

# Genomics and Clinical Virology 2024 Software Installation

## Table of Contents

---

Introduction to Software Installation .....	1
Software list according to course modules .....	1
Installation of Python and Python Installer (Anaconda).....	3
Shell Scripting .....	4
Installing Google Earth .....	8
Additional Resources and Troubleshooting .....	9

---

## Introduction to Software Installation

This guide provides extensive instructions for software installation on Ubuntu/Windows/MacOS systems, applicable to both virtual machine (VM) and host machine setups, as taught and utilized in the Genomics and Clinical Virology 2024 course. For VM installation specifics, please consult the supplementary VM installation guide attached separately. Additionally, valuable troubleshooting links are provided below for additional assistance.

**Note:** The Virtual Machine file (.vdi) contains all the software installed for the course.

---

## Software list according to course modules

Module	Software	Summary	Version	Website
Reference alignment, coverage and stats	Burrow-Wheeler Aligner (BWA)	Software package for mapping short DNA sequences against a large reference genome, such as the human genome.	0.7.17	<a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>
Reference alignment, coverage and stats	Bowtie2	Tool for aligning sequencing reads to long reference sequences.	2.5.2	<a href="https://github.com/BenLangmead/bowtie2">https://github.com/BenLangmead/bowtie2</a>
Reference alignment, coverage and stats/ Variant calling	samtools	Suite of programs for interacting with high-throughput sequencing data.	1.19	<a href="https://www.htslib.org/">https://www.htslib.org/</a>
Reference alignment, coverage and stats	weeSAM	Provides information on breadth and depth coverage of a SAM file.	1.6	<a href="https://github.com/centre-for-virus-research/weeSAM/tree/master">https://github.com/centre-for-virus-research/weeSAM/tree/master</a>

Reference alignment, coverage and stats	Tablet	Lightweight, high-performance graphical viewer for next generation sequence assemblies and alignments.	1.17.08 .17	<a href="https://github.com/cropgeeks/tablet">https://github.com/cropgeeks/tablet</a>
Reference alignment, coverage and stats	iVar	Computational package that contains functions broadly useful for viral amplicon-based sequencing.	1.4.2	<a href="https://github.com/andersen-lab/ivar">https://github.com/andersen-lab/ivar</a>
Reference alignment, coverage and stats	LoFreq	Sensitive variant calling from sequencing data	2.1.5	<a href="https://github.com/CSB5/lofreq">https://github.com/CSB5/lofreq</a>
Reference alignment, coverage and stats	Qualimap	Examines sequencing alignment data in SAM/BAM files and maps reads.	2.3	<a href="https://github.com/refiner-y-platform/qualimap2">https://github.com/refiner-y-platform/qualimap2</a>
Reference alignment, coverage and stats	Prinseq	Helps pre-processing of genomic and metagenomic sequence data in FASTA or FASTQ format.	0.20.4	<a href="https://github.com/uwb-linux/prinseq">https://github.com/uwb-linux/prinseq</a>
Introduction to metagenomics	SPAdes	Small genome assembler uses de novo sequencing	3.15.5	<a href="https://github.com/ablab/spades">https://github.com/ablab/spades</a>
Introduction to metagenomics	Centrifuge	Classifier for metagenomic sequences	1.0.4	<a href="https://github.com/DaehwanKimLab/centrifuge">https://github.com/DaehwanKimLab/centrifuge</a>
Introduction to metagenomics	Seqtk	Tool for processing sequences in the FASTA or FASTQ format	1.4	<a href="https://github.com/lh3/seqtk">https://github.com/lh3/seqtk</a>
Phylogenetic Analysis	Mafft	Align multiple amino acid or nucleotide sequences.	7.52	<a href="https://github.com/GSLBioTech/mafft">https://github.com/GSLBioTech/mafft</a>
Phylogenetic Analysis/Phylogenetics	MEGA-X	Tool for conducting sequence alignment and inferring phylogenetic trees	11.0.13 -1	<a href="https://www.megasoftware.net/">https://www.megasoftware.net/</a>
Phylogenetic Analysis	Modeltest-ng	Tool for selecting the best-fit model of evolution for DNA and protein alignments	0.1.7	<a href="https://github.com/ddarriba/modeltest">https://github.com/ddarriba/modeltest</a>
Phylogenetic Analysis	IQ-TREE	Infer phylogenetic trees by maximum likelihood.	2.2.6	<a href="http://www.iqtree.org/">http://www.iqtree.org/</a>
Phylogenetic Analysis/Phylogenetics	Figtree	Graphical viewer of phylogenetic trees	1.4.4	<a href="http://tree.bio.ed.ac.uk/software/figtree/">http://tree.bio.ed.ac.uk/software/figtree/</a>
Phylogenetic Analysis	Entrez utilities	Interface into the Entrez query and database	16.2	<a href="https://www.ncbi.nlm.nih.gov/books/NBK25497/">https://www.ncbi.nlm.nih.gov/books/NBK25497/</a>
File Formats & QC	FastQC	Quality control checks on raw sequence data	0.12.1	<a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a>
File Formats & QC	Trim Galore	A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to	0.6.10	<a href="https://github.com/FelixKrueger/TrimGalore">https://github.com/FelixKrueger/TrimGalore</a>

		FastQ files, with extra functionality for RRBS data		
Variant calling	unzip	A conda-smithy repository for unzip	6	<a href="https://github.com/conda-forge/unzip-feedstock">https://github.com/conda-forge/unzip-feedstock</a>
De novo	IVA	De novo assembler designed to assemble virus genomes that have no repeated sequences	1.0.11	<a href="https://github.com/sanger-pathogens/iva">https://github.com/sanger-pathogens/iva</a>
De novo	khmer	Software which has command line tools for working with DNA shotgun sequencing data	2.1.2	<a href="https://github.com/dib-lab/khmer">https://github.com/dib-lab/khmer</a>
De novo	quast	Quality Assessment tool, evaluates genome/metagenome assemblies.	5.2.0	<a href="https://github.com/ablab/quast">https://github.com/ablab/quast</a>
De novo	Spades	Open source genome assembler tool for de novo sequencing		<a href="https://github.com/ablab/spades">https://github.com/ablab/spades</a>
De novo	Trimmomatic	Read trimming tool for illumina NGS Data	0.39	<a href="https://github.com/usadellab/Trimmomatic">https://github.com/usadellab/Trimmomatic</a>
Phylodynamics	Tempest	Tool for investigating the temporal signal and 'clocklikeness' of molecular phylogenies.	2.0.3	<a href="https://github.com/beast-dev/Tempest/blob/main/README.md">https://github.com/beast-dev/Tempest/blob/main/README.md</a>
Phylodynamics	Tracer	Program for analysing the trace files generated by Bayesian MCMC runs	1.7.2	<a href="https://github.com/beast-dev/tracer">https://github.com/beast-dev/tracer</a>
Phylodynamics	BEAST (includes BEAGLE, TreeAnnotator and BEAUti plugin)	High-performance library that can perform the core calculations at the heart of most Bayesian and Maximum Likelihood phylogenetics package.	1.10.4.0	<a href="https://github.com/bioDS/beast-phylonco">https://github.com/bioDS/beast-phylonco</a> Also refer for documentation: <a href="https://beast.community/first_tutorial">https://beast.community/first_tutorial</a>
Phylodynamics	Google Earth	Google Earth is a computer program that renders a 3D representation of Earth based primarily on satellite imagery.	7.3.4	<a href="https://maps.google.co.uk/intl/en_uk/earth/">https://maps.google.co.uk/intl/en_uk/earth/</a>

**Note:** The versions might differ as new releases are enrolled by the software company.

## Installation of Python and Python Installer (Anaconda)

### For Windows:

1. Visit the official Python website (<https://www.python.org/>).
2. Navigate to the "Downloads" section and download the latest version of Python for Windows.
3. Run the Installer:
  - Double-click the downloaded installer (e.g., python-3.x.x.exe).
  - Follow the on-screen instructions, selecting "Add Python to PATH" during installation.

4. Verify Installation: Open the Command Prompt and type `python --version` to confirm the Python version.

#### **For macOS:**

1. Install Python using Anaconda: Download the Anaconda distribution from the official website (<https://www.anaconda.com/products/distribution>).
2. Follow the installation instructions for macOS.
3. Verify Installation: In Terminal, type `python --version` to confirm the Python version.

#### **For Linux:**

1. Install Python using Package Manager:
2. Run `sudo apt-get install python3` for Debian/Ubuntu or `sudo yum install python3` for Red Hat/Fedora.
3. Verify Installation: Open Terminal and type `python3 --version` to confirm the Python version.

## **Installing R and RStudio**

#### **For Windows:**

1. Download R Installer: Visit the CRAN website (<https://cran.r-project.org/mirrors.html>).
2. Select a mirror and download the latest version of R for Windows.
3. Run the Installer: Double-click the downloaded installer (e.g., R-4.x.x-win.exe).
4. Follow the on-screen instructions, selecting default settings.
5. Download RStudio from the official website (<https://www.rstudio.com/products/rstudio/download/>).
6. Follow the installation instructions for your operating system.
7. Verify Installation: Open RStudio and check the R version using `version`.

#### **For macOS:**

1. Install Python using Anaconda: Download the Anaconda distribution from the official website (<https://www.anaconda.com/products/distribution>).
2. Follow the installation instructions for macOS.
3. Verify Installation: In Terminal, type `R --version` to confirm the R version.

#### **For Linux:**

1. Install R using Package Manager: Run `sudo apt-get install r-base` for Debian/Ubuntu or `sudo yum install R` for Red Hat/Fedora.
2. Install RStudio: Download RStudio from the official website (<https://www.rstudio.com/products/rstudio/download/>).
3. Follow the installation instructions for your operating system.
4. Verify Installation: Open RStudio and check the R version using `version`.

---

## **Shell Scripting**

#### **Conda Overview:**

Conda is an open-source package management and environment management system that runs on Windows, macOS, and Linux. It simplifies the installation and management of software packages, ensuring dependencies are correctly handled.

#### **Installation Script:**

Ensure you have Conda installed on your system before running the script.

1. Create a Conda Environment (Optional): If you prefer to create a separate environment for these tools, you can do so by running:

```
# Create Conda environment
conda create -n bioinformatics python

# Activate Conda environment
conda activate bioinformatics
```

2. Add Conda Channels: Add necessary channels for accessing bioinformatics packages:

```
# Add necessary Conda channels
conda config --add channels defaults
conda config --add channels bioconda
conda config --add channels conda-forge
```

3. Installations of software and verify:

```
# Install and Verify BWA
conda install bwa
bwa --version

# Install and Verify Bowtie2
conda install bowtie2
bowtie2 --version

# Install and Verify Samtools
conda install samtools
samtools --version

# Install and Verify weeSAM
conda install weesam
weesam --version

# Install and Verify Tablet
conda install tablet
tablet --version
```

```
# Install and Verify iVar
conda install iVar
iVar version

# Install and Verify LoFreq
conda install lofreq
lofreq version

# Install and Verify Qualimap
conda install qualimap2
qualimap --version

# Install and Verify Prinseq
conda install prinseq
prinseq --version

# Install and Verify SPAdes
conda install spades
spades.py --version

# Install and Verify Centrifuge
conda install centrifuge
centrifuge --version

# Install and Verify Seqtk
conda install seqtk
seqtk

# Install and Verify Mafft
conda install mafft
mafft --version

# Install and Verify MEGA-X
conda install mega-x
megacc --version

# Install and Verify Modeltest-ng
conda install modeltest-ng
modeltest-ng --version
```

```
# Install and Verify IQ-TREE
conda install iqtree
iqtree --version

# Install and Verify Figtree
conda install figtree
figtree --version

# Install and Verify FastQC
conda install fastqc
fastqc --version

# Install and Verify Trim Galore
conda install trim-galore
trim_galore --version

# Install and Verify Unzip
conda install unzip
unzip

# Install and Verify IVA
conda install iva
iva --version

# Install and Verify khmer
conda install khmer
interleave-reads.py --version

# Install and Verify Quast
conda install quast
quast.py --version

# Install and Verify Trimmomatic
conda install trimmomatic
trimmomatic --version

# Install and Verify Tempest
conda install tempest
tempest --version
```

```
# Install and Verify Tracer
conda install tracer
tracer --version

# Install and Verify BEAGLE
conda install beagle
beagle --version

# Install and Verify BEAST
conda install beast
beast --version
```

**Note:** In order to install individual software, use the command line command “`conda install softwareName`”. For example: `conda install spades`.

4. If you created a separate environment, deactivate it:

```
# Deactivate Conda environment
conda deactivate
```

**Note:**

- Conda environments provide isolation for different sets of tools, enhancing reproducibility.
- You can customize the script based on your preferences, such as creating a dedicated environment or modifying the channels. Activate the Conda base environment (`conda activate base`) if you choose not to create a separate environment.
- By using Conda, this script streamlines the installation process, ensuring that the specified bioinformatics tools and their dependencies are correctly configured on your system.
- Save this script to a file (e.g., `install_bioinformatics.sh`), make it executable (`chmod +x install_bioinformatics.sh`), and run it in your terminal (`./install_bioinformatics.sh`). Adjust the script as needed for your specific requirements.

---

## Installing Google Earth

### For Windows/Mac:

1. Visit the Google Earth download page ([https://maps.google.co.uk/intl/en\\_uk/earth/](https://maps.google.co.uk/intl/en_uk/earth/)) in your web browser.
2. Click on the "Agree and Download" button to download the installer.
3. Once the download is complete, locate the downloaded file (usually in your Downloads folder) and double-click on it to run the installer.
4. Follow the on-screen instructions provided by the installer. You may be asked to confirm permissions or choose an installation location.
5. After the installation is complete, you should find a Google Earth icon on your desktop or in your Start menu.
6. Double-click the Google Earth icon to launch the application.



**For Linux:**

1. Google Earth for Linux is available as a .deb or .rpm package. Choose the appropriate package for your Linux distribution.
2. Visit the Google Earth download page ([https://maps.google.co.uk/intl/en\\_uk/earth/](https://maps.google.co.uk/intl/en_uk/earth/)) in your web browser.
3. Download the .deb or .rpm package depending on your distribution.
4. Open a terminal and navigate to the directory where the downloaded package is located.
5. Install the package using the package manager for your distribution. For example, on Debian-based systems (like Ubuntu), you can use the following command: `sudo dpkg -i google-earth-pro-stable_current_amd64.deb`
6. Replace "google-earth-pro-stable\_current\_amd64.deb" with the actual name of the downloaded file.
7. Once the installation is complete, you can launch Google Earth from the applications menu or by running google-earth-pro in the terminal.

**Note:** Remember to check the official Google Earth website for any updated installation instructions or changes.

---

## Additional Resources and Troubleshooting

- Ubuntu Documentation: <https://help.ubuntu.com/>
- Ubuntu Community Support: <https://ubuntu.com/support/community-support>
- Conda Installation: <https://conda.io/projects/conda/en/latest/user-guide/install/index.html>
- Conda Documentation: <https://conda.io/projects/conda/en/latest/user-guide/getting-started.html#managing-python>
- Stack overflow: <https://stackoverflow.com/> (Public Q&A platform for debugging)
- Bioinformatics (BioStars) Forum: <https://www.biostars.org/t/Forum/> (General Bioinformatics queries)