# CeMiSt Workshop 2020

* Minimum number of attendees: 10
* Attendees: PhD-students associated with the CeMiSt center
* Dates: 19-20 November 2020
* All days from 08:00-17:00
* Lectures and practical exercises
* Two workshops of one day each (2 days in total)
* Mandatory deliverables at the end of each workshop to gain ECTS points
* Workload is two full days of workshop (18 hours), two days of online self-study in programming languages (15 hours) and five days of project work for the deliverables (37 hours) for 70 hours in total.
* The course ECTS is 2.5
* Course responsible: Mikael Lenz Strube (milst@dtu.dk)

The workshop will have a large practical component based on example data and the students may use their own data in the exercises.

## Topics

* Development of skills in R and Linux
* Whole genome assembly
* Gene prediction and annotation
* Basic statistics
* Multivariate statistics

## Workshops

**Day 1: The theory and practice behind Nanopore sequencing, assembly and analysis**

Responsible: Tue Sparholt Jørgensen

**Day 2: The theory and application of multivariate statistics with an emphasis on biological data.**

Responsible: Mathies Brinks Sørensen

## Learning objectives

**Day 1: The theory and practice behind Nanopore sequencing, assembly and analysis**

Responsible: Tue Sparholt Jørgensen (tuspjo@biosustain.dtu.dk)

* Understand the theory and principles of DNA sequencing, assembly and annotation
* Use Linux-based bioinformatic methods to profile a bacterial genome
* Infer and discuss the biological significance of genomic data

**The theory and application of multivariate statistics with an emphasis on biological data.**

Responsible: Mathies Brinks Sørensen (mabso@kemi.dtu.dk)

* Understand the motivation and basic mathematical principles of multivariate analysis
* Use R-based statistical coding to find and test patterns in biological data
* Conclude and discuss on the biological meaning of multivariate data.

## Deliverables:

A report containing:

1. Assembly, polishing, annotation and biological analysis of two genomes.
2. Analysis of two multivariate datasets using both unsupervised and supervised methods.
3. An appendix containing the source code for both analyses.