Applied Microbial 'Omics

Course Plan

October, 2023

Contents

Preface

The purpose of this document is to share the details of the course "**Applied Microbial 'Omics**". In the following sections you will find the course description, learning objectives, plan, schedule, attendance and grading policies, as well as other key information that is critical for the course attendees to consider.

Course Details

This course is a part of the module "**Applied Molecular Ecology**" (**mar260**) taught by Prof. Dr. A. Murat Eren (Meren) and Prof. Dr. Iliana Baums.

Course Details	
Name	**Applied Microbial 'Omics**
Meeting Location	**W32 1-112**
Number	5.12.263
Type	Seminar/Exercise
Credits	1
Language	English

In addition to this course, the module **mar260** contains three additional components that each participant of this course is expected to also attend,

- **Introduction to Popular 'Omics Strategies** (5.12.262, Lecture, Meren, Course Plan: HTML, PDF)
- **Coastal Conservation in the 'Omics Age** (5.12.260, Lecture, Iliana Baums)
- **Readings/Exercises in Coastal Conservation** (5.12.261, Seminar, Iliana Baums)

Please familiarize yourself with the details of the remaining courses using the appropriate documentation provided for the other components.

Faculty and Communication

Exercises throughout **Applied Microbial 'Omics** will be primarily delivered by Meren. However, additional experts will take part in the design and/or delivery of various sections. The following table lists individuals who will be involved in the course, and their contact information:

Name	Role	Expertise	Contact information
Meren	Professor	Microbial Ecology, Computer Science	meren@hifmb.de
Iva Veseli	Postdoc	Microbial Ecology, Computer Science	iva.veseli@hifmb.de
Jessika Füssel	Postdoc	Microbial Metabolism, Biogeochemistry	jessika.fuessel@uol.de
Florian Trigodet	Postdoc	Microbiology, Bioinformatics	florian.trigodet@hifmb.de

Throughout the semester (and beyond) you can reach out via email with any question to Meren, who should be your first contact for anything related to the course activities unless specified otherwise anytime throughout the semester.

Description and Learning Objectives

Generating hundreds of millions of sequences or tens of thousands of genomes to study naturally occurring microbial communities has become commonplace for many microbiologists. The ability to benefit from the ongoing data revolution demands the new generation of microbiologists to be familiar with the arsenal of 'omics tools that enable in-depth investigations of the new data streams that offer detailed snapshots of the lifestyles of microbes. **The primary aim of this course** is to help its participants gain **hands-on experience** with some of the most popular data types and approaches in microbial 'omics and put some of the strategies detailed in the lecture **Introduction to Popular 'Omics Strategies** in practice.

Applied Microbial 'Omics is a *seminar* that is designed to introduce its participants to 'omics data analyses to answer real-world questions with often simplified datasets. Participants will learn about the practical aspects of working with popular 'omics data types and their contemporary applications. The 'omics data types and strategies that will be covered throughout the semester include **genomics**, **metagenomics**, **metagenomic read recruitment**, **metabolic reconstruction**, **pangenomics**, **phylogenomics**, and **microbial population genetics**.

The learning objectives of the course includes the following:

- To gain experience in the **UNIX shell** and its utility in working with open-source software and large datasets.
- To apply **state-of-the-art 'omics approaches** to various data types to make sense of complex datasets.
- To communicate data analyses through **reproducible bioinformatics workflows**, figures, and written reports.

Prerequisites

Throughout the course we will use anvi'o for 'omics analyses. Anvi'o is an open-source software platform that brings together many aspects of today's cutting-edge computational strategies of data-enabled microbiology, including genomics, metagenomics, metatranscriptomics, pangenomics, meta-pangenomics, phylogenomics, and microbial population genetics in an integrated and easy-to-use fashion through extensive interactive visualization capabilities. Anvi'o is cited over 1,000 times in the literature, and is actively maintained. The latest release of anvi'o, v8 (marie), became available only a few weeks ago. While it is not a prerequisite, it will be most beneficial if the participants have access to personal computers, preferably laptops, that they can bring to the classroom.

The participants of this course are also expected to be familiar with the UNIX shell (also known as the 'terminal environment', or 'command line interface'). If you have no prior experience with the command line interface, that is OK, as you will generate those skills throughout the course as the vast majority of data analyses we will do will take place in the command line interface. Arguably, the exposure to the command line environment and developing a level of mastery of it will be one of the most impactful gains you will have from this course that will help you throughout your professional journey almost regardless of which career path you choose that involves data; so if you are not familiar with the command line environment, see this as an opportunity to invest time into developing some skills in it. You can use some of the following material to familiarize yourself with the command line interface, and Meren will be happy to help you with any questions:

- Beginner's Guide to the Bash Terminal (a video introduction to the Linux command line environment – although Joe Collins is talking about Linux, the topics are relevant to anyone who uses a command line environment and Meren strongly recommends everyone to watch this in its entirety, and try to replicate commands).
- Learning the Shell (a chapter from the open book "*The Linux Command Line*" by William Shotts – Meren highly recommends).

The course will require its participants to read and understand contemporary literature written in English.

Content Delivery

The primary mode of course content delivery will be through data analysis assignments and discussions during the class time related to their implementation. The vast majority of activities will be informed by the **core concepts**, **data types**, and **analysis strategies** explained in the companion lecture **Introduction to Popular 'Omics Strategies**. There will often be extensive discussions over best analysis practices, and the observations we can make through our analyses, to which participants will be encouraged to show **active, verbal participation**. Analysis tasks and data will be available on Stud.IP two weeks before the next session.

Please note that **preparation** and **participation** will play a key role in your success as usual. For an effective learning experience please consider (1) taking a brief look at the new analysis assignment (~1 hour of study), (2) take time early on to try to complete it by yourself (~2-4 hours of study), and (3) ask questions to Meren or others before it is too late.

Attendance Policy

Each participant is expected to attend each lecture in person (unless a legitimate reason for absence that is recognized by the University is in effect). The attendance will be recorded by analysis reports.

Analysis Report Submission Guidelines

Please follow these guidelines carefully to format your reports. Each analysis report should be submitted as Markdown file as well as a PDF document in a single compressed directory. The directory structure and file names should look like this:

```
Eren_AM_EX_01/
├── Eren_AM.md
├── Eren_AM.pdf
├── files/
│   ├── auxiliary-data.db
│   ├── contigs.db
│   └── profile.db
└── images/
    ├── Figure_01.png
    ├── Figure_02.png
    └── Figure_03.png
```

where,

- `Eren_AM` is the 'last name' and 'initials' of the person who prepared the report.
- `EX_01` is the label shared for the exercise in the course plan.
- `Eren_AM.md` is the flat-text report file formatted in *markdown*. Here you can find a syntax guide for markdown, and MacDown is an open-source markdown editor with a WYSIWYG editor, which makes life easier.
- `Eren_AM.pdf` is a PDF document generated from `Eren_AM.md` using pandoc. Once you have your markdown file, you can easily convert it to a PDF document using the following command:

```
pandoc Eren_AM.md --toc --pdf-engine=xelatex -o Eren_AM.pdf
```

- `files` directory can include files necessary to share along with the report (they should be the absolute minimum list of files that are absolutely necessary to reproduce an analysis nicely curated together).
- and `images` directory is to keep images cited from the report.

Once the directory is ready, you can run the following command to compress it into a single archive file and submit the new file:

```
tar -zcf Eren_AM_EX_01.tar.gz Eren_AM_EX_01/
```

Please also consider the following formatting guidelines for your reports:

- Make sure your report is a stand-alone text, so that anyone who reads the report can understand its purpose without having to read the assignment itself.
- Use headers and sub-headers to separate sections from each other – make your report pretty and nicely formatted. For each section briefly explain your analysis strategy and reasoning behind it.
- Use code blocks to separate command lines from free text (for which tEX_0here is a markdown syntax).
- Make your analyses reproducible. So if someone reads your report and runs every command line one by one, they should be able to reproduce your analysis if they are in a directory that contains the original input data.
- Screenshots of key observations are encouraged :) Make sure you have figures to communicate your observations clearly.

Course Plan

Please note that each lecture takes place on **Wednesdays** (bi-weekly), between **10:15 - 12:00**, at **W32 1-112**.

Each session below is

25/10/23 :: Introduction to anvi'o and installation check

The primary purpose of this session is to **discuss the course format**, future exercises, and how to return reports.

We will also **discuss anvi'o and its features**, and will make sure everyone has a **working copy of anvi'o installed** on their computers. Towards the end of this first session those who have attempted yet not been able to install the platform will receive hands-on help. Please try to install the development version of anvi'o on your computers before you come to this exercise session. You will find the installation instructions on https://anvio.org/install

- *Suggested Reading*:
  - Eren AM, et al (2021). Community-led, integrated, reproducible multi-omics with anvi'o. *Nature Microbiology.*

08/11/23 :: EX 01: A genome across metagenomes

The purpose of this exercise is to help you have a direct exposure to individual analysis steps and tools that enables one to recruit reads from metagenomes (essentials of which is covered extensively on the 30/10/23 lecture in the companion course), and profile the read recruitment results to investigate gene distribution patterns of a given population.

Please read the assignment below carefully, and **return your reports by 07/11/23**:

> TBD

- *Key Resources*:
  - A simple read recruitment exercise. The purpose of this tutorial is to walk you through a simple read recruitment experiment. Throughout this exercise you will use a mock dataset to (1) familiarize yourself with commonly used file formats such as FASTA, FASTQ, SAM, and BAM, (2) learn the basic steps of read recruitment through Bowtie2 and samtools, (3) learn how to profile read recruitment results using anvi'o, and (4) familiarize yourself with downstream steps of the analysis of recruited reads that will be very helpful to apply to the assignment above.

22/11/23 :: EX 02: Refining metagenome-assembled genomes

Please read the assignment below carefully, and **return your reports by 21/11/23**:

> TBD

06/12/23 :: EX 03: Pangenomic analysis of Prochlorococcus

Please read the assignment below carefully, and **return your reports by 05/12/23**:

> TBD

20/12/23 :: EX 04: Phylogenomic placement of an unknown microbe

Please read the assignment below carefully, and **return your reports by 19/12/23**:

> TBD

17/01/24 :: EX 05: Comparative microbial metabolism

Please read the assignment below carefully, and **return your reports by 17/01/23**:

> TBD

31/01/24 :: EX 06: Population genetics of a plasmid

Please read the assignment below carefully, and **return your reports by 30/01/23**:

> TBD

Evaluation and Grading

The evaluation of the attendee performance in this course (along with all the other three in the module "**Applied Molecular Ecology**" (**mar260**) will be based on two items to be returned by each attendee individually: (1) a **research pre-proposal** (which will provide the basis for the full proposals due at the end of class) and a final **research proposal**.

Writing research proposals provide one with a critical skill that will be useful regardless of the profession one chooses to pursue after an undergraduate education. The primary purpose of a research proposal is to persuade others that *your* idea (in this case science) is worthy of committing resources to, and you are the best person to implement it. The final research proposals will be graded based on their ability to address the following questions:

- Is the proposal contains **a novel research objective** that is relevant to the "**Applied Molecular Ecology**" (**mar260**) module content?
- Does it make a **strong case** given what is **known** and what is **unknown**?
- Does it adequately and accurately **cite the existing literature**?
- Does it suggest the use of methods that are **relevant** and **effective** to address the research question?

- Does the investigator demonstrate their **ability to use the proposed methods**?
- Does the proposal includes a discussion of **expected outcomes**, **potential risks**, and how to **mitigate** risks?
- Does it include a **reasonable budget** and a **meaningful timeline** to carry out the proposed research objectives?
- Does the proposal written in a **clear**, **concise**, and **accurate** manner that is expected of scientific work?

Upon the submission of the pre-proposal, smaller teams of students will serve as a "peer-review panel" and discuss each pre-proposal. The names of participants who wrote these proposals will not be known to the class. The panel will take place in class for all participants. The purpose of the panel is to provide feedback on how well the proposal was able to convey ideas. Participants are expected to improve their short pre-proposals based on the feedback they received from the peer-review panels and submit their full proposal as their final project.

The participants will be provided with detailed instructions on how to come up with a proposal idea, what resources are available to carry out the projects, how to format the pre-proposal and final proposals and how to evaluate them in panel.

It is challenging to think of a topic for research that can be done in a short amount of time with limited resources. Therefore, the course directors (Meren and Iliana) will meet with each of the course participants individually to discuss their proposal idea, tell them whether it is feasible, and help them with resources. To initiate this discussion, you will hand in a one-paragraph research idea early in the semester.

The grading scale for this module is as follows:

Grade	Threshold
1.0	95%
1.3	90%
1.7	85%
2.0	80%
2.3	75%
2.7	70%
3.0	65%
3.3	60%
3.7	55%
4.0	50%

Examination Policy

Please find all relevant university policies here: https://uol.de/studiengang/pruefungen/umweltwissenschaften-fach-bachelor-136

Academic Integrity

All UOL policies regarding ethics and honorable behavior apply to this. Academic integrity is the pursuit of scholarly activity free from fraud and deception and is an educational objective of this class. All University policies regarding academic integrity apply to this course. Academic dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating of information or citations, facilitating acts of academic dishonesty by others, having unauthorized possession of examinations, submitting work of another person or work previously used without informing the instructor, or tampering with the academic work of other students. For any material or ideas obtained from other sources, such as the text or things you see on the web, in the library, etc., a source reference must be given. Direct quotes from any source must be identified as such.

Disability Access Statement

UOL welcomes students with disabilities and students with care obligations for their children or close relatives into the University's educational programs. In order to receive consideration for reasonable accommodations, you must contact the *Prüfungsausschuss*. Please let Meren and/or Iliana know at the beginning of the semester what accommodations were approved for you.