ETH zürich

551-1119-00L Microbial Community Genomics

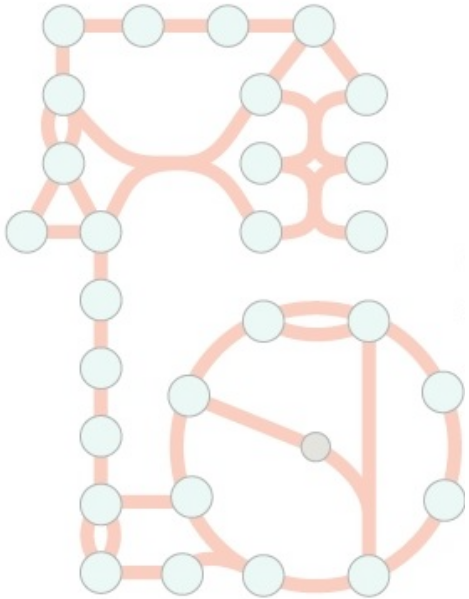
Lecture:
Tools for metabolic modeling in microbial communities

Samuel Miravet-Verde 8-Nov-23

D BIOL

Lecture content:

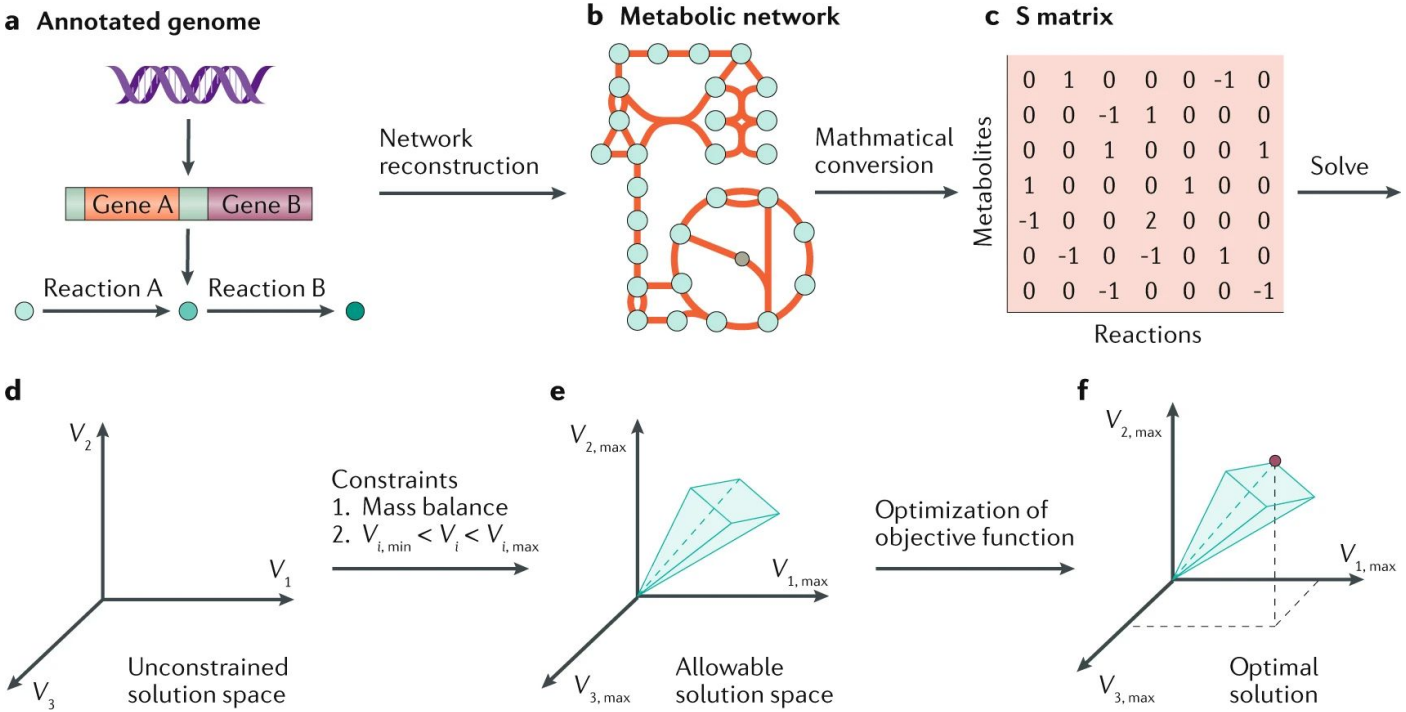
- Introduction to Genome-scale Metabolic models (GEMs)
- Building GEMs
- Evaluation of GEMs
- Exploration of GEMs
- Optimization of GEMs
- Visualization of pathways
- Exploring metabolic interaction in the community
- Introduction for Hands-on tutorial
- Conclusions



GEMs as a tool in microbiology research:

Genome-scale metabolic models (GEMs) provide a representation of an organism's metabolism, encompassing all known metabolic reactions and associated genes. This allows:

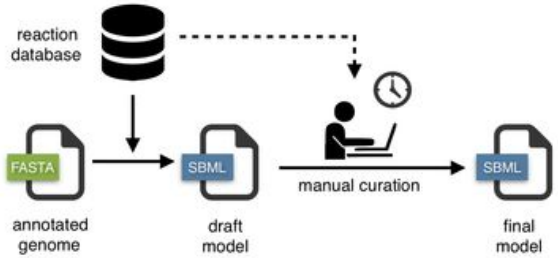
- to understand and simulate the biochemical processes within a cell
- predict of cellular behavior under different conditions, such as:
  - growth rates
  - nutrient availability
  - genetic perturbations
  - analysis of metabolic pathways and flux distributions
  - identification of potential drug targets
- GEMs can be integrated with omics data to enable the study of metabolic responses to environmental changes, diseases, and the design of biotechnological applications.



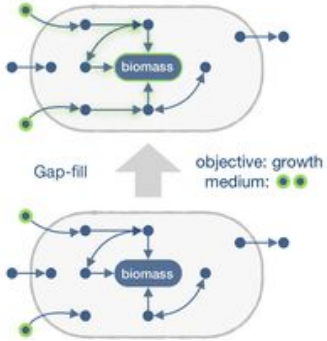
Adapted from: [Reconstructing organisms in silico: genome-scale models and their emerging applications | Nature Reviews Microbiology]

Building GEMs (bottom-up)

In the traditional (bottom-up) approach, a draft model is automatically generated from the genome of a given organism (by homology/orthology against annotated genes), followed by extensive manual curation



Adapted from: [Fast automated reconstruction of genome-scale metabolic models for microbial species and communities]



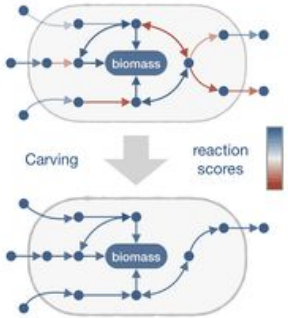
New reactions are iteratively added to the network by gap-filling. This process is context dependent, i.e. it requires specifying the environmental conditions (growth medium) and the expected phenotype (usually biomass formation or growth)

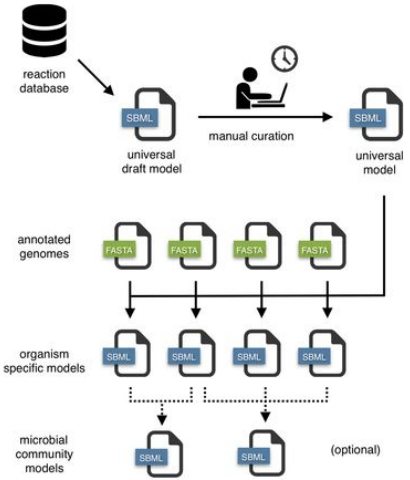
Building GEMs (top-down)

1. A universal model is generated and curated.
2. This model is used as a template for organism-specific model generation (carving)
   a. Identify (by homology/orthology) reactions present in the given organism.

This does not require manual intervention and can be parallelizable to generate large numbers of models.

This can be applied to the generation of microbial community models by merging single-species models.





Building GEMs (top-down)

cdanielmachado/carveme: CarveMe: genome-scale metabolic model reconstruction (github.com)

CarveMe is a python-based tool for genome-scale metabolic model reconstruction

Input: genes (nucleotide or amino acids)

Output: .xml and .tsv file with genes associated to reactions

BiGG Models

- By default it uses DIAMOND against BiGG:
  - composed of >70 highly curated metabolic models
  - 58 of these are from *Escherichia coli*

Do you see any limitation here?

- Alternatively:

eggNOG-mapper (embl.de)

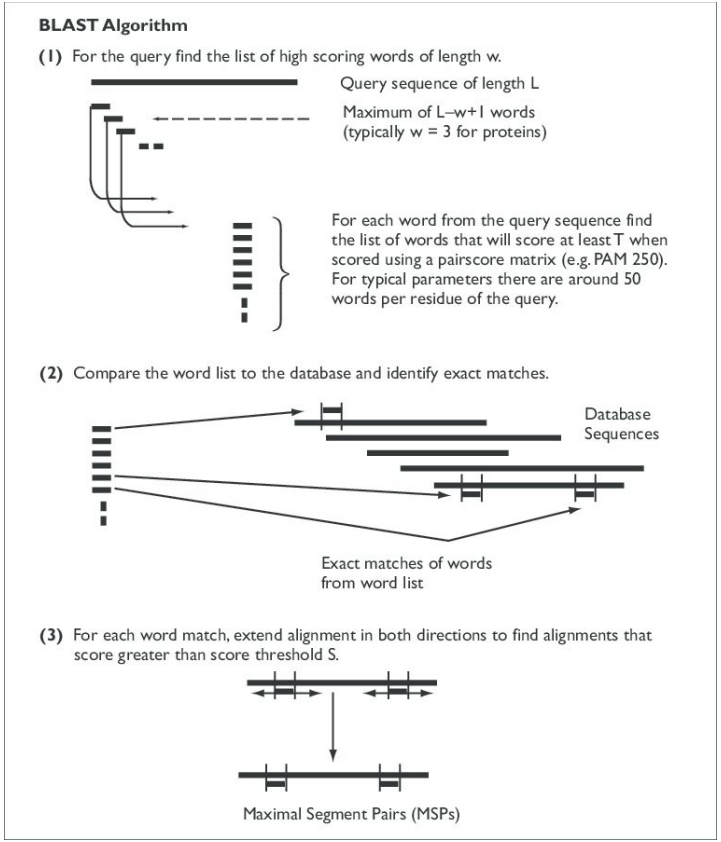
- Precomputed Orthologous Groups (OGs)

diamond

bbuchfink/diamond: Accelerated BLAST compatible local sequence aligner. (github.com)

DIAMOND is a sequence aligner for protein and translated DNA searches, designed for high performance analysis of big sequence data. The key features are:

- Pairwise alignment of proteins and translated DNA at 100x-10,000x speed of BLAST.

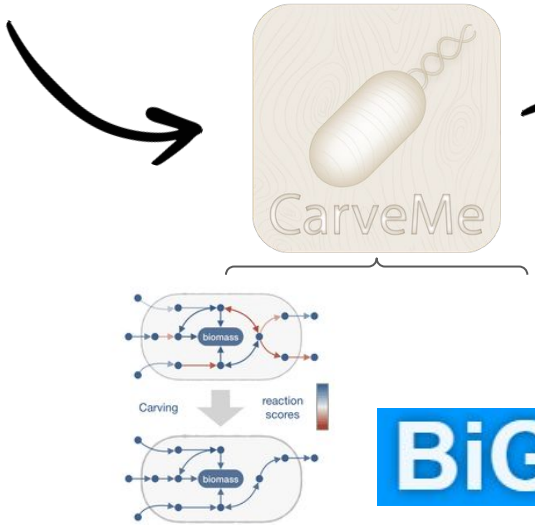


Building GEMs (top-down)

FASTA Sequence File

```
>generic|ENSMUSP00000107433|Erp29|ER protein 29
MAAAAGVSGAASLSPLLSVLLGLLLLFAPHGGSGLHTKGALPLDTVTFYKSRLLLGP

>generic|ENSMUSP00000120715|Rps2|ribosomal protein S2
MADDAGAAGGPGGPGGPGLGGRGGFRGGFGSGLRGRGRGRGRGRGRGARGGKAEDKEWIPVTKLGRLVKDMKIKSLEEIY
LFSLPIKESEIIDFFLGASLKDEVLKIMPVQKQTRAGQR
```

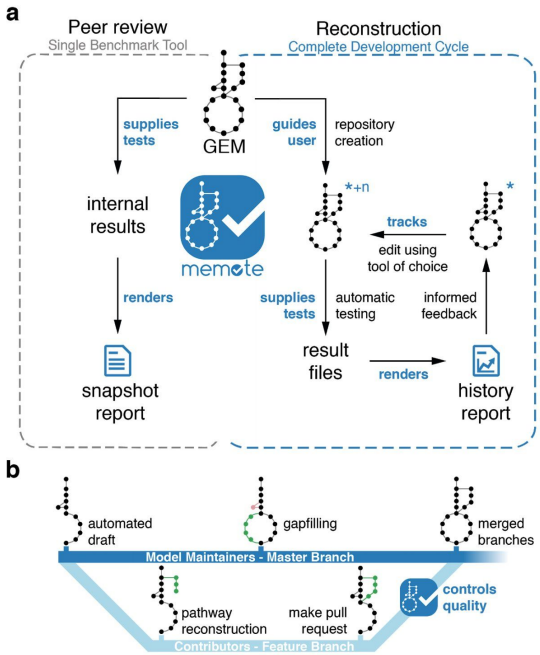


BiGG

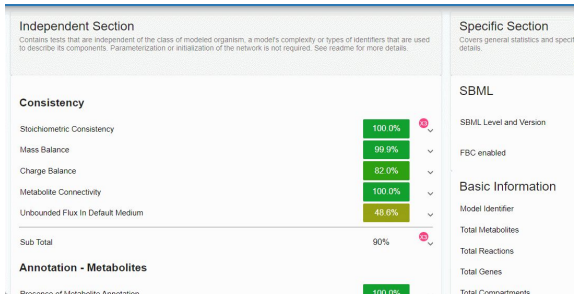
SBML Systems Biology Markup Language

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml level="2" version="1" xmlns="http://www.sbml.org/sbml/level2">
  <model name="example">
    <listOfCompartments>
      <compartment id="cell" size="1e-15" />
    </listOfCompartments>
    <listOfSpecies>
      <species id="A" compartment="cell"
                      initialConcentration="0.6" />
      <species id="B" compartment="cell"
                      initialConcentration="0.3" />
    </listOfSpecies>
    <listOfParameters>
      <parameter id="k" value="1" />
    </listOfParameters>
    <listOfReactions>
      <reaction id="AinB">
        <listOfReactants>
          <speciesReference species="A" />
        </listOfReactants>
        <listOfProducts>
          <speciesReference species="B" />
        </listOfProducts>
        <kineticLaw>
          <listOfParameters>
            <parameter id="k" value="0.1">
          </listOfParameters>
          <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
              <times/>
              <ci>cell</ci>
              <ci>k</ci>
              <ci>A</ci>
            </apply>
          </math>
        </kineticLaw>
      </reaction>
    </listOfReactions>
  </model>
</sbml>
```

Evaluation of GEMs



- MEMOTE serves as a benchmark tool generating a comprehensive, human-readable report, which quantifies the model's performance



  - Snapshot → report on a model
  - Diff → To compare models
  - History → version control of changes in a model

- For model reconstruction, MEMOTE helps users to create a version-controlled repository for the model and to activate continuous integration

[MEMOTE for standardized genome-scale metabolic model testing | Nature Biotechnology]

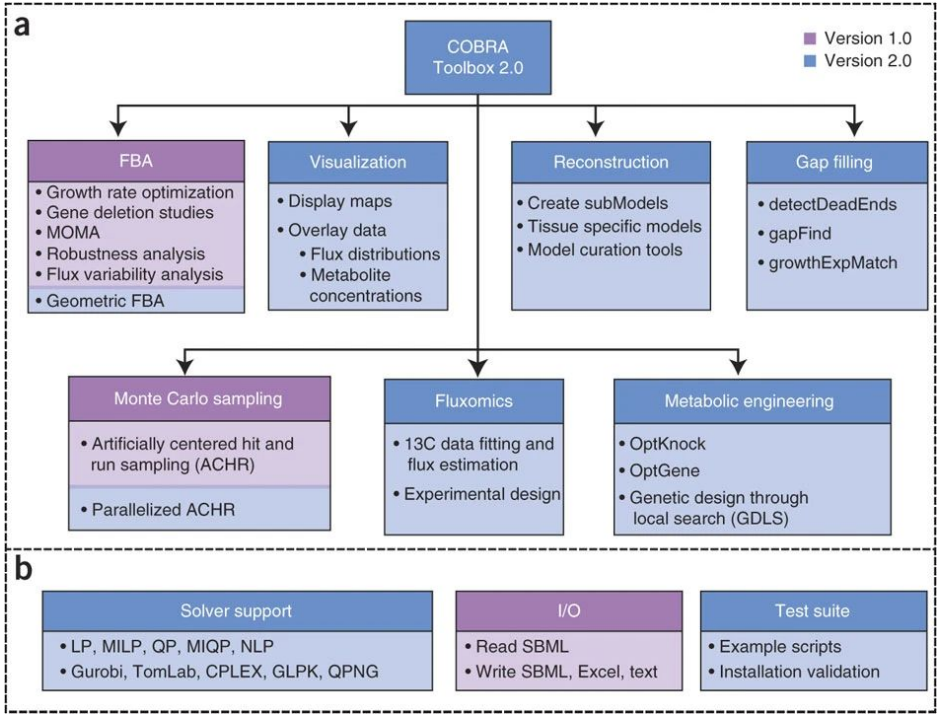
Exploring GEMs

cobrapy

models fluxes algorithms

Complete downstream analyses from a model:

- Inspection
- gap filling
- FBA
- dynamic FBA, etc



Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0 | Nature Protocols

Exploring GEMs

```
G = read_sbml_model('./test_files/aa_seqs.xml')   # Read model
G.reactions[50] # Access reactions
```

```
1  G.reactions[50]
✓ 0.0s
```

Reaction identifier	3M2OPLOXRD
Name	3-Methyl-2-oxopentanoate:lipoamide oxidoreductase(decarboxylating and...
Memory address	0x7f23a80646d0
Stoichiometry	3mop_c + h_c + lpam_c --> 2mbdhl_c + co2_c (S)-3-Methyl-2-oxopentanoate + H+ + Lipoamide C8H15NOS2 --> S-(2-Methylbutanoyl)-dihydrolipoamide + CO2 CO2
GPR	MALA_SAMN05422137_METAG_scaffold_51_gene_40
Lower bound	0.0
Upper bound	1000.0

Exploring GEMs

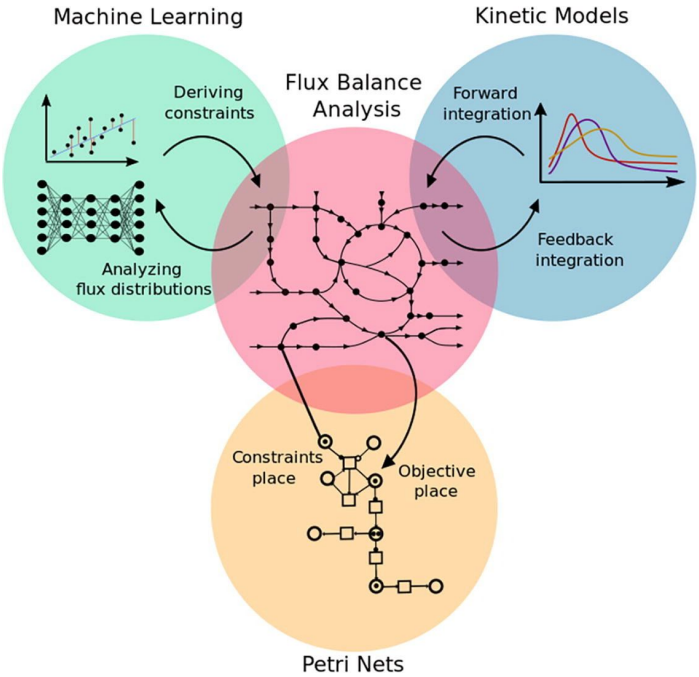
```
G.metabolites.get_by_id("atp_c") # Get by metabolite
```

```
1  G.metabolites.get_by_id("atp_c")
2
```

0.0s Python

Metabolite identifier	atp_c
Name	ATP C10H12N5O13P3
Memory address	0x7f23ad543650
Formula	C10H12N5O13P3
Compartment	C_c
In 368 reaction(s)	2AGPEAT181, ADSK, TRPabc, PE120abcpp, HISHISabcpp, PTRCabc, GLCabc, ASPK, CHOLSabc_1, CYSabc, TSULabcpp, HEMEti, MELIBabc, PG160abcpp, PTRCabcpp, ZN2abcpp, PFK, PTPATi, UAMAGS, 4HPTNCOAK, HISabc,...

Optimization of GEMs - FBA



Advances in flux balance analysis by integrating machine learning and mechanism-based models - ScienceDirect

- Flux balance analysis is the main tool for predicting flux distributions in genome-scale metabolic models. This enables:
  - Modeling context-specific network behavior
  - Growth estimation
  - Gene deletion impact
  - etc

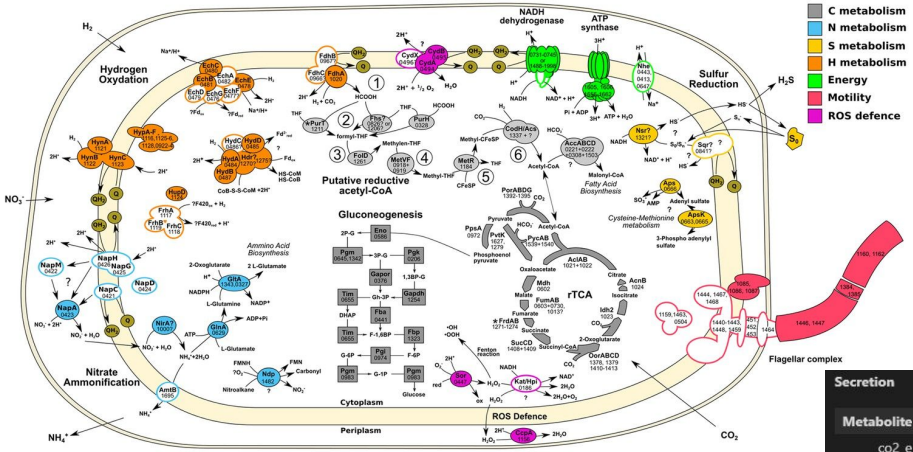
```
1  solution = G.optimize()
2  solution.objective_value
3  G.summary()
```

✓ 0.2s

Objective

1.0 Growth = 61.339459346031354

Optimization of GEMs



- FBA is based in:
  - Uptake of required metabolites
  - Media used to grow the cell

Secretion

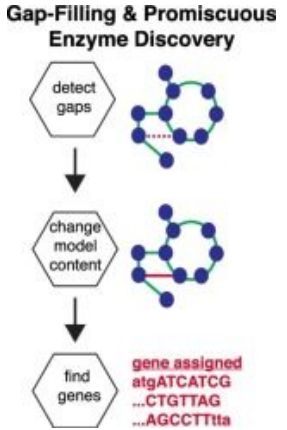
Metabolite	Reaction	Flux	C-Number	C-Flux
co2_e	EX_co2_e	-1000	1	10.41%
h2s_e	EX_h2s_e	-536.3	0	0.00%
h_e	EX_h_e	-471.3	0	0.00%
inost_e	EX_inost_e	-843.5	6	52.69%
nh4_e	EX_nh4_e	-1000	0	0.00%
no2_e	EX_no2_e	-359.9	0	0.00%
ocdcea_e	EX_ocdcea_e	-55.95	18	10.48%
oxa_e	EX_oxa_e	-1000	2	20.82%
pi_e	EX_pi_e	-951.9	0	0.00%
thym_e	EX_thym_e	-75.61	5	3.94%
ura_e	EX_ura_e	-39.69	4	1.65%

Uptake

Metabolite	Reaction	Flux	C-Number	C-Flux
5mcsn_e	EX_5mcsn_e	0.01981	5	0.00%
LalaDgluMdapDala_e	EX_LalaDgluMdapDala_e	6.134	18	0.90%
acgam1p_e	EX_acgam1p_e	12.27	8	0.80%
ade_e	EX_ade_e	12.53	5	0.51%
arg__L_e	EX_arg__L_e	337.1	6	16.55%
asn__L_e	EX_asn__L_e	14.79	4	0.48%
ca2_e	EX_ca2_e	0.3193	0	0.00%
chor_e	EX_chor_e	8.465	10	0.69%
cl_e	EX_cl_e	0.3193	0	0.00%
cobalt2_e	EX_cobalt2_e	0.006134	0	0.00%
cu2_e	EX_cu2_e	0.04349	0	0.00%
fe2_e	EX_fe2_e	0.4119	0	0.00%
fe3dcit_e	EX_fe3dcit_e	0.4789	12	0.05%
fol_e	EX_fol_e	0.04104	19	0.01%
glcur_e	EX_glcur_e	843.5	6	41.41%
glu__D_e	EX_glu__D_e	11	5	0.45%
glyc3p_e	EX_glyc3p_e	1000	3	24.55%
gua_e	EX_gua_e	14.88	5	0.61%
h2o_e	EX_h2o_e	94.09	0	0.00%
hishis_e	EX_hishis_e	2.906	12	0.29%
indole_e	EX_indole_e	3.487	8	0.23%
k_e	EX_k_e	11.97	0	0.00%
lys__L_e	EX_lys__L_e	21.05	6	1.03%
met__L_e	EX_met__L_e	9.427	5	0.39%

Optimization of GEMs - Gap Filling

Reaction gap filling is a computational technique for proposing the addition of reactions to genome-scale metabolic models to permit those models to ‘run’ correctly

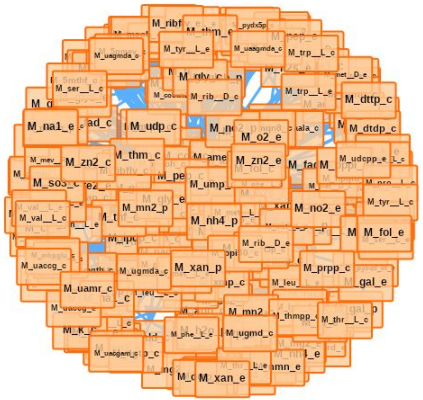


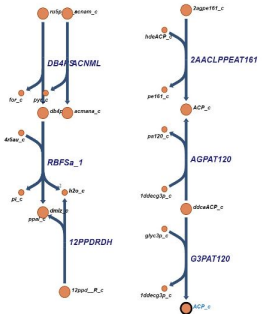
Application examples:

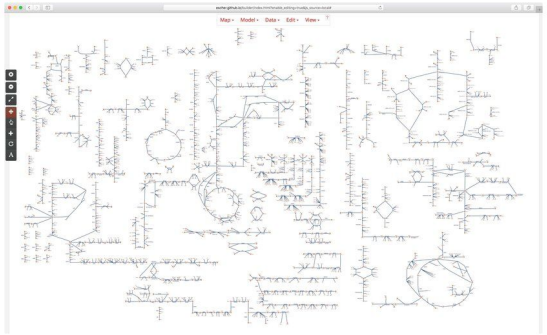
- We know our model organism grows in LB media but the model reports it is not possible:
  - Guide a better gene annotation of the model
    - A gene orphan in function? → enzyme discovery
    - Multifunctional enzymes?
- What does the model organism needs to grow in M9 media?
  - Supplement the media with metabolite X to grow
  - Engineer the model (e.g., adding an enzyme) to metabolize specific compounds

Visualization of pathways

- SBMLDiagrams
  - Welcome to SBMLDiagrams's documentation! — SBMLDiagrams 0.0.1 documentation (sys-bio.github.io)
- Escher Welcome to the documentation for Escher — Escher 1.7.3 documentation
  - Uses JSON files (from cobra)
  - Allows to build from reaction/gene/metabolite and organize
- NetworkX NetworkX — NetworkX documentation → useful for custom networks



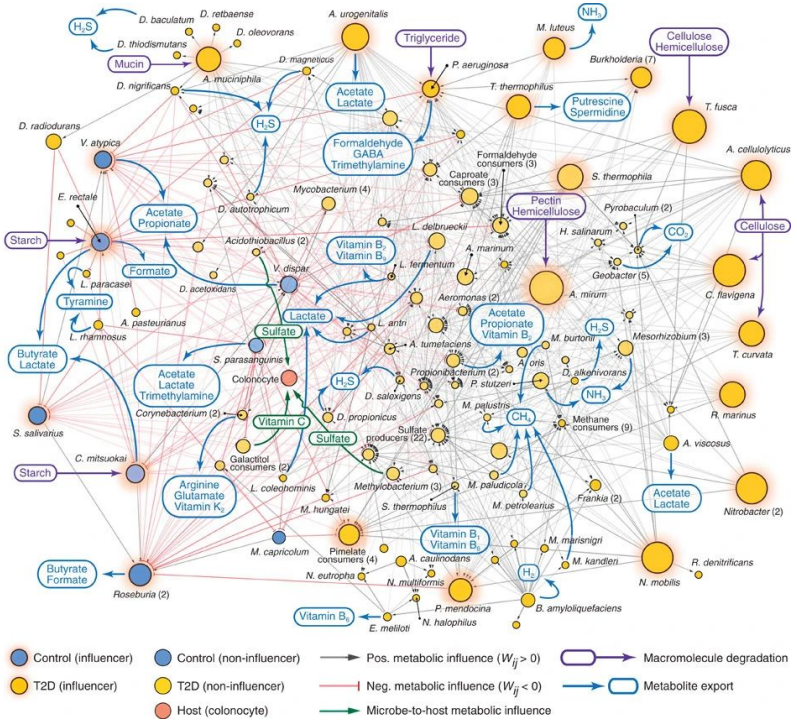




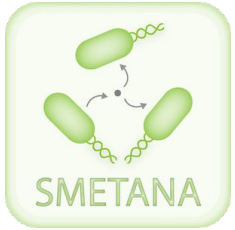
Exploring GEMs

- In the population context, a microbe will interact with the community
  - In some cases, it will be essential for a bacterium to take metabolites from its environment to survive
  - Compounds secreted can be taken into advantage by other individuals
- GEMs can be combined to evaluate these interactions

Global metabolic interaction network of the human gut microbiota for context-specific community-scale analysis | Nature Communications



Metabolic interaction



cdanielmachado/smetana: SMETANA: a tool to analyse interactions in microbial communities (github.com)

- Main concepts:
  - Global level:
    - MRO (metabolic resource overlap): calculates how much the species compete for the same metabolites.
    - MIP (metabolic interaction potential): calculates how many metabolites the species can share to decrease their dependency on external resources.
  - Individual level
    - SCS (species coupling score): measures the dependency of one species in the presence of the others to survive
    - MUS (metabolite uptake score): measures how frequently a species needs to uptake a metabolite to survive
    - MPS (metabolite production score): measures the ability of a species to produce a metabolite
    - SMETANA: the individual smetana score is a combination of the 3 scores above, it gives a measure of certainty on a cross-feeding interaction (species A receives metabolite X from species B).

Metabolic interactions

Global mode just provides the MIP and MRO values for the whole community:

community	medium	size	mip	mro
all complete	8	8	0.600907029478458	

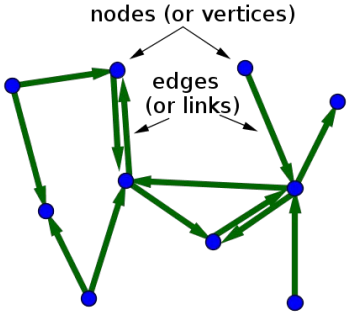
Detailed mode

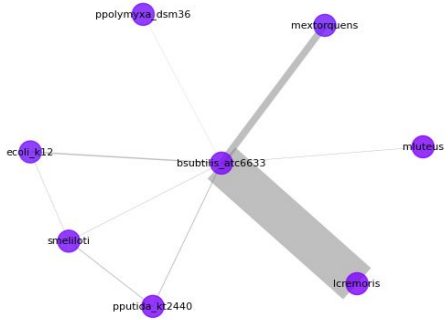
community	medium	receiver	donor	compound	scs	mus	mps	smetana
all	minimal	bsubtilis_atc6633	ecoli_k12	M_alaala_e	1.0	0.01	1	0.01
all	minimal	bsubtilis_atc6633	ecoli_k12	M_arg__L_e	1.0	0.01	1	0.01
all	minimal	bsubtilis_atc6633	ecoli_k12	M_asp__L_e	1.0	0.01	1	0.01
all	minimal	bsubtilis_atc6633	ecoli_k12	M_cellb_e	1.0	0.04	1	0.04
all	minimal	bsubtilis_atc6633	ecoli_k12	M_cgly_e	1.0	0.73	1	0.73
all	minimal	bsubtilis_atc6633	ecoli_k12	M_cu2_e	1.0	1.0	1	1.0
all	minimal	bsubtilis_atc6633	ecoli_k12	M_cys__L_e	1.0	0.13	1	0.13
all	minimal	bsubtilis_atc6633	ecoli_k12	M_fe2_e	1.0	0.01	1	0.01
all	minimal	bsubtilis_atc6633	ecoli_k12	M_fe3_e	1.0	0.97	1	0.97
all	minimal	bsubtilis_atc6633	ecoli_k12	M_fe3pyovd_kt_e	1.0	0.02	1	0.02
all	minimal	bsubtilis_atc6633	ecoli_k12	M_g3pe_e	1.0	0.22	1	0.22
all	minimal	bsubtilis_atc6633	ecoli_k12	M_g3pg_e	1.0	0.43	1	0.43
all	minimal	bsubtilis_atc6633	ecoli_k12	M_glyald_e	1.0	0.05	1	0.05
all	minimal	bsubtilis_atc6633	ecoli_k12	M_glyc3p_e	1.0	0.03	1	0.03
all	minimal	bsubtilis_atc6633	ecoli_k12	M_glyc_e	1.0	0.01	1	0.01

Metabolic interactions

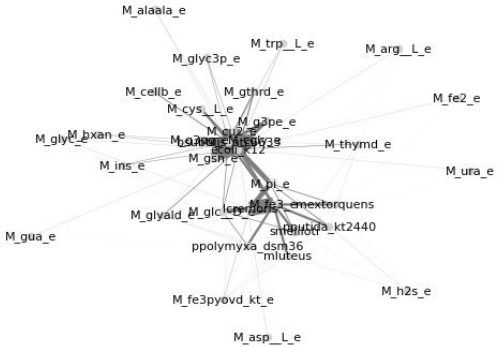
Custom NetworkX scripts to visualize (examples with a 8-members community):

- Networks (nodes and edges) as main representation

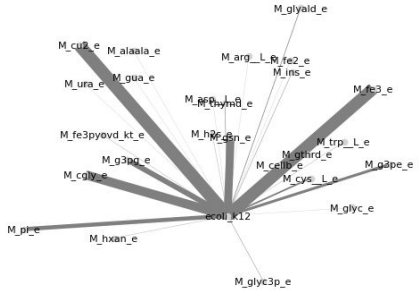




P(interaction) between each species



Donor, receiver and compounds as nodes, P(interaction) as edge weight

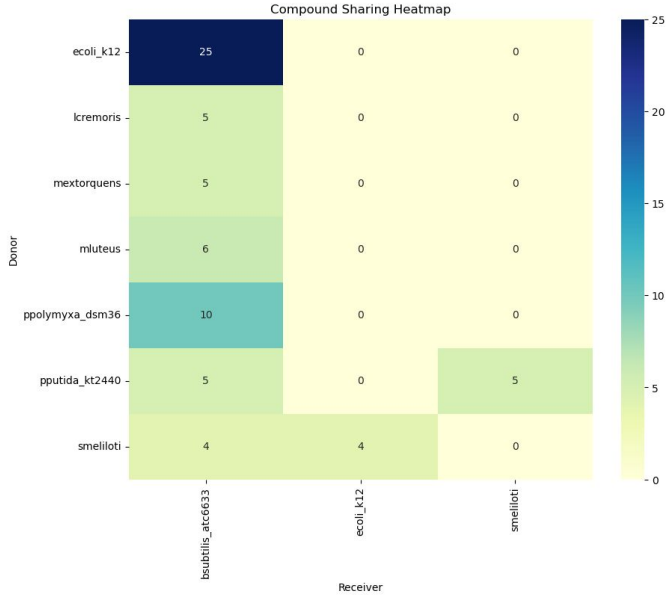


Compounds provided to *B. subtilis* by *E. coli,* edge size by probability

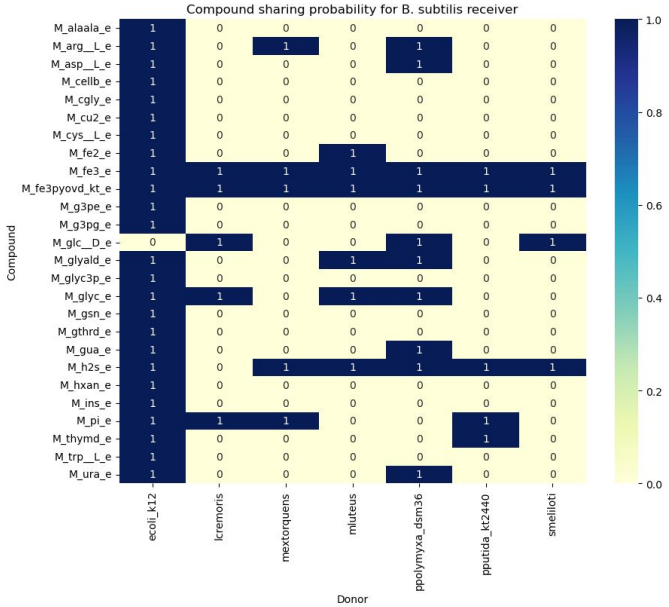
Metabolic interactions

- Heatmaps are also a good way to evaluate interactions:

Do you see any limitation here?

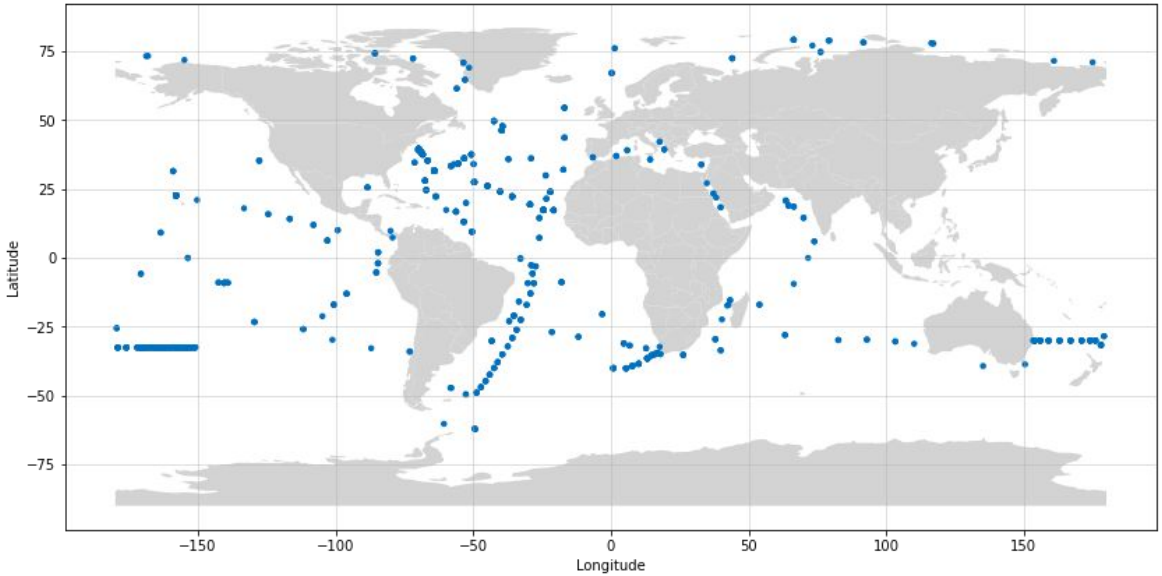


Number of compounds shared in the community by each individual



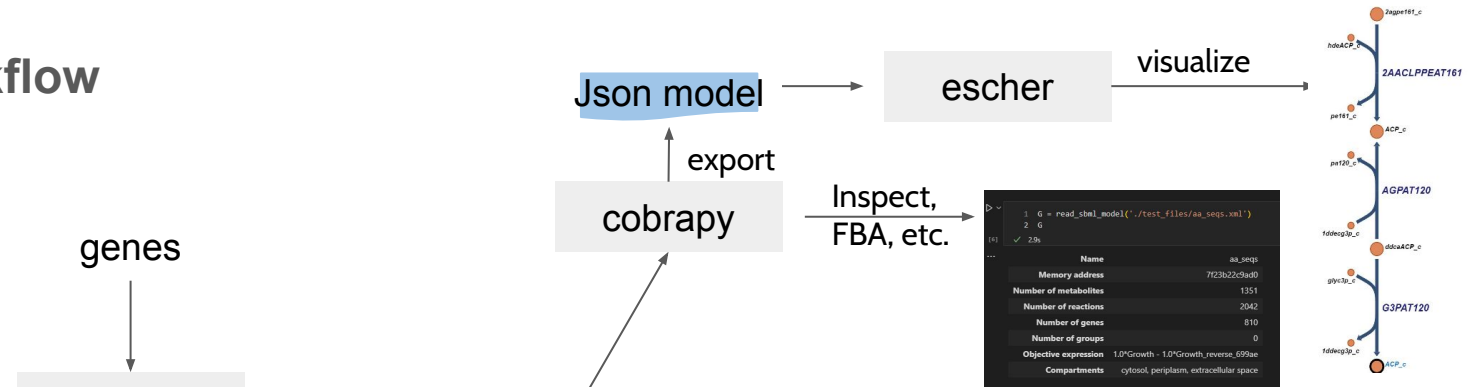
Map of compounds received by *B. subtilis*

Data: the ocean microbiomics database

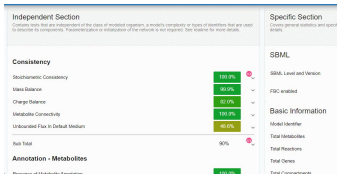


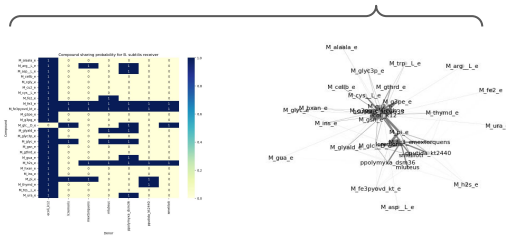
- 1,038 samples
- 34,815 genomes
- ~8,300 taxonomic groups

Workflow


genes
carveme
build
xml model
memote
evaluate
Html report
Json model
export
cobrapy
escher
visualize
Inspect,
FBA, etc.
N models
SMETANA
Metabolic
Interaction
scores
2AACLPPEAT161
AGPAT120
G3PAT120
4CP_c
ddcaACP_c
ACP_c

1 G = read_sbml_model('./test_files/aa_seqs.xml')
2 G
2.9s
Name aa_seqs
Memory address 7f23b22c9ad0
Number of metabolites 1351
Number of reactions 2042
Number of genes 810
Number of groups 0
Objective expression 1.0*Growth - 1.0*Growth_reverse_699ae
Compartments cytosol, periplasm, extracellular space


Independent Section
Specific Section
Consistency
SBML
Basic Information
Annotation - Metabolites



Conclusions

- Genome-scale metabolic models (GEMs) provide a representation of an organism's metabolism, encompassing all known metabolic reactions and associated genes.
- Several tools developed to build, assess, explore and exploit models for research purposes
- Still, several assumptions are taken:
  - Bias towards known gene annotations
  - Bias towards curated models:
    - *E. coli* will be easy and accurate to model
    - Alternative metabolism can be difficult to model
      - Even more if we consider most of these microbes cannot be cultured!
- GEMs can help to identify the conditions required to grow specific organisms
- Metabolic interactions can be predicted highlighting dependencies between individuals in a community such as microbiomes in the gut, skin, soil, ocean, etc.
  - These models cannot represent species abundance