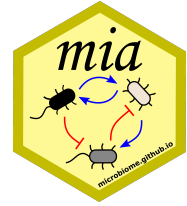


Orchestrating Microbiome Analysis with Bioconductor



This proposed talk will introduce how microbiome data science is conducted in R/Bioconductor. It is well known and has a good reputation for providing high-quality, open-source methods which is due to the fact that all the packages must fulfill the rather strict Bioconductor standards. All the packages undergo peer-review and automatic testing to ensure the robustness and correctness of the methods. In addition to being a platform for bioinformatics software, Bioconductor is a worldwide community of developers and practitioners which is a unique and important feature of Bioconductor. In the beginning of the talk, the Bioconductor is briefly introduced.

The software design in Bioconductor is based on data containers. They allow creation of modular and efficient workflows. Previously, and still today, phyloseq is a popular data container in the microbiome field. However, it was designed almost 15 years ago for 16S data. Today's – and the future's – microbiome data science has different requirements than a decade ago. In many studies, researchers integrate different omics, and the datasets have grown past then, which make phyloseq a suboptimal solution. The TreeSummarizedExperiment data container extends the common SummarizedExperiment data container to the microbiome field. By utilizing shared data container across different fields in Bioconductor, users can more easily integrate datasets from metagenomics to metabolomics. Additionally, this design choice fosters collaboration between fields and catalyzes innovations. The second part of the talk will introduce the idea of data containers.

To analyze and visualize the data in TreeSummarizedExperiment format, we have developed mia (Microbiome Analysis) family of packages. These provide common operations in microbiome data science such as tools for wrangling the data and calculating diversity. The toolset is further enriched by individual developers. In addition to methods, data from multiple different databases can be directly retrieved in TreeSummarizedExperiment format. The main part of the talk will show an example workflow where we fetch data from an open database and apply common data operations.

Bioconductor has a strong emphasis on documentation. Following the Bioconductor's best principles, we have developed an online book called Orchestrating Microbiome Analysis with Bioconductor (OMA). The idea of the book is not only to show how to apply methods but also to disseminate the best practices in the field. These best practices are formed through collaborative development which is why we welcome everyone to contribute to the discussion in our discussion forums which also provide support for users. The talk will be ended by introducing OMA and the discussion forums. Our aim is to encourage people to contribute to this project. There are many ways to contribute, and we believe that collaborative development is the best way to advance this rapidly evolving field.

Relevant links:

Project homepage: <https://microbiome.github.io/>

OMA: <https://microbiome.github.io/OMA/docs/devel/>