Nucleotide Sequence Analysis

Class Workbook

Miha Prajs

May 8, 2025

## Theory

## Environment Preparaion

Before we can start analyzing sequences we must set up the environment. In new terminal window using “weget” command we will download Conda Package Manager installation script from official Anaconda repository. Then we run downloaded script using “bash” command followed by script name. Then we will set channels where we will download needed packages from using “conda config --append channels” command. Now it’s time to set working environment with “conda create -n” followed by a environment’s name and activating it with “conda activate” and it’s name. Then we can install all the necessary packages using “conda install”. It is also recommended that you regularly update installed packages[[1]](#footnote-21).

# Installation of Miniconda   
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86\_64.sh  
bash ./Miniconda3-latest-Linux-x86\_64.sh  
  
conda config --append channels defaults  
conda config --append channels bioconda  
conda config --append channels conda-forge  
  
# Environment preparation  
conda create -n anz  
conda activate anz  
   
# Installation of necessary tools  
conda install emboss entrez-direct sra-tools  
  
# Update instruction  
conda update --all # follow the instructions in terminal window

After the initial set up is complete we can start downloading and analyzing sequences.

## Exercise 1

**Task explanation**

# Installing tools  
conda install fastqc multiqc  
  
# Downloading sequences  
prefetch SRR30833064  
fasterq-dump SRR30833064  
  
prefetch SRR30833065  
fasterq-dump SRR30833065  
  
# Checking quality of the sequences  
fastqc SRR30833064  
fastqc SRR30833065  
  
#

**Result explanation**

## Exercise 2

**Task explanation**

# CODE

**Result explanation**

## Exercise 3

**Task explanation**

# CODE

**Result explanation**

1. Please note that in this workbook we will only use bash code [↑](#footnote-ref-21)