Microbiome data science with R/Bioconductor

Oulu, June 2024

2024-10-20

Contents

Chapter 1

Overview

1.1 Contents and learning goals

Contents and learning goals: This course provides an introduction to microbiome data science with R/Bioconductor, a popular open source environment for scientific data analysis. A special emphasis is given to multi-omic data integration methods. After the course you will know how to organize multiple data sources into a coherent framework, implement reproducible data science workflows, and approach common data analysis tasks by utilizing available documentation and R tools. Whereas the primary focus is on microbiome research, the covered data science methods are generally applicable and we will discuss links with other application domains such as transcriptomics, metabolomics, and single cell sequencing.

Target audience: MSc students, PhD, postdoctoral, and other researchers who wish to learn new skills in statistical programming and data analysis. Academic students and researchers from Finland and abroad are welcome and encouraged to apply.

Teaching material: We will follow open online documentation created by the course teachers, primarily the Orchestrating Microbiome Analysis (OMA) book. The training material walks you through the standard steps of omics data analysis covering data access, exploration, analysis, visualization, and reproducible workflows. Preparatory material and video clips, and online support are available before the course. All teaching materials are shared openly.

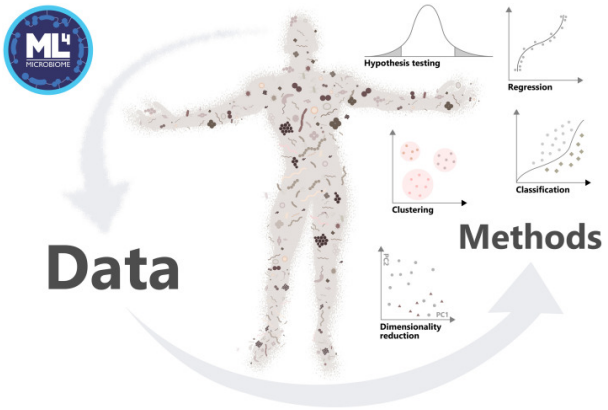

Figure 1.1: Figure source: Moreno-Indias *et al.* (2021) *Frontiers in Microbiology* 12:11.

1.2 Schedule

Venue: University of Oulu. December 18-20, 2024 (Wed-Fri). The course is organized in a live format; no remote option available.

Costs: Registration is free. Participants are expected to cover their own travel and accommodation.

Accommodation: Housing tips can be found at https://visitoulu.fi/en/arrival-overnight/.

Schedule: Contact teaching daily between 9am – 5pm, including lectures, demos, practicals, and breaks. For a detailed schedule, see Section 2. The course can be extended by an independent assignment (details to be agreed with the main teacher).

1.3 How to apply

- Send a brief motivation letter to Anna Kaisanlahti anna.kaisanlahti@oulu.fi
- Applications from local students, and applications sent before Nov 15 will be given priority
- The course has maximum capacity of 20 participants.

1.4 Teachers and organizers

Teachers: Leo Lahti is the main teacher and Professor in Data Science at the University of Turku, and a certified Carpentries Instructor. PhD researcher *Tuomas Borman* is a co-teacher and main developer of the data science framework used in the course. *Anna Kaisanlahti* (Oulu) is a course assistant, and Docent *Justus Reunanen* is the course coordinator. The course is organized by Health and Biosciences Doctoral Programme (HBS-DP) University of Oulu Graduate School, Research Unit of Translational Medicine, University of Oulu. We thank the Finnish IT Center for Science (CSC) supports the course by providing cloud computing services.

This is a Bioconductor course Drnevich et al. (2024) and we follow the best practices recommended by Software carpentries.

1.5 Code of Conduct

The Bioconductor community values an open approach to science that promotes the

- sharing of ideas, code, software and expertise
- open collaboration and community contributions
- diversity and inclusivity
- a kind and welcoming environment

More details on its enforcement are available here.

1.6 Acknowledgments

Citation: We thank all developers and contributors who have contributed open resources that supported the development of the training material. Kindly cite the course material as Tuomas Borman and Leo Lahti (2024).

Contact: Refer to https://microbiome.github.io.

License and source code:

Chapter 2

Program

The course takes place daily from 9am – 5pm (CEST), including coffee, lunch, and short breaks. Most of the time will be dedicated to practical exercises, complemented by short lectures and demos.

We expect that participants will prepare for the course in advance. Instructions will be sent to the registered participants. Online support is available.

The material follows open online book created by the course teachers, Orchestrating Microbiome Analysis, which supports R/Bioconductor framework for multi-omic data integration and analysis.

Figure source: Moreno-Indias *et al.* (2021) Statistical and Machine Learning Techniques in Human Microbiome Studies: Contemporary Challenges and Solutions. Frontiers in Microbiology 12:11.

2.1 Day 1 - Open data science

Reproducible workflows with R/Bioconductor and Quarto

Morning

10-11 Coffee, Welcome & Practicalities

11-12 Learning environment (CSC RStudio notebook and reproducible reporting with Quarto)

12-13 Lunch break

Afternoon

13-14 Lecture: open data science

14-16 Working with data containers and workflows

16-17 Q & A

2.2 Day 2 - Tabular data analysis

Morning

9-10 Lecture: analysis & visualization of *tabular data* (single omics)

10-12 Data wrangling, exploration, and summaries

12-13 Lunch break

Afternoon

13-14 Univariate data analysis and visualization

14-16 Multivariate data analysis and visualization

16-17 Q & A

Evening

Course dinner (optional; own cost)

2.3 Day 3 - Multi-assay data integration

Morning

9-10 Lecture: analysis & visualization of *multi-assay data* (multi-omics)

10-12 Multi-assay data analysis and visualization

12-13 Lunch break

Afternoon

13-15: Advanced methods (e.g. time series, machine learning, simulation)

15-16: Summary and wrap-up

16-17: Q & A

Chapter 3

Venue

3.1 Tips for visiting Oulu

Tourist info/city website can be found here.

3.2 Arrival to Oulu

Airplane: Airport is located in Oulunsalo, approx. 15km distance away from Oulu city center and the course venue. From the airport it is possible to take a buss (lines 8 and 9) or taxi to Oulu.

Train: Train station is located close to the city center (address: Rautatienkatu 11). The train operator, schedules and tickets are available through VR website.

3.3 Public transport

- General information is available on the ouka website
- Routes and timetables
- When traveling by bus you can buy your ticket directly from the bus via contactless payment with debit/credit card, or in advance through either mobile ticket application or ticket machine:
  - **Contactless payment**. You can use your debit or credit card (Visa, Visa Electron, Mastercard and Eurocard) or mobile (Google Pay and Apple Pay) to pay for your fare. You can use contactless card or device to pay for your own travels only.
  - **Mobile ticket** (application named Waltti Mobiili): Install Waltti Mobiili application and add your debit/credit card into it, make sure you have enabled online payments. Available payment methods include Visa, Visa Electron or Mastercard. You do not need to register to buy tickets. When you make your first purchase, choose region "Oulu". Show the ticket to the driver when you board.
  - At the **Waltti ticket machines**, you can buy single tickets as well as add more value and seasons to your Waltti travel card. You can find a Waltti ticket machine in example at the following locations:
    - Valkea Shopping Centre, Kesäkatu (next to the ATM)
    - OYS, N-entrance (formerly A3: 1) - photo
    - Oulu Airport, Arrivals Hall (next to screens)

3.4 Electric boards and bikes

Below are listed companies offering electric boards and bikes in Oulu area with their websites with instructions for use and rental:

- Voi
- Tier
- Lime

3.5 Accommodation

Hotel information in Oulu area can be found here.

Below suggestion for hotels in Oulu city centre (via booking.com):

- Best Western Hotel Apollo
- De Gamlas Hem Hotel & Restaurant
- Forenom Aparthotel Oulu
- Radisson Blu Hotel, Oulu
- Scandic Oulu City
- Original Sokos Hotel Arina Oulu

Chapter 4

Getting started

4.1 Checklist

- Test your login to the CSC R environment (and optionally, set up your own R); see below for details
- Watch short online videos on microbiome data science with R/Bioconductor.
- Test reproducible reporting Quarto
- Read about the selected demo data: HintikkaXOData. You can also try to load it in R and explore the data structure.
- Online support: for installation and other issues, join us at Gitter.
- Optionally: check the reading tips below

4.2 Study material

The course will use material from OMA online book, which is currently in its beta version. We encourage to familiarize with the content of the book and try some examples already before the course starts but this is optional.

4.2.1 Lecture slides

The slides and other teaching material will be shared at a later point through OMA.

4.2.2 Basics of R/Bioconductor

We expect that participants have previous experience with R/Bioconductor. To refresh your R/Bioconductor skills, you can check additional study material in OMA

4.3 Setting up R environment

4.3.1 Your own computer

Setting up your own computer and using local R installation will allow you to continue using the same tools after the course. In order to do this, you can follow the OMA setup checklist.

The course organizers can provide limited online support for installations via Gitter. However, installation is more demanding in some systems.

We also provide the CSC notebook for the course (see below); this comes with all necessary R packages and data pre-installed for ease of use.

4.3.2 CSC Notebook

We provide a temporary access to a cloud computing environment that readily contains the available software packages. Instructions to access the environment will be sent to the registered participants.

1. Read the instructions
2. Go to the CSC notebook frontpage
3. Login
   a. Haka login
      - If you have a Finnish university account, you should be able to login with Haka
      1. Press **Login** button from the frontpage
      2. Press **Haka** button
      3. Select right organization
      4. Enter login information
   b. CSC login
      - You can create a CSC account by following the instructions
      1. Press **Login** button from the frontpage
      2. Press **CSC** button
      3. Enter login information
   c. Special login
      - For those who cannot login with Haka or CSC account
      1. Contact Giulio on gitter if you are not able to login
      2. We give you a guest account
      3. Press **Special Login** button from the frontpage (below the **Login** button)
      4. Enter login information (username goes to **email** slot)
4. Join workspace
   a. Press **Join workspace** button (Top right corner)
   b. Enter the **Join Code** (Check your email)
5. Start session
   a. Press **ON** button
6. You can save files to **my-work** directory. They are kept stored even when the session is closed. **shared** folder is shared with all participants.

Bibliography

Drnevich, J., Tan, F. J., Almeida-Silva, F., Castelo, R., Culhane, A. C., Davis, S., Doyle, M. A., Holmes, S., Lahti, L., Mahmoud, A., Nishida, K., Ramos, M., Rue-Albrecht, K., Shih, D. J. H., Gatto, L., and Soneson, C. (2024). Learning and teaching biological data science in the Bioconductor community.

Tuomas Borman and Leo Lahti (2024). *Multi-omic data science with R/Bioconductor*.