Metage2Metabo PostaViz: a Python package for exploring, visualising, and comparing the metabolic potential of microbial communities

Léonard Brindel [1] **and Clémence Frioux** [1]¶

1 Inria, Univ. Bordeaux, INRAE, F-33400 Talence, France ¶ Corresponding author

Summary

Microbial communities consist of up to thousands of distinct microbial populations, each characterized by its genomic DNA sequences, and all sharing a habitat and environmental conditions. The word *microbiome* describe the holistic concept associated all the previous components (Berg et al., 2020). Characterizing the populations in samples and understanding both their roles of and interactions within the microbiome require a combination of experimentation, high-throughput data acquisition and computational models (Klitgord & Segrè, 2011). The role of microbial populations can be abstracted by the study of their metabolism, encompassing all biochemical reactions that cells may perform. The genome encodes the genetic information associated to the metabolism, and can therefore be queried to estimate the metabolic capabilities of microorganisms, represented as metabolic networks. Because such metabolism is highly redundant, depends on the environment, and because of the size of microbial communities, computational models are needed and predict, from the networks, the possible behaviours of and interactions within microbial populations in given environmental conditions. Comparing the outcomes of such predictive models in multiple samples further increases the difficulty of integrating results into actionable hypotheses. Visualisation and integration of metadata can help perform such comparison. They need to be as customisable as possible to facilitate exploration by end-users.

Statement of need

Metabolic modelling is widely used to determine the roles of microorganisms in microbial communities (Cerk et al., 2024): what molecules they can produce, which interactions are likely to happen depending on the environmental conditions… Many models rely on integer linear programming optimisations (García-Jiménez et al., 2021) and some alternatives use Boolean abstractions (Frioux et al., 2018) (Belcour et al., 2020) or probabilistic approaches (Bernstein et al., 2019) to provide predictions on community behaviours. Key questions are scalability to large communities, and also integration of these predictions when analysing many samples or community compositions. Each microorganism is abstracted by a collection of biochemical reactions it may perform according to its genomic information, thus forming a network connecting transformed molecules. A community of microorganisms is therefore represented as a collection of such networks, referred to as genome-scale metabolic networks. A metabolic model will provide predictions on the possible behaviour of microorganisms and communities in defined simulation conditions.

`Metage2Metabo-PostaViz` (M2M-PostAViz) is a Python package that performs analyses on the predictions generated by the metabolic-modelling tool Metage2Metabo (M2M) (Belcour et al., 2020). M2M screens the metabolic potential of a microbial community represented as a collection of genome-scale metabolic networks. When working with cohorts of hundreds or

42 thousands of samples, one has to run the tool as many times as there are samples, then analyse
43 the results of the model. The tool's outputs are, for each community, several data frames
44 describing the role of each microorganism with respect to the whole community's functions.
45 Properly comparing all samples requires combining all the outputs, and taking into account
46 sample metadata describing individuals lifestyle or clinical information for instance.

47 `M2M-PostaViz` integrates all such data and provides a visualisation interface that permits
48 exploration through custom plot generation and statistical tests. The underlying data treatment
49 was optimised in order to deal with large numbers of samples without impeding user experience.
50 `M2M-PostaViz` notably permits a pre-treatment and storage of the data such that future
51 exploration can be launched in a computationally efficient manner. Exploration is performed
52 at several levels: molecules (metabolites) that may or may not be producible across samples,
53 microorganisms that may have different behaviours across samples depending on interactions
54 with other community members, or more general overviews of the community functions. The
55 tool works as a local web-based application.

Outlook

Data integration and storage

58 Each sample has to be run in M2M prior analysis. Inputs to the application are a set of
59 directories generated by M2M for each sample, composed of several tabulated and json files.
60 Additional inputs include relative abundance of microbes in each sample, used to weigh the
61 predictions, metadata associated to the samples and possibly metadata related to molecule
62 descriptions and microbial taxonomy of the corresponding metabolic networks present in the
63 communities.

64 Reading and integrating all data is computationally demanding when considering several
65 hundreds or several thousands of samples (and M2M outputs directories). Taking into account
66 that users are likely to explore the same data across several runs of the application, efforts
67 were done to efficiently store the required and processed data such that only the first run takes
68 time and the future ones directly load pre-processed data. The parquet format is used to store
69 all this information as a database limiting the use of memory for accesses.

Application content

71 The application opens as a multi-tab browser page where the first one is an overview of the
72 data that summarises it and enables several first analyses customisable by variables of metadata.
73 Two tabs provide analyses centered of microbe roles and molecules respectively (see below)
74 and a last one summarises the metadata and permits customising variable types to fine-tune
75 analyses and plots.

Exploration of microorganism roles

77 The second tab of the application focuses on the role of microorganisms in the production
78 of metabolites across samples. The same species, and thus the same metabolic network can
79 appear in several samples, but behave differently because of interactions with other microbial
80 populations. In addition, microbial species can be grouped according to their taxonomy,
81 enabling to consider not only the role of a metabolic network but the role of all those falling
82 in taxonomic groups of different levels (phylum, family, genera…). Again, sample metadata
83 variables can refine the analysis, enabling users to compare groups or filter certain samples.

Exploration of metabolite production across samples

85 The third tab of the application is dedicated to the analysis the metabolites, that can also be
86 groups of families if a proper ontology is provided. Focusing on metabolites enables a targeted

analysis in order to compare samples or groups of samples on specific metabolic functions.

Acknowledgements

References

Belcour, A., Frioux, C., Aite, M., Bretaudeau, A., Hildebrand, F., & Siegel, A. (2020). Metage2Metabo, microbiota-scale metabolic complementarity for the identification of key species. *eLife*, *9*, e61968. https://doi.org/10.7554/elife.61968

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C. C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G. H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J. A., Maguin, E., Mauchline, T., McClure, R., … Schloter, M. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, *8*(1), 103. https://doi.org/10.1186/s40168-020-00875-0

Bernstein, D. B., Dewhirst, F. E., & Segrè, D. (2019). Metabolic network percolation quantifies biosynthetic capabilities across the human oral microbiome. *eLife*, *8*, e39733. https://doi.org/10.7554/elife.39733

Cerk, K., Ugalde-Salas, P., Nedjad, C. G., Lecomte, M., Muller, C., Sherman, D. J., Hildebrand, F., Labarthe, S., & Frioux, C. (2024). Community-scale models of microbiomes: Articulating metabolic modelling and metagenome sequencing. *Microbial Biotechnology*, *17*(1), e14396. https://doi.org/10.1111/1751-7915.14396

Frioux, C., Fremy, E., Trottier, C., & Siegel, A. (2018). Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. *Bioinformatics*, *34*(17), i934–i943. https://doi.org/10.1093/bioinformatics/bty588

García-Jiménez, B., Torres-Bacete, J., & Nogales, J. (2021). Metabolic modelling approaches for describing and engineering microbial communities. *Computational and Structural Biotechnology Journal*, *19*, 226–246. https://doi.org/10.1016/j.csbj.2020.12.003

Klitgord, N., & Segrè, D. (2011). Ecosystems biology of microbial metabolism. *Current Opinion in Biotechnology*, *22*(4), 541–546. https://doi.org/10.1016/j.copbio.2011.04.018