# Final assignment

The final assignment will let you use the tools you have learned on the course. You can use your own data, if you have any, or the data we supply you. Hand-in and approval of the assignment is necessary to pass the course and, in turn, get ECTS-points.

Code must be supplied in an appendix. Important commands can preferably be incorporated inline in the text.

Hand in-date is 20th of December.

Part 1 - WGS:

* Assemble, polish, and annotate 2 genomes.
* For each, report the following:
  + Genome assembly workflow
  + Genome completeness/topology
  + Organism and closest ANI%
  + Read statistics (including coverage and read N50, GC%)
  + Genome statistics (including size, number of genes, number of annotated genes, number of rRNA genes, Number of biosynthetic gene clusters)
  + 1-2 compounds likely produced by the strain
  + Explain why you used the workflow you did
  + If you have analyses specific to your research interests, feel free to include them as well.

Part 2 – Multivariate statistics

* Analyze two datasets using multivariate methods
* First, you should do an explorative analysis of your choosing (unsupervised method)
  + Argue for the choice of method
  + State/show how you have selected number of components, linkage method etc.
  + Explain the biological meaning, if any, of your observations
* Secondly, use a supervised model to fit the data
  + Define your model (preferably with mathematics, but written words are also fine), describe what you want to model with your data.
  + Report your training and testing error
  + Discuss your model and what could be improved – compare it with other methods in the course and/or other statistical methods in general.
  + Explain the biological meaning, if any, of your models