cMFA for multi-omics data integration in microbial community models

Sthyve TATHO, Inrae/Inria - Bordeaux **Simon LABARTHE**, Inrae - Bordeaux
Valentina BALDAZZI, Inrae/Inria - Sophia Antipolis

Understanding microbial community functions is challenging due to complex interactions and assembly mechanisms ; however, advances in sequencing have enabled the collection of multi-omics data, including population counts and metabolomic or metatranscriptomic data. Our main objective is to develop a mathematical model capable of integrating time series of multiomics data at a community scale.

We introduce the community metabolic flux analysis (cMFA) method, which generalizes metabolic flux analyses (MFA) [1], using a list $\mathcal{M}$ of time series data of experimentally measured production and consumption rates of metabolites and microorganism growth $R_{m,t}$. We aim to infer, for each member of the microbial community, the intracellular distribution of metabolic fluxes. $\nu_{m,t}^{(b)}$ by solving the inference problem :

find, for $t \geq 0$ and for N_b micro-organisms, the rate vectors $(\nu^{\hat{(1)}}(t), \cdots, \nu^{(\hat{N}_b)}(t))$ such that

$$(\nu^{\hat{(1)}}(t), \cdots, \nu^{(\hat{N}_b)}(t)) := \underset{\substack{S^{(b)} \cdot \nu^{(b)} = 0, \text{ for } b = 1, \cdots N_b \\ C_{min}^{(b)} \leq \nu^{(b)} \leq C_{max}^{(b)}, \text{ for } b = 1, \cdots N_b}}{\arg\min} \sum_{m \in \mathcal{M}} \|R_{m,t} - \sum_{b \in \mathcal{B}_m} \mathcal{X}_t^{(b)} \nu_{m,t}^{(b)}\|_2^2 + \lambda \mathcal{R}(\nu) \quad (1)$$

Where $\mathcal{X}_t^{(b)}$ is the density of microorganisms b at time, t and $\mathcal{B}_m$ is the list of microorganisms metabolizing m. This is a high-dimensional constrained linear regression problem, informed by mass conservation constraints ($S^{(b)} \cdot \nu^{(b)} = 0$) and metatranscriptomic data, encoded in the penalty term $\mathcal{R}(\nu)$. The difficulty here is in accurately inferring latent internal rates from a few observations of exchange fluxes.

We evaluated the cMFA method on synthetic data from dynamic models of increasingly complex microbial communities, based on metabolic models of different mutants of *Escherichia coli* using dynamic flux balance analysis (dFBA) [2]. Synthetic metatranscriptomic data were obtained from internal metabolic fluxes in the dynamic model. Different regularization terms $\mathcal{R}$ were tested, including different levels of sparsity, for the selected penalty weight λ. To evaluate the robustness of the method, multiple benchmarks were tested. These included assessments of the robustness of the method to data noise, incomplete meta-transcriptomic data, and inaccurate prior knowledge of metabolic import rates. Currently, we are working with real data and expanding the study to a larger microbial community.

Keywords : Dynamic systems, applied mathematics, inference, biological systems

[1] L. Gerken-Starepravo, X. Zhu, B. A. Cho, F. Vega-Ramon, O. Pennington, E. Antonio del Río-Chanona, K. Jing, D. Zhang. *An miqp framework for metabolic pathways optimisation and dynamic flux analysis*. Digital Chemical Engineering, **2**, 100011, 2022. doi :https://doi.org/10.1016/j.dche.2022.100011.

[2] R. Mahadevan, J. S. Edwards, F. J. Doyle. *Dynamic flux balance analysis of diauxic growth in escherichia coli*. Biophysical journal, **83(3)**, 1331–1340, 2002.

Contact : sthyve.tatho@inria.fr