Combinatorial gap-filling for genome-scale metabolic networks using taxonomy and pathway information

Clémence FRIOUX 1,2,3, Sylvain PRIGENT 4, Anne SIEGEL 2,3,1
1 Inria équipe DYLIS 263 avenue du Général Leclerc, 35042 Rennes, Cedex, France
2 CNRS UMR 6074, Rennes, France
3 IRISA, Université de Rennes 1, Rennes, France
4 Chalmers University of Technology, Göteborg, Sweden

**CONTEXT**

- Metabolic networks modeling
  - Compounds and reactions
  - Stoichiometric or Topological (graph-based)

2A + 3B → 2C or A + B → C

- Genome-scale metabolic network reconstruction
  2 steps:
  1) building a **draft** from annotation and orthology
  2) **gap-filling** to make the network able to produce targets (T) from seeds (S1, S2, S3)

**TOPOLOGICAL GAP-FILLING OF METABOLIC NETWORKS**

- Draft network \( R_{\text{draft}} \)
- Scope principle to assess producibility of compounds
  - A metabolite \( m \in M \) is producible if \( m \in \text{scope}_{x_w} \)
  
  \[
  m \in M_{\text{prod}} \quad \text{or} \quad m \in \text{product}(r) \text{ s.t. } r \in R_{\text{draft}}
  \]

- Selection of reactions from a database \( R_s \) such that target compounds (i.e., metabolites known to be produced by the organism) can be reachable from seed metabolites (e.g., growth medium)

How to reconstruct high-quality metabolic networks, particularly for exotic species with few available data?

**TAPATH GAP-FILLING METHOD**

**I. Principle**

Use taxonomy information to select relevant pathways for the production of targets

A pathway is a set of reactions associated to a biological purpose (e.g., synthesis of a compound) and organisms or taxa

**II. Methods**

Answer Set Programming (ASP)
- Programming paradigm
  - Logical expression of problem (constraints)
    - e.g., definition of pathways and reactions
    - distance to studied organism
    - path(y,orthology, organism)
    - reaction(PERSTIVH-RKEN, orthology)
    - reaction \( R_1, R_2 \)
    - pathway \( p \in P \)

**III. Data & distance calculation**

- **Data**
  - Metacyc database pathways are associated to taxa (ETR3)
  - Lineages of ETRs and studied organisms are obtained through NCBI taxonomy browser

- **Distance**
  - Distance of chosen pathways
  - Minimize \( d(P_1, P_2) \in P_{\text{E}} \)
  - Minimize number of chosen pathways
  - Minimize \( h(P) \in P_{\text{E}} \)

**IV. Optimizations**

- Maximize number of produced targets
  - Minimize \( d(P_1, P_2) \in P_{\text{E}} \)
- Minimize number of reactions that do not belong to pathways
  - Minimize \( h(R_s) \in R_{\text{E}} \)

**EXAMPLE OF APPLICATION**

*Acidithiobacillus ferrooxidans* strain Wenelen\(^b\)
- Proteobacterium
- Involved in copper bioleaching activity
- Releases soluble metals of economic interest

Metabolomic study
- Inorganic compounds as primary energy source: \( \text{Fe}^{2+}, \text{SO}_4^{2-}, \text{S}_2\text{O}_3^{2-} \)
- 37 seed compounds
- Identification of metabolites required for biomass production
- 80 target compounds

**Tapath run**
- Union, intersection and enumeration of solutions
- Selected pathways are all associated to bacteria taxa

<table>
<thead>
<tr>
<th>solution</th>
<th>1 solution</th>
<th>intersection</th>
<th>union of 12 solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td># reactions</td>
<td>23 to 24</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td># pathways</td>
<td>7</td>
<td>4</td>
<td>11</td>
</tr>
</tbody>
</table>

**IN PROGRESS**

- Scaling-up of the analyses
  - Run tapath on 3600 completions
  - 40 degraded Escherichia coli networks
  - 90 targets compounds sets

- Comparison of the results with the ones from
  - another topological gap-filling tool: Meneco
  - a stoichiometric gap-filling tool: GapFill or Mirage\(^d\)

- Develop new metrics to assess biological relevance

**REFERENCES**

\(^a\) Progress in clasp Series 3, Gebser et al, LPMN, 2015

\(^b\) Global transcriptional responses of Acidithiobacillus ferrooxidans Wenelen under different sulfate minerals. Latane et al, Bioresource Technology, 2016

\(^c\) Optimization based automated curation of metabolic reconstructions. Kumar et al, BMC Bioinformatics, 2007


**CONCLUSION**

Tapath = decision support tool

- Deciphering taxonomically relevant pathways to gap-fill genome-scale metabolic networks and allow the topological production of targets from seeds.