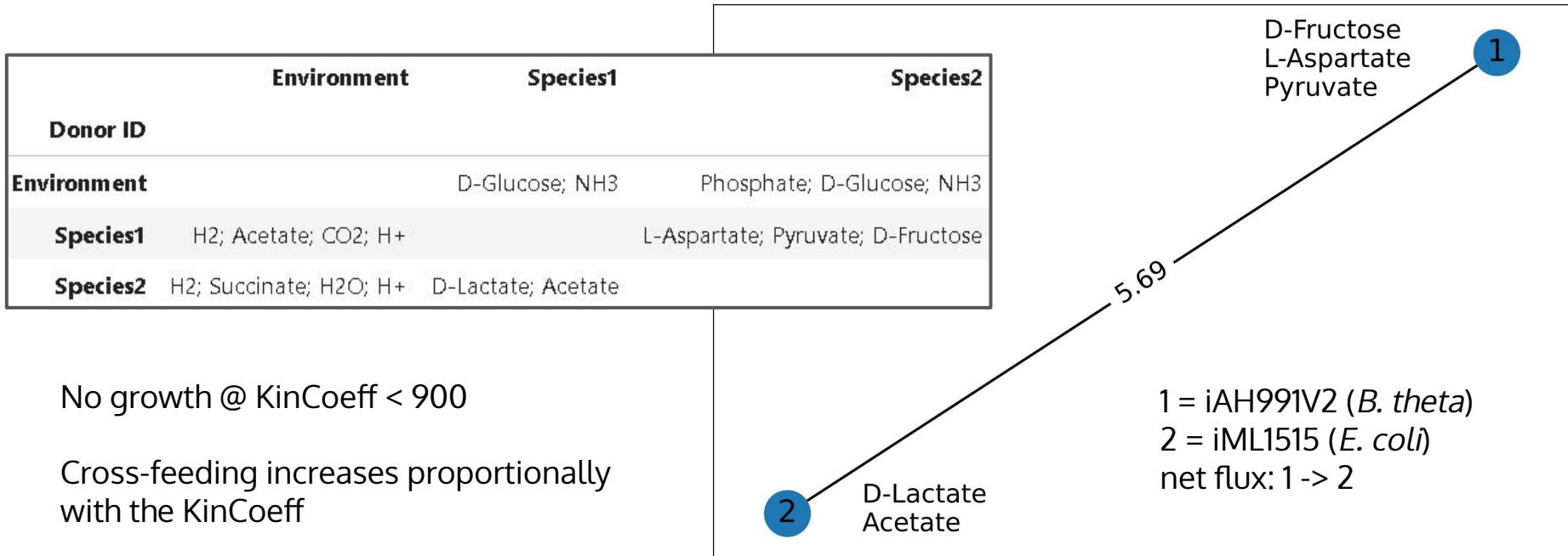


Environment	Species1	Species2
Donor ID		
Environment	D-Glucose; NH3	Phosphate; D-Glucose; NH3
Species1	H2; Acetate; CO2; H+	L-Aspartate; Pyruvate
Species2	H2; Succinate; H2O; H+	D-Lactate; Acetate



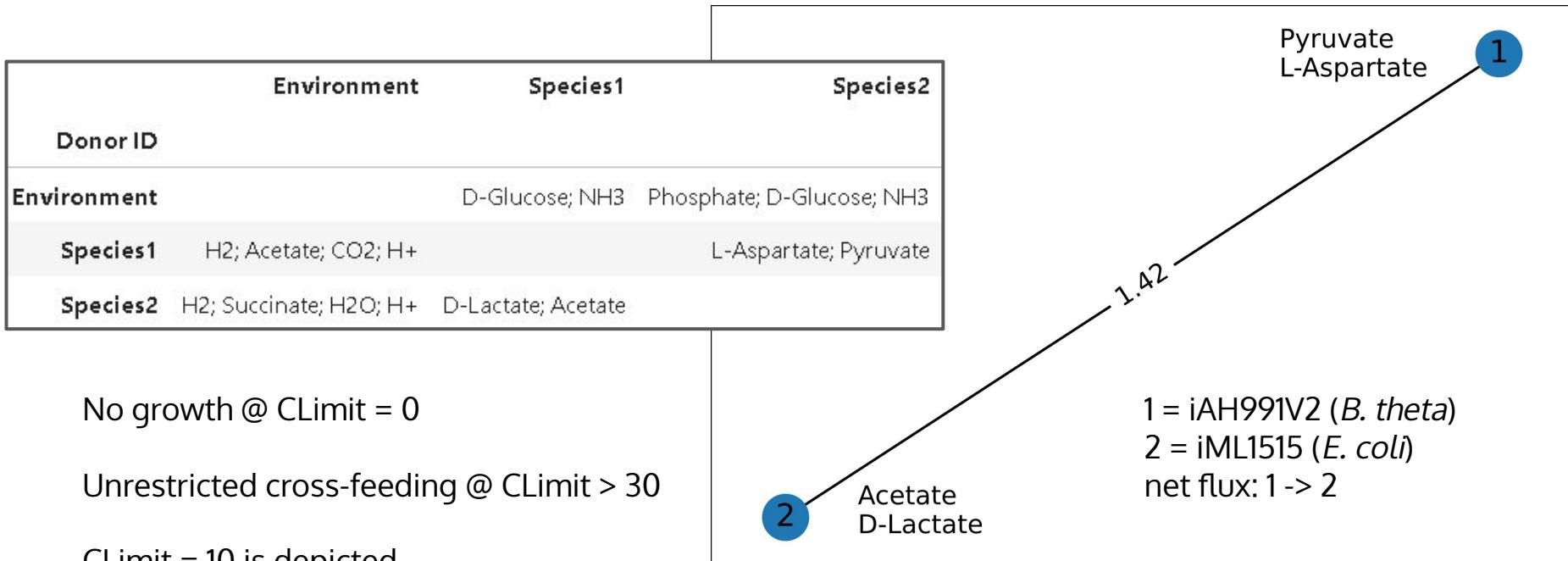
MSSteadyCom+CommKinetics

$$\sum_{rxn}^R (rxn_{forward} + rxn_{backward}) = KinCofef * bioRxn_{forward}$$



MSSteadyCom+ElementUptake

$$\sum_{exRxn}^{EX} (totalEle_{exRxn} * (exRxnExpr_{forward} \oplus exRxnExpr_{backward})) = eleLimit$$

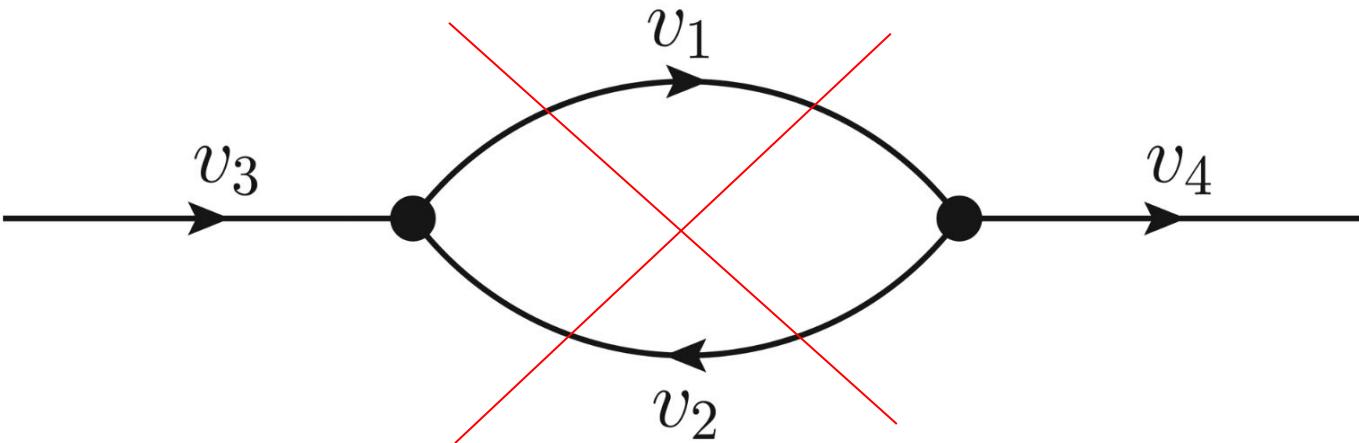


SimpleThermo

$$-1000 * revbin_i + forv_i \leq 0$$

$$1000 * revbin_i + revv_i \leq 1000$$

$$0 \leq max_abs_energy * revbin_i - |min_energy| * dgbinR_i + max_energy * dgbinF_i + \sum_{i,j}^{I,J} (n_{i,j} * \Delta G_j) \leq max_abs_energy$$

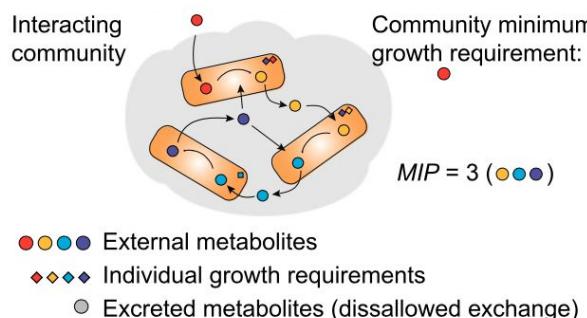
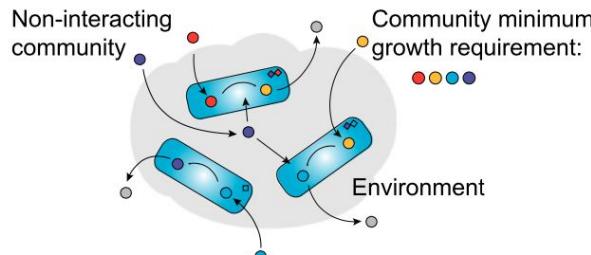


- Thermodynamically infeasible pathways are eliminated by estimating the Gibbs free energy for all reactions, and only permitting flux directions that correspond to negative free energy

MSSMETANA

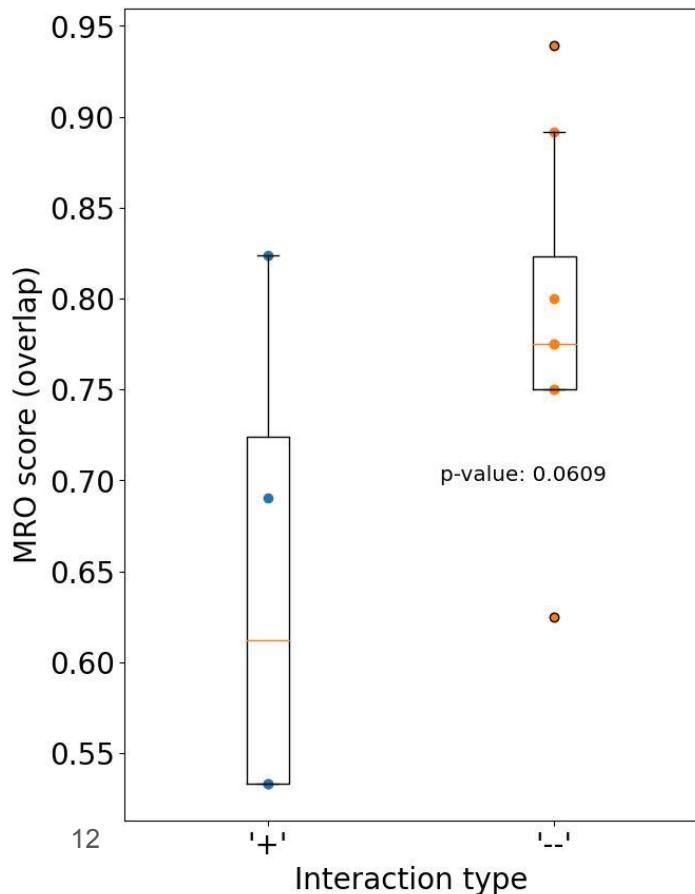
$$MRO = i, j \in N | i \neq j \left(\frac{|M_i \cap M_j|}{|M_i|} \right)$$

$$MIP = M_{\text{non-interacting}} - M_{\text{interacting}}$$

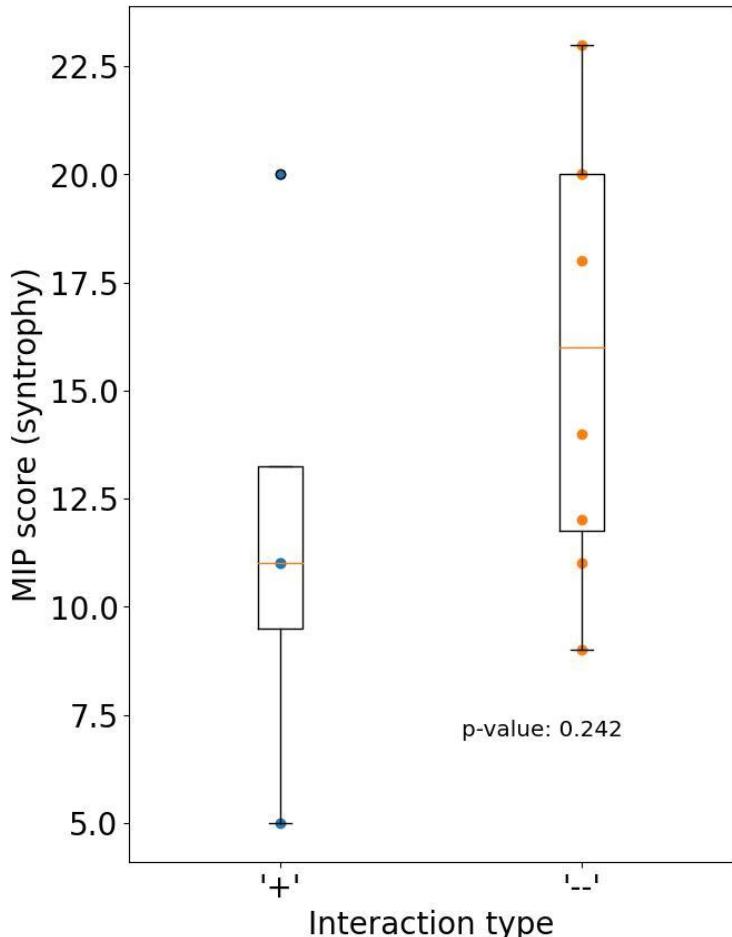


- **Species METabolic interaction ANAlysis**
- A set of **5 scores (& 1 superscore)** that quantitatively describe exchanges of community members
 - **MRO (metabolic resource overlap)** = % overlap of metabolic requirements of members
 - **MIP (metabolic interaction potential)** = the # of compounds that may be sourced through syntrophy
 - **MU (metabolite uptake)** = the fraction of FBA solutions that include each influx
 - **MP (metabolite production)** = the set of metabolites the each member can contribute via syntrophy
 - **SC (species coupling)** = the fraction of FBA community solutions that include a member interaction
 - **sметана** = an abstracted superscore of the SC, MU, and MP scores
- ModelSEEDpy + COBRAKBase model objects
 - [example Notebook](#)

All v. all for 10 community pairs



- 1) MIP cannot capture direction
- 2) Competition may be more consequential
<https://www.science.org/doi/10.1126/science.abn5093>



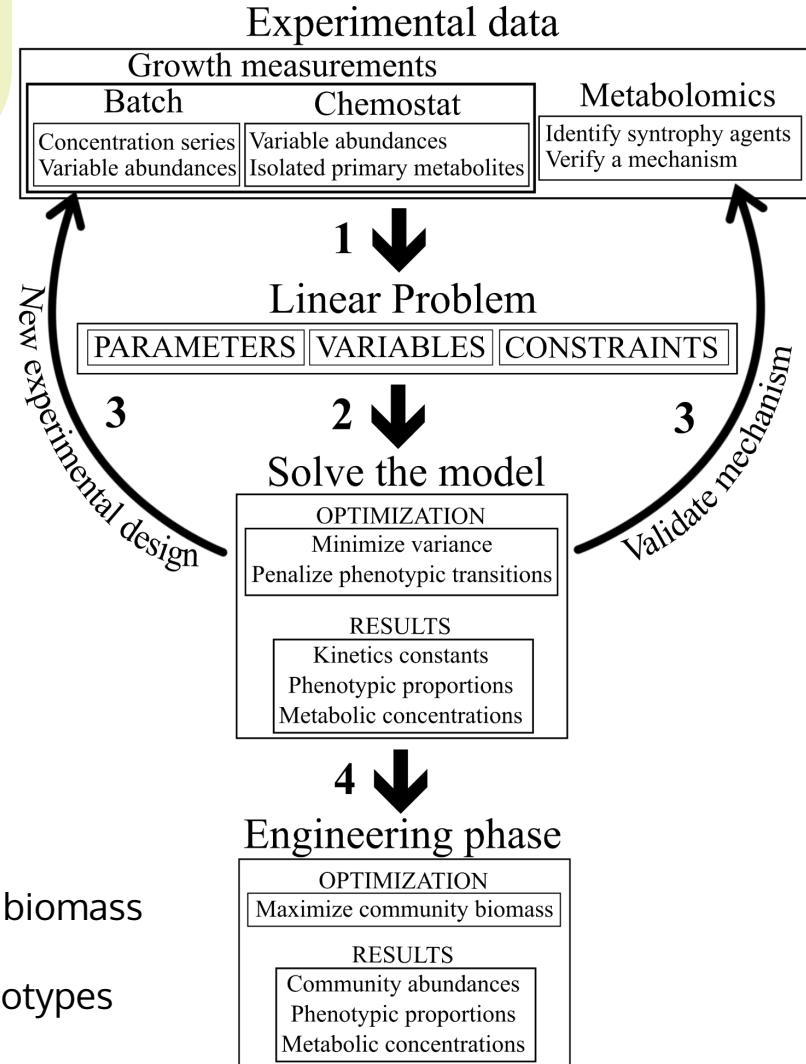
U.S. DEPARTMENT OF
ENERGY



Office of Science

MSCommPhitting

- COMP Poster session (Hall F-H, #826, 7pm)
[Freiburger et al., *bioRxiv*, 2022,
<https://doi.org/10.1101/2022.12.15.520667>]
- Defines community metabolic phenotypes
- Predicts phenotype abundances by fitting data
 - Global fitting algorithm ensures best results
 - Accepts myriad datatypes
 - Growth (fluorescent protein and OD)
 - BIOLOG
 - Metabolomics
- Time-resolved predictions
 - Biomass abundances for all phenotypes
 - Conversion constants from experimental signals to biomass
 - Media concentrations
 - Kinetic growth rate coefficients for the growth phenotypes

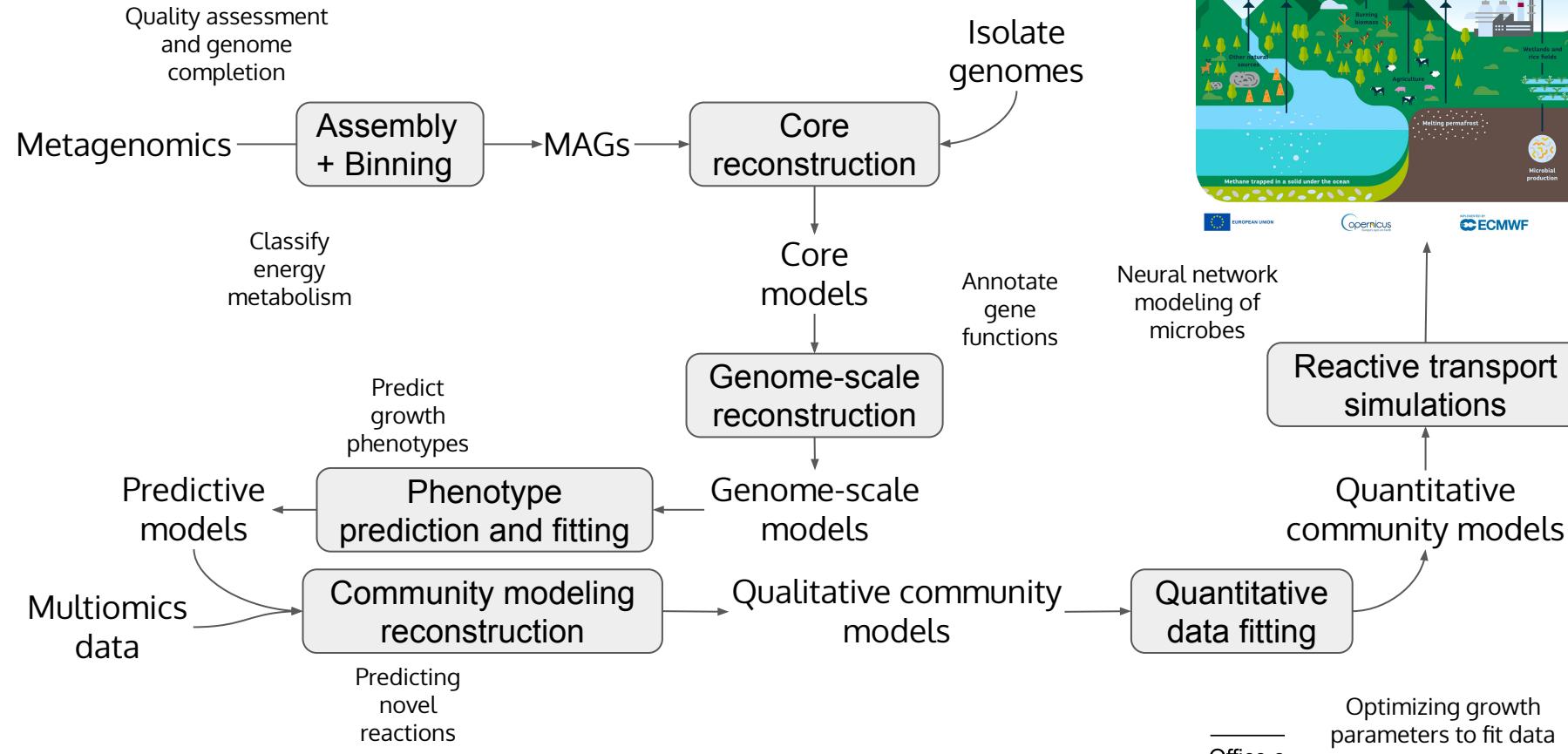


Future work

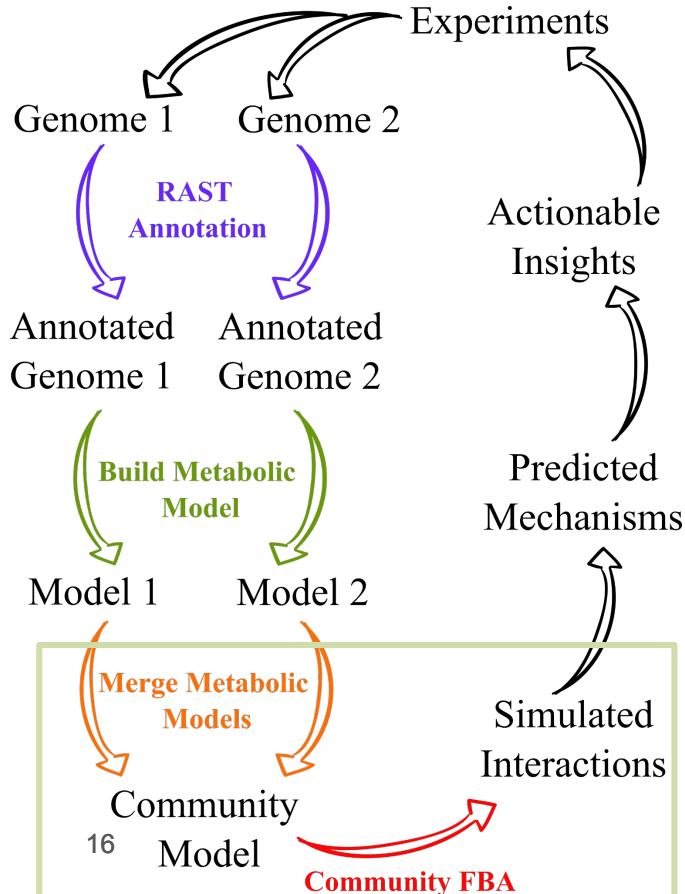
- Additional development
 - MSSmetana
 - Constrain the MRO & MIP to specified media conditions
 - Develop and implement original scores
 - antiSMASH (Priya Ranjan at Oak Ridge Lab)
 - Functional complementarity, Costless metabolites, etc
 - MSSteadyCom
 - Abundance variability analysis
 - MSCommPhitting
 - Evaluate large, ecological, communities
 - GROW and ENGIMA systems with collaborators
 - Curate other packages
 - Biofilm growth, CASINO, Resource Balance Analysis, dFBA
- Further validation with experimental data from our collaborators
- Wrap these API packages into a point-click KBase Apps



Enabling mechanistic understanding of environmental ecology



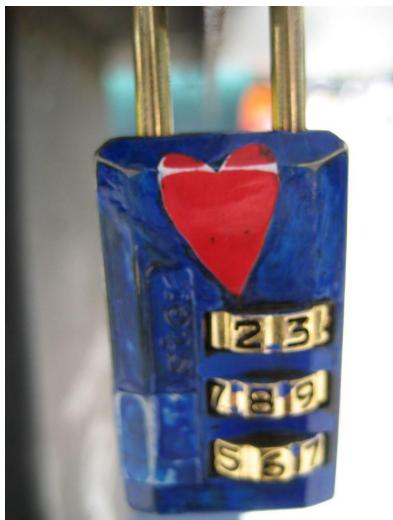
Metabolic modeling in KBase



- The **Community FBA** tools consist of numerous packages that probe various aspects of community dynamics
 - **MSSMETANA**
 - `build_from_species_models`
 - **MSSteadyCom**
 - **MSCompatibility**
- Complementary fitting App that discerns community growth parameters from myriad data: growth, BIOLOG, metabolomics
 - **MSCommPhitting**
- Colored arrows denote programmatic steps via KBase Apps
- Black arrows denote manual tasks and data processing

Chemical reaction dynamics “flux”

Flux balance analysis (CobraPy) – linear programming toward a directive



$$\begin{array}{c} \text{1} & \text{2} & \text{3} \\ \text{A} & \boxed{-a} & \boxed{-a} & \boxed{0} \\ \text{B} & \boxed{-b} & \boxed{0} & \boxed{0} \\ \text{C} & \boxed{c} & \boxed{0} & \boxed{-c} \\ \text{D} & \boxed{d} & \boxed{-d} & \boxed{0} \\ \text{Y} & \boxed{0} & \boxed{y} & \boxed{0} \\ \text{Z} & \boxed{0} & \boxed{z} & \boxed{-z} \\ \text{growth} & \boxed{0} & \boxed{0} & \boxed{\text{growth}} \end{array} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0 \quad [\text{steady state}]$$

- 1 $aA + bB \rightarrow cC + dD$
- 2 $dD + aA \leftrightarrow yY + zZ$
- 3 $cC + zZ \rightarrow \text{growth}$

$$\frac{d[A]}{dt} = 0 = v_1 * -a + v_2 * -a_2 + v_3 * 0$$

maximize growth = maximize v_3

'EX_cpd00007_e0',
'EX_cpd0018_e0',
'EX_cpd0023_e0',
'EX_cpd0027_e0',
'EX_cpd0028_e0',
'EX_cpd0030_e0',
'EX_cpd0033_e0',
'EX_cpd0034_e0',
'EX_cpd0035_e0',
'EX_cpd0039_e0',
'EX_cpd0041_e0',
'EX_cpd0046_e0',
'EX_cpd0048_e0',
'EX_cpd0051_e0',
'EX_cpd0054_e0',
'EX_cpd0058_e0',
'EX_cpd0060_e0',
'EX_cpd0063_e0',
'EX_cpd0065_e0',
'EX_cpd0066_e0',
'EX_cpd0069_e0',
'EX_cpd0091_e0',
'EX_cpd0099_e0',
'EX_cpd00107_e0',
'EX_cpd00119_e0',
'EX_cpd00126_e0',
'EX_cpd00129_e0',
'EX_cpd00149_e0',
'EX_cpd00156_e0',
'EX_cpd00161_e0',
'EX_cpd00184_e0',
'EX_cpd00205_e0',
'EX_cpd00215_e0',
'EX_cpd00218_e0',
'EX_cpd00220_e0',
'EX_cpd00239_e0',
'EX_cpd00254_e0',
'EX_cpd00322_e0',
'EX_cpd00381_e0',
'EX_cpd00393_e0',
'EX_cpd00644_e0',
'EX_cpd00654_e0',
'EX_cpd00793_e0',
'EX_cpd00654_e0',
'EX_cpd00793_e0',
'EX_cpd10516_e0',
'EX_cpd10516_e0'

CF402 & BC15 pair

CF402 (44
nutrients)

Overlapping
media (18
nutrients)

'EX_cpd00007_e0',
'EX_cpd00028_e0',
'EX_cpd00030_e0',
'EX_cpd00034_e0',
'EX_cpd00048_e0',
'EX_cpd00058_e0',
'EX_cpd00063_e0',
'EX_cpd00099_e0',
'EX_cpd00149_e0',
'EX_cpd00156_e0',
'EX_cpd00205_e0',
'EX_cpd00220_e0',
'EX_cpd00254_e0',
'EX_cpd00322_e0',
'EX_cpd00393_e0',
'EX_cpd00644_e0',
'EX_cpd00654_e0',
'EX_cpd00793_e0',
'EX_cpd10516_e0',
'EX_cpd10516_e0'

BC15 (20
nutrients)

'EX_cpd00007_e0',
'EX_cpd00028_e0',
'EX_cpd00030_e0',
'EX_cpd00034_e0',
'EX_cpd00048_e0',
'EX_cpd00058_e0',
'EX_cpd00063_e0',
'EX_cpd00099_e0',
'EX_cpd00149_e0',
'EX_cpd00156_e0',
'EX_cpd00205_e0',
'EX_cpd00220_e0',
'EX_cpd00254_e0',
'EX_cpd00322_e0',
'EX_cpd00355_e0',
'EX_cpd00393_e0',
'EX_cpd00654_e0',
'EX_cpd00793_e0',
'EX_cpd10516_e0',
'EX_cpd15603_e0'

$$MRO = i, j \in N | i \neq j \left(\frac{|M_i \cap M_j|}{|M_i|} \right)$$

CF402 from BC15's perspective:
MRO = 18/20 = 90% overlap
A big nutritional competitor

BC15 from CF402's perspective:
MRO = 18/44 = 41% overlap
Not a big nutritional competitor

10-member community

- Experimentally examined
 - Index is the spot and the header is the lawn for the respective community
-

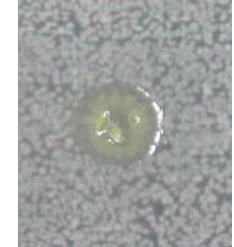
(-)



(+)



()



		YR343	CF313	CF142	AP49	GM17	BC15	CF402	AP07	YR139	*	BT03
Genus	Strain											
<i>Pantoea</i>	YR343										ND	
<i>Pseudomonas</i>	GM17	-	-	-	-	-	-	-	-	ND	-	
<i>Sphingobium</i>	AP49			-	-						ND	
<i>Rhizobium</i>	CF142										ND	
<i>Variovorax</i>	CF313			-							ND	-
<i>Bacillus</i>	BC15										-	ND
<i>Caulobacter</i>	AP07	-									ND	
<i>Duganella</i>	CF402				+	+					ND	
<i>Streptomyces</i>	YR139	-	-						-		ND	
<i>Paraburkholderia</i>	BT03					-					ND	

Assessment of community pairs

<https://pubmed.ncbi.nlm.nih.gov/33995895/>

Table 2

Pairwise interaction screen results. Strain designations across top of table indicate lawn of microbes spread on R2A agar plate and designations on left indicate cells spotted on center of lawn. + indicates a positive interaction while - indicates an antagonistic interaction. Empty cells indicate no obvious colony phenotype change.

Genus	Strain	YR343	GM17	AP49	CF142	CF313	BC15	AP07	CF402	YR139*	BT03
<i>Pantoea</i>	YR343		77%:0	100%:0		+ -	86%:0 -	82%:0 -	82%:5 -	ND	82%:0 -
<i>Pseudomonas</i>	GM17		-		-	- -	- -	- -	- -	ND	-
<i>Sphingobium</i>	AP49				86%:0	+				ND	
<i>Rhizobium</i>	CF142					76%:0				ND	95%:0
<i>Variovorax</i>	CF313				-				90%:4 -	ND	-
<i>Bacillus</i>	BC15		90%:0							ND	
<i>Caulobacter</i>	AP07		-		39%:3 -	41%:4 +				ND	
<i>Duganella</i>	CF402					+	+			ND	
<i>Streptomyces</i>	YR139		-	-				-		ND	
<i>Paraburkholderia</i>	BT03			-						ND	

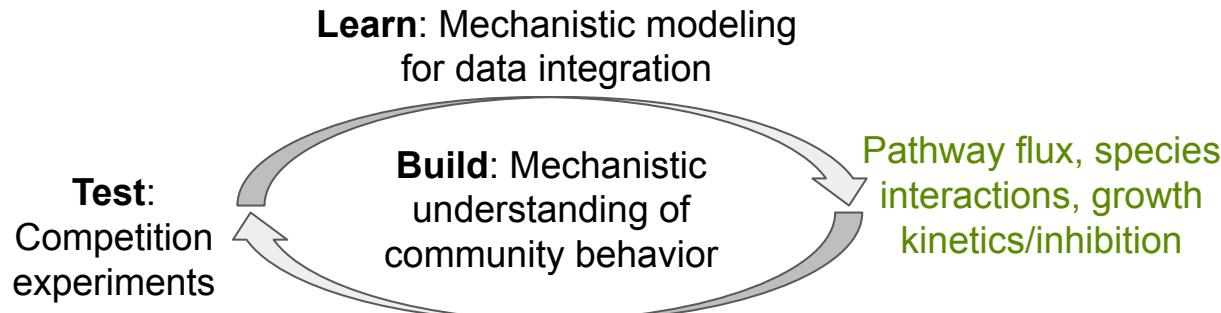
MRO:MIP

MRO = competition
MIP = syntrophy

cardinality

Fitting phase

Design/Build/Test/Learn cycle



Design: New experimental designs to reduce model uncertainty and test hypotheses

- Time-resolved species phenotypes, with full chemical detail and transparency of active pathways
- Numerical parameters of uncertainty suggest new data and experiments to reduce uncertainty
- Model-driven experimental design invites applications in *self-driven automated labs*

Optimization Approach to Community Simulation

Indices

Time, Trial, Species,
Phenotype, Metabolite

Variables

Species biomasses [g/L]
Metabolite concentrations [M]
Phenotype abundances
Growth Km/vmax

Constraints

Metabolite mass balance
Biomass mass balance
Phenotype transitions

Fitting objective: minimize deviation from the fit to experimental data
Design objective: variable (e.g. maximize community growth)

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 .
cvt & $cvcf$	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.
$cvmmin$	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.
$cvt_{k,t,j}$ & $cvcf_{k,t,j}$	The quantity of strain k biomass that transitions to and from the stationary phase, respectively, at an experimental datum.



U.S. DEPARTMENT OF
ENERGY



KBase
DOE Systems Biology Knowledgebase

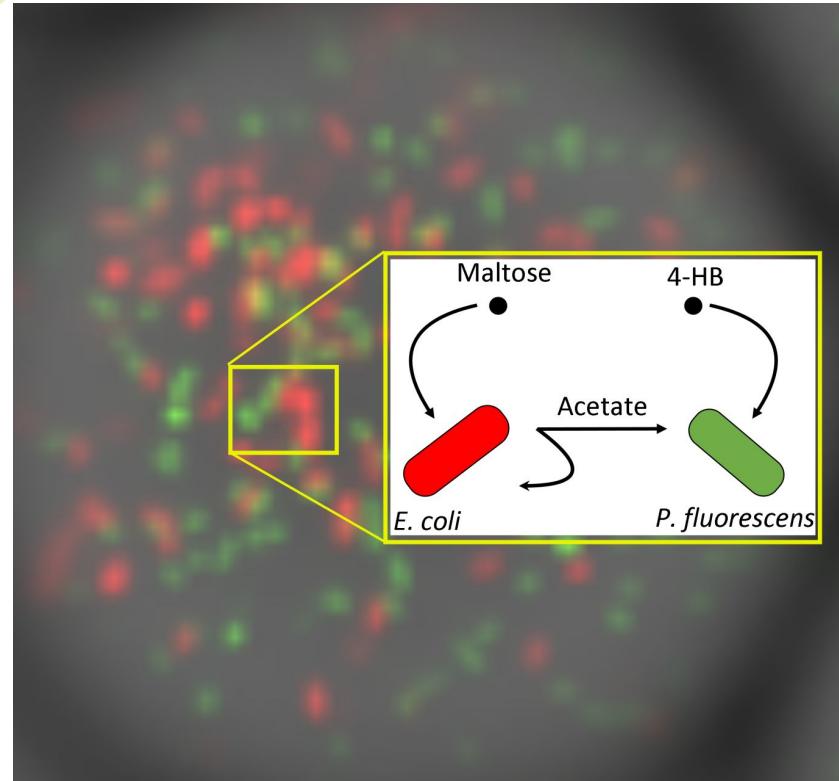
Office of Science

Phenotyping logic

1. Select a carbon source for the sought phenotype
2. Metabolic adjustments
 - a. Hydrogen consumption is prohibited
 - b. Oxygen consumption is constrained to equal that of the carbon source
 - c. A minimal biomass growth is defined
3. The total influx of carbon sources are minimized, with less penalty for consuming the carbon source
4. Known excreta for the carbon source through the organism can be maximally excreted
5. Apply pFBA to acquire the most efficient metabolic pathway
6. Save the vector of non-zero fluxes as the flux phenotype profile for the given carbon source

2-member community

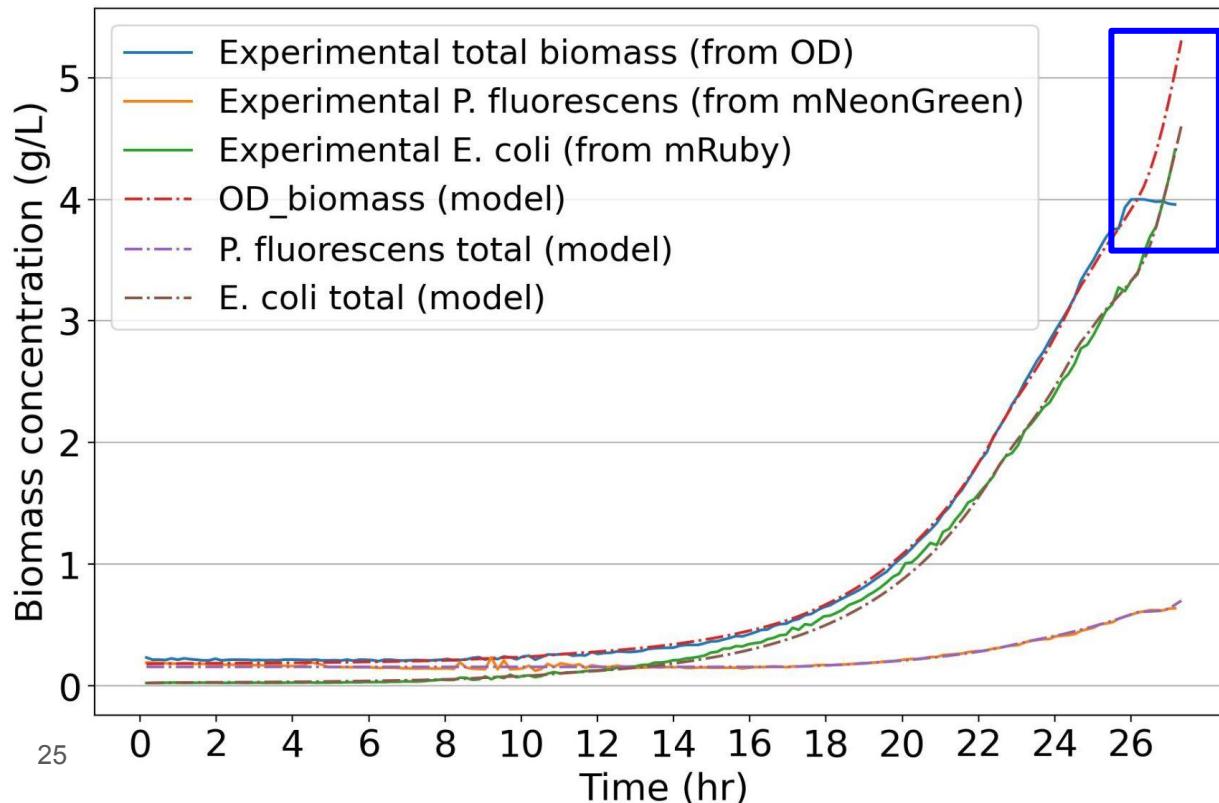
- *E. coli* and *P. fluorescens*
 - Experimentally convenient
 - Vital cross-feeding (acetate) in maltose
 - Competitive dynamics
- Analytically examined
 - Fluorescence
 - BIOLOG
 - Metabolomics
- Explored for engineering
 - Gene knockouts
 - Acetate production adjustments



Growth data - fit

1:20 *E. coli* : *P. fluorescens*

80% [maltose] reduction; 0.5 mM final [Acetate] (per metabolomics)



Simple kinetics is flawed by assuming saturated uptakes

-We are integrating full Michaelis Menten kinetics into the model

Fit quality varies between trials

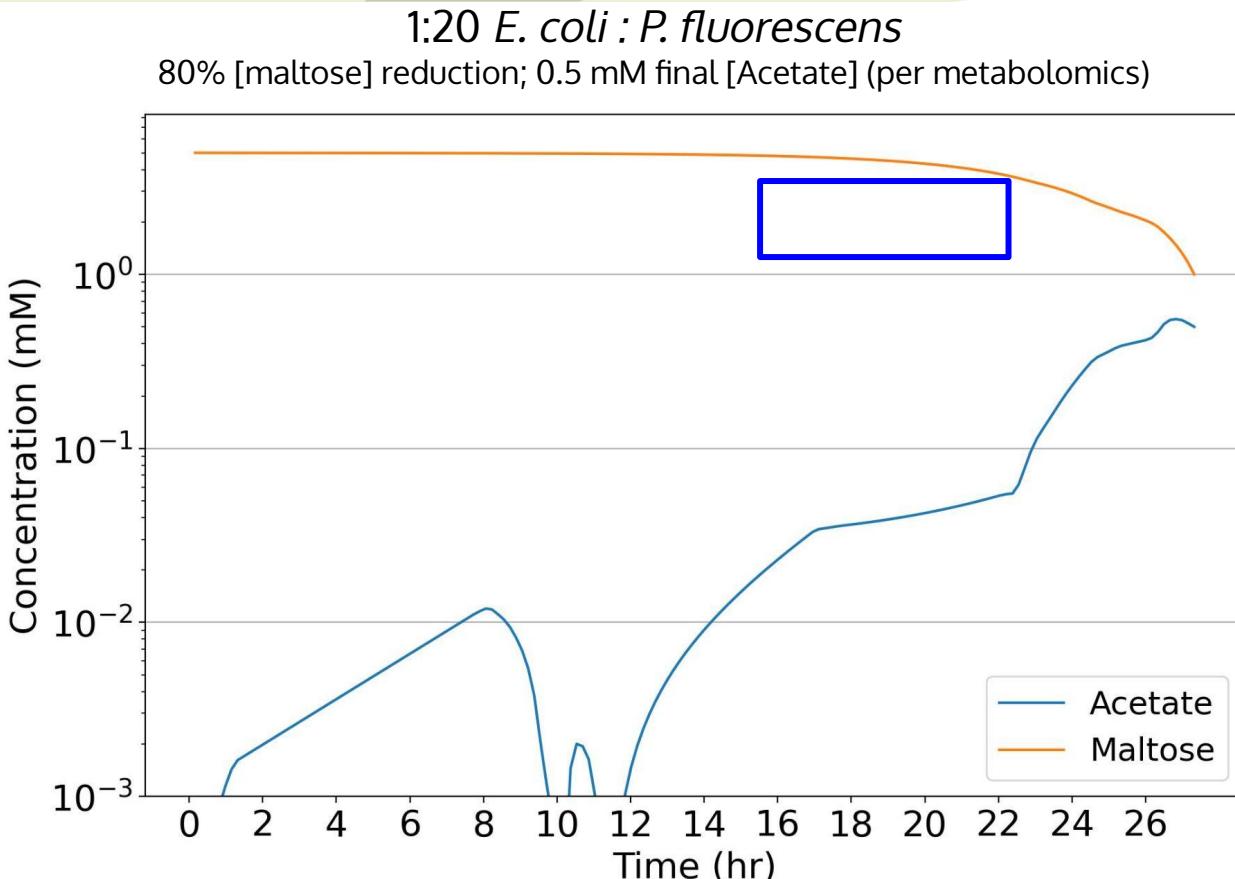


U.S. DEPARTMENT OF
ENERGY

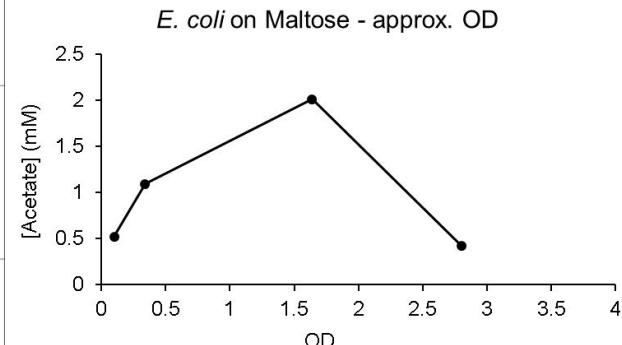


Office of Science

Growth data - concentrations



Simple kinetics further hinders recapitulation of metabolomics [Acetate], per incorrect acetate kinetics



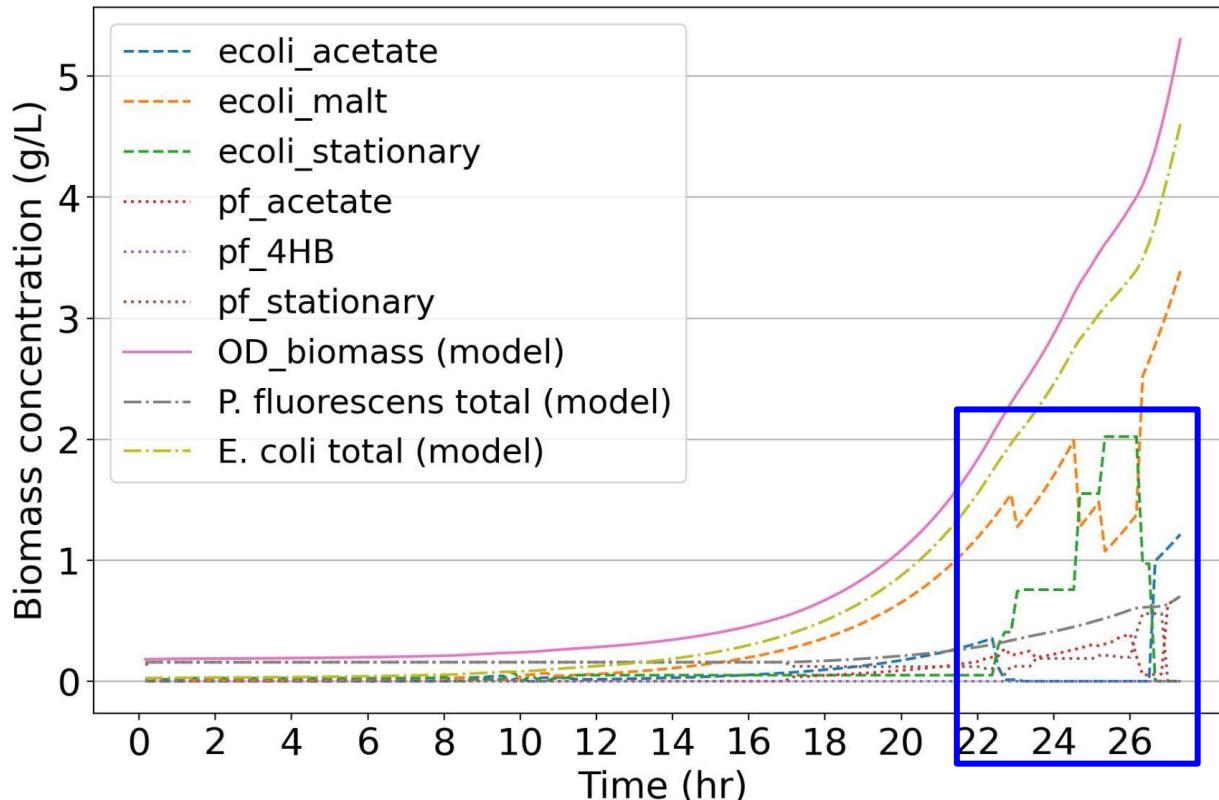
U.S. DEPARTMENT OF
ENERGY

KBase
DOE Systems Biology Knowledgebase

Growth data - phenotypes

1:20 *E. coli* : *P. fluorescens*

80% [maltose] reduction; 0.5 mM final [Acetate] (per metabolomics)



By assuming saturation kinetics, the model can only fit the later portion of the growth curve by moving in and out of stationary phase

-We are exploring preventing this behavior with additional constraints

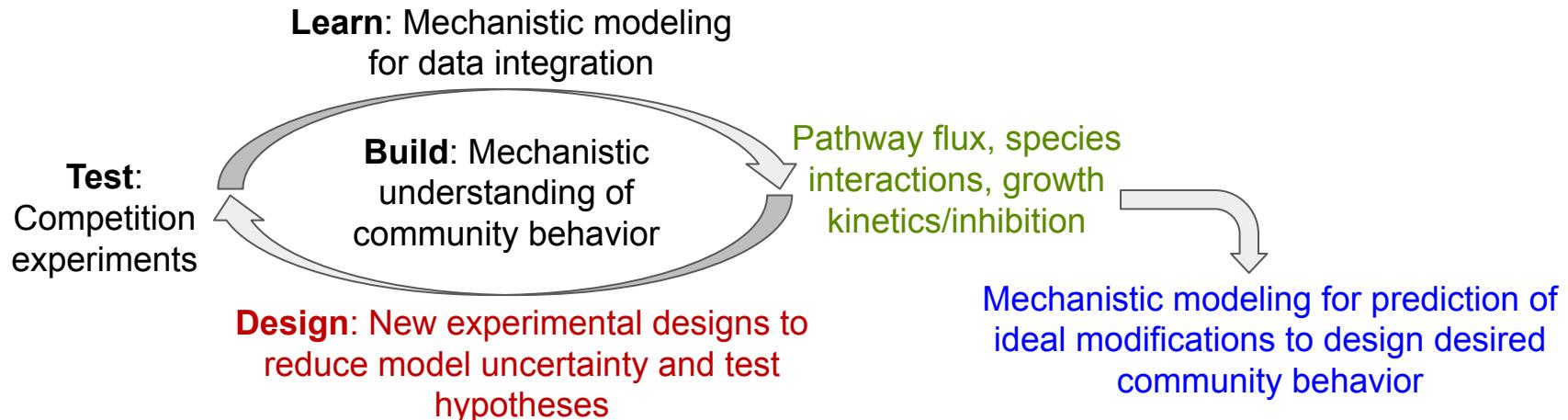
Another issue relates to problems with fluorescence as a measure of biomass as fluorescence production increases in stationary phase

-Phenotype-specific fluorescence OD conversion would prevent this

Phenotype growth kinetics

The growth kinetics of each phenotype are predicted by our model (table at left). The use of linear kinetics introduces error with phenotypes that do not grow in the media, and should be corrected by implementing Michaelis-Menten kinetics to our model.

Fitting phase + Engineering phase Design/Build/Test/Learn cycle



- Time-resolved species phenotypes, with full chemical detail and transparency of active pathways
- New data and experiments can be suggested to reduce uncertainty
- Model driven experimental design invites applications in self-driven automated labs
- The **Engineering phase** rapidly simulates community behaviors from strain or system modifications

Engineering phase

Design:

- Objectives:
 - Biomass ratios (e.g. high EC low PF)
 - Maximize desired byproduct titre/yield/flux
- Input = initial culture conditions
- Output = initial inoculation ratio and phenotype expression over time (includes consideration of potential phenotypes that could arise from strain engineering)

Prediction:

- Objective:
 - Maximize total species biomass
- Input = Environmental conditions and inoculation ratios
- Output = Species composition, metabolite concentration, and phenotype expression over time

Solve the model

OPTIMIZATION

Minimize variance

Penalize phenotypic transitions

RESULTS

Kinetics constants

Phenotypic proportions

Metabolic concentrations



Engineering phase

OPTIMIZATION

Maximize community biomass

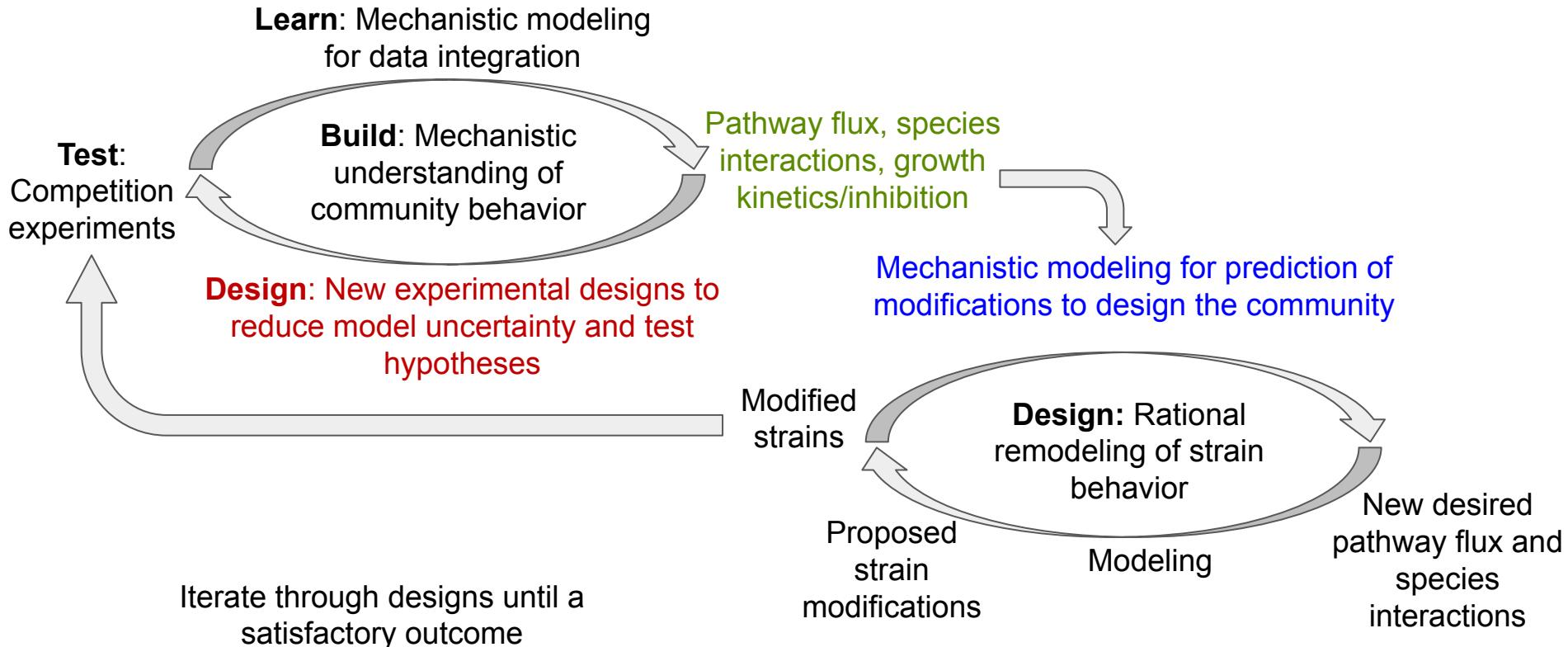
RESULTS

Community abundances

Phenotypic proportions

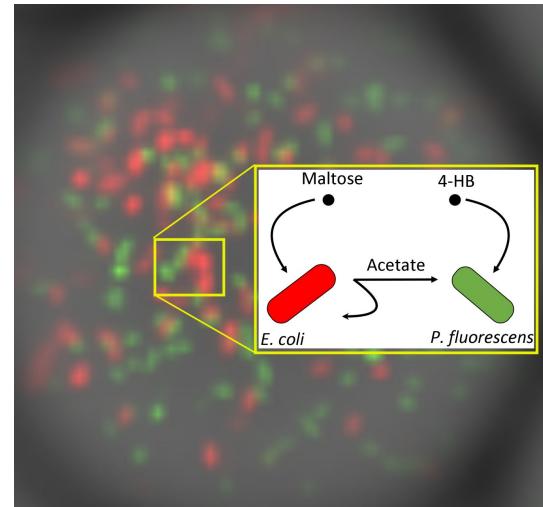
Metabolic concentrations

MSCommPhitting Design/Build/Test/Learn cycles

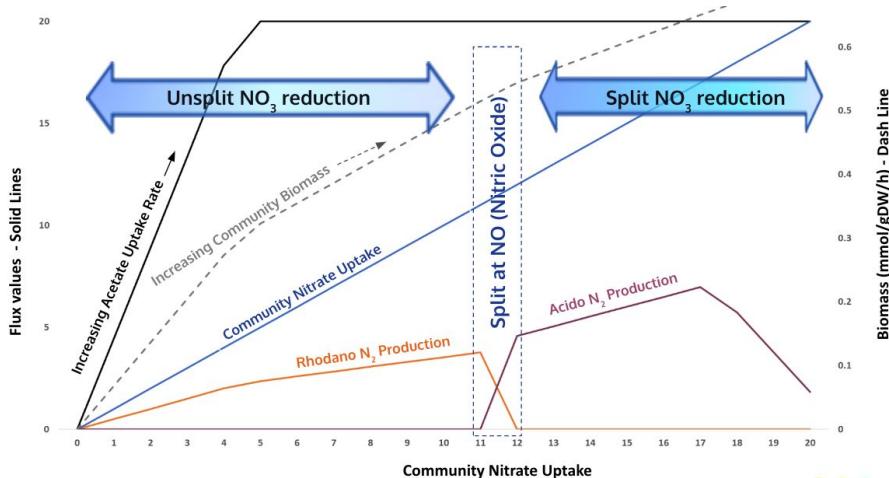


Recent Applications

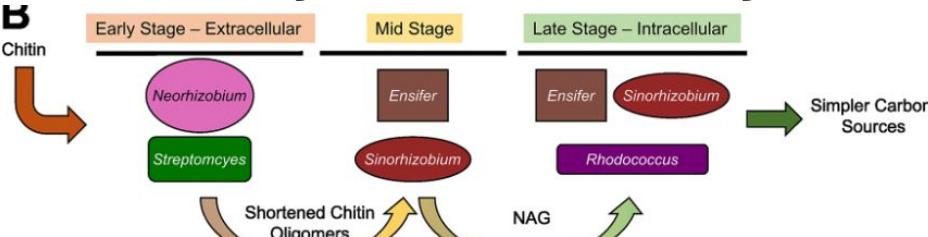
ANL: Predicting metabolic competition



ENIGMA: Predicting interactions involving nitrate cycle



PNNL: Predicting interactions in chitin degradation



ORNL: Predicting effective synthetic communities

Genus	Strain	YR343	YR139	GM17	AP49	CF142	BC15	AP07	CF402	YR139	BT03
<i>Pantoea</i>	YR343				+	+				ND	
<i>Pseudomonas</i>	GM17			-	-	-	-	-	-	ND	-
<i>Sphingobium</i>	AP49				-	+				ND	
<i>Rhizobium</i>	CF142					+				ND	
<i>Variovorax</i>	CF313				-					ND	-
<i>Bacillus</i>	BC15								-	ND	
<i>Caulobacter</i>	AP07				-					ND	
<i>Duganella</i>	CF402					+	+			ND	
<i>Streptomyces</i>	YR139			-	-			-		ND	
<i>Paraburkholderia</i>	BT03				-					ND	