Jena School for Microbial Communication
Graduate School of Excellence

PhD project title: Integrative analysis of metabolomic and transcriptomic data to assess the biosynthetic capacity of organisms in microbial communities

PhD Student: Mahnoor Zulfiqar
Supervisors: Prof. Dr. Christoph Steinbeck

Abstract:
The Metabolome of the microbial consortia reflects the set of species within the consortia and the state of these communities. The prime focus of the Microverse Cluster is to shape and determine the consortia of microbes through their endogenous and exogenous collection of Metabolites. These metabolic entities are known to be involved in the regulation of certain molecular pathways and also serve as molecular switches supervising the outcome of the communities.

The challenge within this realm of research is that most of the metabolites exuded and sorted among the neighboring interacting communities of the consortia remain unknown. From potentially thousands of metabolite signals from organisms, less than 30% are known and are part of chemical databases. Rest of 70% still remain unknown. However, certain signals propose their existence and suggest a set of physicochemical properties, their molecular formula and high resolution mass. Cheminformatics analysis of these unknown entities leads to another daunting insight and that is the presence of many structural/ constitutional isomers. So, for example, Sesquiterpene, with a molecular formula as simple as $C_{15}H_{22}O$, has more than 174 Million structural isomers and all of these isomers can be potential candidates to be the part of the metabolite space of living organisms. And out of the 174M compounds only 6500 or 0.004% of the structures are known. Further more, the constitutional space of sesquiterpene, clusters into a group of compounds with a high molecular similarity which makes it even more complicated. With respect to the Quantitative Structure Property Relationship (QSPR) analysis, this results in high spectral similarity and it becomes difficult to predict the true compound that is being utilized by the consortia [1]. Previously, we have demonstrated that living organisms have a limited fraction of theoretically possible chemical structures within the biochemical repertoire [2]. Methods to predict the biochemical repertoire of living organisms to explore the interactable spaces among the communities, has been investigated but is limited to specialized cases like polyketides.

Here, we propose to examine the Multi-partner consortia of bacteria and algae model strains by Pohnert Group via the integration of metabolomic and transcriptomic data in order to predict the virtual metabolome of the consortia. Through the applications of enzymatic reactions and their influence on the core metabolome of the organisms, we will be able to predict their biochemical repertoire. Since these model systems are developed on the basis of not only the bilateral and multilateral interactions but also depends on the external stressors such as bacteria, so we can also suggest their influence on the transcriptomic response and in turn the transcriptomic impact on the consortia metabolome as well. This routine however, will not only be limited to aforementioned model systems but can also be examined on other model and non-model consortia, as for example investigated in the Küsel and the Baldwin group.

In this PhD project, we lay the foundation of future research on the investigation of communications between the complex consortia of living organisms, to initiate iterative cycles of theoretical analyses and experimental investigations that uncover common functional principles in microbial communities across various habitats. The project constitutes of development of customized algorithms for quantitative data analysis and predictive modeling (Objective 1). To make inference regarding the understanding of communications among various microbial communities through integrative data analysis for model system (Objective 2). Insights to the molecula pathways involved in the on-demand production of metabolites which leads to the identification of principles for organization, functional aspects, stability and dynamics of the microbial communities (Objective 3).

References:

[1] Jayaseelan KV, Steinbeck C: Building blocks for automated elucidation of metabolites: natural product-likeness for candidate ranking. BMC Bioinformatics 2014, 15:234.
[2] Jayaseelan KV, Moreno P, Truszkowski A, Ertl P, Steinbeck C: Natural product-likeness score revisited: an open-source, open-data implementation. BMC Bioinformatics 2012, 13:106.