DTU

DTU Health Technology
Bioinformatics

Metagenomics and Microbiome Analysis: Introduction to course

Gisle Vestergaard
Associate professor
Section of Bioinformatics
Technical University of Denmark
gisves@dtu.dk

Menu

- Who are we?
- What is this course about?
- Course timeline
- Exam and evaluation
- Project
- Pre-test

Course responsible

Gisle Vestergaard

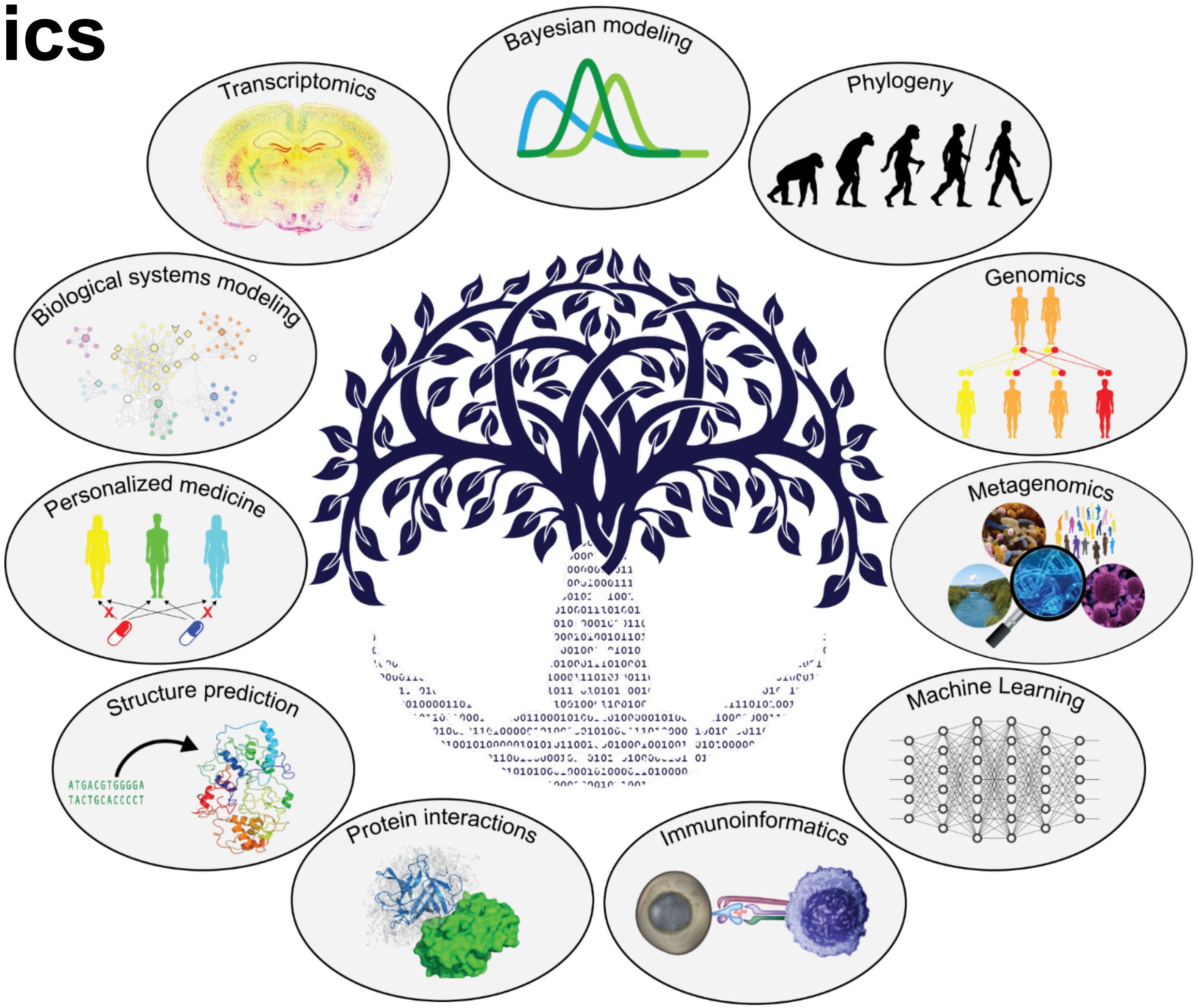
Associate Professor, Section of Bioinformatics

Marlene Danner Dalgaard

Senior researcher, Head of DTU Multi-AssayCore (DMAC)

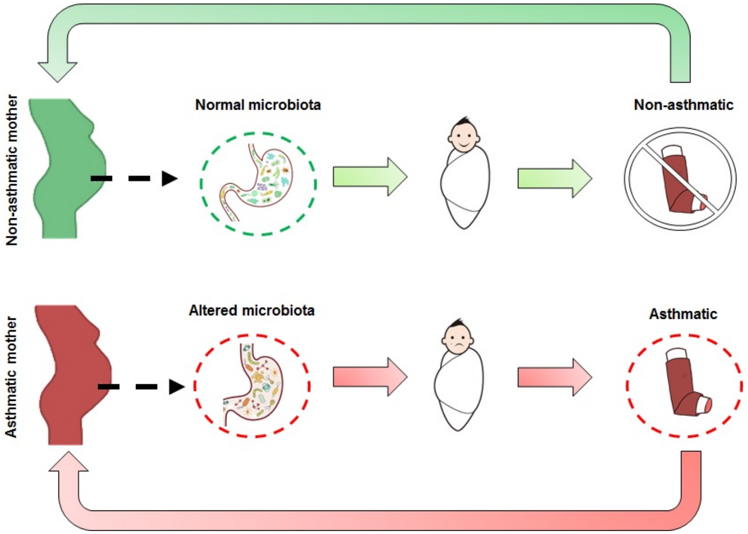
Section of Bioinformatics

Solving biological problems using a wide variety of state-of-the-art informatics solutions



Who are we?

- PhD in Biochemistry from Copenhagen University
- Postdoc at Helmholtz Zentrum München on a Humboldt Research Fellowship
- Assistant Professor at Copenhagen University in 2018
- Associate Professor at DTU February 2019
- Taught NGS in München and at Copenhagen University
- Also organizer of the January course “Next-Generation-Sequencing Analysis”



Science & Environment

Atacama's lessons about life on Mars

By Jonathan Amos
BBC Science Correspondent

27 February 2018

The Atacama has been in a hyperarid state for millions of years

Even in the driest places on Earth there is life eking out an existence, it seems.

Scientists have examined the soils in those parts of the Atacama desert that may not see any rains for decades.

Who are we?

- MSc in Molecular Biology and Chemistry
- PhD in Human Biology: Testicular Cancer
- Senior Researcher, Head of DMAC: Multiassay Core Facility
- 5th year course responsible

Who are you?

What stage?

- Master students
  - Systemsbiology
  - Biotechnology
  - Quantitative biology
- A PhD student
- A handful of externals

Feedback

- Eighth time course is running
- First time I am running the course
- We are still improving
- Please give us feedback !
  - Please do the evaluation at DTU Inside

Learning objectives

- Explain the concept of metagenomics and its applications opposed to single genomics
- Discuss a possible metagenomics experimental design; from sampling to sequencing
- Explain basic metagenomic data generation
- Explain the advantages and limitations of microbiome data
- Perform basic metagenomics analysis
- Explain the basic principles of microbiome ecology
- Explain the basic principles of quantitative count based data analysis
- Perform basic quantitative count based data analysis
- Perform calculations of basic ecological measures
- Present project work orally and on a poster

Course material

- No text book required
- Material will be distributed during the course on Campus net and especially the wiki

http://teaching.healthtech.dtu.dk/36636/index.php/36636:Course_plan_autumn_2019

Online learning

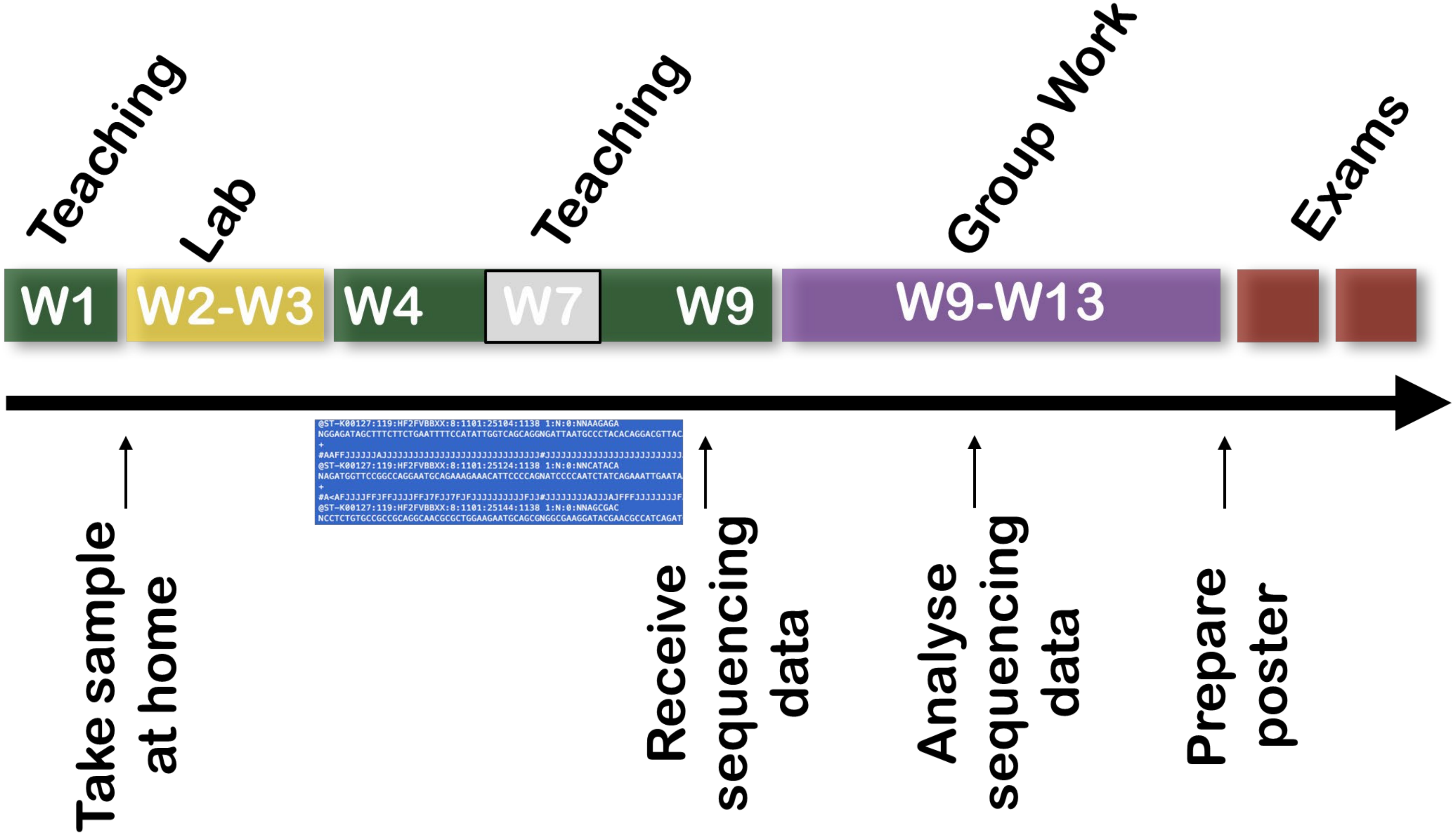
WILL NOT BE PART OF THE CURRICULUM

Great recent overview paper: ”Best practices for analysing microbiomes”
https://www.nature.com/articles/s41579-018-0029-9.pdf

Coursera course: ”Gut Check: Exploring Your Microbiome”
https://www.coursera.org/learn/microbiome

- Videos, reading and quizzes

Course timeline



```
@ST-K00127:119:HF2FVBBXX:8:1101:25104:1138 1:N:0:NNAAGAGA
NGGAGATAGCTTTCTTCTGAATTTTCCATATTGGTCAGCAGGNGATTAATGCCCTACACAGGACGTTAC
+
#AAFFJJJJJJAJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ#JJJJJJJJJJJJJJJJJJJJJJJJJJ
@ST-K00127:119:HF2FVBBXX:8:1101:25124:1138 1:N:0:NNCATACA
NAGATGGTTCCGGCCAGGAATGCAGAAAGAAACATTCCCCAGNATCCCCAATCTATCAGAAATTGAATA
+
#A<AFJJJJFFJFFJJJJFFJ7FJJ7FJFJJJJJJJJJJFJJ#JJJJJJJJAJJJAJFFFJJJJJJJJF
@ST-K00127:119:HF2FVBBXX:8:1101:25144:1138 1:N:0:NNAGCGAC
NCCTCTGTGCCGCCGCAGGCAACGCGCTGGAAGAATGCAGCGNGGCGAAGGATACGAACGCCATCAGAT
```

YOUR microbial metagenomic study

- Today design this years microbial metagenomic study
- Limitations have to be considered: time limitation, effect, cohort size, feasibility
- Collect samples until next weeks laboratory exercise
- Previous years this has always been a students stool used for an intervention study

Wet lab exercise

- 2 days of laboratory work at Room 006 in Building 208 and Laboratory Room 001 also in Building 208
- Sampling, purification and extraction of DNA, QC
- Library preparation and sequencing
- Don’t worry if you lack wet-lab experience!!

Exams & grading

- **Oral exam/Group Work:**
  - Each group will prepare a poster and present the poster for a group of examiners
  - Afterwards each group member will be individually questioned about the course objectives and your project (~ 5-10min) that you have prepared as a group
- **2 hours Written exam without help**
- **Deliverables:**
  - Quizzes and assignments

Oral exam / Group work

- **Oral exam/Group Work:**
  - Each group will prepare a poster and present the poster for a group of examiners
  - Afterwards each group member will be individually questioned about the course objectives and your project (~ 5-10min) that you have prepared as a group

- You will have the chance to assign the groups
- Questions?

40%

Written exam

- 2 hours written exam without help
- Questions?

40%

Deliverables

- See program webpage for overview and dates
- Quizzes and assignments graded as passed/not passed
- Questions?

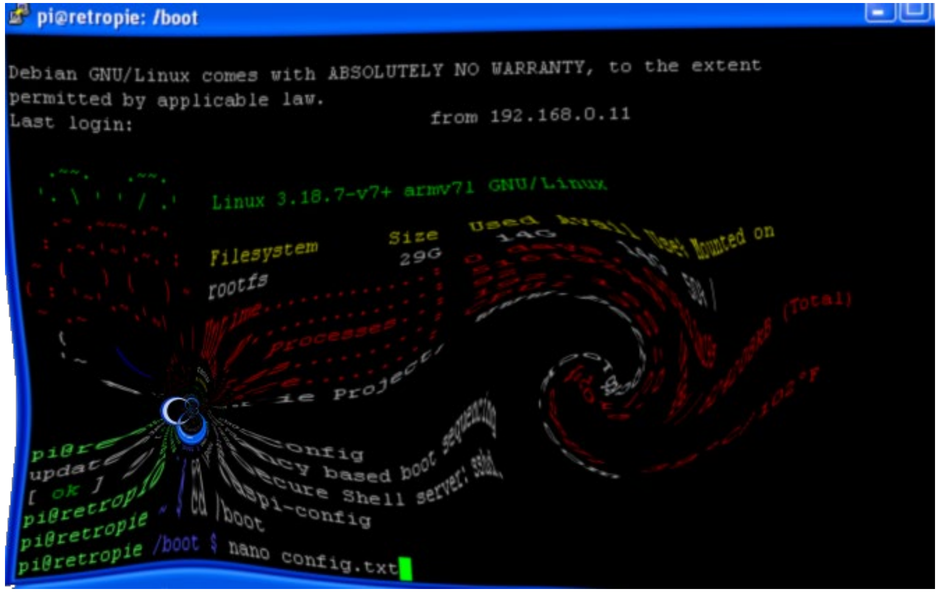
Deliverable	Date	Description
I	04/09-2019	Class Room presentation in groups: Intervention study
II	09/09-2019 23:59	Hand-in: Lab protocol Powersoil Kit
III	17/09-2019 23:59	Hand-in: Lab protocol Nextflex DNAseq kit
IV	30/09-2019 23:59	Hand-in: Sequencing technology report
V	02/10-2019	Class Room Quiz: NGS I
VI	30/10-2019	Class Room Quiz: NGS I+II & Quantitative metagenomics I+II

20%

Cloud computing

- The course has moved to the Cloud!
- Danish National Supercomputer for Life Science (Computerome) located at DTU Risø
- 16048 cores, 92 Tb RAM an 3Pb storage
- We have 2 nodes
  – Each has 28 cores and 128 Gb RAM

Why shell terminal?



```
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login:                                  from 192.168.0.11

                  Linux 3.18.7-v7+ armv7l GNU/Linux

pi@retropie /boot $ nano config.txt
```

- The command-line is not as dangerous as it looks…trust us!
- Almost all tools for NGS analysis are command line only
- Generally more efficient/flexible, you can play around with the tools/data
- They can be pipelined, ie. analyzing 100 files in windowed mode is a pain ...
- Alternative approaches: Galaxy, CLC-workbench, Geneious

Course essentials

- Course wiki:
  - http://teaching.healthtech.dtu.dk/36636/index.php/36636:Course_plan_autumn_2019

- Teaching Wednesdays in Room 062, Building 208, EXCEPT:
  - Week 2 & 3 wet lab exercises at Room 006 in Building 208 and Laboratory Room 001 also in Building 208

- Form groups for project work to begin 23.10.2019

- Poster exam 04.12.2019

- Written exam 19.12.2019

Course advice

- Learn principles of how to analyse metagenomics data
- Use principles for your project
  - This includes formulating a project that can be answered with the tools and data you have available
- Use the cloud servers for group analysis and coordinate data heavy analysis across all groups (no need for all groups to perform assembly of the same data)
- Form groups with people from other backgrounds…diversity IS an advantage
- …and again please feel free to answer questions and give us feedback!!

Course breakdown – Week 1

- Course introduction
- Pre-test
- General introduction to metagenomics
- Introduction to course study design
- Agreement on course study…YOUR study!

Course breakdown – Week 2 & 3

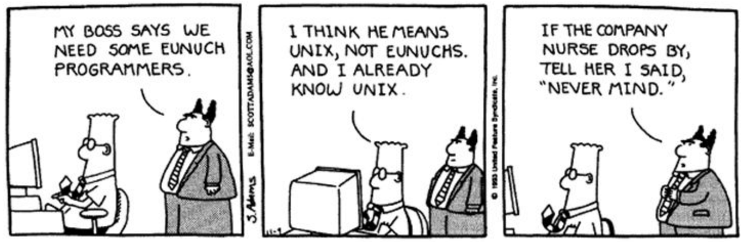

W1 W2-W3 W4 W7 W9 W9-W13

- Sample purification
- DNA isolation
- Quality control
- Fragmentation
- Create sequencing library

Course breakdown – Week 4

W1 W2-W3 W4 W7 W9 W9-W13

- Introduction to UNIX
- Log on to Computerome
- Introduction to Next-Generation Sequencing
- Next-Generation Sequencing basics



Course breakdown – Week 5-9

W1 W2-W3 W4 W7 W9 W9-W13

- R
- 16s rRNA analysis
- Quantitative metagenomics
- De novo assembly & Binning
- Human gut microbiome
- Association studies
- Group formation and project formulation

Course breakdown – Week 9-13

- Group work on projects

Course breakdown – Exams


W1 W2-W3 W4 W7 W9 W9-W13

- Oral presentation of project posters
  - – Group work, < 4 per group, and present results on a poster
  - – try to form diverse groups in terms of skills
  - – Analyse your own data!
  - – supplement with existing data (e.g. MetaHIT, SRA…)
  - – don’t run too large datasets (computational and time restrictions)

Pre-test

- 10 min individually
- discuss in plenum and summarise
- not used for grading, it will give us an idea of your level