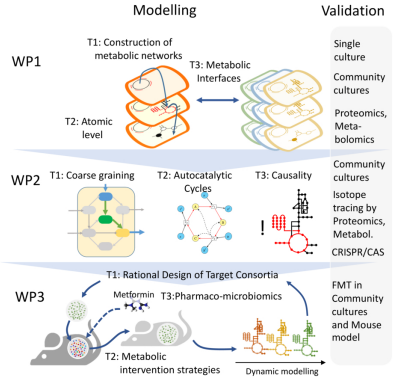
Comparison of Metabolic Networks

Thomas Gatter

10.02.2025

Leipzig University

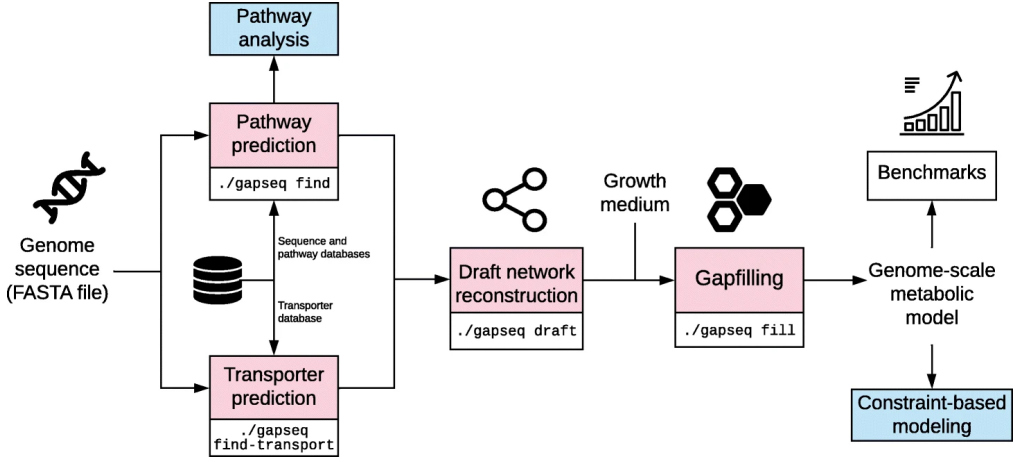
The Ultimate Goal



Model the interactions of **SIHUMix** organisms:

A. cacae L1-92 (DSM 14662), B. thetaiotaomicron (DSM 2079), B. longum (NCC2705), B. producta (DSM 2950), C. butyricum (DSM 10702), C. ramosum (DSM 1402), E. coli (K-12 MG1655), L. plantarum (DSM 20174)

Reconstruction Pipeline of GapSeq



[1] Zimmermann et al. (2021)

SBML Reaction Parsing

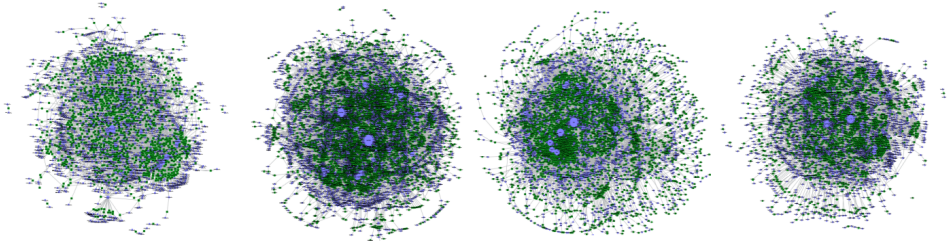
We want to compare the following networks (for each species):

- CarveMe[2] prediction (gapfilled by medium)
- GapSeq[3] prediction (gapfilled by medium)
- AGORA2 (hand-curated) [4] based on KBase prediction[5]

Question: How different are they?

[1] Machado et al. (2018), [2] Zimmermann et al. (2021), [3] Heinken et al. (2023), [4] Arkin et al. (2018)

Option A: Manual Comparison



Networks of different reconstructions of the same organism (E. coli)

[1]SAMMI (https://bioinformatics.mdanderson.org/Software/SAMMI/)

Option B: Experimental Validation

Compare how well models fit to lab measurements:

- Growth rates
- Compound uptake and secretion rates
- Isotope tracing
- ...

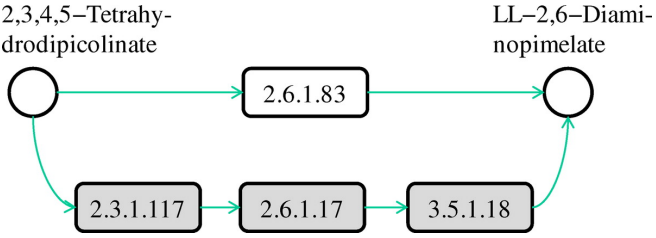
Option C: ID Matching/Matching of Chemical Structure

```
<species metaid="meta_M_h2o_e" sboTerm="SBO:0000247" id="M_h2o_e"
      name="H2O H2O" compartment="C_e" hasOnlySubstanceUnits="false"
      boundaryCondition="false" constant="false" fbc:charge="0"
      fbc:chemicalFormula="H2O">
  <notes>
    <html xmlns="http://www.w3.org/1999/xhtml">
      <p>FORMULA: H2O</p>
      <p>CHARGE: 0</p>
      <p>SBOTerm: SBO:0000247</p>
    </html>
  </notes>
  <annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dcterms="
      <rdf:Description rdf:about="#meta_M_h2o_e">
        <bqbiol:is>
          <rdf:Bag>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-109276"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-113518"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-113519"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-113521"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-141343"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-1605715"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-189422"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-2022884"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-29356"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-351603"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-5278291"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-5668574"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-5693747"/>
            <rdf:li rdf:resource="https://identifiers.org/reactome/R-ALL-8851517"/>
            <rdf:li rdf:resource="https://identifiers.org/kegg.compound/C00001"/>
            <rdf:li rdf:resource="https://identifiers.org/kegg.compound/C01328"/>
```

SBML provides no common ID scheme and no chemical structures!

Tools use different databases / versions!

Option C: ID Matching/Matching of Chemical Structure



Most existing software relies purely on KEGG, e.g. Chen and Hofestädt (2004), Hatzimanikatis et al. (2005), Li et al. (2008), Ay et al. (2011), Abaka et al. (2013), Alberich et al. (2014), Cocco et al. (2021)

Idea: translate to chemical structures whatever we can → abstract → compare

Image: Ay et al. (2011))

How to parse an SBML?

Supported Paths:

- SMBL Compounds → StructRecon → Best InChI * → SBML Reactions
- Seed Reaction → Seed Compounds → InChIKey → PubChem → InChI
- Rhea Reaction → CHEBI Compounds → InChI
- MetaNetX Reaction * → MetaNetX Compounds * → InChI
- KEGG Reaction → KEGG Compounds → Mol → RDK to InChI
- reactome Reaction → reactome Compounds * → CHEBI Compounds → InChI
- BiGG Reaction * → BiGG Compounds * → InChIKey → PubChem → InChI
  or → MetaNetX Compounds * → InChI

* = potential ID links to other databases/structures

For each sbml reaction entry we get list of structurally resolved reactions.

How to process?

Problems...

- directions inconsistent $\rightarrow$ educt, product sides inconsistent
- missing structures (for some databases)
- variation in charges, tautomers, steorechemistry etc
- long query times (10+ online DB requests per reaction)
- networking errors with APIs

Two metrics for equality...

How to process?

When are two reactions identical?

Full Identity

- number of compounds without structure identical
- structures identical
- either direction matches

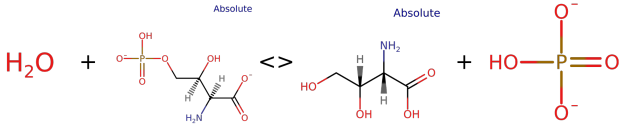
Simplified and Unified

- compounds without structure ignored
- presence of water and protons ignored
- all structures simplified by FICTS [2]
- either direction matches
- subsets matched $\rightarrow$ represented by larger

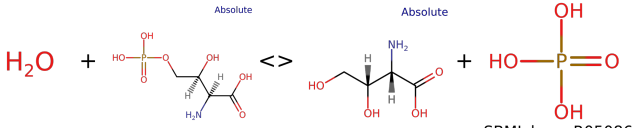
[2] Eriksen et al. (2023)

Example

Fully Identical

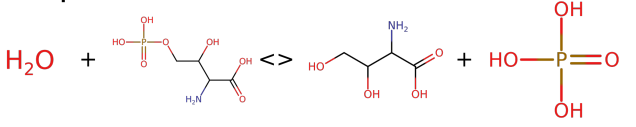


bigg:4HTHRS,rhea:30659,metanetx:MNXR95024,seed:rxn03446



SBML,kegg:R05086

Simplified



SBML,kegg:R05086,bigg:4HTHRS,rhea:30659,
metanetx:MNXR95024,seed:rxn03446

Some example stats

CarveMe reconstruction E. coli with cimIV.

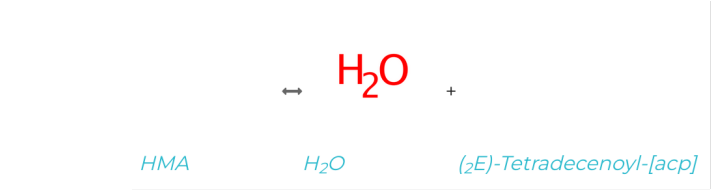
Different reactions per SBML reaction

	Full	Simplified
# Reactions	901	901
mean	3.36	1.2
min	1	1
25%	3	1
50%	3	1
75%	4	1
max	7	3

Database IDs per compacted reaction

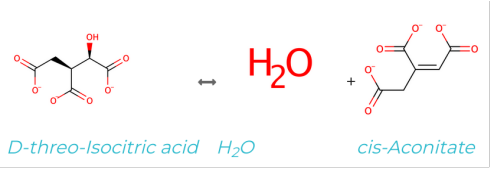
	Full	Simplified
# Reactions	901	901
mean	2.26	7.22
min	1	1
25%	1.2	3
50%	1.5	6
75%	2	6
max	42	126

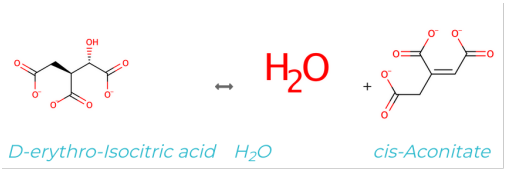
The old annotation problem...

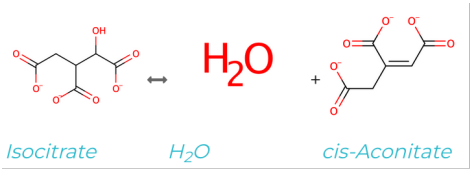


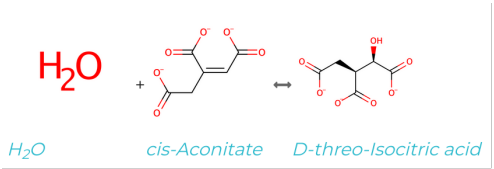
“Duplicates” within same SBML…

e.g. stereochemistry, direction



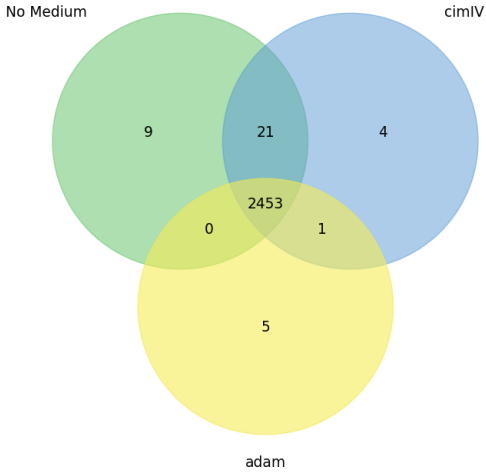






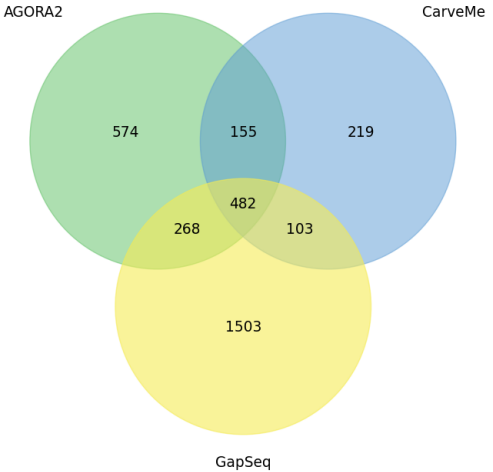
Now how similar are our reconstructions?

E. coli GapSeq reconstructions using no medium, cimIV and adam

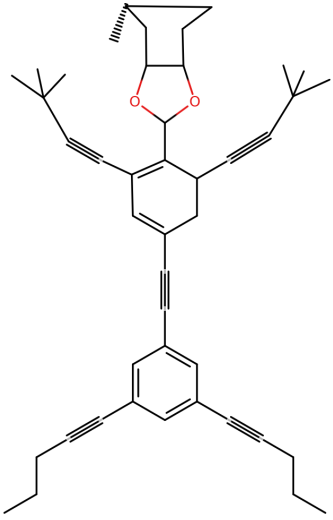


Now how similar are our reconstructions?

E. coli reconstructions of CarveMe, GapSeq and AGORA2 using cimIV



Oh well...



Thank You!

Peter Stadler

Natasha Jorge

Jobs Jobs Jobs!

We have 4 PhD positions available!

- *"RNA-seq-based screening for active self-cleaving ribozymes and transcriptome-wide assessment of their regulatory functions"* (Now!)
- for SPP 2363 *"Utilization and Development of Machine Learning for Molecular Applications"* (Summer/Autumn)
- *"Learning to Explain: Inference of Chemical Reaction Mechanisms"* (Autumn/Winter)
- *"Improving integration of multimodal omics data for precision metabolic health"* (Likely soon)

References

Abaka, G., Bıyıkoğlu, T., and Erten, C. (2013). CAMPways: constrained alignment framework for the comparative analysis of a pair of metabolic pathways. *Bioinformatics*, 29(13):i145–i153.

Alberich, R., Llabrés, M., Sánchez, D., Simeoni, M., and Tuduri, M. (2014). Mp-align: alignment of metabolic pathways. *BMC Systems Biology*, 8:1–16.

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., Glass, E. M., Chivian, D., Gunter, D., Weston, D. J., Allen, B. H., Baumohl, J., Best, A. A., Bowen, B., Brenner, S. E., Bun, C. C., Chandonia, J.-M., Chia, J.-M., Colasanti, R., Conrad, N., Davis, J. J., Davison, B. H., DeJongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J. N., Fang, G., Faria, J. P., Frybarger, P. M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J., Haun, H. L., He, F., Jain, R., Joachimiak, M. P., Keegan, K. P., Kondo, S., Kumar, V., Land, M. L., Meyer, F., Mills, M., Novichkov, P. S., Oh, T., Olsen, G. J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S. S., Price, G. A., Ramakrishnan, S., Ranjan, P., Ronald, P. C., Schatz, M. C., Seaver, S. M. D., Shukla, M., Sutormin, R. A., Syed, M. H., Thomason, J., Tintle, N. L., Wang, D., Xia, F., Yoo, H., Yoo, S., and Yu, D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*, 36(7):566–569.

Ay, F., Kellis, M., and Kahveci, T. (2011). SubMAP: Aligning Metabolic Pathways with Subnetwork Mappings. *Journal of Computational Biology*, 18(3):219–235.

Chen, M. and Hofestädt, R. (2004). PathAligner: Metabolic Pathway Retrieval and Alignment. *Applied Bioinformatics*, 3(4):241–252.

Cocco, N., Llabrés, M., Reyes-Prieto, M., and Simeoni, M. (2021). MetNet: A two-level approach to reconstructing and comparing metabolic networks. *PLOS ONE*, 16(2):e0246962.

Eriksen, C. A., Andersen, J. L., Fagerberg, R., and Merkle, D. (2023). Reconciling Inconsistent Molecular Structures from Biochemical Databases. In Guo, X., Mangul, S., Patterson, M., and Zelikovsky, A., editors, *Bioinformatics Research and Applications*, volume 14248, pages 58–71. Springer Nature Singapore, Singapore. Series Title: Lecture Notes in Computer Science.

Hatzimanikatis, V., Li, C., Ionita, J. A., Henry, C. S., Jankowski, M. D., and Broadbelt, L. J. (2005). Exploring the diversity of complex metabolic networks. *Bioinformatics*, 21(8):1603–1609.

Heinken, A., Hertel, J., Acharya, G., Ravcheev, D. A., Nyga, M., Okpala, O. E., Hogan, M., Magnúsdóttir, S., Martinelli, F., Nap, B., Preciat, G., Edirisinghe, J. N., Henry, C. S., Fleming, R. M. T., and Thiele, I. (2023). Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. *Nature Biotechnology*, 41(9):1320–1331.

Li, Y., De Ridder, D., De Groot, M. J., and Reinders, M. J. (2008). Metabolic pathway alignment between species using a comprehensive and flexible similarity measure. *BMC Systems Biology*, 2(1):111.

Machado, D., Andrejev, S., Tramontano, M., and Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*, 46(15):7542–7553.

Zimmermann, J., Kaleta, C., and Waschina, S. (2021). gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biology*, 22(1):81.