

Towards integrated rhizosphere microbial community modeling through bottom-up COMETS genome-scale metabolic simulations and top-down data analysis

Ilija Dukovski^{4*} (dukovski@bu.edu), Alan Pacheco⁴, Dileep Kishore⁴, Elysha Sameth⁴, Jing Zhang⁴, Nathaniel Borders⁴, Peter Kim³, Kateryna Zhalnina¹, Peter Andeer¹, **Daniel Segrè**⁴ (dsegre@bu.edu), Adam Deutschbauer^{1,2}, and Trent Northen¹.

¹Lawrence Berkeley National Laboratory, Berkeley; ²University of California, Berkeley; ³Sandia National Laboratory, Livermore; ⁴Boston University, Boston.

<http://mCAFEs.lbl.gov> , <https://runcomets.org>

Project Goals: Understanding the interactions, localization, and dynamics of grass rhizosphere communities at the molecular level (genes, proteins, metabolites) to predict responses to perturbations and understand the persistence and fate of engineered genes and microbes for secure biosystems design. To do this, advanced fabricated ecosystems are used in combination with gene editing technologies such as CRISPR-Cas and bacterial virus (phage)-based approaches for interrogating gene and microbial functions *in situ*—addressing key challenges highlighted in recent DOE reports. This work is integrated with the development of predictive computational models that are iteratively refined through simulations and experimentation to gain critical insights into the functions of engineered genes and interactions of microbes within soil microbiomes as well as the biology and ecology of uncultivated microbes. Together, these efforts lay a critical foundation for developing secure biosystems design strategies, harnessing beneficial microbiomes to support sustainable bioenergy, and improving our understanding of nutrient cycling in the rhizosphere.

In order to disentangle and control the inter-microbial processes that shape rhizosphere communities, we have been developing multiple complementary computational approaches that can be iteratively refined through comparisons with experimental measurements. At the heart of our efforts to reach a detailed mechanistic understanding of how intracellular circuits shape microbe-microbe and microbe-plant interactions, is the use of genome-scale stoichiometric reconstructions of microbial metabolism, simulated using dynamic flux balance analysis (dFBA). In particular, we have been enhancing our comprehensive, open-source modular software for Computation Of Microbial Ecosystems in Time and Space (COMETS). COMETS extends dFBA to generate simulations of multiple microbial species in spatially structured environments. Notably, in COMETS, interactions between species arise as an emergent property of the metabolic dynamics of individual species. The newest version of COMETS (available at runcomets.org)¹ incorporates numerous innovative components that will greatly facilitate the construction of rhizosphere community models. These include a more accurate biophysical model of microbial biomass expansion, newly designed modules for evolutionary dynamics and extracellular enzyme activity, user-friendly Python and MATLAB interfaces, as well as comprehensive documentation and tutorials. Using COMETS, we have started constructing a

geometrically and molecularly accurate *in silico* version of a microfluidics device that creates rhizosphere relevant gradients (the μ EcoFAB) This COMETS model will be used to predict the environmental dependence of community growth within the μ EcoFAB. We have further used COMETS to implement an evolutionary algorithm that can search through vast spaces of environmental conditions to find those able to support growth and maintenance of communities with desired taxonomic composition or functional properties². We envisage that this approach will identify community-level interventions (which can be achieved through CRISPR-Cas editing tools) likely to affect the community in a desired way.

In parallel to advancing the mechanistic approach summarized above, we have developed new data-driven methods for analyzing high-throughput genomics data. These allow us to generate co-occurrence networks that can be compared with mechanistic predictions. In particular, we have revisited the set of steps necessary for translating 16S ribosomal RNA (16S rRNA) amplicon datasets into networks of inter-microbial associations - a process that is prone to statistical artifacts. By performing a careful analysis of each step in this pipeline, we found that certain choices of algorithms and parameters affect the co-occurrence network more dramatically than others. We further generated an estimate of the most accurate and robust tools for inferring co-occurrence networks based on comparison with mock and synthetic datasets. Our new pipeline (available at <https://github.com/segrelab/MiCoNE>)³, can therefore provide a default avenue for generating standardized co-occurrence networks, with the option of easily exploring the variations induced by alternative choices of tools and parameters. This pipeline can be used for integrating multiple datasets, as well as for generating networks that will help us identify community hubs on which to further focus predictive modeling of the outcome of rhizosphere perturbations.

References:

1. I. Dukovski, D. Bajić, J. M. Chacón, M. Quintin, J. C.C. Vila, S. Sulheim, A.. R Pacheco, D. B. Bernstein, W. J. Rieh, K. S. Korolev, A. Sanchez, W. R. Harcombe, D. Segrè: Computation Of Microbial Ecosystems in Time and Space (COMETS): An open source collaborative platform for modeling ecosystems metabolism: arXiv 2020, <https://arxiv.org/abs/2009.01734>
2. A. Pacheco, D. Segrè: An evolutionary algorithm for designing microbial communities via environmental modification, BioRxiv (2020), <https://doi.org/10.1101/2020.11.25.398644>
3. D. Kishore, G. Birzu, Z. Hu, C. DeLisi, K Korolev, D. Segrè: Inferring microbial co-occurrence networks from amplicon data: a systematic evaluation, BioRxiv (2020), <https://doi.org/10.1101/2020.09.23.309781>

Funding statement: This material by m-CAFEs Microbial Community Analysis & Functional Evaluation in Soils, (m-CAFEs@lbl.gov) an SFA led by Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231