



SWEDISH UNIVERSITY OF AGRICULTURAL SCIENCES
Department of Plant Biology
Syllabus

Analysis of High Throughput Sequencing RNA-Seq Data I

3,0 Credits

Finalized by: Research school Organism Biology, Mattias Thelander chair, 2023-10-13

Valid from: Second half-year 2023 (2023-07-01)

Level within study regulation: Third cycle

Subject

Bioinformatics

Entry requirements

Enrolled as a PhD student in Biology or with equivalent competence. The course is primarily for SLU PhD students but will be open also for researchers if space allows.

Examination formats

Pass / Failed. The requirements for passing are described in the course assessment criteria which are contained in a supplement to the course syllabus.

Objectives

The aims of this course are to:

- familiarize the participants with advanced RNA-Seq data analysis methodologies and how biological knowledge can be gained from these by illustrating different analyses approaches
- allow the participants to acquire computational competences about latest RNA-Seq analytical approaches and related statistical methods
- expose the participants to modern computing technologies, as the computer hands-on of the course will be performed using cloud computing resources
- develop the participants' awareness to critically reflect onto RNA-Seq analysis strengths and weaknesses
- enable participants to perform the analysis of their own data (instead of the exemplary dataset otherwise available)

The corresponding Learning Outcomes are:

- develop basic skills to navigate and interact with the Unix Command Line Interface (CLI)
- develop basic skills to interact with the RStudio Integrated Development Environment (IDE)
- reproduce basic analysis in the R programming language
- summarise High Throughput Sequencing (HTS) technologies, past and present
- report advantages and limitations of HTS for expression profiling
- describe the HTS data pre-processing
- perform the HTS data pre-processing
- describe the pseudo-alignment principles
- compare (pseudo-)alignment methods
- perform the pseudo-alignment
- list statistical concepts of importance for RNA-Seq
- perform the biological Quality Assessment (QA), also known as Exploratory Data Analysis (EDA)
- describe the principle of Differential Expression (DE)
- enumerate the statistical concepts of DE
- list the assumptions associated with DE
- perform the differential expression analysis

Content

High Throughput Sequencing (HTS) and in particular Next Generation Sequencing (NGS) have revolutionized the way we conduct gene expression analysis. In comparison to micro-array-based methods, NGS has unleashed an almost unlimited power to perform gene expression analysis, for a similar price and with faster, more comprehensive, more efficient, and more reproducible characteristics. Rapidly, the data generation rate has exceeded the analytical capabilities and data analysis has become the major bottleneck in gene expression studies. This course, aimed at advanced students, has the objective to help addressing this problem by helping participants (i) to critically assess the challenges faced in the HTS field, (ii) to develop efficient communication skills/strategies with bioinformaticians, and (iii) to learn to assess, analyse and interpret HTS data.

The course runs over two weeks: In the first week, the students are introduced to the Unix and R environments to give attendees the necessary skills for the second week and future bioinformatics work. The second week covers the analysis of High Throughput RNA-Seq data. The course will provide three ECTs. We strongly recommend to pair this course with course "Analysis of High Throughput Sequencing RNA-Seq Data II - apply to your own data, 1.0 Credit", which will have attendants apply all the knowledge from this course to their own

data.

Week 1 will have active teaching on two days, and extensive, but directed, self-study of literature and exercises for the rest of the week. Supervisors will be available for questions and guidance. Week 2 will have active teaching and exercises every day of the week.

The course will mix lectures, interactive lectures, computer hands-on session, literature review, etc. The course will use various teaching environments, including meetings as a classroom, but also as smaller groups. In addition, asynchronous virtual environment will be used, such as virtual classrooms, flipped classrooms, etc. to offer the participants the possibility to review and deepen their understanding of the course material at their own pace and to discuss among peers.

Other information

Teaching on this course will be provided by the SLU Bioinformatics Infrastructure (SLUBI), namely by Nicolas Delhomme, Iryna Shutava, Adnan Niazi, Abu Bakar Siddique, Lizel Potgieter and Amrei Binzer-Panchal. The course is organized on behalf of the SLU research school Organism Biology. There is no tuition fee. Participants are expected to bring their own laptop for the practical computer exercises.

The course will be conducted online only. However, if students wish so, they can organise themselves in groups and join using one of the teaching facilities available at SLU.

The maximum number of participants is 20. SLU-registered PhD students are prioritized over other researchers but otherwise admission will be on a first come, first serve basis, provided that the minimal requirements are met.

If you have any question, please contact the course leader or the slack channel [#organism-biology-research-school-rna-seq-course](#) of the [slubi-workspace.slack.com](#) workspace.