

Software in Genomics and **How to Install them**

Instructor

Abdullah Al Nahid



Brought to you by
Community of Biotechnology

Part of **COB Genomics**
Learning Program 2022

Virtual Environment

- create an isolated environment for programming projects
- avoid conflicts between multiple versions of the same or different software
- can be imagined as folders where all of your project-specific packages and tools are
- just like folders, you can create and delete them once a project is completed
- eliminates the hassle of painful installation process
- save disk space of your computer
- can be created with multiple tools such as **conda**, virtualenv or pyenv

Conda

1. Package, dependency and virtual environment management for any language
2. Runs on Windows, Linux, MacOS
3. Quick installation
4. 7500+ Python, R, C/C++ based packages
5. Easy to update packages
6. Remove when the project is finished
7. Saves disk space and your time

***Although conda works on all operating systems, most bioinformatics software only work on Linux and MacOS

Install:Conda

Anaconda

1. Complete data science toolkit with conda packages
2. Takes ~3GB disk space with more than intended packages

Download

<https://www.anaconda.com/products/individual#windows>

Miniconda

1. Minimal conda installer with only necessary default packages
2. Takes ~400MB disk space
3. Install any package later when needed

Download

<https://docs.conda.io/en/latest/miniconda.html>

Conda Commands

See all environments

```
conda info -e
```

Create an environment

```
conda create -n <env-name> python=3.8 -y
```

Activate an environment

```
conda activate <env-name>
```

Cheat sheet

<https://docs.conda.io/projects/conda/en/latest/user-guide/cheatsheet.html>

Deactivate current environment

```
conda deactivate
```

Delete an environment

```
conda env remove -n <env-name>
```

Install a package

```
conda install <package-name> -y
```

Conda in Google Colab (For non-Linux users)

Setup conda in Google Colab

```
!pip install condacolab  
import condacolab  
condacolab.install()
```

Run Conda

```
!conda
```

Google Colab

<https://colab.research.google.com>

Create an environment

```
!conda create -n <env-name> python=3.8 -y
```

Activate Environment

```
!source activate <env-name>
```

Deactivate Environment

```
!source deactivate
```

***Every tool might not work in Colab

fastQC

A quality control tool for high throughput sequence data

Installation

```
conda install -c bioconda fastqc -y
```

Installation URL

<https://anaconda.org/bioconda/fastqc>

trimmomatic

A flexible read trimming tool for Illumina Next Generation Sequencing data

Installation

```
conda install -c bioconda trimmomatic -y
```

Installation URL

<https://anaconda.org/bioconda/trimmomatic>

BWA

Burrows–Wheeler Aligner (BWA) is a software package for mapping low-divergent sequences against a large reference genome, such as the human genome

Installation

```
conda install -c bioconda bwa -y
```

Installation URL

<https://anaconda.org/bioconda/bwa>

samtools

Tools for dealing with SAM, BAM and CRAM files

Installation

```
conda install -c bioconda samtools -y
```

Installation URL

<https://anaconda.org/bioconda/samtools>

bedtools

A powerful toolset for genome arithmetic

Installation

```
conda install -c bioconda bedtools -y
```

Installation URL

<https://anaconda.org/bioconda/bedtools>

SeqKit

A cross-platform and ultrafast toolkit for FASTA/Q file manipulation

Installation

```
conda install -c bioconda seqkit -y
```

Installation URL

<https://anaconda.org/bioconda/seqkit>

megahit

An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph

Installation

```
conda install -c bioconda megahit -y
```

Installation URL

<https://anaconda.org/bioconda/megahit>

quast

Quality Assessment Tool for Genome Assemblies

Installation

```
conda install -c bioconda quast -y
```

Installation URL

<https://anaconda.org/bioconda/quast>

Thank you