

# Long read *de novo* assembly exercise

---

## Instructions

---

Document your work well.

Add Assembly and BUSCO results here:

[Google sheets hyperlink](#)

## Set up

---

1. **Set up a new directory structure** for a *de novo* genome sequencing projects
2. **Create a private github repository** for the project. All directories and file except sequence files should be included.  
The repo should have a readme file with a detailed description of what you have done in the project.
3. **Add dagahren as a collaborator.**

**Note! It is important that you throughout your project think of ways of checking what you did.**

**For example check that the download was complete or that a software ran to completion without errors.**

Use top to follow the process and see what the commands are doing.

## Dataset

---

The dataset is a HiFi Pacbio reads from the yeast *Saccharomyces cerevisiae*.

The data can be downloaded from NCBI Sequence Read Archive (SRA)

Download the reads for the following page or use sra-tools:

[SRR13577846](#)

Make the Data directory readonly!

## Computational resources

---

**When using the Course server, max number of CPUs (threads) to be used is 10!**

## Software

---

Make a separate directory for each run & investigate

- Quality assessment using fastqc
- *de novo* assembly (separate dirs if you try more than one!)
- Quast quality assessment
- BUSCO assessment

**Make sure that you are documenting all work you do in the README file.**

**Note the version of each software you use**

**Time the run of each command so that you can report how fast the assembly was.**

**Remember to commit your changes and push to Github**

Pick an assembler suitable for PacBio Sequencing data or choose freely from publications (but give the proper reference):

For example you can use:

**Hifiasm**

<https://github.com/chhylp123/hifiasm>

<https://hifiasm.readthedocs.io/en/latest/index.html>

## Quality assessment

---

Use the following tools to assess the sequences and the genome assembly using:

- FastQC
- Quast
- BUSCO
- MultiQC
- Other?