mar622 - Profile Module R programming for (meta)-genomic sequence analysis

Module label: Profile Module R programming for (meta)-genomic sequence analysis
Module code: mar622
Credit points: 6.0 KP
Workload: 180 h
Used in course of study:
- Master's Programme Microbiology > Mastermodule
Contact person:
- Liliana Cristina Moraru

Module responsibility:
- Liliana Cristina Moraru

Authorized examiners:
- Liliana Cristina Moraru

Entry requirements:
The course „Introduction in sequencing and sequence analysis“. Previous programming experience is not required.

Skills to be acquired in this module:
DNA sequencing has become a routine method in microbiology research. Most of the times, sequence analysis requires knowledge of a programming language. One of the programming languages most used for this purpose is R. The course will cover the following topics:
- I. programming in R using an integrated development environment (RStudio)
- II. working with strings (stringr package)
- III. working with lists and data frames (readr and dplyr package)
- IV. sequence analysis (seqinr, Bioconductor packages: Biostrings, GenomicRanges, Decipher)
- V. (meta)-genomic and data visualization (ggplot2, Gviz)
- VI. Creating sequence / metadata databases
- VII. Accessing and mining sequence / metadata databases though R based web applications (Shiny, DT and Shinyjs packages)
- VIII. reporting in R (Rmarkdown and Knitr packages)
- IX. managing code (Roxygen2 package)
- X. microbial genome annotation using R

Module contents:
R programming for (meta)-genomic sequence analysis will run over a two weeks period. A single, introductory lecture will be offered within the first day of the course. Then, the course will be structured in programming exercises which cover all topics at point 12. The exercises are designed to exemplify the use R programming within the framework of microbial (meta)-genome analysis. In addition to the teacher–student sessions, the students will work on individual projects. Each student will receive a short microbial genome (e.g. viral genome), and will analyze it by building custom, self-programmed pipelines. The output from the individual projects will consist in a analysis report prepared in Rmarkdown and Knitr packages. The report will include both the R code and the genome analysis results.

Reader’s advisory:
will be announced

Links:

Language of instruction: English
Duration (semesters): 1 Semester
Module frequency: once a year
Module capacity: 15
Modulelevel: ---
Modularart: Wahlimodul / Opportunity
Lern-Lehrform / Type of program: Seminar and computer lab, 2 continuous weeks
Vorkenntnisse / Previous knowledge:

Examination:
Time of examination: Announced during the course.
Type of examination: Written protocol (80%) and class participation (20%).

Course type:
- Seminar: 2.00 SWS, SuSe, 28 h
- Practical: 2.00 SWS, SuSe, 28 h

Total time of attendance for the module: 56 h