# Installing the workflow; DNA2A-seq-analyzer

**The workflow was tested on a Linux (Ubuntu) environment.**

## 1. Install Miniconda and Mamba

**Important: First you have to install Miniconda3!; workflow tested with: Conda 23.11.0**

```
> wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
> chmod +x Miniconda3-latest-Linux-x86_64.sh
> ./Miniconda3-latest-Linux-x86_64.sh
```

**Important: If you do not have Mamba installed in your conda environment!; workflow tested with: Mamba 1.5.6**

```
> conda install -c conda-forge mamba
```

## 2. Create the environment and install the workflow

**Create a folder for your workflow**

```
> mkdir -p path/to/project-workdir
> cd path/to/project-workdir
```

**Download the workflow**

```
If you don't have the workflow yet, you can download it from GitHub.
Otherwise, if you already have the data, you can skip the following step.

Download and unzip the workflow from github
You can also download the workflow as a ZIP file and then unpack it in your project folder.

Download repo from GitHub
> git clone https://github.com/dusti1n/DNA2A-seq-analyzer.git

To install the environment with all packages, navigate to the project folder.
It is important that you are in the parent folder.
This folder contains, for example, the folder: config and workflow.
The file (environment.yaml) is also located in the folder.
```

**Create an env and install all required packages**

```
> conda env create -n snakenv -f environment.yaml
> conda activate snakenv


Important: If you have Mamba you can also use the following command!

> mamba env create -n snakeworkflow -f environment.yaml
> mamba activate snakemake
```

## 3. Install all packages and start the workflow

**Set configfile (config.yaml) parameters; open configfile.yaml**

```
# Set the path to the folder where your sample folders are located! #{***}
# The sample folders must have the following structure!
# Example: smpl_01/ont/ontfile.fastq.gz; smpl_01/illumina/illuminafile_1.fastq; smpl_01/illumina/illuminafile_2.fastq;
- set sample_path: /path/to/datasets/

# Set the project name for your samples; Example: drosophila_samples; #{***}
# This will create a subfolder in 'results' with the project name you entered.
- set project_name: saccharomycetes_smpls

# Set save_dict: true to save a JSON file (Dictonary) with your samples!
- set save_dict: true

# Canu; Set genomeSize; genomeSize={params.canu_genome_size}; #{***}
# Determine the size of your genome!
- set canu_genome_size: "12m" (Example "12m" for Saccharomyces cerevisiae)
```

```
# BUSCO; lineage; extra=config["busco_lineage"]; #{***}
# For a specific eukaryotic organism; https://busco.ezlab.org/list_of_lineages.html
- set busco_lineage: "saccharomycetes_odb10"

# FASTQC; Set memory; 8GB should be used as a minimum! #{***}
# Example(use of 8GB): fastqc_memory: "8192";
- set fastqc_memory: "8192"

# THIS PART IS IMPORTANT IF YOU ALSO WANT TO USE ILLUMINA DATA!
# Set illumina_data: true; If you want to use Illumina data!
# Set illumina_data: false; If you don't want to use Illumina data!
- set illumina_data: true

# Pilon; Set Memory for JAVA heap Space; 32G = 32 Gigabyte; #{***}
# 8GB should be used as a minimum! Example(use of 8GB): pilon_memory: "8G";
- set pilon_memory: "32G"
```

**It is important that you first load all your samples with the Python script (import_samples.py), then you can start the workflow!**

**Python script (import_samples.py) creates an automatic database for all your samples!**

```
> python workflow/scripts/import_samples.py
```

**Snakemake starts the workflow with all available cpu cores and installs all necessary packages!**

```
> snakemake --cores all --use-conda
# The workflow creates a results folder. The folder contains all analyzed samples.


Optional Commands
# Snakemake creates a flowchart for the workflow
> snakemake --dag | dot -Tpng > dag.png
```