# Software Installation Instructions:

You are going to need the following software to run this analysis.

## FastQC

FastQC is a convenient software to check the quality of your reads. It is a java based application. So you need to make sure your system contains java. Check the install.txt file for instructions once you download the zip file and unzip it.

|  |  |
| --- | --- |
| **Task** | **Code** |
| Download FastQC | wget <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/fastqc_v0.11.5.zip> |
| Unzip | unzip fastqc\_v0.11.5.zip && cd FastQC |
| Change permissions | chmod 755 fastqc |
| Creates symbolic link in the local bin folder. Now you can just use the command fastqc from any directory | ln -s ${HOME}/FastQC/fastqc ${HOME}/bin/fastqc |

## Sickle

Go to sickle github page: <https://github.com/najoshi/sickle> and follow the instructions to install it.

## Mothur

|  |  |
| --- | --- |
| **Task** | **Code** |
| Download mothur | wget <https://github.com/mothur/mothur/releases/download/v1.39.4/Mothur.linux_64.zip> |
| Create a new folder | mkdir mothur |
| Navigate into the folder | cd mothur |
| Unzip | unzip ../Mothur.linux\_64.zip |
| Add mothur to Path file. | export PATH=$PATH:${HOME}/mothur |

# Reference Databases

## Silva

|  |  |
| --- | --- |
| **Task** | **Code** |
| Download v123 | wget https://www.mothur.org/w/images/b/be/Silva.nr\_v123.tgz |

## MiDAS

I have attached the MiDAS database in the folder.

# Create Custom Alignment Database

We will customize out Silva database file to the V4 region of the 16S rRNA gene. Run the following code in mothur

|  |  |
| --- | --- |
| **Task** | **Code** |
| Run in-silico PCR in mothur | pcr.seqs(fasta=silva.nr\_v123.align, start=13862, end=23444, keepdots=F, processors=5) |

# Preprocess Sequences

In this section of protocol we will perform the following steps:

1. Change the name of the files.
2. Decompress .fastq.gz files into fastq files.
3. Create symbolic links in a separate fastq folder for quality checking by combining all the forward reads and combining all the reverse reads.
4. Create FASTQC reports.
5. Perform sickle filtering on the fastq files in the sample folders, output files with same file name but with a prefix of "q".
6. Create a batch file for mothur in the bash script.
7. Create symbolic links for the quality-filtered files in a separate folder outside of the native folder organization.

The folder organization for using this code is : create a folder called mothur in your home directory. Inside that transfer the project folder from Illumina basespace to the mothur folder. Inside the project folder, there should be folders for each sample within which there are the forward and reverse reads.

|  |  |
| --- | --- |
| **Task** | **Code** |
|  | Navigate first to the project directory before running this script |
| Run this in the command line. Preprocess sequences | preprocess\_sequences.sh |

# Mothur Analysis

Run the steps in mothur\_batch\_v4\_script.txt. You can either do it in interactive mode or just run it batch mode using the mothur\_batch\_v4\_job.sh script.

## Run mothur in batch mode

Before running the script in batch mode, I would suggest benchmarking the code using a small set of data like we did at the workshop to make sure all the steps in mothur\_batch\_v4\_script.txt work properly.

|  |  |
| --- | --- |
| **Task** | **Code** |
| Run mothur in batch mode | sbatch mothur\_batch\_b4\_job.sh |