



Building metabolic models of bacterial isolates

Christian Diener, Gibbons Lab



from the **2021 ISB Virtual Microbiome Series**

CC-BY-SA

gibbons.isbscience.org

gibbons-lab

[@thaasophobia](https://twitter.com/thaasophobia)



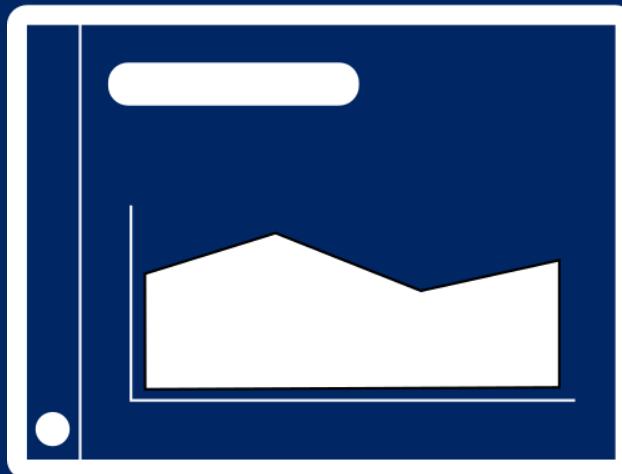
Let's get the slides first (use your computer, phone, TV, fridge)

https://gibbons-lab.github.io/isb_course_2021/models



Quick reminder 🕒

Presentation



logic
explanations
links

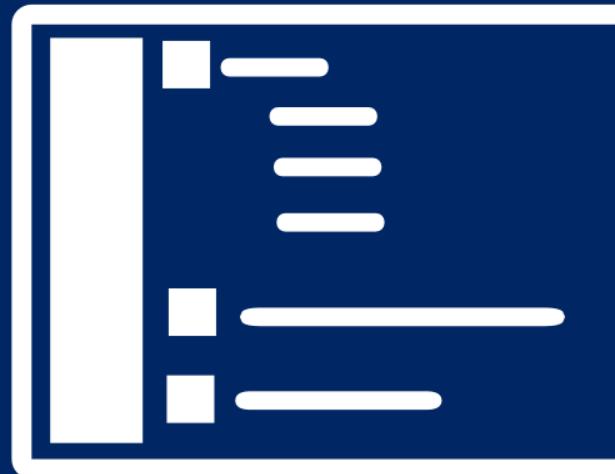
Click me to open the notebook!

Notebook



technical aspects
materials
visualizations

Chat



support
Q&A

Functional analyses

Tries to predict what the microbiome **does** from sequencing data.

Uses gene/transcript/protein/metabolite abundances (metagenomics, metatranscriptomics, proteomics or metabolomics).

Gene content provides insights into metabolic **capacity** or **potential**.

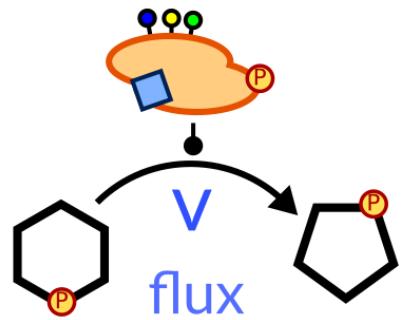


Genes and metabolite abundances are cool and all, but what you should really care about are metabolic fluxes*

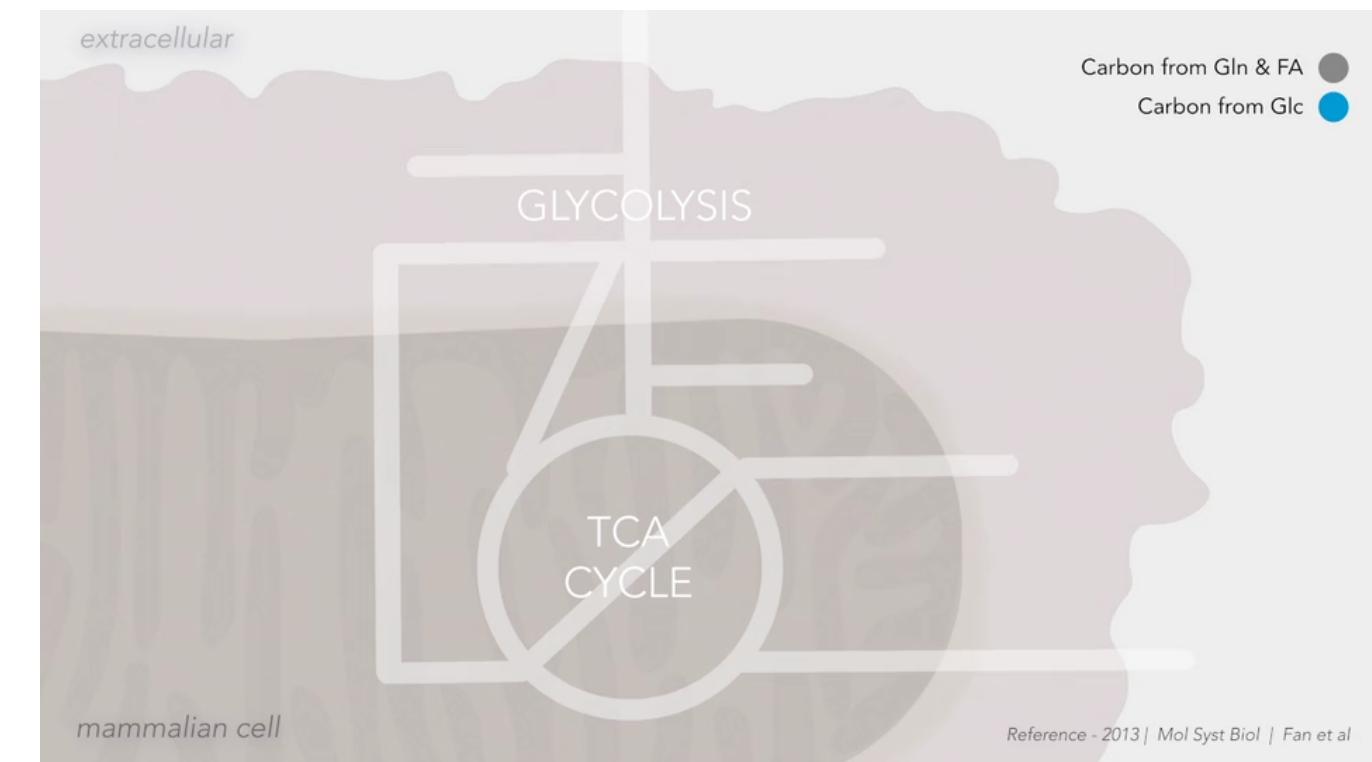
hot take 🔥



Fluxes



- rate of mass conversion
- unit is mmol/(gDW·h)
- costly to measure
 - longitudinal metabolomics
 - targeted temporal ^{13}C or ^{15}N



video courtesy of [S. Nayak](#) and [J. Iwasa](#)

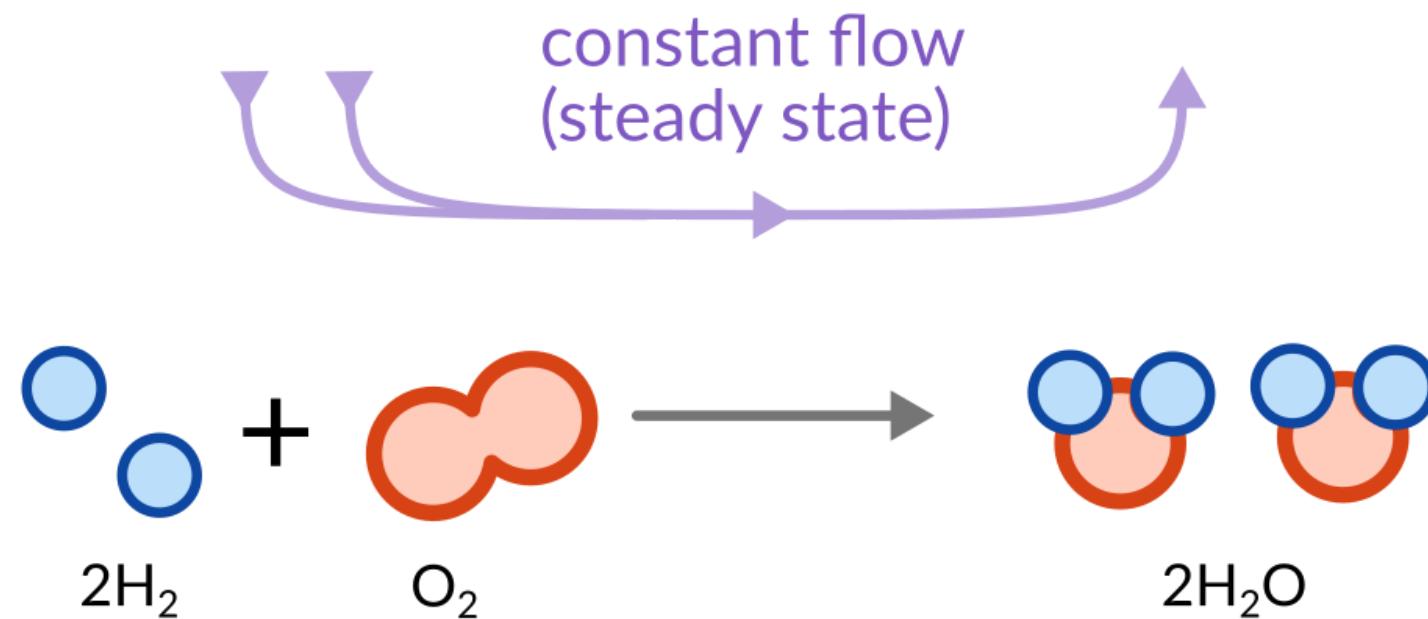


Flux Balance Analysis(FBA)

Can we infer the most likely fluxes in a biological system, given a set of environmental constraints, if we know all available metabolic reactions?



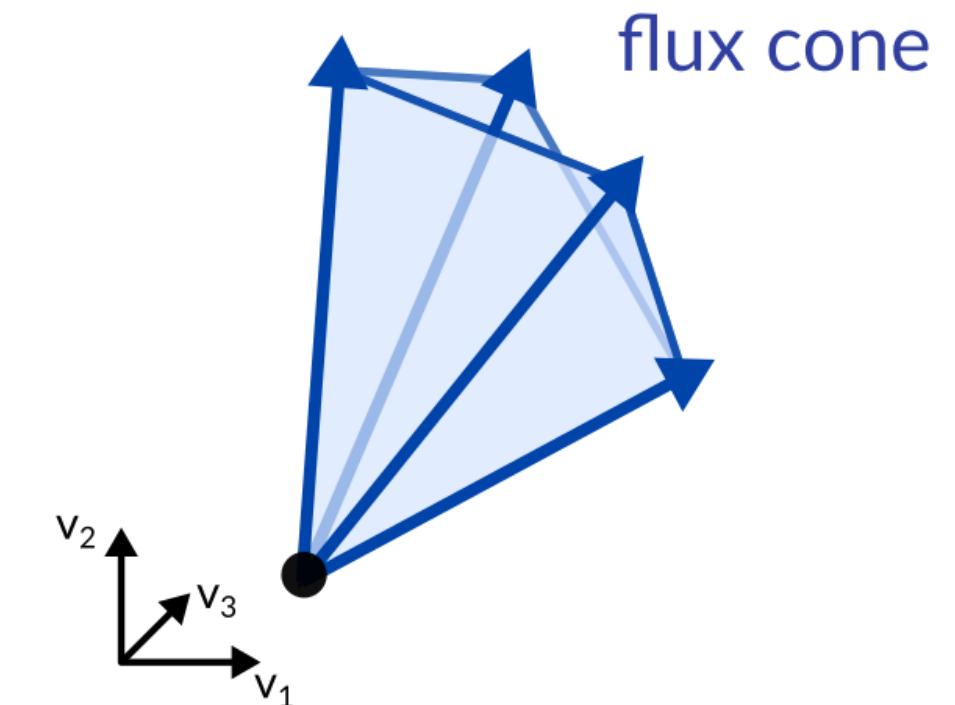
The flux cone



balance equations

$$2 \cdot v_{\text{H}_2\text{O}} - 2 \cdot v_{\text{H}_2} - v_{\text{O}_2} = 0$$

$$v_{\text{H}_2\text{O}} \geq 0$$



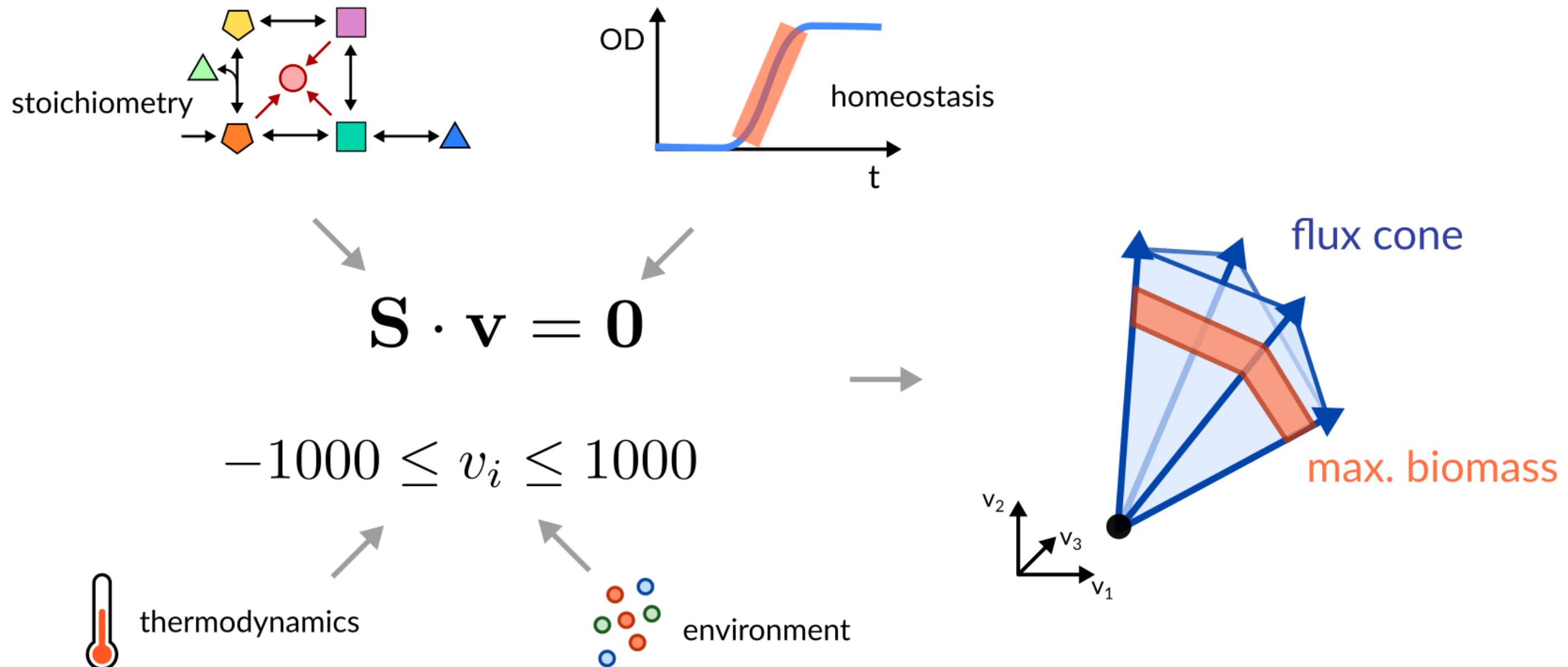
$$\mathbf{S} \cdot \mathbf{v} = 0$$

$$-1000 \leq v_i \leq 1000$$

The goal of FBA is to **reduce** the overall flux space to a **biologically relevant** one.



Genome-scale metabolic modeling



How do we get from sequencing data to metabolic reactions?

Metabolic reactions are catalyzed by an organism's **enzymes**, which are encoded in its **genes**.

So what we need is a **genome**.



Genome Assembly with De Bruijn graphs

tgaacgctgcg — original sequence
tgaacgct
gaacgctg
aacgctgg
acgctggc
cgctggcg] k-mers



De Bruijn graph

connect to k-mers if they have a shifted overlap

tgaacgct
gaacgctg
aacgctgg
acgctggc
cgctggcg



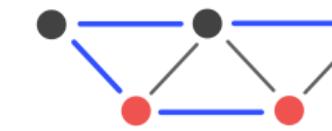
assembly is a traversal of the graph

tgaacgctgcg
tgaacgct
gaacgctg
aacgctgg
acgctggc
cgctggcg

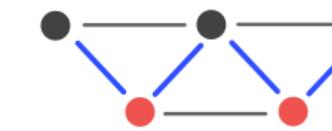


De Bruijn graph

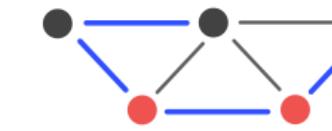
?



?



?

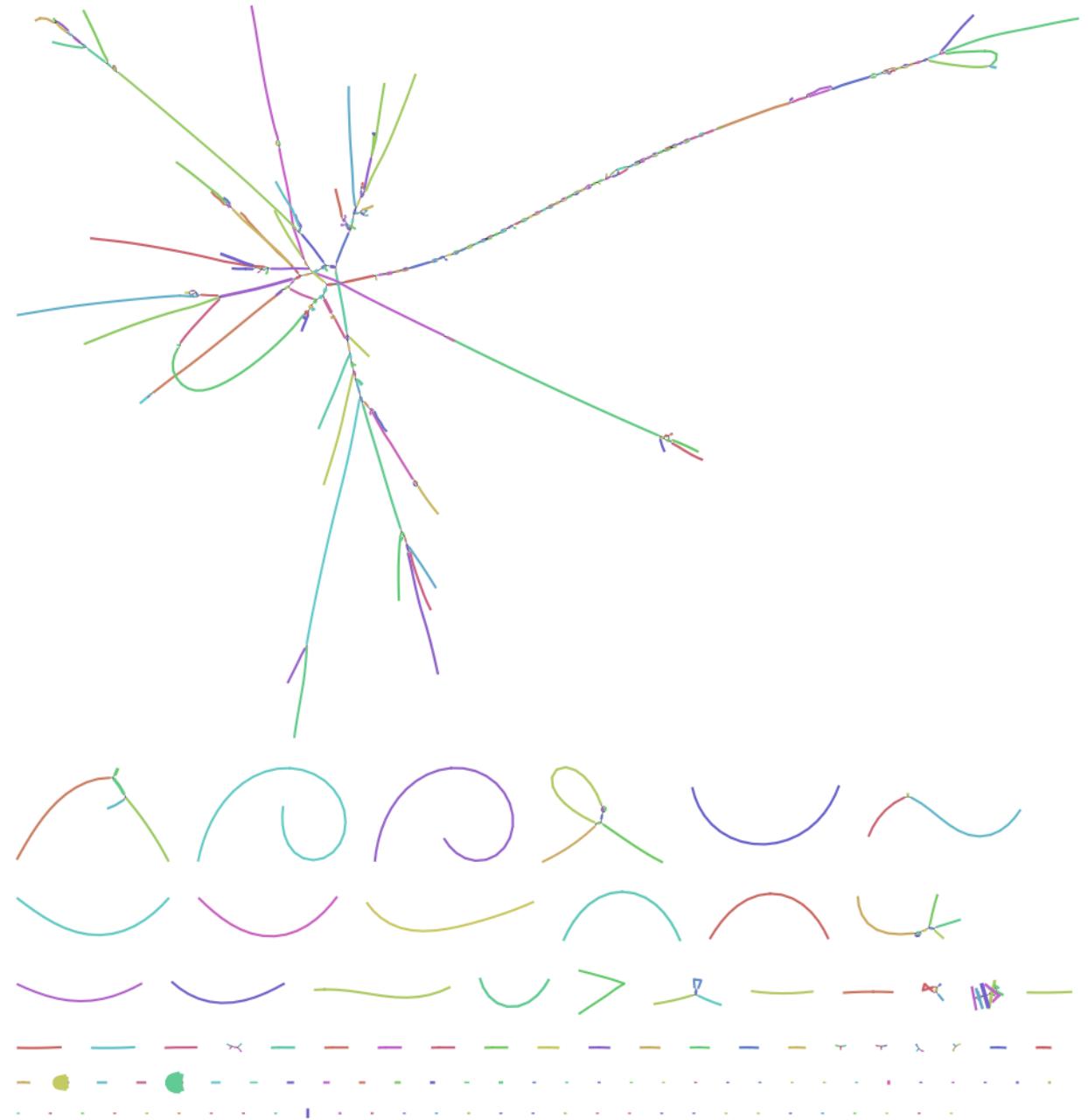


repeat regions can lead to loops in the graph
(no unique assembly possible)

fix: multiple k-mer sizes



Not so straight-forward



Finding genes

Hyatt *et al.* BMC Bioinformatics 2010, **11**:119
<http://www.biomedcentral.com/1471-2105/11/119>



SOFTWARE

Open Access

Prodigal: prokaryotic gene recognition and translation initiation site identification

Doug Hyatt^{1,2*}, Gwo-Liang Chen¹, Philip F LoCascio¹, Miriam L Land^{1,3}, Frank W Larimer^{1,2}, Loren J Hauser^{1,3}

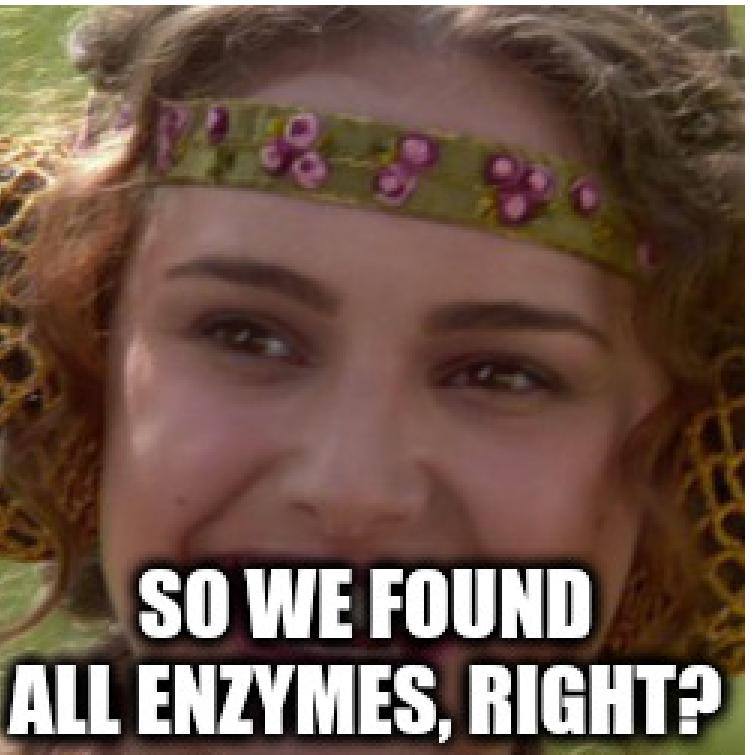
- finds open reading frames (ORFs)
- identifies ribosomal binding sites and spacers
- **really good** at identifying proteins

<https://github.com/hyattpd/Prodigal>





I FOUND ALL
PROTEINS IN THE GENOME



SO WE FOUND
ALL ENZYMES, RIGHT?



imgflip.com



SO WE FOUND
ALL ENZYMES, RIGHT?

Genome-Scale Metabolic Model Reconstruction

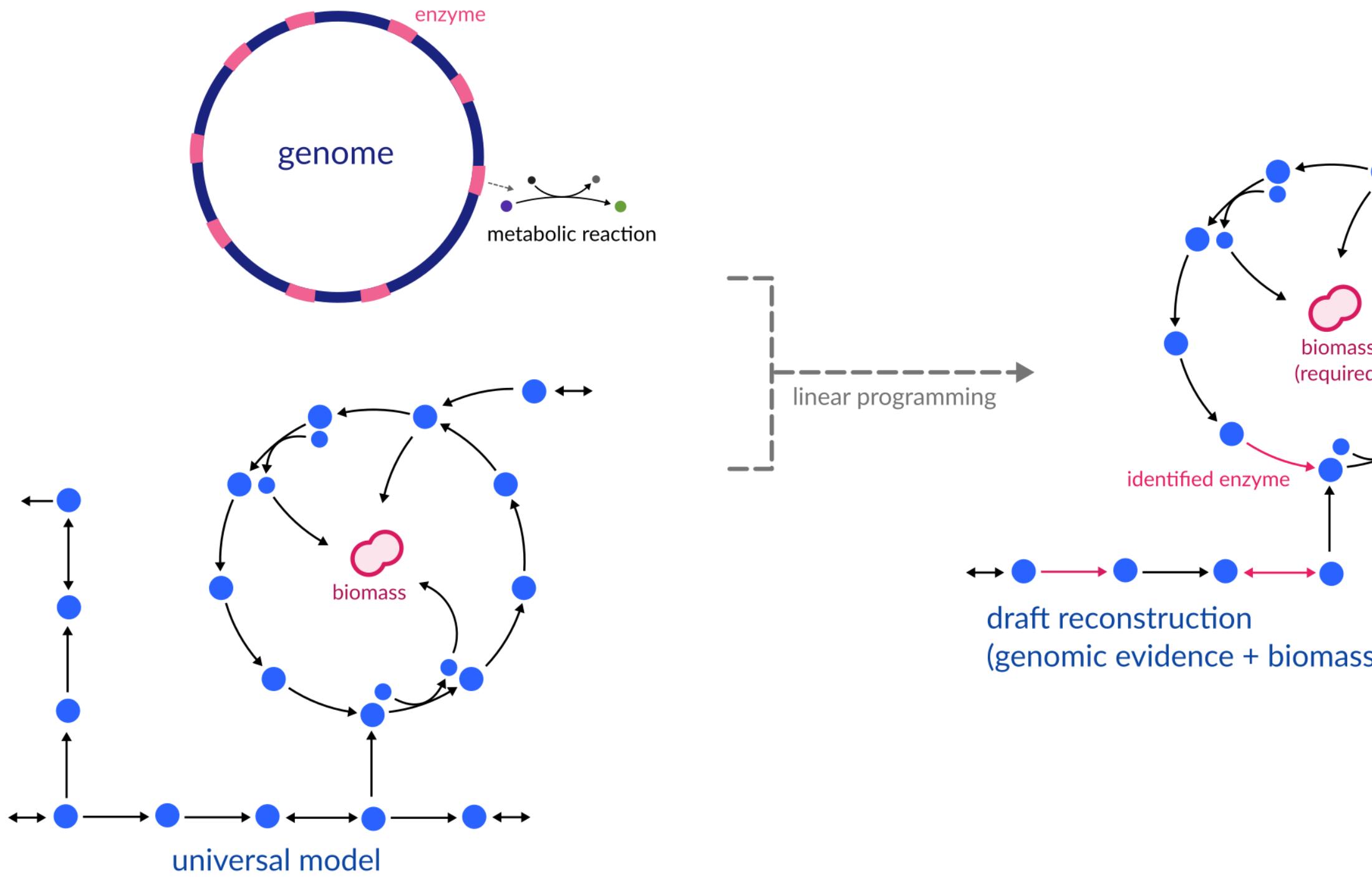
When mapping to reference databases, we will only ever identify a small set (~1/3) of all enzymes and transporters in a given genome.

We need to fill in the critical gaps in metabolic pathways by imposing **functional requirements** on the model.

What's a reasonable requirement that you'd attribute to a living bacterium?



General strategy





Curated reconstructions

Curation is the process of adding or removing reactions to/from the model based on experimental evidence.

Basic

- check structural quality ([MEMOTE](#))
- gap-filling on a standard growth medium (LB, M9, ...)
- example: [carveME EMBL GEMs](#) - 5587 strains

Stringent

- growth on various carbon sources (e.g. "likes maltose, but not glucose")
- known metabolic conversion (e.g. "produces indole from tryptophan")
- strain-specific biomass composition
- example: [AGORA](#) - 818 strains from human gut



	carveME	ModelSEED/Kbase	gapseq
speed	😊	😔	😢
sensitivity	😔	😢	😊
model quality	😊	😔	😊
free solver	😢	😔	😊
easy install	😊	😊	😢
easy to use	😔	😊	😢
many media	😔	😊	😊*
SBML standard	😊	😊	😊

Limitations

- unknown enzymes/pathways are never captured
- dependency on a universal model
- hard to formulate growth media
- growth objective may not always apply (e.g. toxicity, human tissues)

The BIO-ML collection

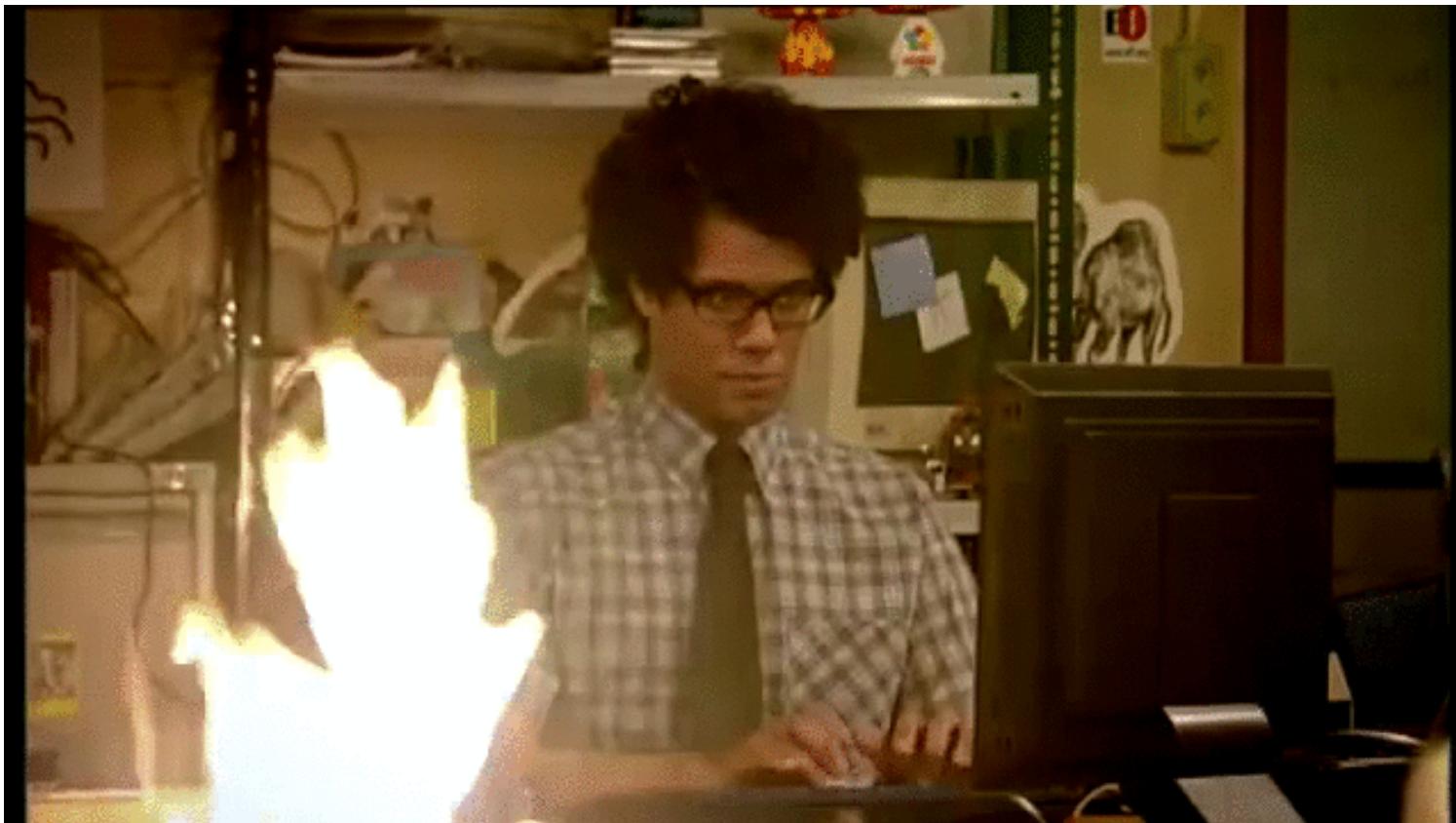
Sequenced 3,632 bacterial isolates from the human gut at the Broad Institute.

Each participant will get a randomly assigned assembly from ~980 of those.

<https://doi.org/10.1101/2019.05.24.809059>



Your turn



Let's go from an isolate genome assembly to a genome-scale metabolic model.

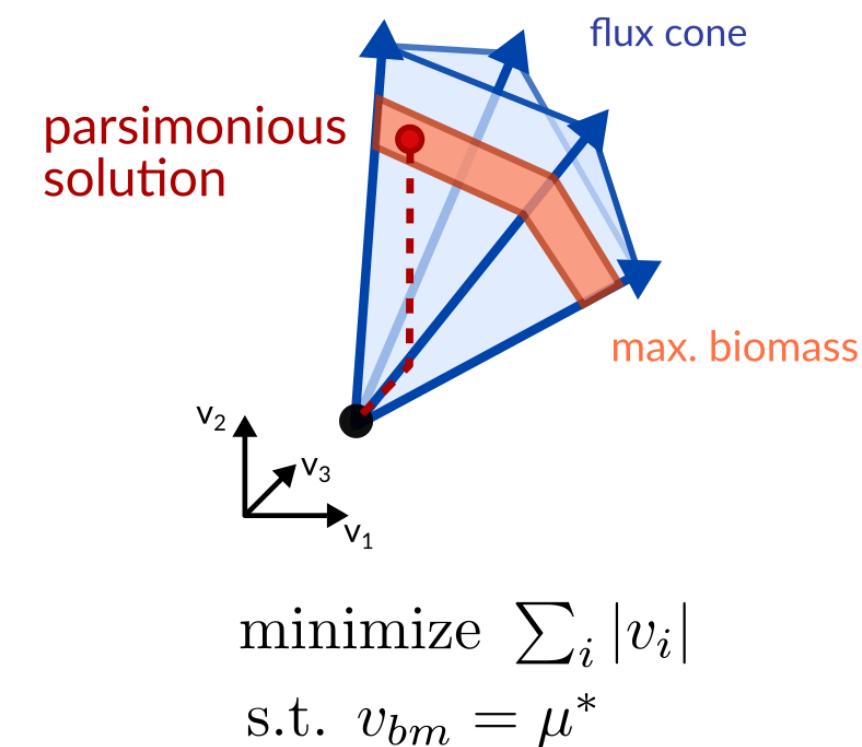
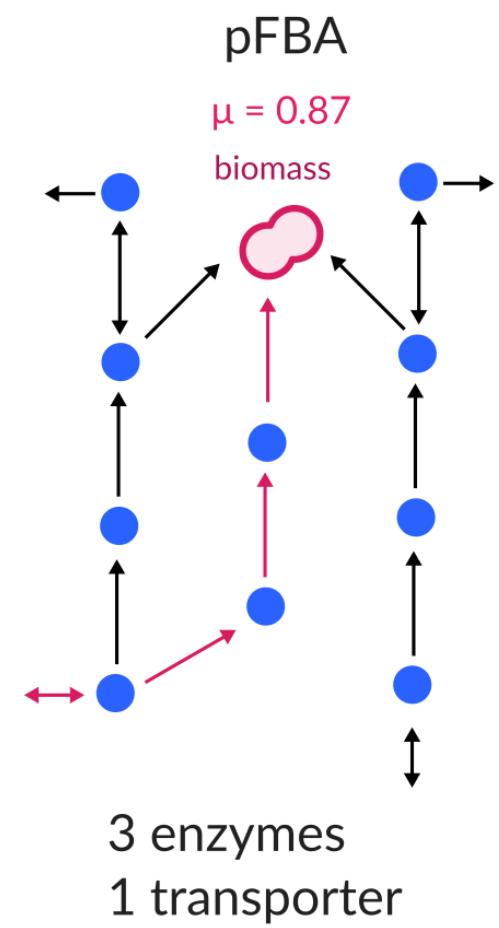
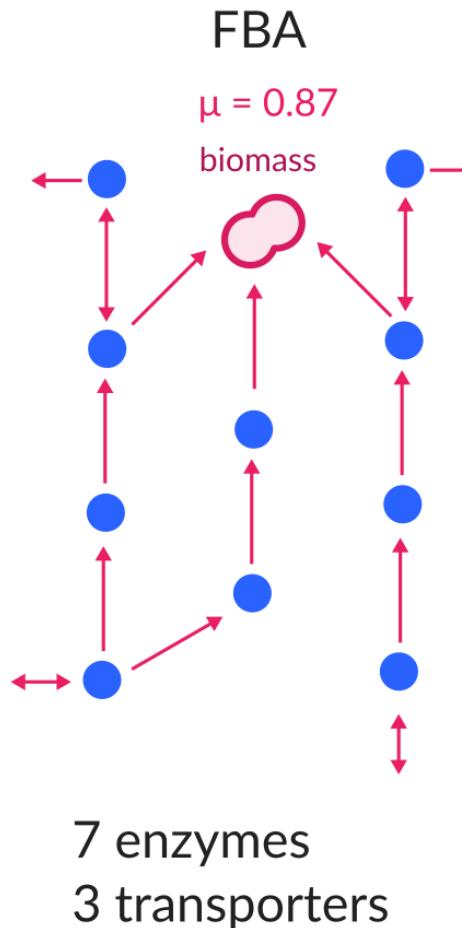
Optimal growth rate \neq unique fluxes

Even though FBA yields a unique estimate of the maximum growth rate, it does not give us a unique flux solution.

Can we constrain the flux space even further?



Selecting biologically relevant fluxes via parsimony



- Bacteria do not like to produce more enzymes than necessary.
- Zero flux = no enzyme required.
- Parsimony reproduces experimental fluxes in *E. coli* very well.

FBA results strongly depend on the **environmental conditions**. Finding uptake constraints is **critical** for realistic results.

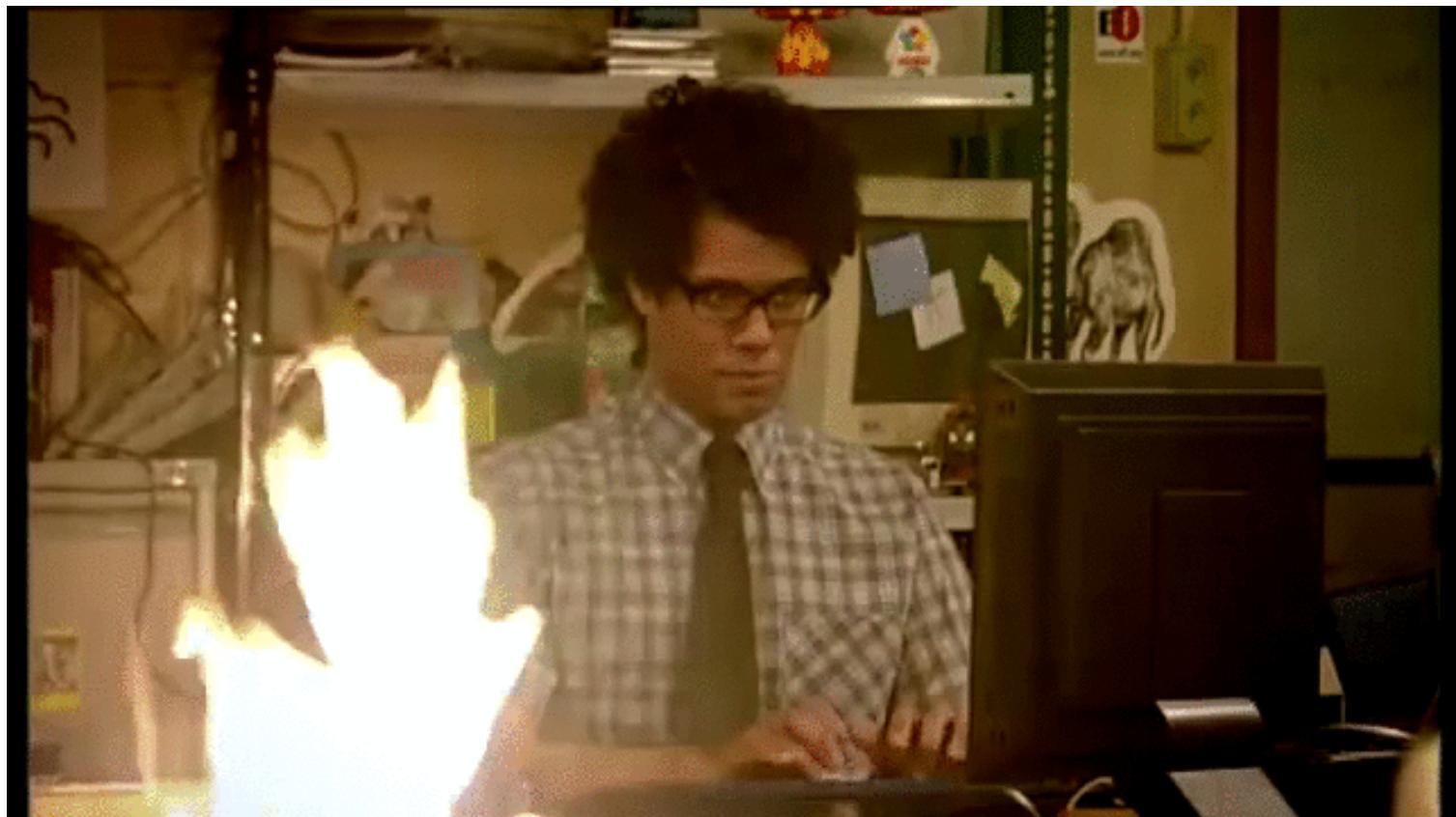
Quantitative > qualitative >> unconstrained

Uptake and secretion fluxes are much easier to measure than internal fluxes (e.g. you can use longitudinal metabolomics).



Your turn

Let's estimate the fluxes for our models.



Break



Reverse ecology: inferring environmental interactions with metabolic models

Christian Diener, Gibbons Lab



from the **2021 ISB Virtual Microbiome Series**

CC-BY-SA gibbons.isbscience.org gibbons-lab.org [@thaasophobia](https://twitter.com/thaasophobia)



Reverse ecology

Ecology is the study of the relationships between living organisms and their physical environment.

Reverse Ecology is the study of an organism's genome as a means to estimate its ecological niche.



Wait, why did we just stumble into Reverse Ecology?

Systems Biology lingo:

All strains were simulated under the exact same environmental conditions, but they each showed different uptake rates for different sets of metabolites to obtain their maximum growth rate.

Ecology lingo:

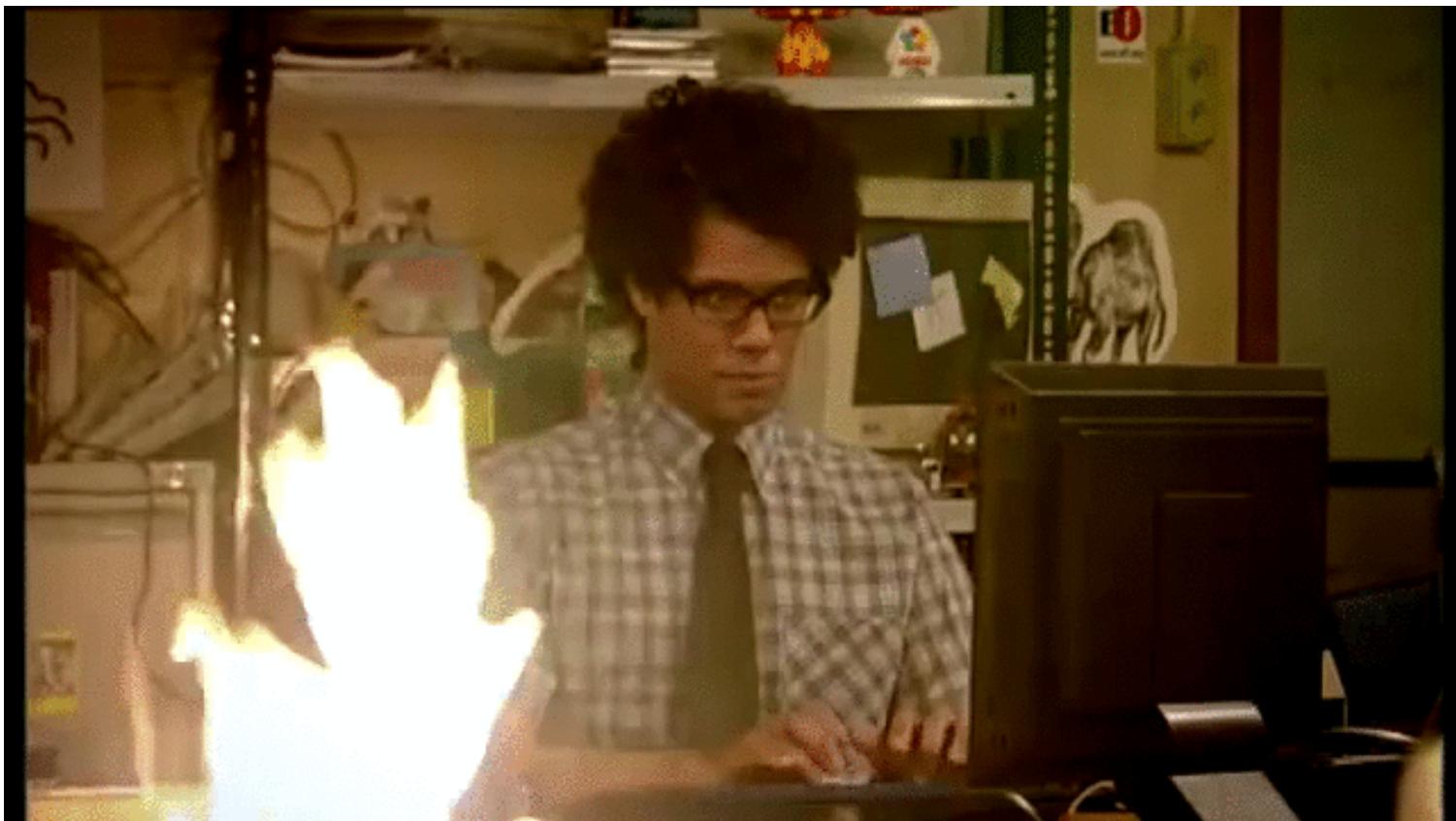
We estimated the realized niche of bacterial strains grown in isolation under an optimal fitness objective.

What would it mean ecologically for two strains to inhabit very similar metabolic niches, or very different ones? Think about the human gut, where many of these strains may be present at the same time.



Your turn

Let's combine **all** the results.



And we are done 🙌

ISB team

Joey Petosa
Allison Kudla
Christian Diener
Sean Gibbons
Priyanka Baloni
Tomasz Wilmanski
Noa Rappaport
Alex Carr
Audri Hubbard
Renee Duprel
Joe Myxter
Thea Swanson

Thanks!

