
METGENO v3.0

USER GUIDE



Acknowledgment:

This work has been supported by Sabz-Qalam via grant Bioinfo-2019-010.



Sabz-Qalam is a privately held company and its Department of Information Technology currently working in IT consulting, software development, testing and support.

Department of Information Technology
Sabz-Qalam
6 Lodge Road, Old Anarkali
Lahore, Pakistan

Contents

<i>About Metgeno</i>	1
Authors	1
Affiliation	1
Abstract	1
Purpose of this guide	1
Intended Audience	1
License	1
How this Guide is Organized	2
Contact	2
 <i>Chapter-1: Software Installation</i>	3
 <i>Chapter-2: Microsoft Windows Support</i>	18
Requirements:	18
Download:	18
Installation:	18
 <i>Chapter-3: macOS Support</i>	28
Requirements:	28
Download:	28
Installation:	28
 <i>Chapter-4: Selection Criteria</i>	35
 <i>Chapter-5: Packaged Tools</i>	43
 <i>Chapter-6: Software Pipeline</i>	47
 <i>References</i>	49

About Metgeno

Authors

Bilal Wajid^{1,2}, Ali Zafar², Sharoze Meraj², Faria Anwar³, Mustafa Kamal AlShawaqfeh⁴ and Hasan Iqbal²

Affiliation

¹Department of Electrical Engineering, University of Engineering and Technology, Lahore, Pakistan.

²Department of Information Technology, Sabz-Qalam, Lahore 54000, Pakistan.

³Department of Emergency, Mayo Hospital, Lahore, Pakistan.

⁴Department of Electrical Engineering, German Jordanian University, Jordan.

Abstract

The main purpose of this document is to provide a detailed description of the Metgeno installation and usage. The primary reason behind making Metgeno is to provide a free platform for downloading and installing Metagenomics tools on Ubuntu systems. Metgeno is a GUI based software package which installs Metagenomics tools and their respective dependencies.

Purpose of this guide

This is a guide for all the users who are either unfamiliar with the working and execution of the tools included in Metgeno or are new to the Ubuntu operating system and having trouble to understand the interface.

Intended Audience

This document is intended for users who install Metgeno and want to run pipeline analysis on data. Some familiarity with basic Linux operating system commands on behalf of user is assumed. Also, we assume that your hardware meets the requirements for installing Metgeno.

License

This program is free software: you can redistribute it and/or modify it under the terms of the GNU Affero General Public License v3.0 as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU Affero General Public License for more details: <https://www.gnu.org/licenses/agpl-3.0>

How this Guide is Organized

This guide is organized into sections grouped according to the intended use by the user:

- + *About Metgeno* describes the details of authors, document's purpose and intended audience
- + *Software Installation Process* (Chapter 1) describes how to download and install Metgeno
- + *Virtualization* (Chapter 2) and (Chapter 3) describe a step by step procedure to configure a virtual OS for Metgeno on Windows and MacOS respectively
- + *Selection Criteria* (Chapter 4) has table for Software Selection Criteria
- + *Packaged Tools* (Chapter 5) has table for all the Included Tools
- + *Software Pipeline* (Chapter 6) contain brief definition of software pipeline including necessary diagrams to facilitate learning of the field and has included all the packaged tools with respective categories

Contact

If you have general, non-technical or technical question:

- + Send an email to sharozemehraj@gmail.com or hasaniqbal777@gmail.com

Chapter-1: Software Installation

1. Install Ubuntu operating system of the recommended version on your system.
2. Download **Metgeno** from the website:
<https://github.com/sharozem/Metgeno-bin>
3. Metgeno can also be downloaded from the git command through terminal but for that your system should have installed **git**, if it is not installed you can simply install it by typing the following command in terminal:

```
$ sudo apt install git
```

```
$ git clone https://github.com/sharozem/Metgeno-bin.git
```

Note: Proper Internet Connection is required to download and install Metgeno.

4. Then the files of Metgeno will be downloaded and available in Home menu.

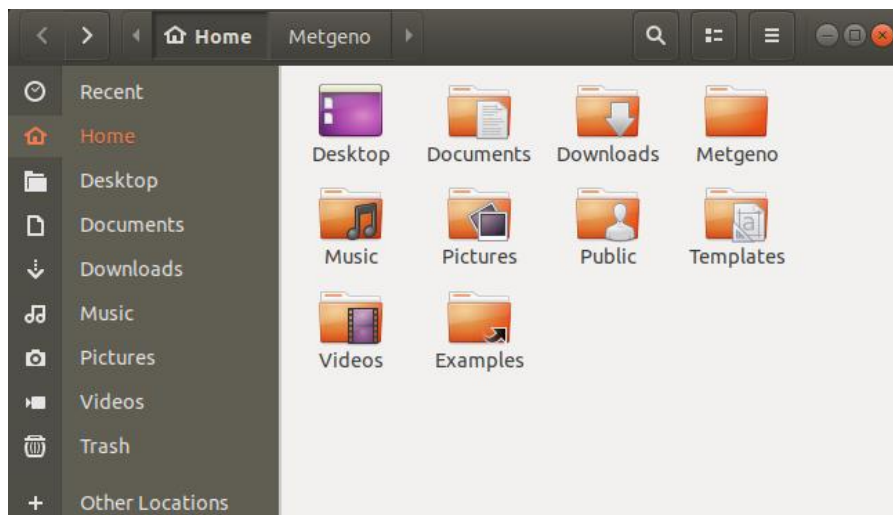


Figure 1.1: Downloaded Files of Metgeno

5. Now run the following commands in terminal:

```
$ cd Metgeno
```

```
$ chmod +x Metgeno_setup_enUS
```

```
$ ./Metgeno_setup_enUS
```


6. If the package manager of your linux distribution is busy the installation will terminate. Wait for the background processes to finish or restart the system to reset all the programs running in the background (recommended).

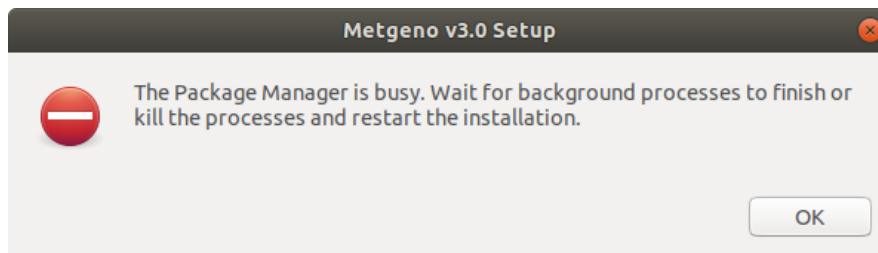


Figure 1.2: Busy Package Manager

7. Enter your superuser credentials to start Metgeno Installation.

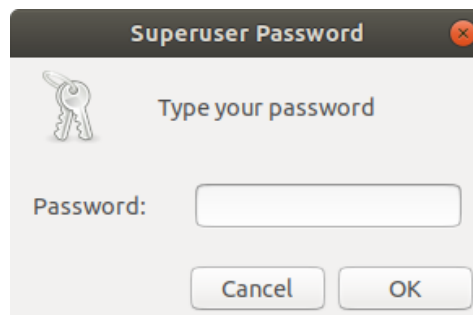


Figure 1.3: Password Screen

8. If the superuser password is incorrect the installation will terminate.



Figure 1.4: Incorrect Password Warning Screen

9. Click **OK** to proceed and confirm the internet connection.



Figure 1.5: Internet Connection Checking Screen

10. Installation is terminated if there is no internet connection.



Figure 1.6: Internet Connection Failed Screen

11. Installation wizard of Metgeno will start.



Figure 1.7: Splash Screen

12. Select you want to install or exit Metgeno.

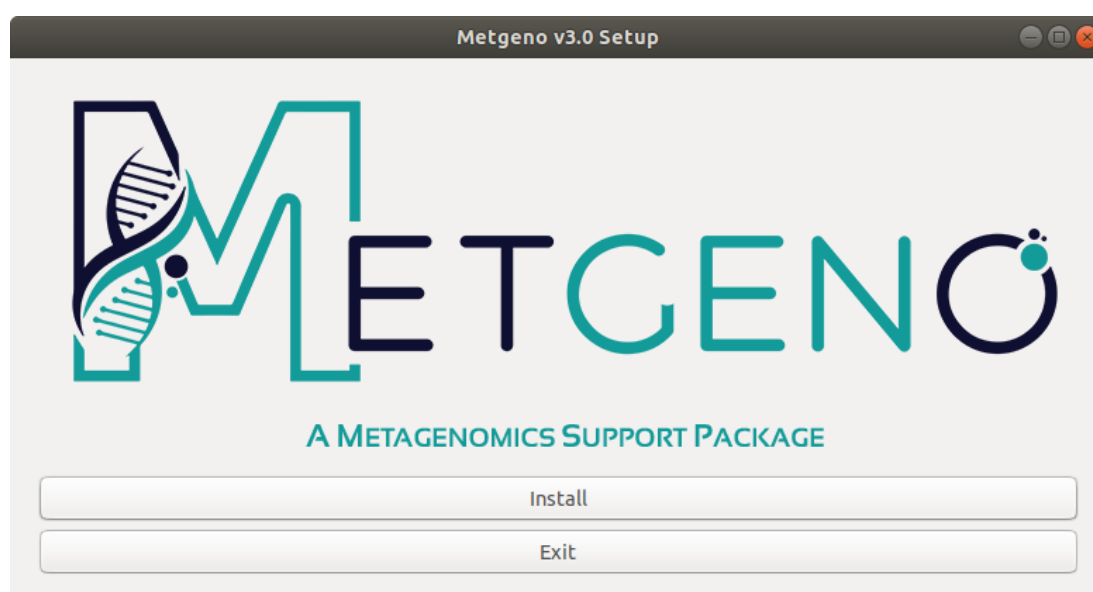


Figure 1.8: Installer Selection Screen

13. Click **Next** to proceed with the installation.

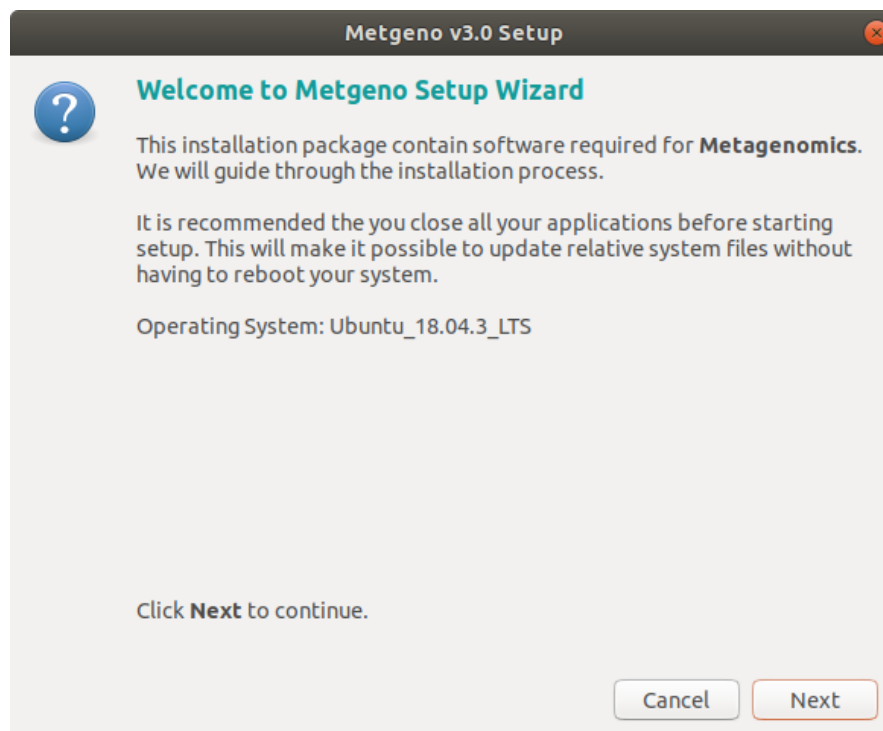


Figure 1.9: Installation Welcome Screen

14. Metgeno pipeline detail screen is displayed. It has the information about which software you want to install. Click **Next** to Proceed.

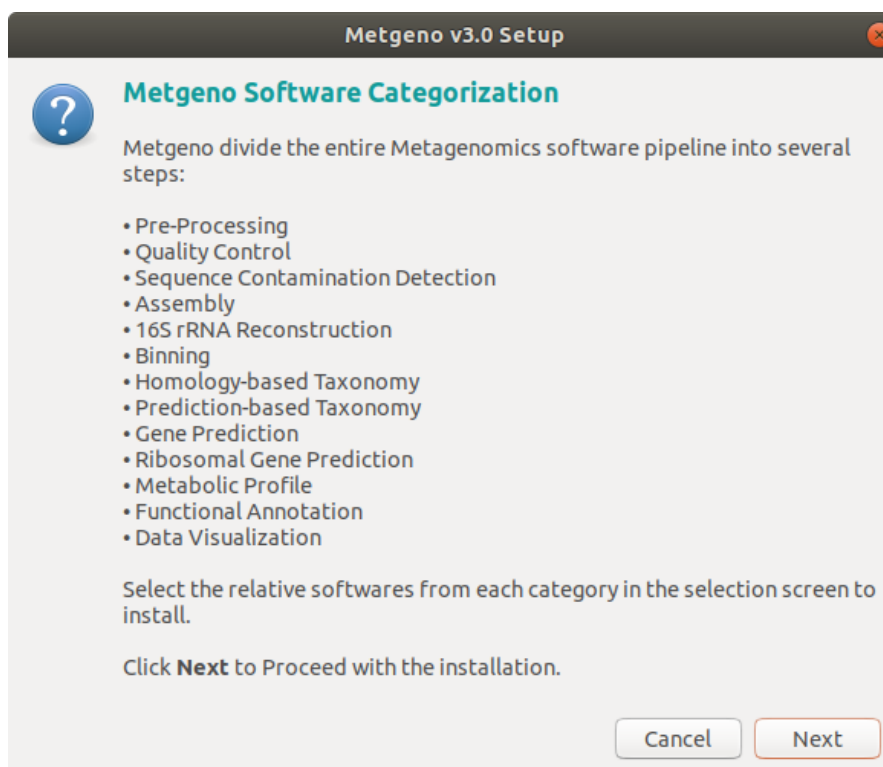


Figure 1.10: Pipeline Information Screen

15. Selection screen for Pre-processing Tools is displayed. Select the required tools and Click **Next**.

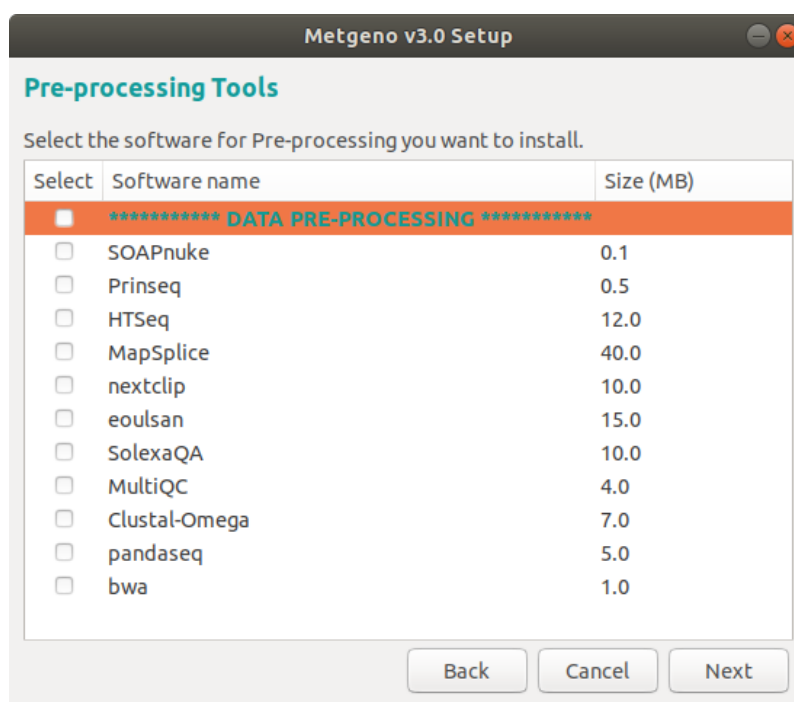


Figure 1.11: Pre-processing Tools Selection Screen

16. Selection screen for Quality Control Tools is displayed. Select the required tools and Click **Next**.

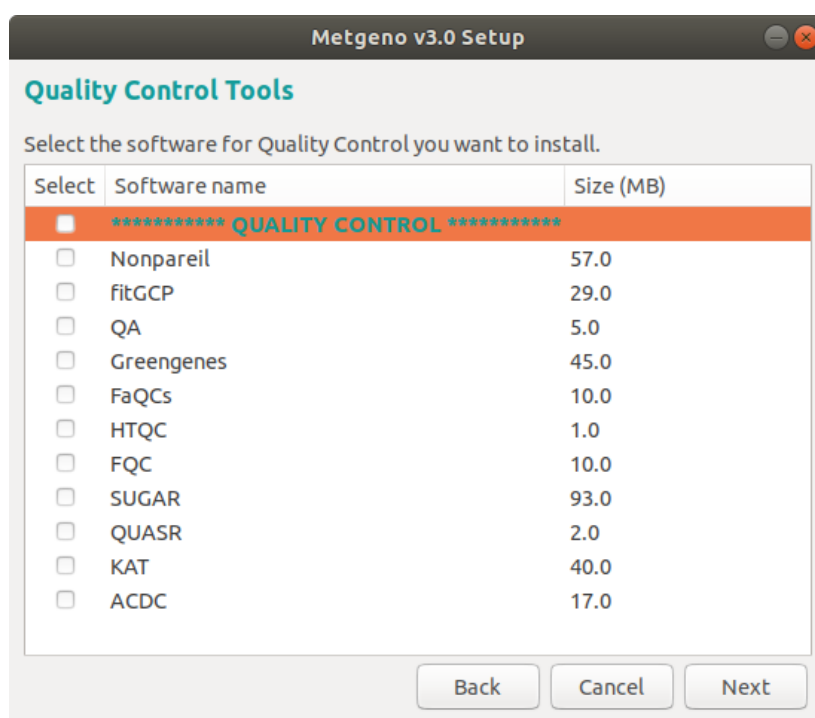


Figure 1.12: Quality Control Tools Selection Screen

17. Selection screen for Sequence Contamination Detection Tools is displayed. Select the required tools and Click **Next**.

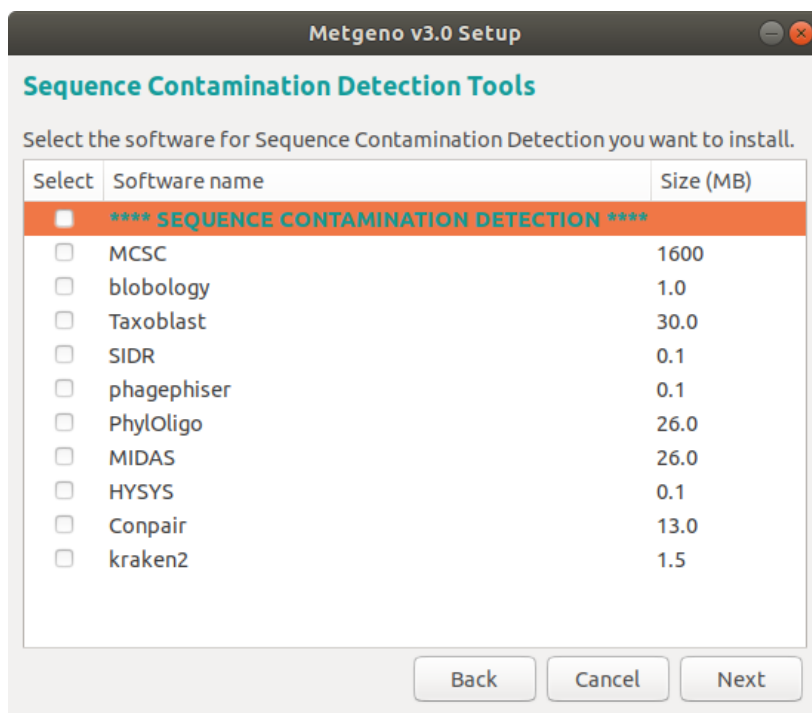


Figure 1.13: Sequence Contamination Detection Tools Selection Screen

18. Selection screen for Assembly Tools is displayed. Select the required tools and Click **Next**.

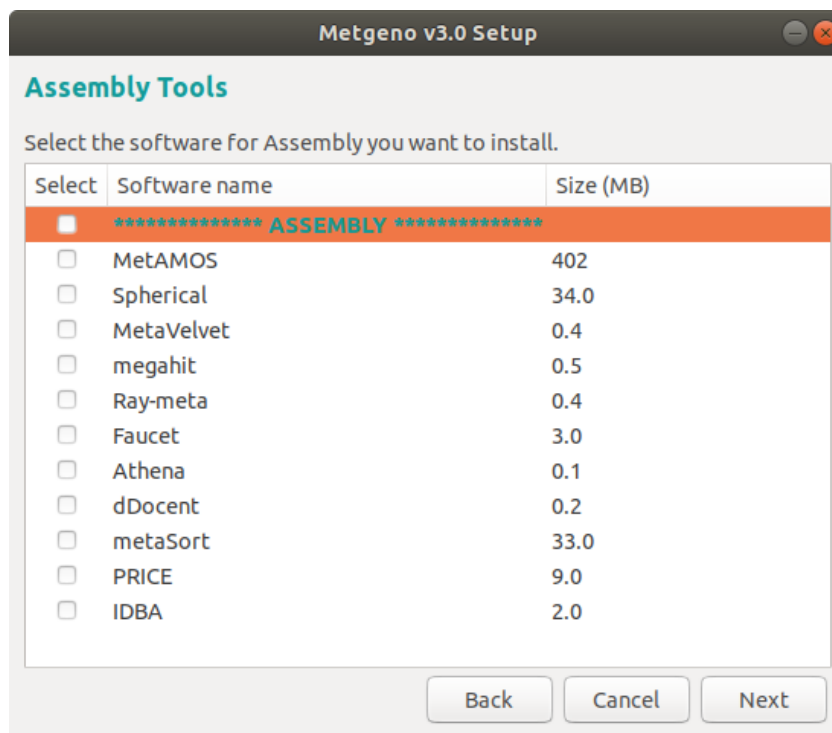


Figure 1.14: Assembly Tools Selection Screen

19. Selection screen for 16S rRNA Reconstruction Tools is displayed. Select the required tools and Click **Next**.

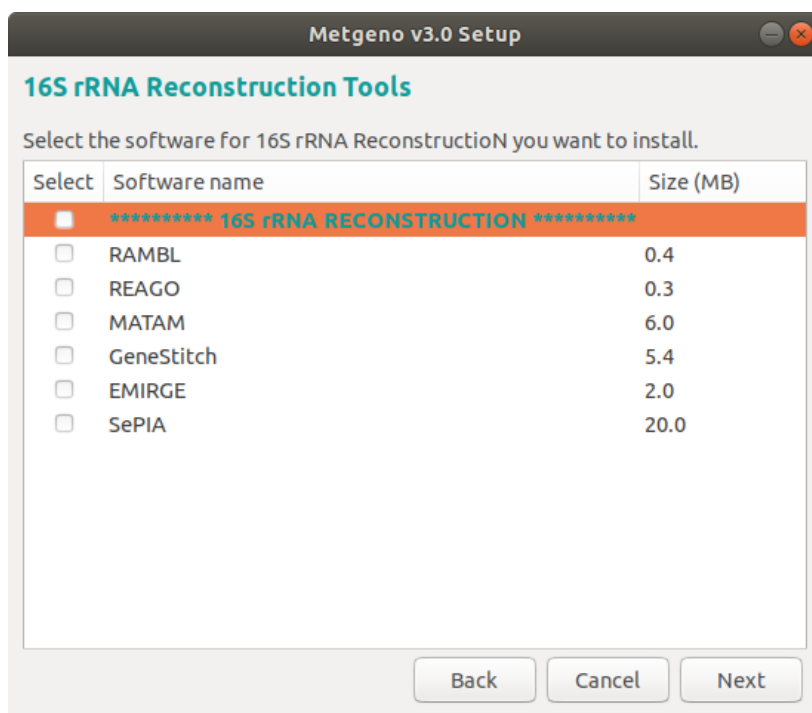


Figure 1.15: 16S rRNA Reconstruction Tools Selection Screen

20. Selection screen for Binning Tools is displayed. Select the required tools and Click **Next**.

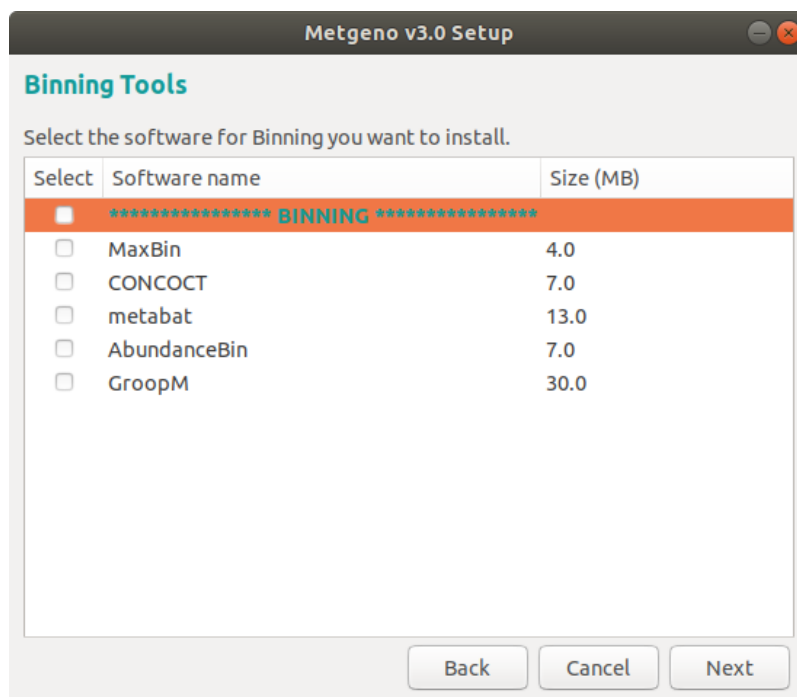


Figure 1.16: Binning Tools Selection Screen

21. Selection screen for Homology-based Taxonomy Tools is displayed. Select the required tools and Click **Next**.

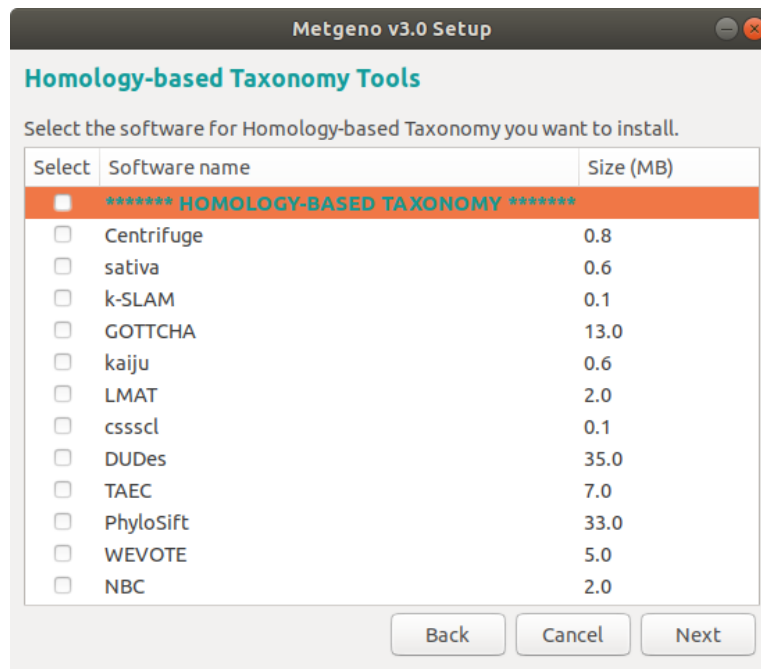


Figure 1.17: Homology-based Taxonomy Tools Selection Screen

22. Selection screen for Prediction-based Taxonomy Tools is displayed. Select the required tools and Click **Next**.

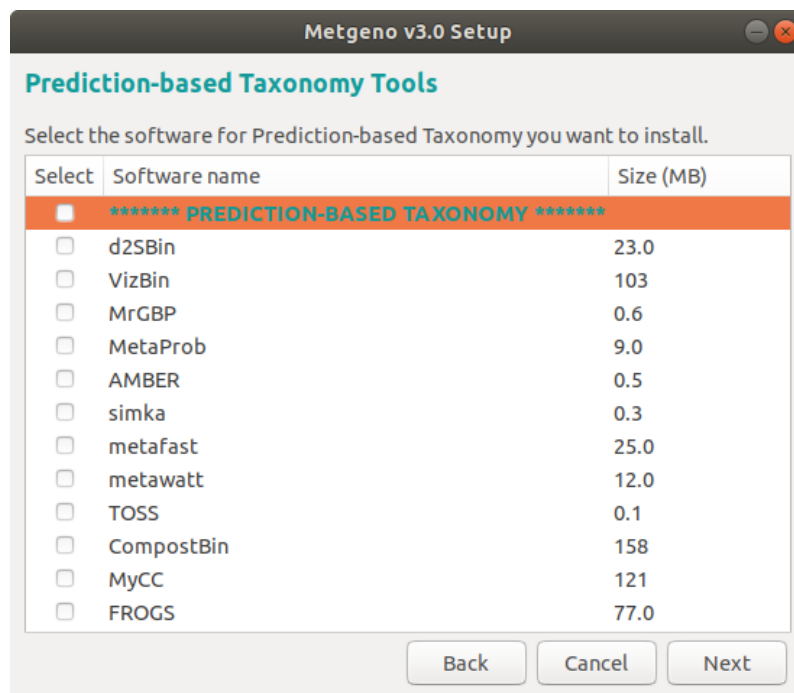


Figure 1.18: Prediction-based Taxonomy Tools Selection Screen

23. Selection screen for Gene Prediction Tools is displayed. Select the required tools and Click **Next**.

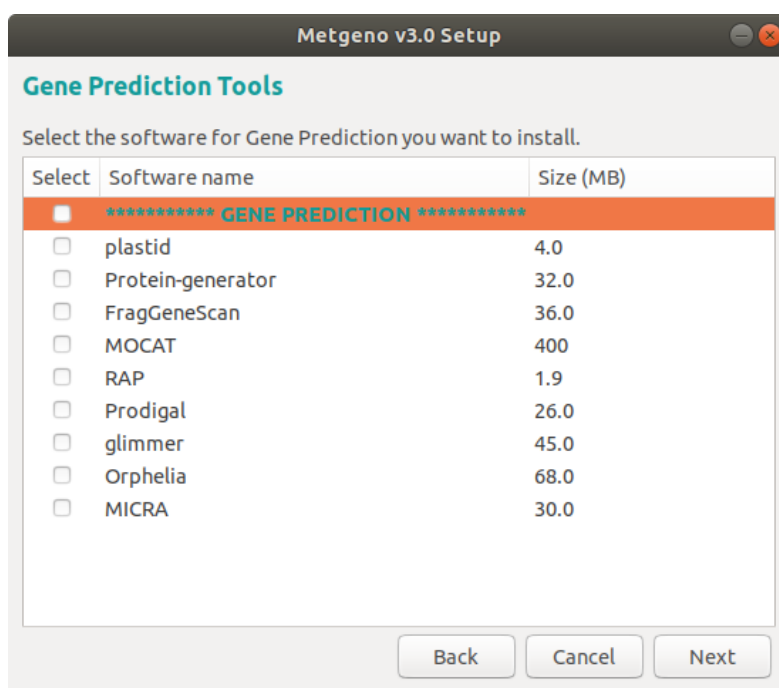


Figure 1.19: Gene Prediction Tools Selection Screen

24. Selection screen for Ribosomal Gene Prediction Tools is displayed. Select the required tools and Click **Next**.

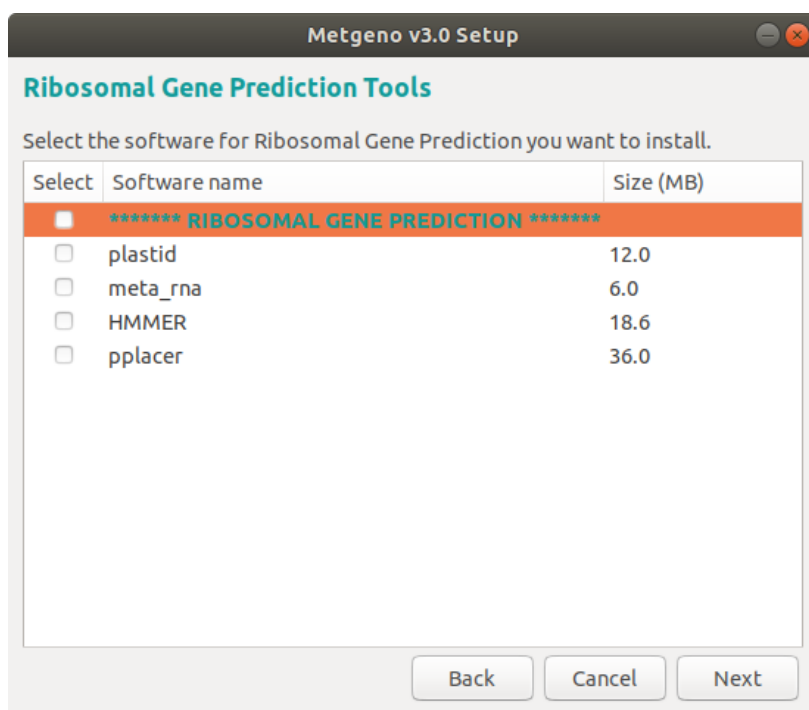


Figure 1.20: Ribosomal Gene Prediction Tools Selection Screen

25. Selection screen for Metabolic Profiling Tools is displayed. Select the required tools and Click **Next**.

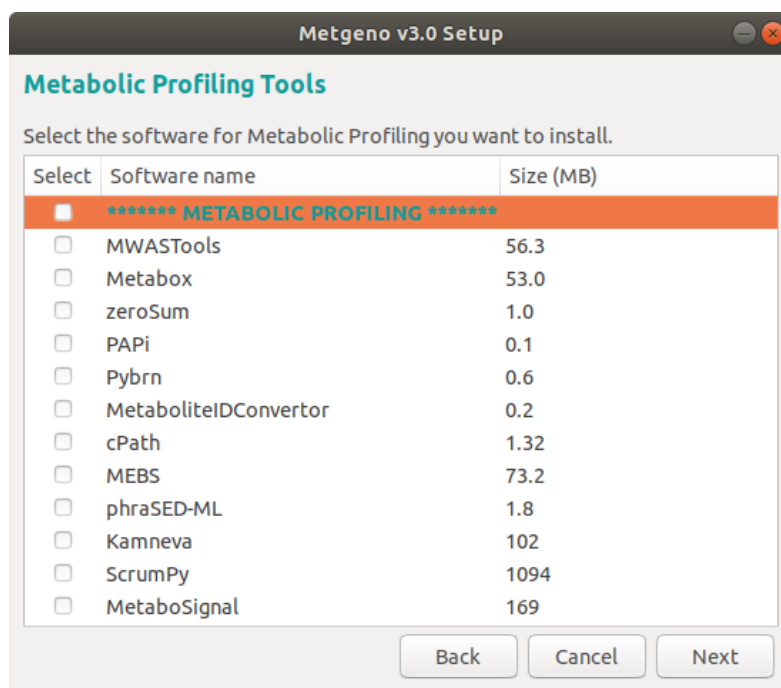


Figure 1.21: Metabolic Profiling Tools Selection Screen

26. Selection screen for Functional Annotation Tools is displayed. Select the required tools and Click **Next**.

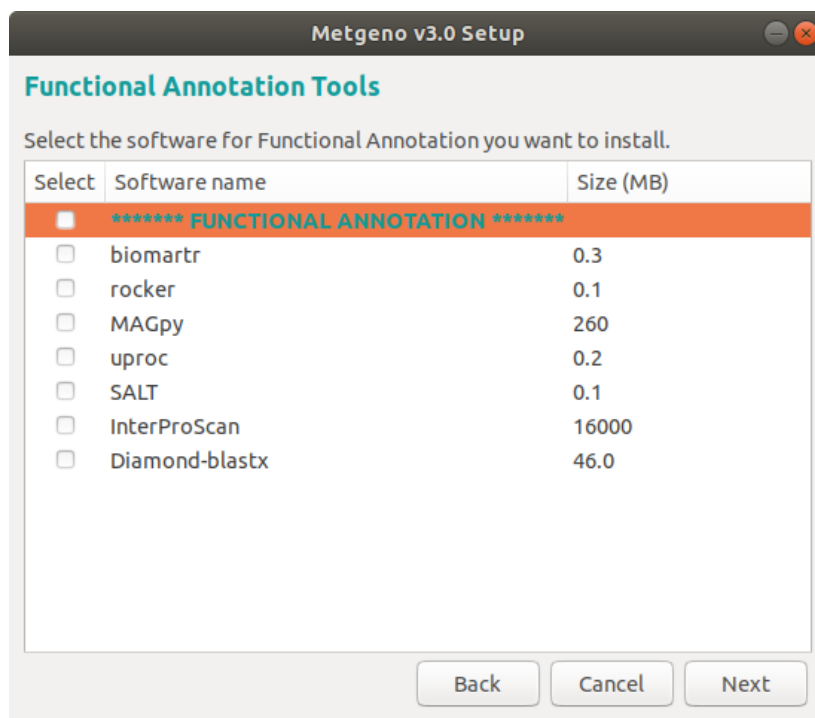


Figure 1.22: Functional Annotation Tools Selection Screen

27. Selection screen for Data Visualization Tools is displayed. Select the required tools and Click **Next**.

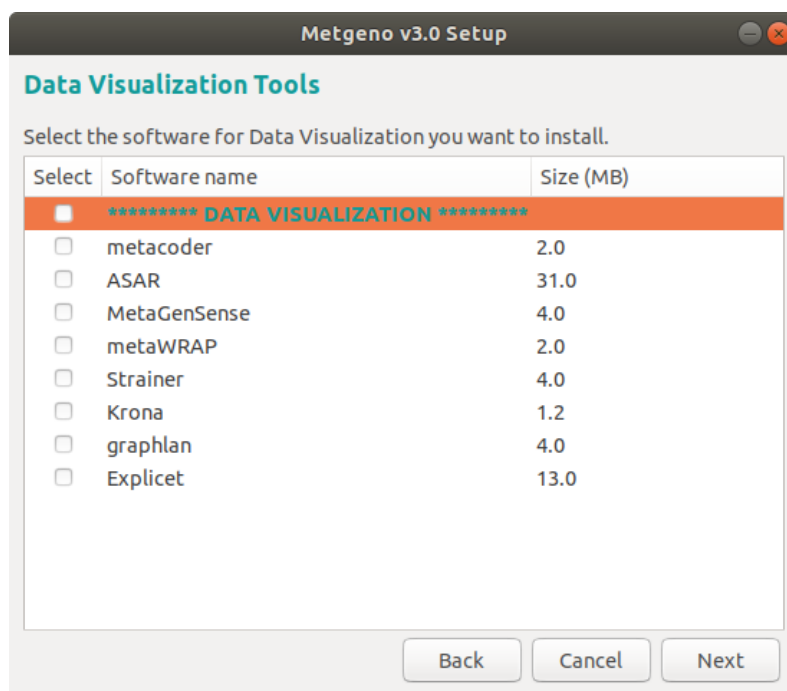


Figure 1.23: Data Visualization Tools Selection Screen

28. Click **Next** to proceed with the installation.

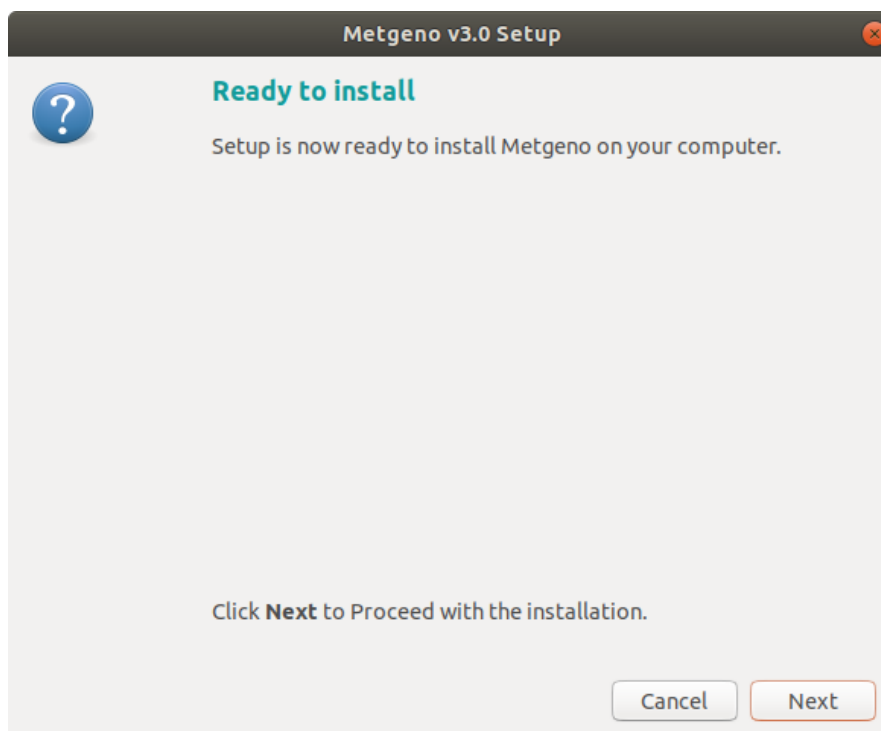


Figure 1.24: Ready to Install Screen

29. If no software tool is selected the installation will Finish.

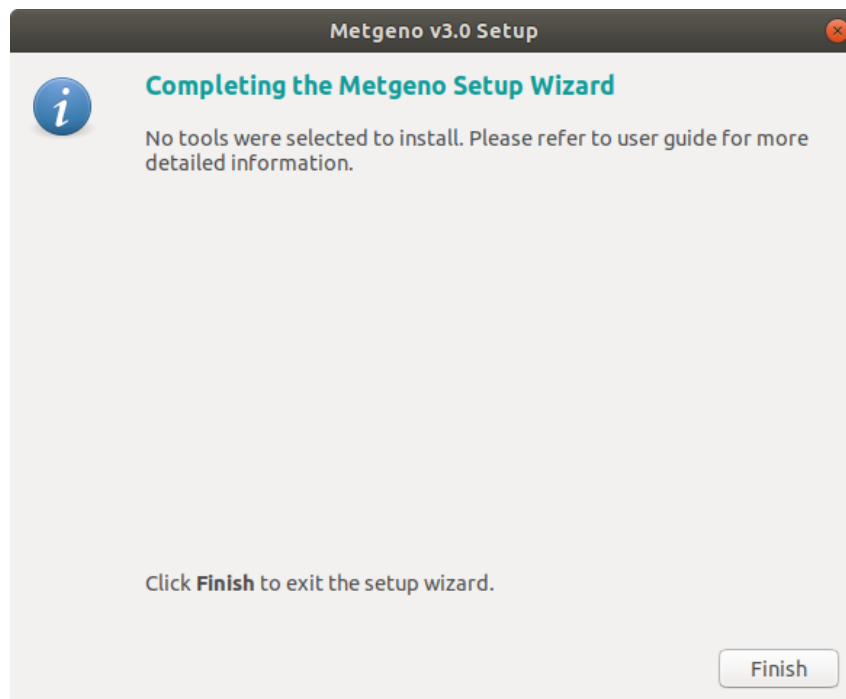


Figure 1.25: No Tools Selected Warning

30. Dependencies related to software will start installing.

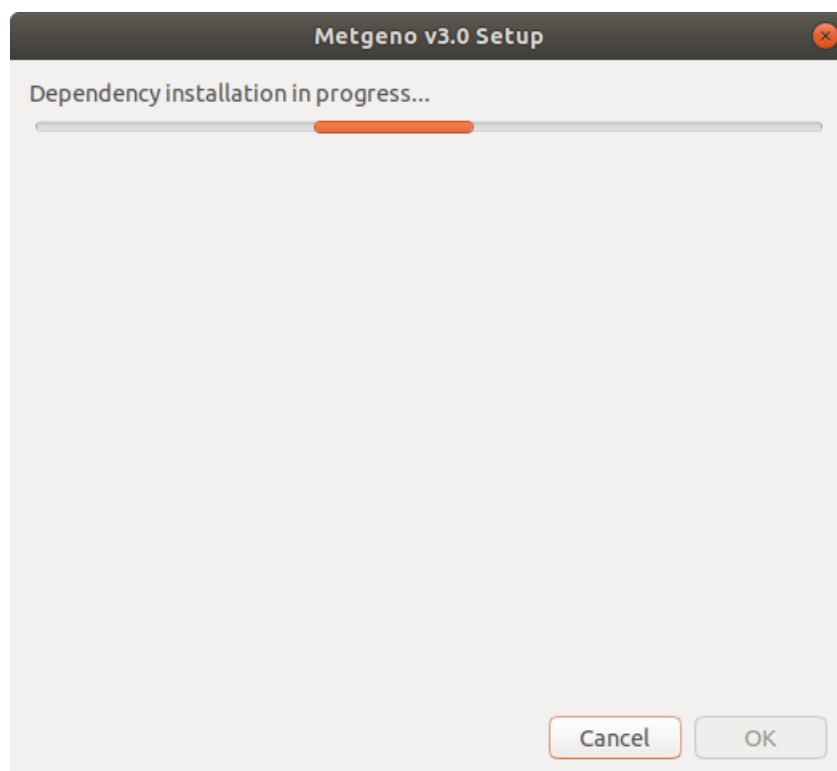


Figure 1.26: Dependencies Installation Screen

31. All the dependencies which are installed are displayed at the end of the installation. Click **OK**.

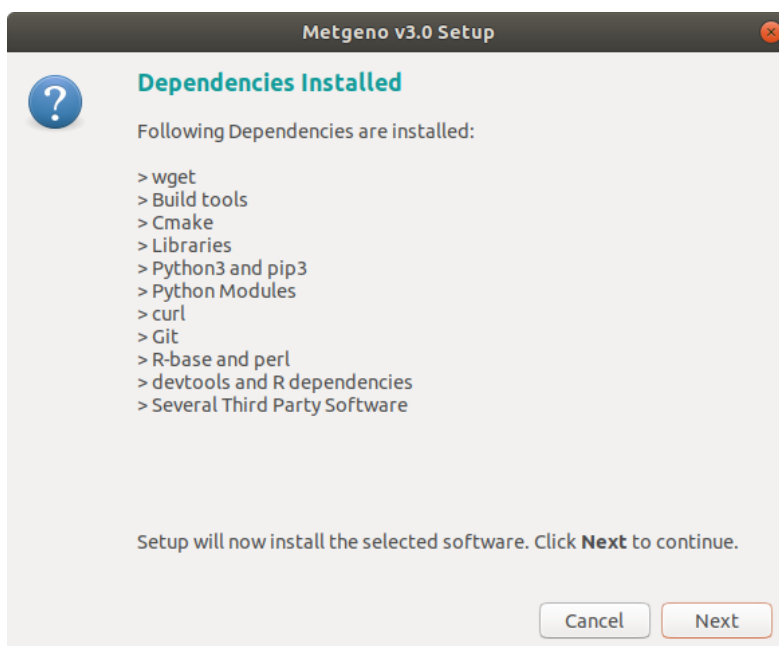


Figure 1.27: Dependencies Installation Finished Screen

32. Tool installation will continue. Some tools install as standalone installation will be called automatically.

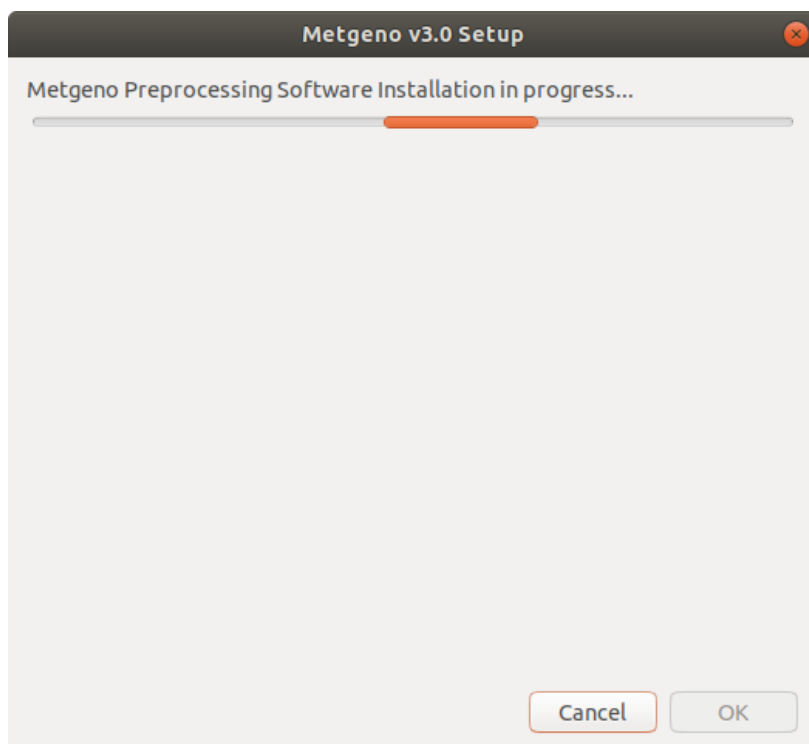


Figure 1.28: Downloading Screen

33. Metgeno will notify you if no software is selected from a category.

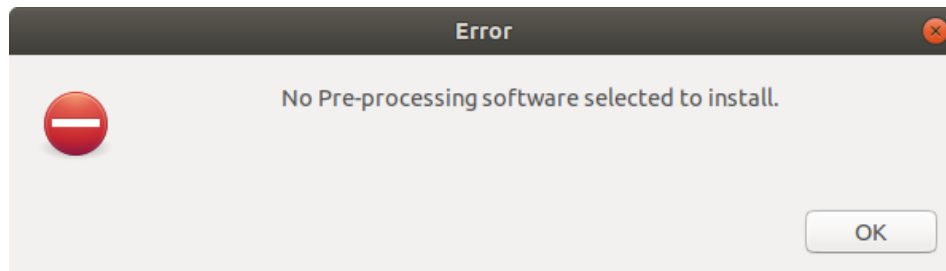


Figure 1.29: Warning Screen

34. Installation of Metgeno is now finished. Click **Finish** to use the tools.

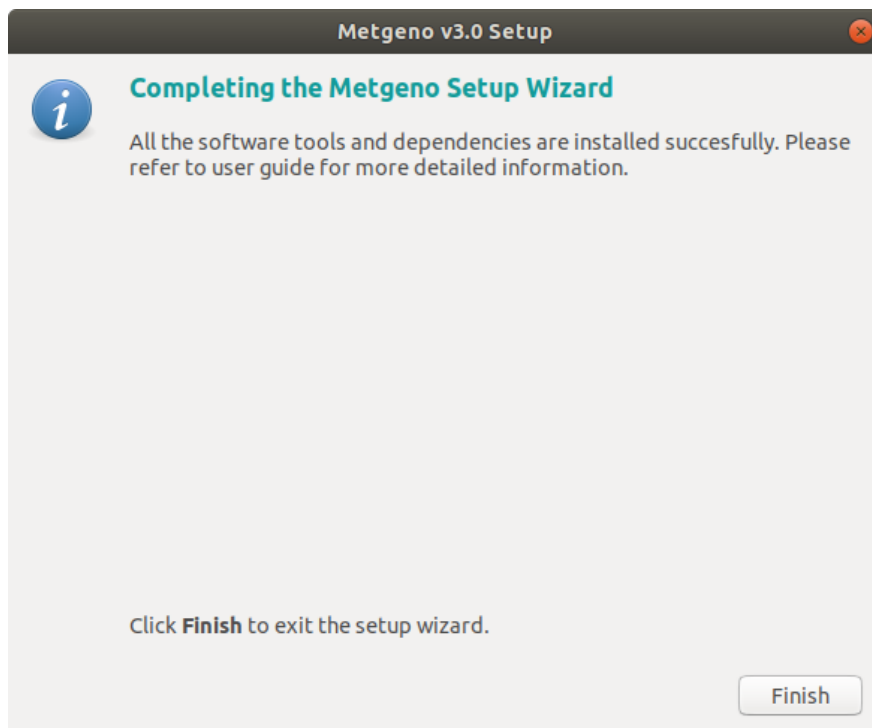


Figure 1.30: Finishing Installation Screen

35. At the end software will show the successfully installed tools and following window will appear:

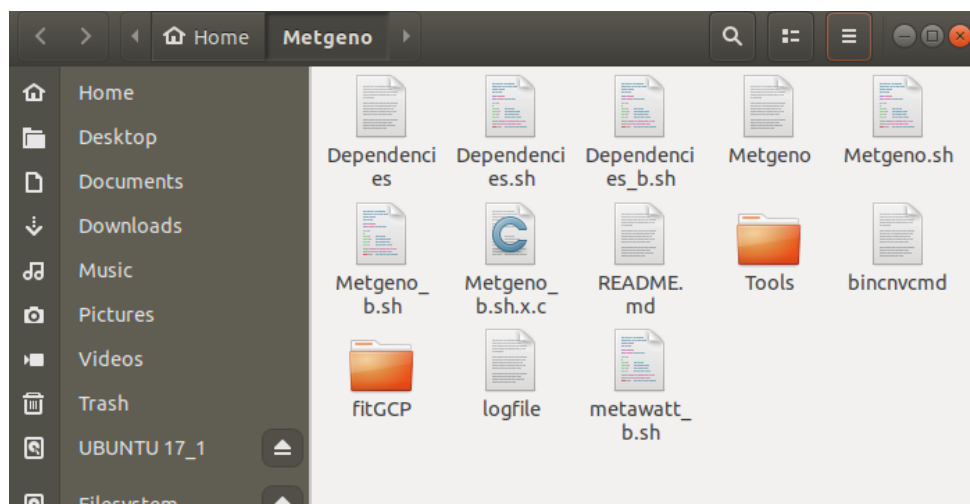


Figure 1.31: Installed Tools showing Screen

Note: See the README file of every software to know *How to use it?*

Chapter-2: Microsoft Windows Support

If you don't want to dual boot Linux/Ubuntu with windows you can use Linux in a virtual machine which gives you the option to try Linux within Windows without making any changes in your Windows system.

VirtualBox is a free and open source virtualization software from Oracle. It enables you to install other operating systems in virtual machines. It is advised that your systems should have at least 4GB RAM to get a decent performance from the virtual operating system.

Requirements:

- ✚ Good internet connection to download software and Linux ISO. You can also use some other computer with internet connection to download these files
- ✚ Windows system with at least 12 GB of free space
- ✚ Windows system with 4GB of recommended RAM. It can work with less RAM as well, but your system will start to lag while using Linux in the virtual machine

Download:

- ✚ Go to the website of Oracle VirtualBox and get the latest stable version
- ✚ Next, you need to download the ISO file of the Linux distribution. You can get this image from the official website of the Linux distribution you are trying to use

Installation:

1. To installing VirtualBox just click on the downloaded file, it is like installing the regular software on Windows.
2. **Start VirtualBox** and click on the **New** symbol.

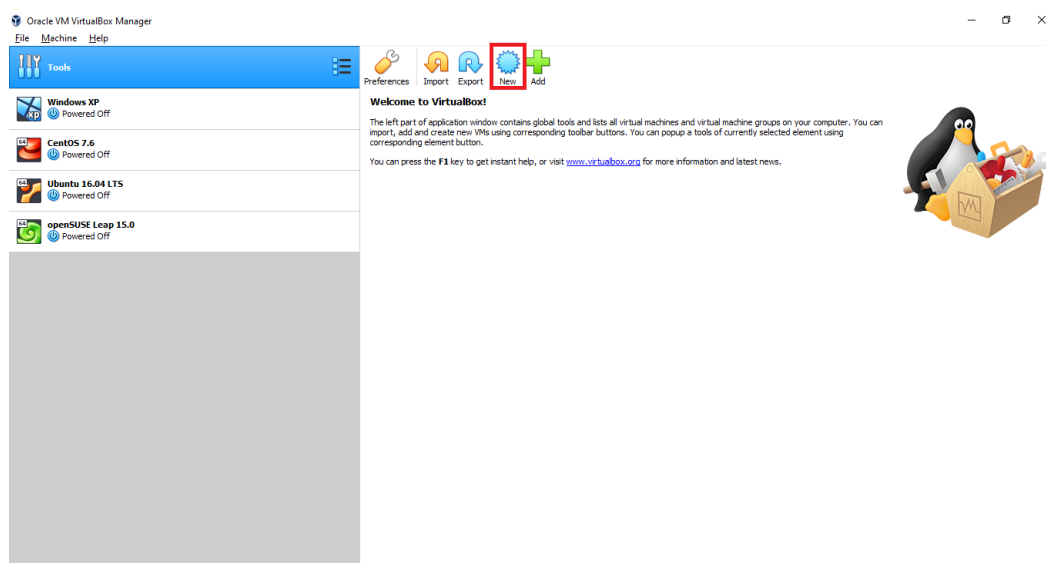
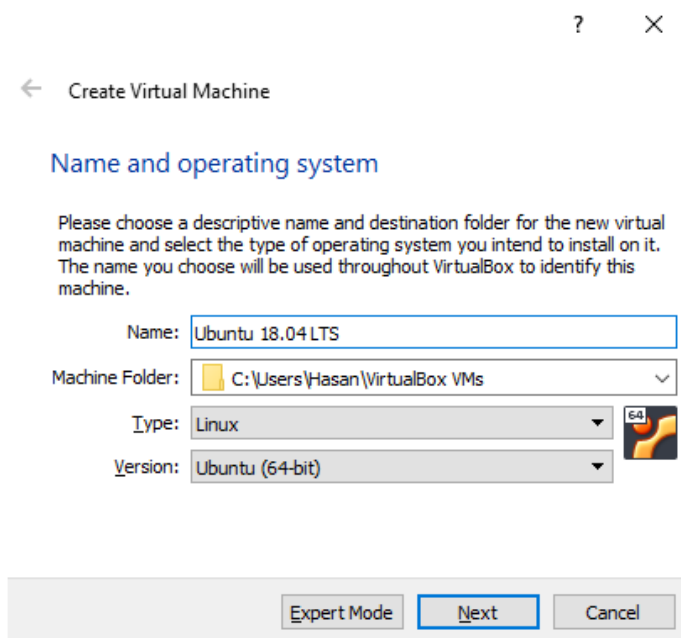
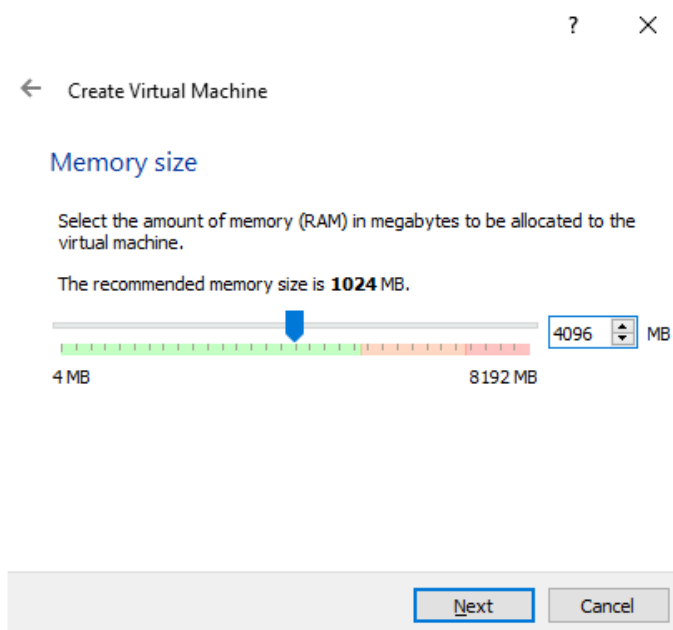


Figure 2.1: New Virtual OS

3. Give the virtual OS a relevant **Name**. Select the **Type** (Linux) and **Version** (Ubuntu 64-bit) and Click **Next**.

*Figure 2.2: Assigning Name*

4. Allocate RAM to the virtual OS. Following system has 8GB of RAM so 4GB of RAM is allocated. You can use more RAM if your system has enough extra RAM.

*Figure 2.3: Setting Memory Size*

5. **Create** a virtual disk. This works as the hard disk of the virtual Linux system. This is where the virtual system will store its files. Click **Create**.

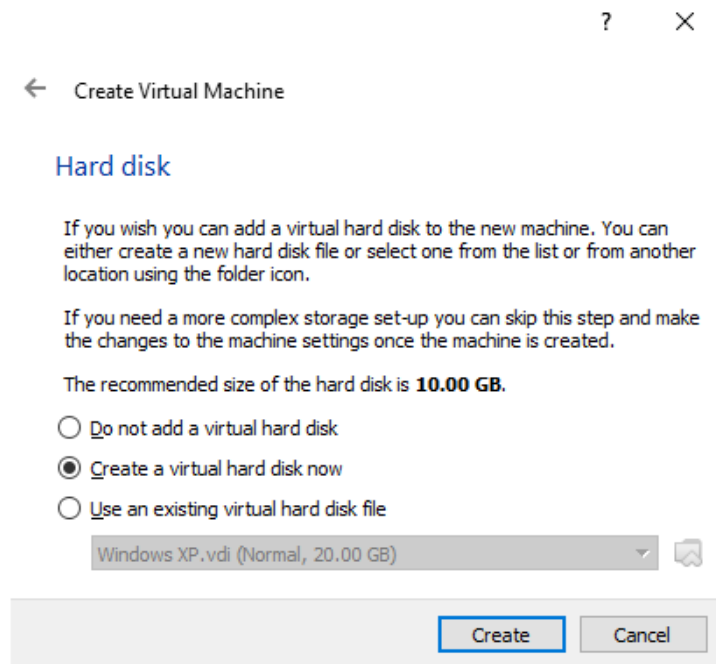


Figure 2.4: Creating a Virtual Hard disk

6. Select **VDI file type** here (recommended). Click **Next**.

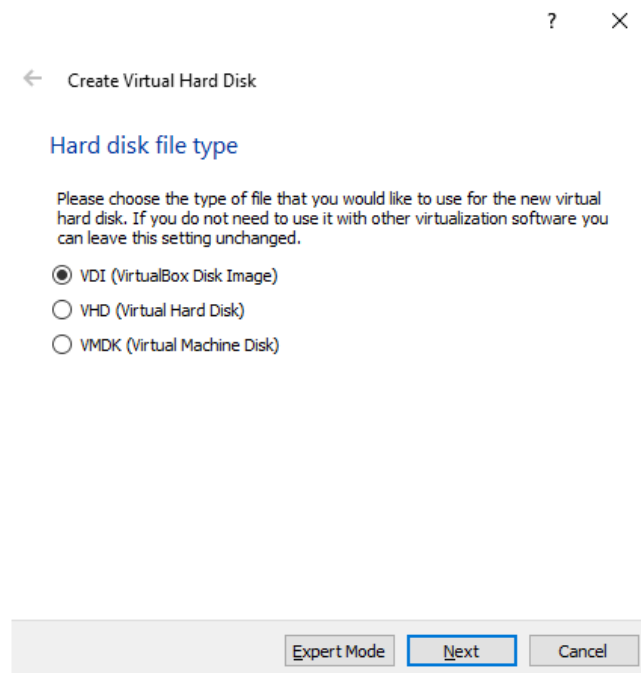


Figure 2.5: Hard disk file type

7. You can choose either of Dynamically allocated or Fixed size option for creating the virtual hard disk. Choose ***Dynamically allocated*** (recommended). Click **Next**.

Note: Dynamic allocation is allocated as time passes and data is increased whereas fixed is allocated instantly.

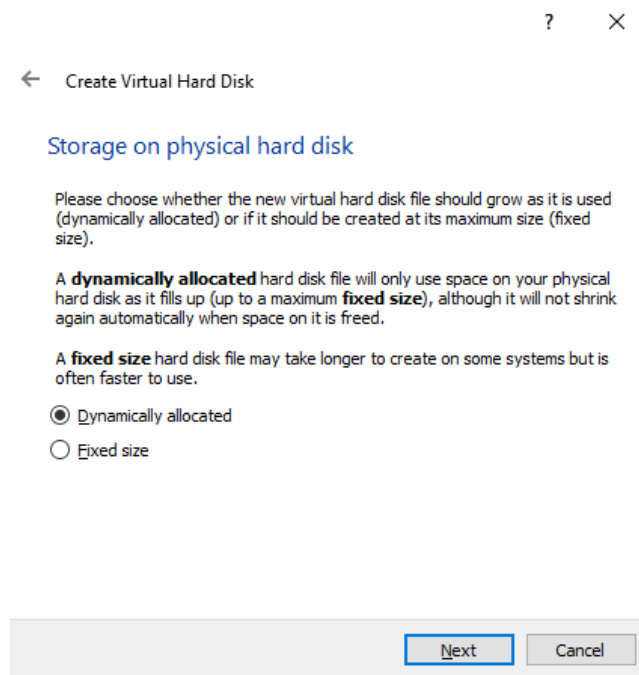


Figure 2.6: Storage type

8. Select the ***Hard Disk size*** (recommended size: 100 GB). Click **Create**.

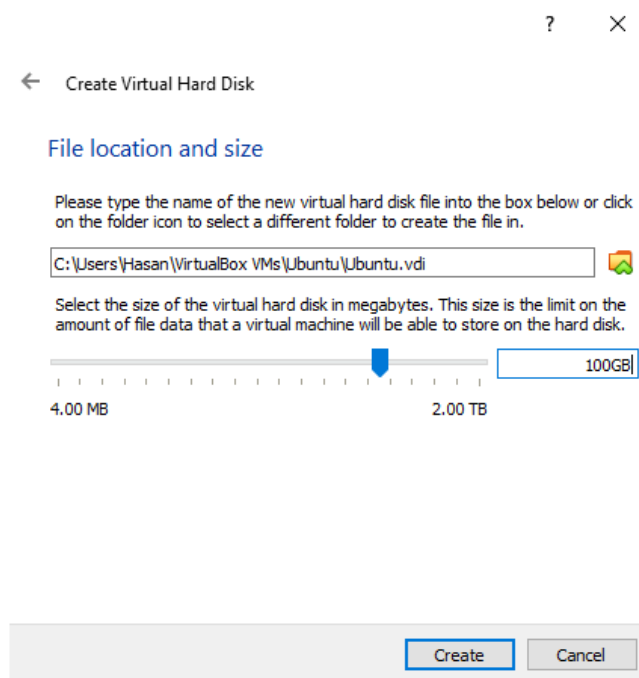


Figure 2.7: File location and Size of Virtual Hard disk

9. Now, Select **Settings** to assign the image file of respective OS to VB.

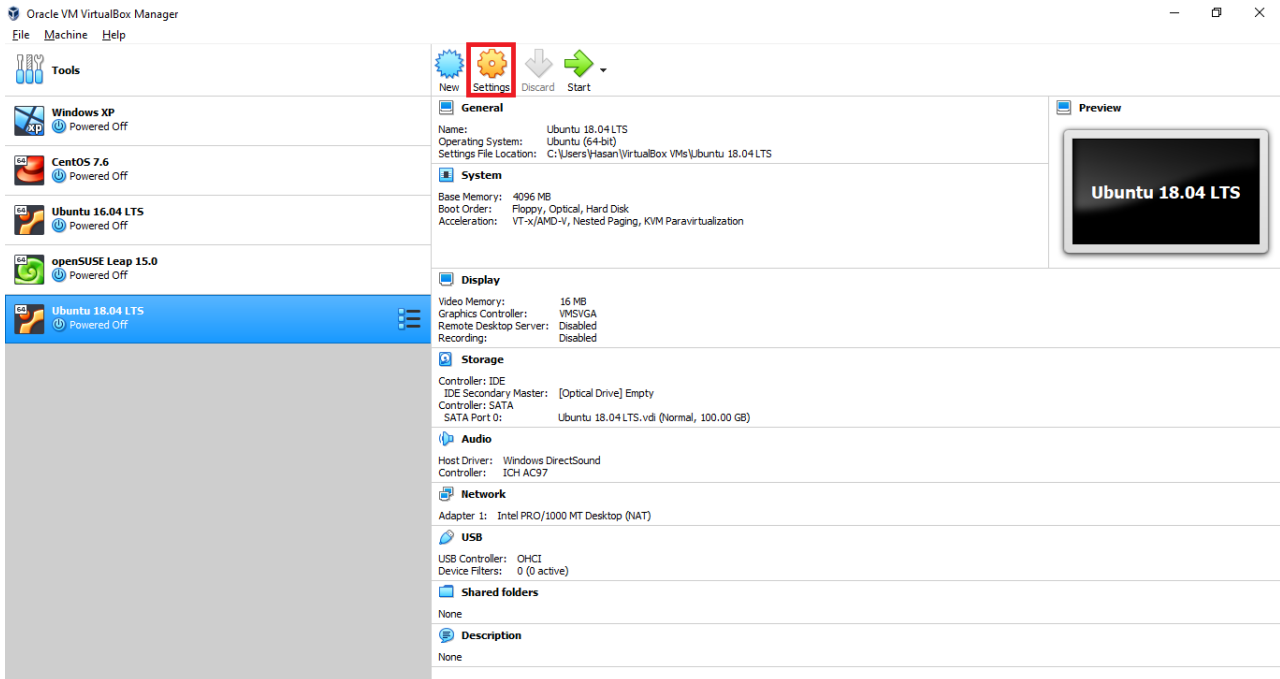


Figure 2.8: Select Settings

10. Select **General** → **Advanced**. Now, select the **Shared Clipboard** and **Drag'n'Drop** option to **Bidirectional**.

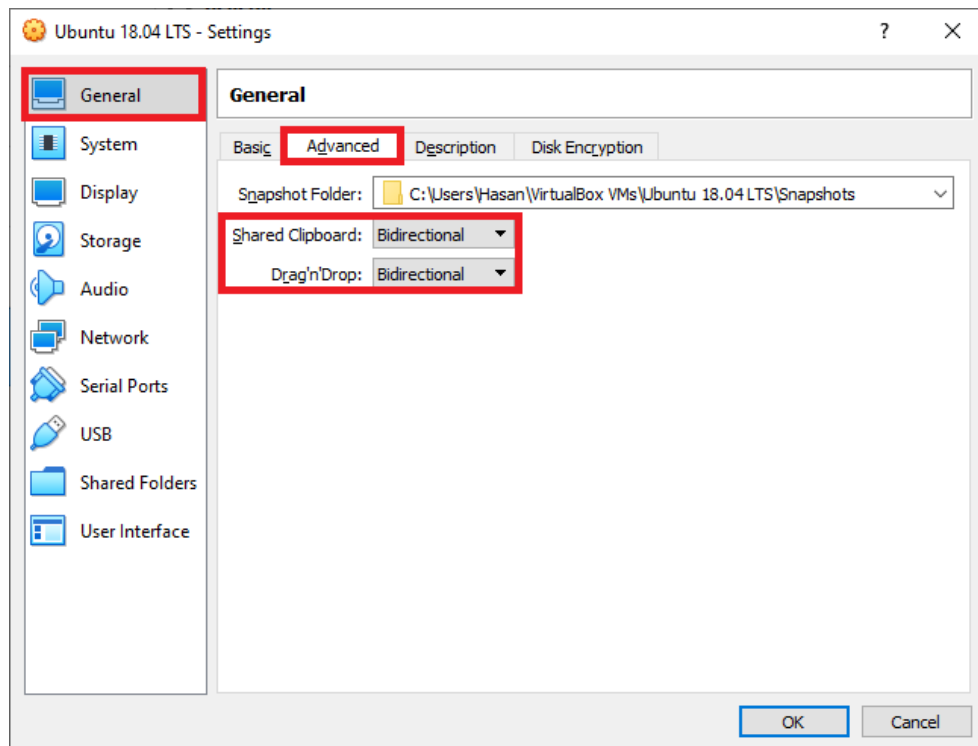


Figure 2.9: Advanced Settings

11. Select **Settings** to assign the image file of respective OS to VB. Select **Storage** → **Controller: IDE** → **Empty**. Now, in the **Attributes** tab, click on **New Disk** and provide the path of downloaded image file of Ubuntu OS. Click **OK**.

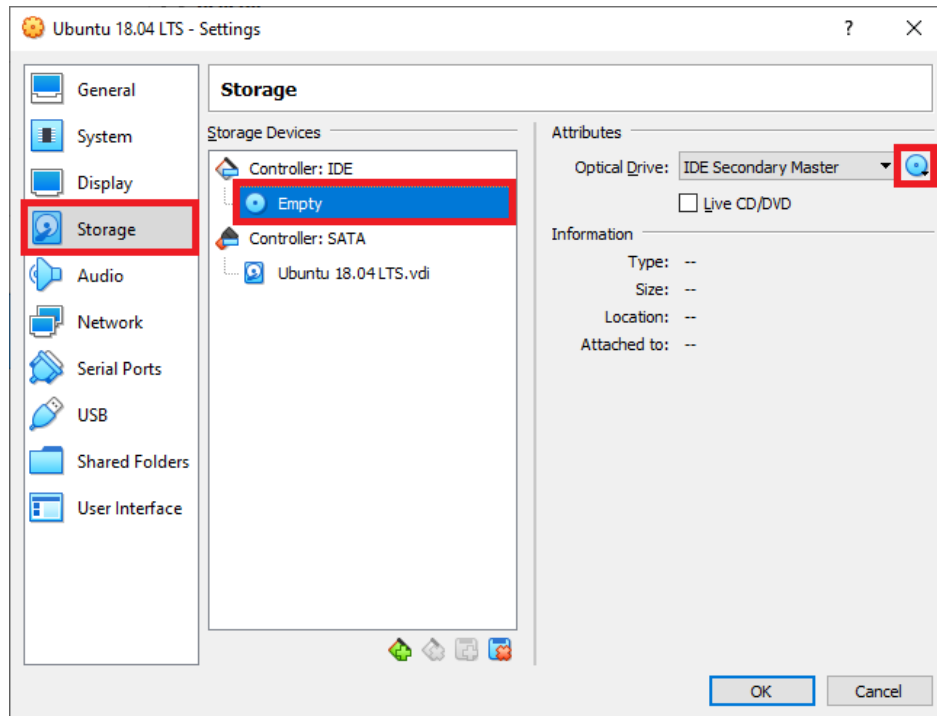


Figure 2.10: Providing Image file

12. Once everything is in place, it's time to boot that ISO and install Linux as a virtual operating system. Click **Start**.

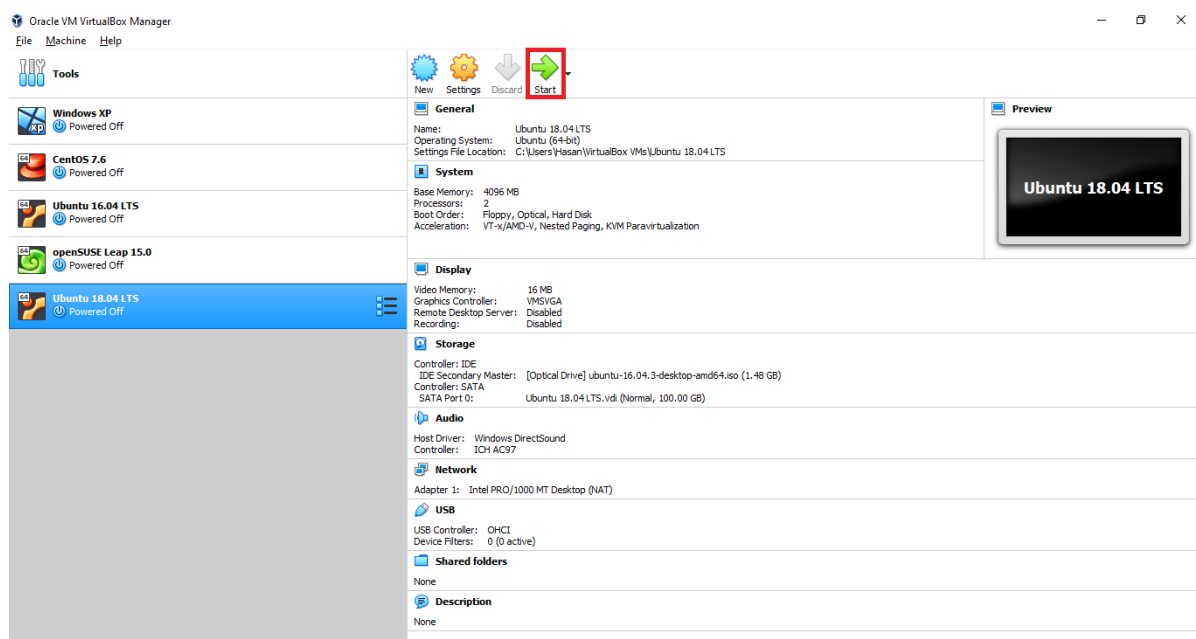


Figure 2.11: Starting Virtual OS

13. Virtual OS will boot into Linux Installation process. You should be presented with the option to install it. Click **Install Ubuntu**.

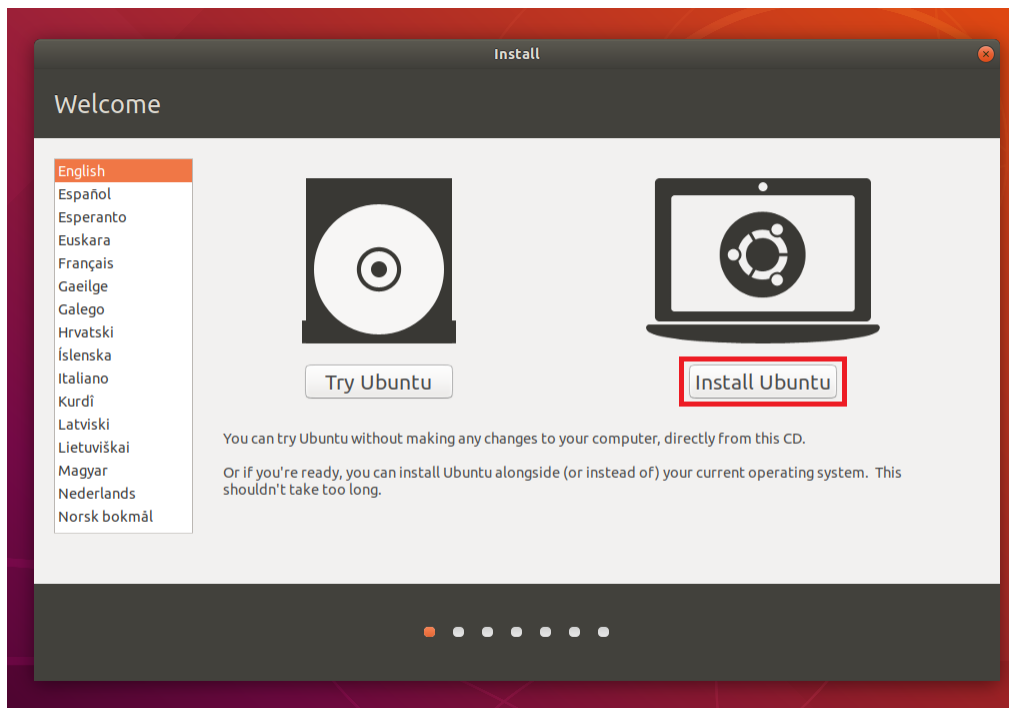


Figure 2.12: Installing Ubuntu

14. Continue with *Normal Installation*.

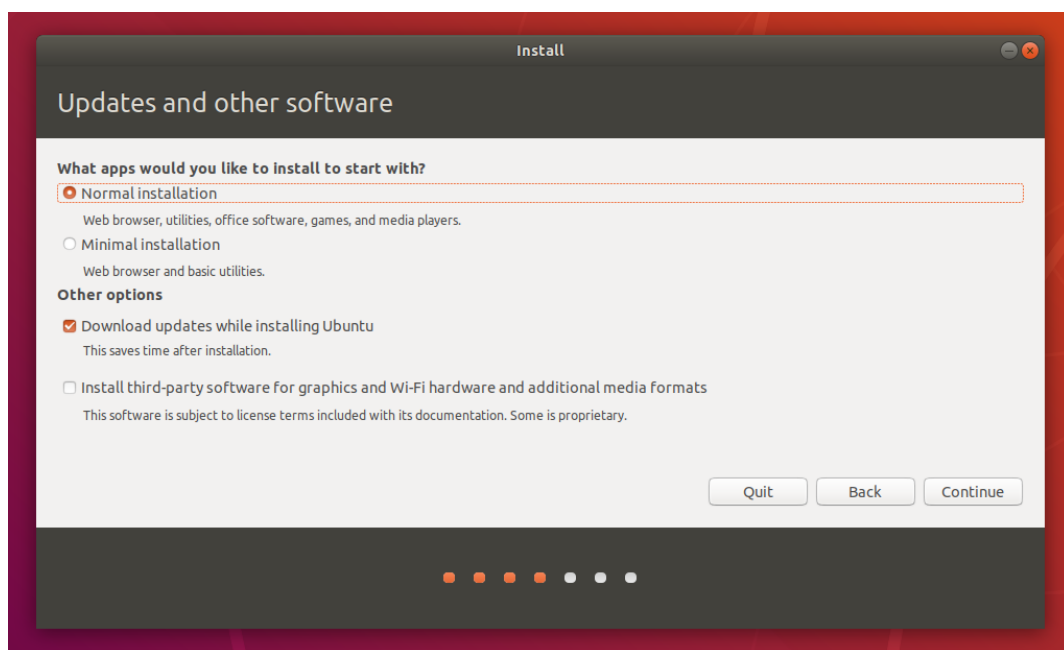


Figure 2.13: Update Screen

15. Select **Erase disk and install Ubuntu**. Don't worry. It won't delete anything on your Windows operating system. You are using the virtual disk space of 100GB that we created in previous steps. It won't impact the real operating system.

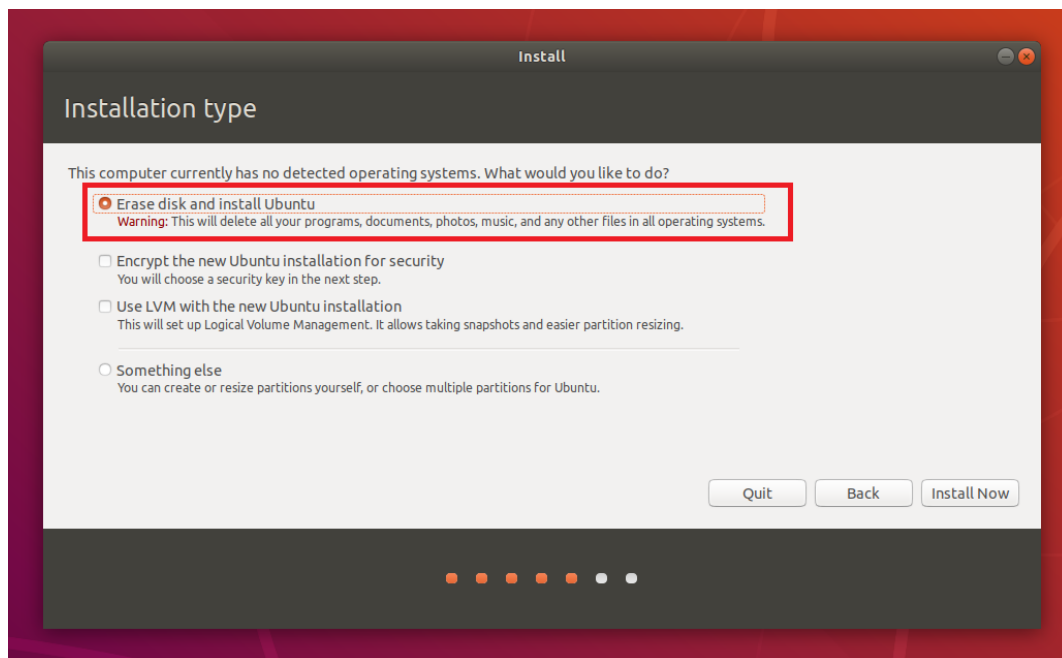


Figure 2.14: Installation type Screen

16. After that click on **Continue**.

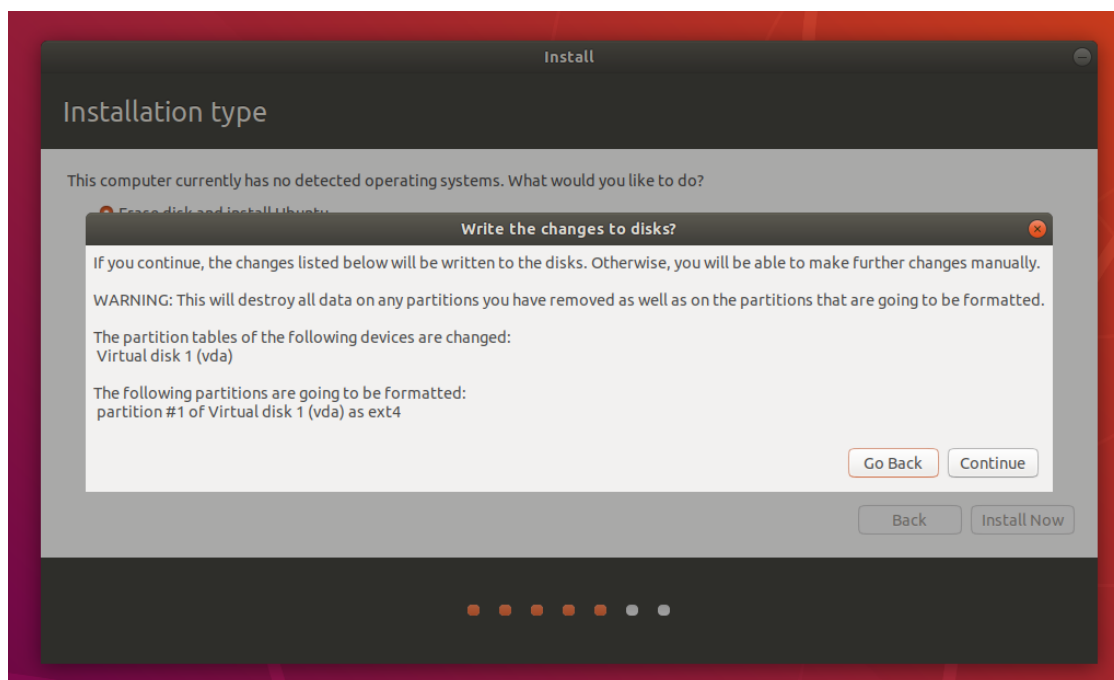
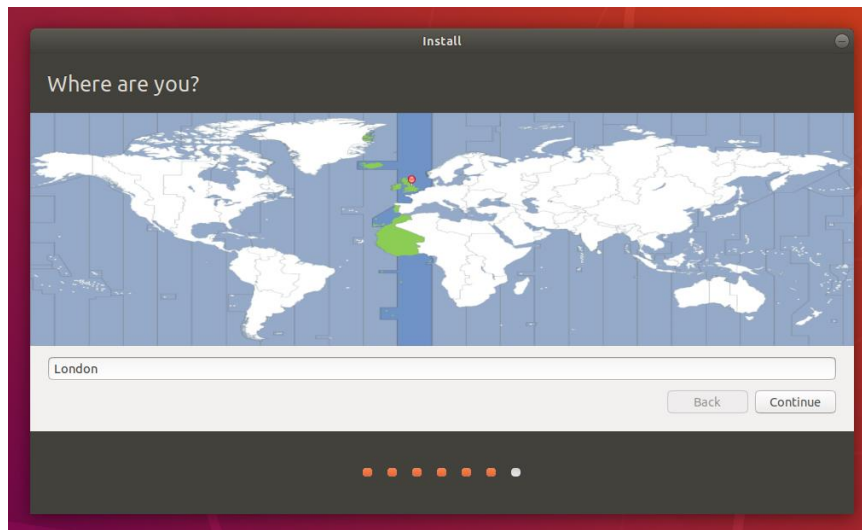
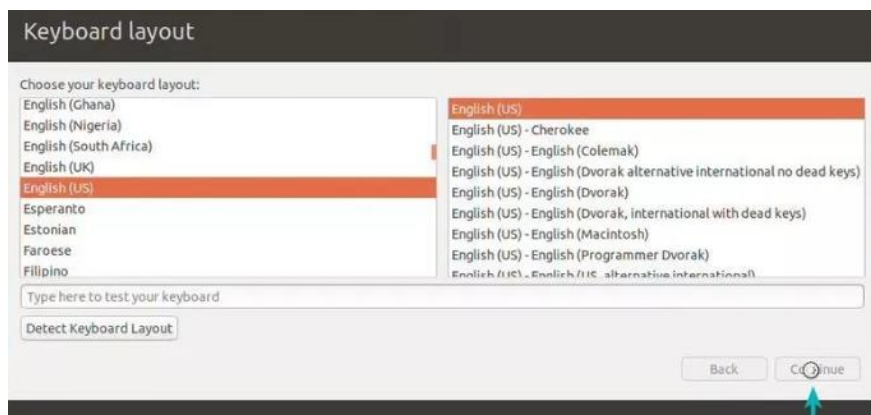
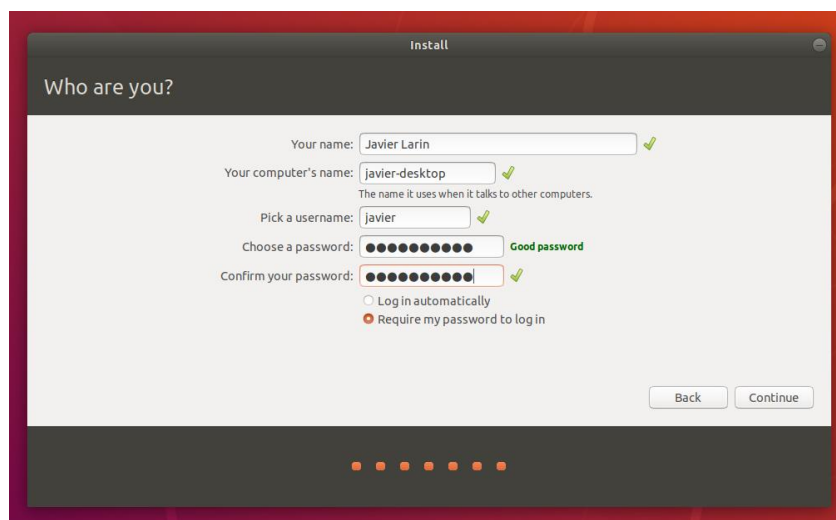


Figure 2.15: Confirmation Screen

17. Select your *Current Location* and *Continue*.*Figure 2.16: Location selection screen***18. Select *keyboard layout* and *Continue*.***Figure 2.17: Selecting Keyboard Layout***19. Fill your Info and click *Continue*.***Figure 2.18: Filling Name and Password*

20. You are almost done. It should take 10-15 minutes to complete the installation.

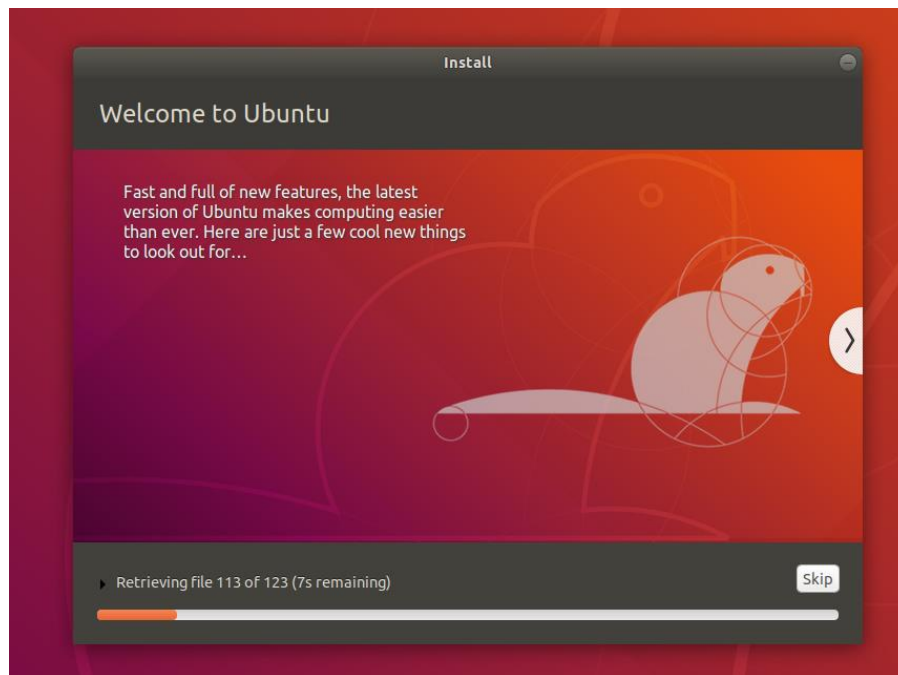


Figure 2.19: Installing Ubuntu

21. Installation is Complete. Click **Restart Now**.

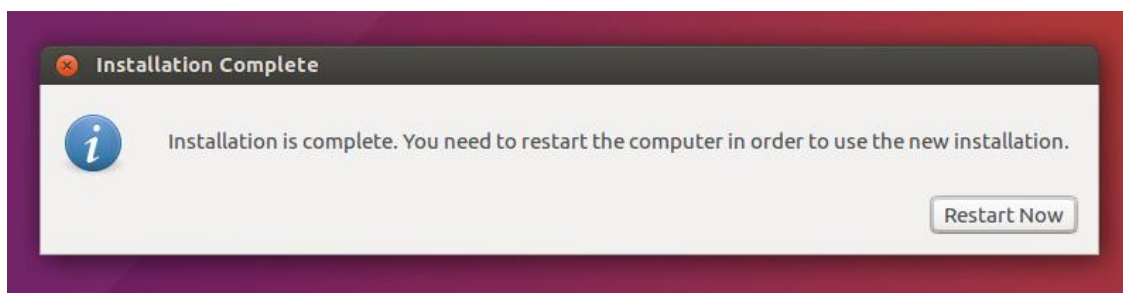


Figure 2.20: Installation Completion Screen




22. After that you'll be able to install the software package by following the steps discussed in chapter 1.

Chapter-3: macOS Support


You can use Metgeno using Linux in a virtual machine which gives you the option to try Linux within the OS without making any changes in your macOS.

VirtualBox is a free and open source virtualization software from Oracle. It enables you to install other operating systems in virtual machines. It is advised that your systems should have at least 2GB RAM to get a decent performance from the virtual operating system.


Requirements:

-  Good internet connection to download software and Linux ISO. You can also use some other computer with internet connection to download these files
-  macOS with at least 10 GB of free space
-  macOS with minimum 2GB of RAM

Download:

-  Go to the website of Oracle VirtualBox and get the latest stable version or you can download it from the following link:

<https://download.virtualbox.org/virtualbox/6.1.10/VirtualBox-6.1.10-138449-OSX.dmg>

-  Next, you need to download the ISO file of the Linux distribution. You can get this image from the official website of the Linux distribution you are trying to use or you can download it from the following link:

<http://releases.ubuntu.com/18.04.3/ubuntu-18.04.3-desktop-amd64.iso>

Installation:

1. To installing VirtualBox in macOS host just click on the downloaded dmg file.
2. Start VirtualBox and click on the **New** symbol.



Figure 3.1: New Virtual OS

3. Give the virtual OS a relevant **Name**. Select the **Type** (Linux) and **Version** (Ubuntu 64-bit). Allocate RAM to the virtual OS. Following system has 16GB of RAM so 2GB of RAM is allocated. You can use more RAM if your system has enough extra RAM. Create a virtual disk. This works as the hard disk of the virtual Linux system. This is where the virtual system will store its files. Click **Create**.

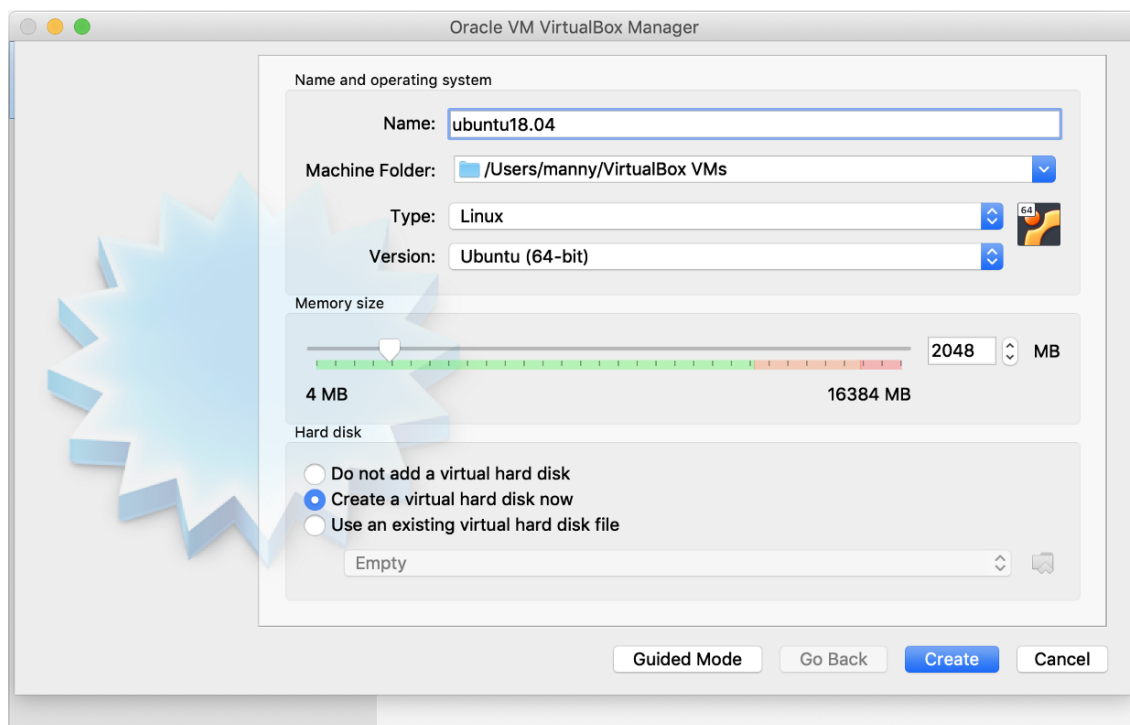


Figure 3.2: Assigning Information

4. Create Select **VDI file type** here (recommended) and Select the **Hard Disk size** (recommended size: 100 GB). You can choose either of Dynamically allocated or Fixed size option for creating the virtual hard disk. Choose **Dynamically allocated** (recommended). Click **Create**.

Note: Dynamic allocation is allocated as time passes and data is increased whereas fixed is allocated instantly.

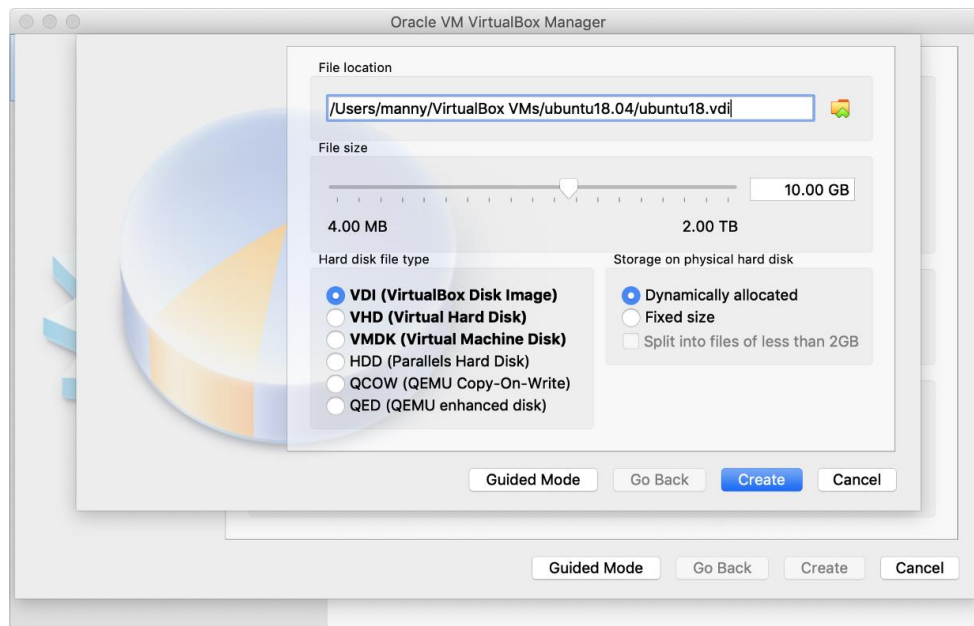


Figure 3.3: Hard disk file type

5. Once your machine settings are setup, you should see the following:

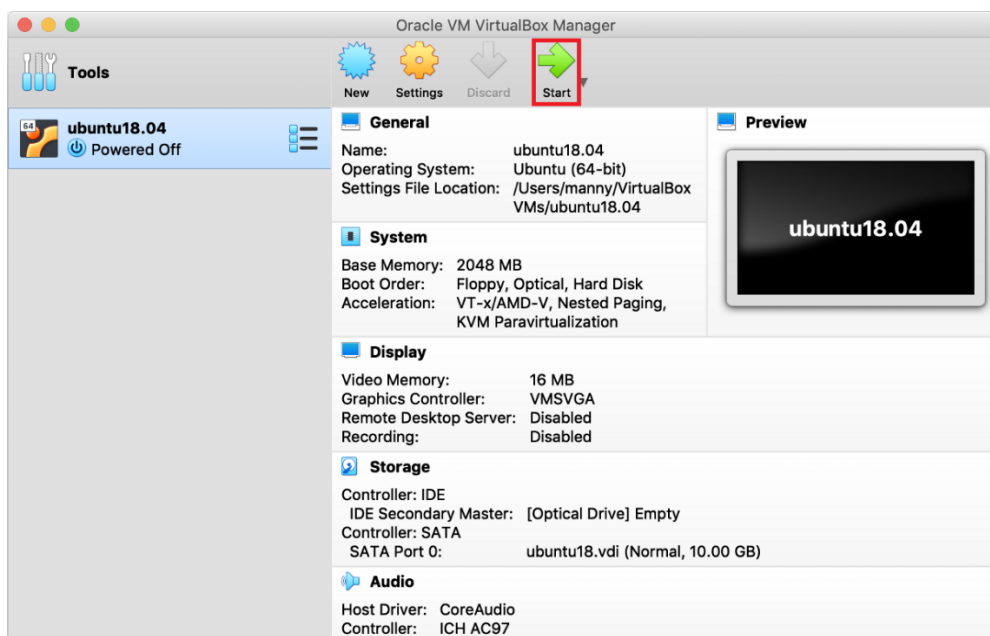


Figure 3.4: Splash Screen

6. To assign the image file of respective OS to VB. Select the machine and just click the Green Start Arrow Icon. We'll soon be prompted with the following, where we'll now select our file, and click **Start**.

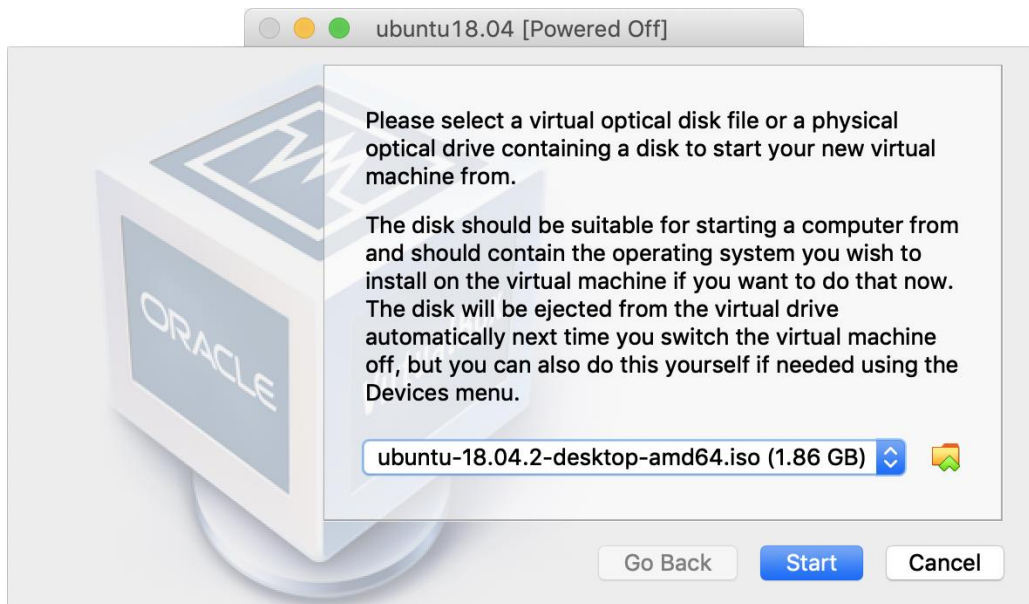


Figure 3.5: Assigning Image file

7. Virtual OS will boot into Linux Installation process. You should be presented with the option to install it. Click **Install Ubuntu**.

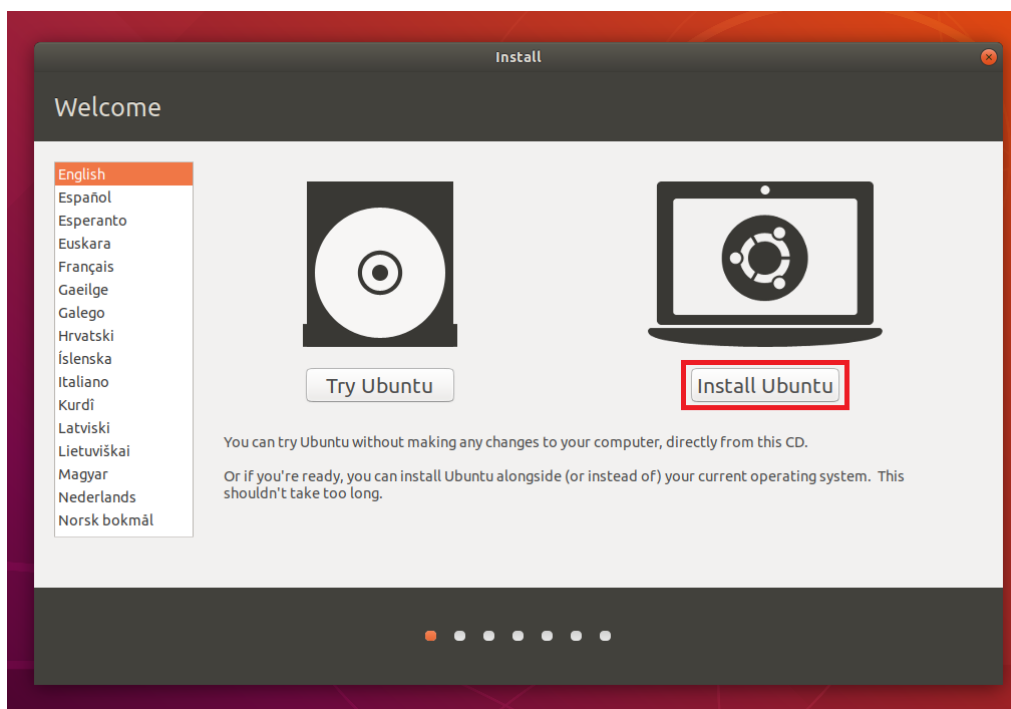


Figure 3.6: Installing Ubuntu

8. Continue with *Normal Installation*.

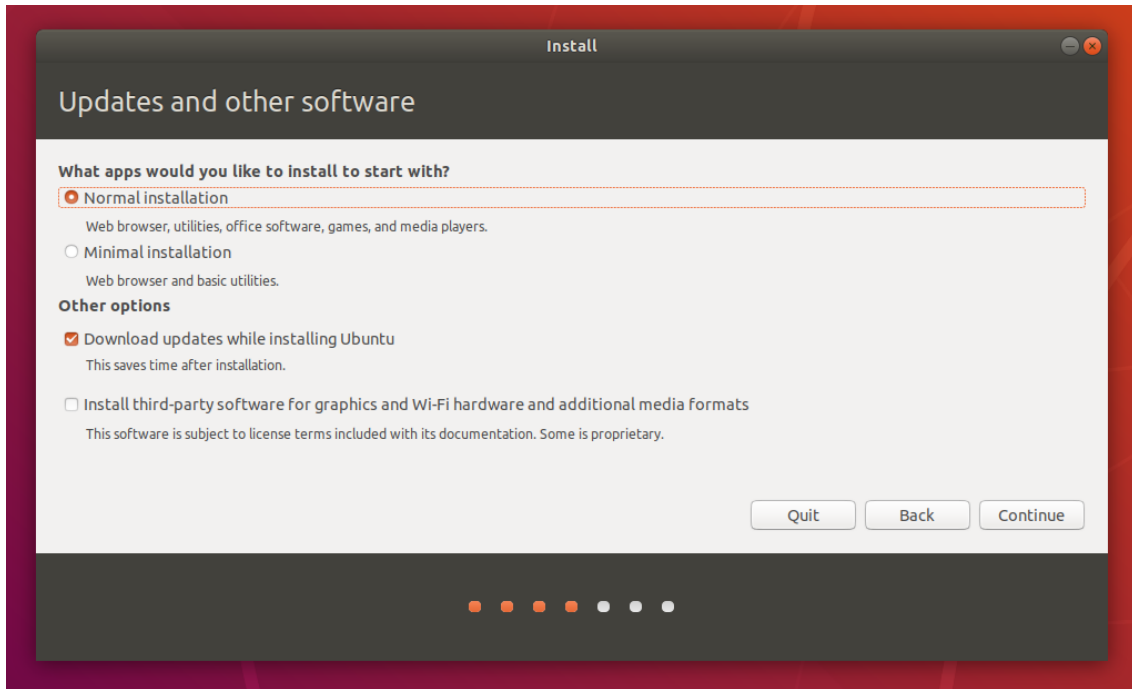


Figure 3.7: Update Screen

9. Select *Erase disk and install Ubuntu*. Don't worry. It won't delete anything on your macOS. You are using the virtual disk space of 10 GB that we created in previous steps. It won't impact the real operating system.

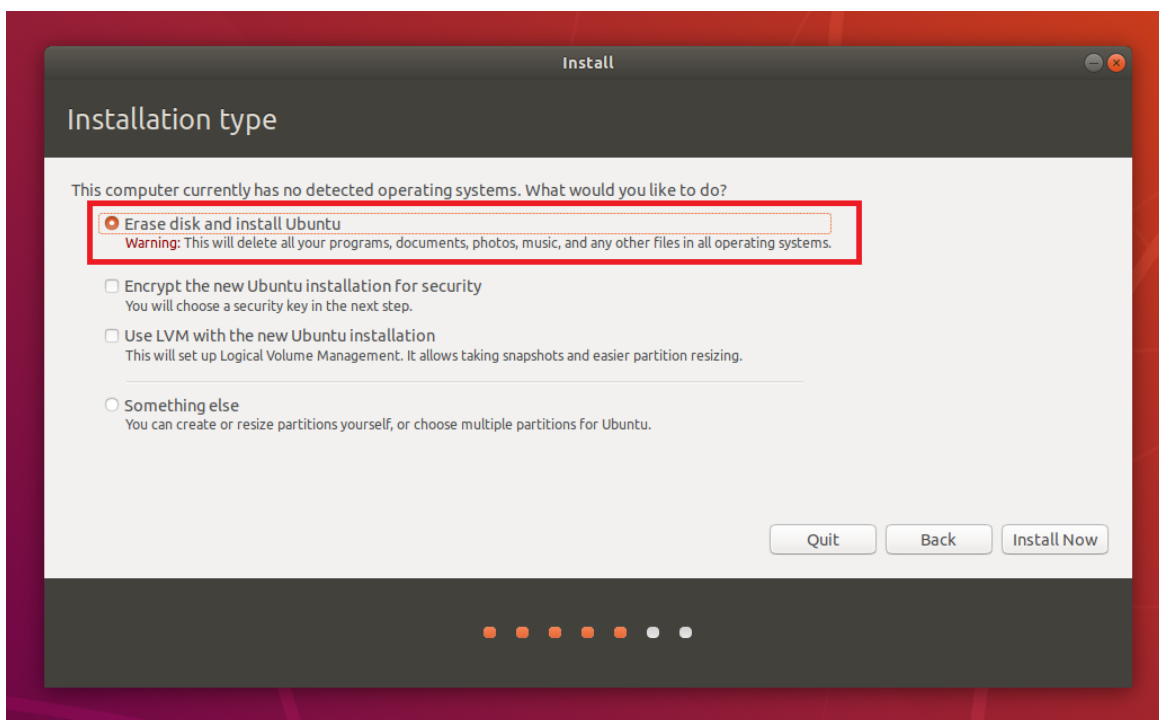


Figure 3.8: Installation type Screen

10. After that click on **Continue**.

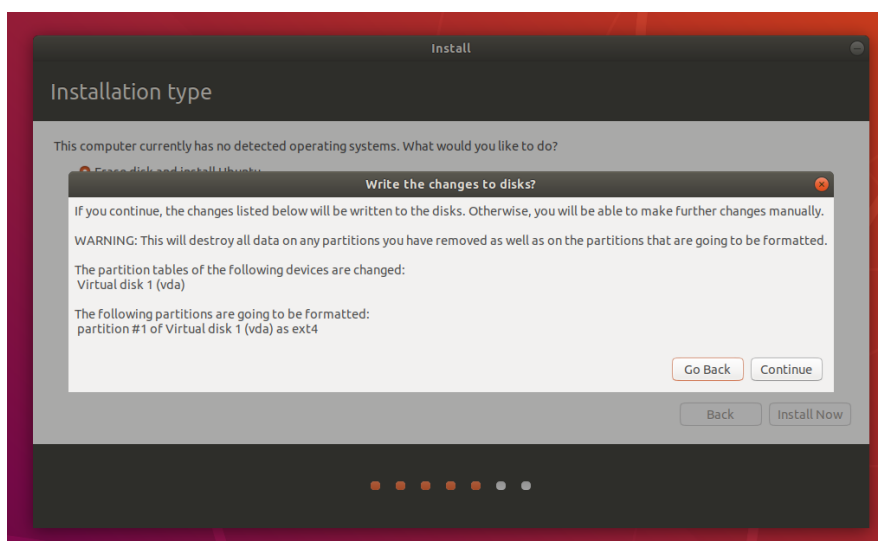


Figure 3.9: Confirmation Screen

11. Select your **Current Location** and **Continue**.

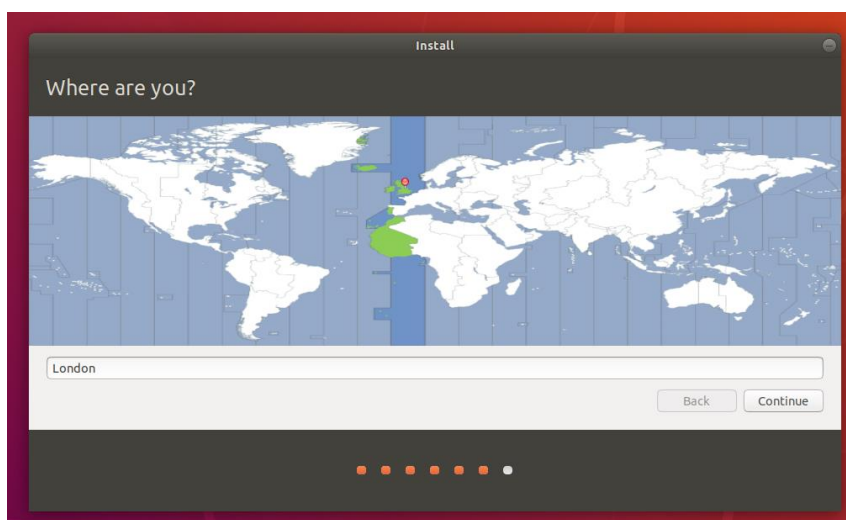


Figure 3.10: Location selection screen

12. Select **keyboard layout** and **Continue**.

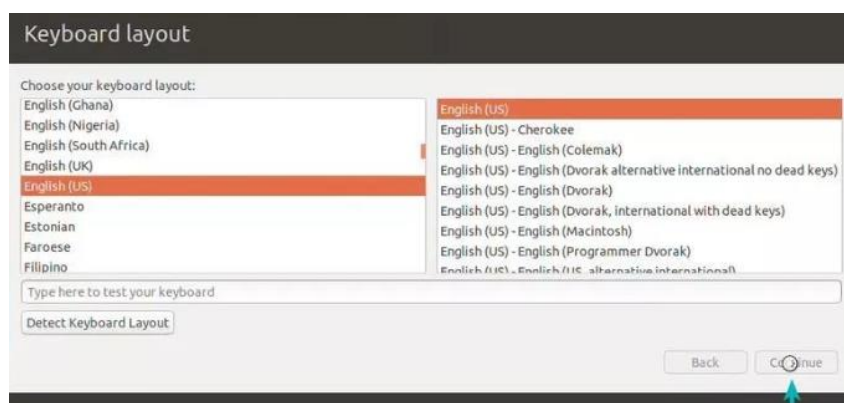


Figure 3.11: Selecting Keyboard Layout

13. Fill your Info and click **Continue**.

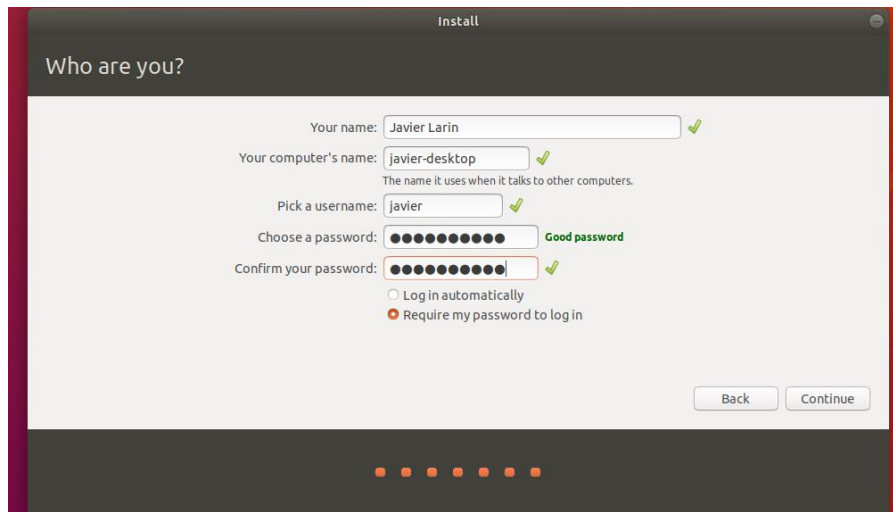


Figure 3.12: Filling Name and Password

14. You are almost done. It should take 10-15 minutes to complete the installation.



Figure 3.13: Installing Ubuntu

15. Installation is Complete. Click **Restart Now**.

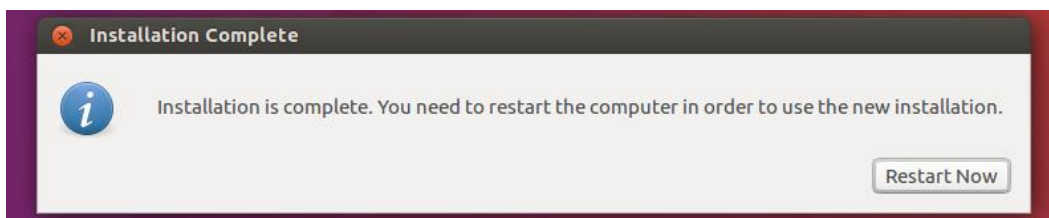


Figure 3.14: Installation Completion Screen

16. After that you'll be able to install the software package by following the steps discussed in chapter 1.

Chapter-4: Selection Criteria

Note: The Selection criteria employed in choosing the tools is that all the included tools are (1) Desktop-based i.e. they do not require an active internet connection for execution. All the tools are (2) Linux based, (3) Published after 2012 and have (4) Freeware License. Also, MATLAB based tools are not included in this package because it is a paid software. Most of the tools are in their stable version and are regularly maintained according to the website *omictools*. However, detailed selection criterion is shown in the **Table I**.

TABLE I. Metgeno (Selection Criteria)

<i>Sr. No.</i>	<i>Software Tool Name</i>	<i>Pub. Date (after 2012)</i>	<i>Linux Based</i>	<i>Offline</i>	<i>Free</i>
Pre-processing					
1	MultiQC [1]	✓	✓	✓	✓
2	HTSeq [2]	✓	✓	✓	✓
3	RobiNA [3]	✓	✓	✓	✓
4	MapSplice [4]		✓	✓	✓
5	Hercules [5]	✓	✓	✓	✓
6	QuickNGS [6]	✓	✓	✓	✓
7	SAMStat [7]		✓	✓	✓
8	NextClip [8]	✓	✓	✓	✓
9	Eoulsan [9]	✓	✓	✓	✓
10	wapRNA [10]		✓	✓	✓
11	oneChannelGUI [11]		✓	✓	✓
12	BBT [12]	✓	✓	✓	✓
13	RseqFlow [13]		✓	✓	✓
14	aRNApipe [14]	✓	✓	✓	✓
15	NGS-Bits [15]	✓	✓	✓	✓
16	NGS-pipe [16]	✓	✓	✓	✓
17	ST Pipeline [17]	✓	✓	✓	✓
18	AlmostSignificant [18]	✓	✓	✓	✓
19	SOAPnuke [19]	✓	✓	✓	✓
20	miARma-Seq [20]	✓	✓	✓	✓
21	CoVaCS [21]	✓	✓		
22	NGS QC Toolkit [22]	✓	✓	✓	✓
23	RNA-QC-Chain [23]	✓	✓	✓	✓
24	miRPursuit [24]	✓	✓	✓	✓
25	SolexaQA [25]		✓	✓	✓
26	mubiomics [26]	✓		✓	✓
27	OncoRep [27]	✓	✓	✓	✓

28	S-MART [28]		✓	✓	✓
29	ReQON [29]	✓	✓	✓	✓
30	TRAPLINE [30]	✓		✓	✓
31	PyroTrimmer [31]	✓	✓	✓	✓
32	Rnnotator [32]		✓	✓	✓
33	TRUFA [33]	✓	✓		
34	BIGpre [34]		✓	✓	✓
35	QuaCRS [35]	✓	✓		
36	preassemble [36]		✓	✓	✓
37	PRINSEQ [37]	✓	✓	✓	✓
38	Pyrocleaner [38]		✓	✓	✓
39	PANDaseq [39]	✓	✓	✓	✓
40	Clustal Omega [40]	✓	✓	✓	✓
Quality Control					
1	Greengenes [41]				✓
2	Nonpareil [42]	✓	✓	✓	✓
3	CorQ [43]	✓	✓	✓	✓
4	FQC [44]	✓		✓	✓
5	QA [45]		✓	✓	✓
6	FaQCs [46]	✓	✓	✓	✓
7	AfterQC [47]	✓	✓	✓	✓
8	HTQC [48]	✓	✓	✓	✓
9	SUGAR [49]	✓	✓	✓	✓
10	KAT [50]	✓	✓	✓	✓
11	FastQ_brew [51]	✓	✓	✓	✓
12	QUASR [52]	✓	✓	✓	✓
13	NGS-QC Generator [53]	✓	✓	✓	✓
14	NGS-eval [54]	✓	✓		
15	mistagging [55]	✓	✓	✓	✓
16	ShortRead [56]		✓	✓	✓
17	sigQC [57]	✓	✓	✓	✓
18	StatsDB [58]	✓	✓	✓	✓
19	G-CNV [59]	✓	✓	✓	✓
20	fitGCP [60]	✓	✓	✓	✓
21	Pheniqs [61]	✓	✓	✓	✓
22	QASDRA [62]	✓	✓	✓	✓
23	Fast-GBS [63]	✓	✓	✓	✓
24	ClinQC [64]	✓	✓	✓	✓
25	SeqAssist [65]	✓	✓	✓	✓
26	PathoQC [66]	✓	✓	✓	✓
27	Zseq [67]	✓	✓	✓	✓
28	MuffinInfo [68]	✓	✓	✓	✓
29	A-Game [69]	✓	✓		
30	MT-Toolbox [70]	✓	✓	✓	✓

31	StreamingTrim [71]	✓	✓	✓	✓
32	FASTQC [72]	✓	✓	✓	✓
33	ACDC [73]	✓			✓
Sequence Contamination Detection					
1	SqClean [74]		✓	✓	✓
2	Vecuum [75]		✓	✓	✓
3	PhylOligo [76]	✓	✓	✓	✓
4	AFS [77]	✓	✓	✓	✓
5	MCSC [78]	✓	✓	✓	✓
6	DeconSeq [79]		✓	✓	✓
7	Kraken [80]	✓	✓	✓	✓
8	Conpair [81]	✓	✓	✓	✓
9	ContEst [82]		✓	✓	✓
10	decontam [83]	✓	✓	✓	✓
11	SIDR [84]	✓	✓	✓	✓
12	MIDAS [85]	✓	✓	✓	✓
13	Cookiecutter [86]	✓			✓
14	Blobology [87]	✓	✓	✓	✓
15	PhagePhisher [88]	✓	✓	✓	✓
16	Taxoblast [89]	✓	✓	✓	✓
17	HYSYS [90]	✓	✓	✓	✓
Assembly					
1	Orione [91]	✓	✓		
2	MetAMOS [92]	✓	✓	✓	✓
3	dDocent [93]	✓	✓	✓	✓
4	Ray Meta [94]	✓	✓	✓	✓
5	TaxMan [95]	✓	✓		
6	DRAP [96]	✓	✓	✓	✓
7	MetaVelvet [97]	✓	✓	✓	✓
8	IDBA-UD [98]	✓	✓	✓	✓
9	MetaSort [99]	✓	✓	✓	✓
10	Snowball [100]	✓	✓	✓	✓
11	Meta-IDBA [101]		✓	✓	✓
12	PRICE [102]	✓	✓	✓	✓
13	Phrap [103]		✓	✓	✓
14	BBAP [104]	✓	✓	✓	✓
15	BBMAP [105]	✓	✓	✓	✓
16	MEGAHIT [106]	✓	✓	✓	✓
17	Xander [107]	✓	✓	✓	✓
18	InteMAP [108]	✓	✓	✓	✓
19	Genovo [109]		✓	✓	✓
20	MetaStorm [110]	✓	✓		
21	Faucet [111]	✓	✓	✓	✓

22	Athena [112]	✓	✓	✓	✓
23	MetaFinisherSC [113]	✓	✓	✓	✓
24	MetaVelvet-SL [114]	✓	✓	✓	✓
25	Mapsembler [115]	✓	✓	✓	✓
26	Spherical [116]	✓	✓	✓	✓
27	BinSanity [117]	✓	✓	✓	✓
28	Xgenovo [118]	✓	✓	✓	✓
29	SAT-Assembler [119]	✓	✓	✓	✓
30	DIME [120]	✓	✓	✓	✓
16S rRNA Reconstruction					
1	Virtual Metagenome [121]	✓	✓		
2	RAMBL [122]	✓	✓	✓	✓
3	SePIA [123]	✓	✓	✓	✓
4	EMIRGE [124]		✓	✓	✓
5	MATAM [125]	✓	✓	✓	✓
6	REAGO [126]	✓	✓	✓	✓
7	TRFMA [127]		✓		
8	GeneStitch [128]	✓	✓	✓	✓
Binning					
1	BUSCO [129]	✓	✓	✓	✓
2	CONCOCT [130]	✓	✓	✓	✓
3	MaxBin [131]	✓	✓	✓	✓
4	MetaCompass [132]	✓	✓	✓	✓
5	MetaCluster-TA [133]	✓	✓	✓	✓
6	CAMI [134], [135]	✓	✓	✓	✓
7	ALE [136]	✓	✓	✓	✓
8	MegaGTA [137]	✓	✓	✓	✓
9	MetaBAT [138]	✓	✓	✓	✓
10	AbundanceBin [139]		✓	✓	✓
11	CheckM [140]	✓	✓	✓	✓
12	MetaQuast [141]	✓	✓	✓	✓
13	GroopM [142]	✓	✓	✓	✓
Homology-based Taxonomy					
1	MetaPhlAn [143]	✓	✓	✓	✓
2	MEGAN [144]	✓	✓	✓	✓
3	Pyrotools [145]	✓	✓	✓	✓
4	Taxonomer [146]	✓	✓		
5	Centrifuge [147]	✓	✓	✓	✓
6	mOTU [148]	✓	✓	✓	✓
7	CARMA [149]		✓	✓	✓
8	Treephyler [150]		✓	✓	✓
9	SATIVA [151]	✓	✓	✓	✓

10	MGmapper [152]	✓	✓	✓	✓
11	CSSSCL [153]	✓	✓	✓	✓
12	TaxSOM [154]		✓	✓	✓
13	Kaiju [155]	✓	✓	✓	✓
14	WebMGA [156]		✓	✓	✓
15	MetaPhyler [157]		✓	✓	✓
16	k-SLAM [158]	✓	✓	✓	✓
17	LMAT [159]	✓	✓	✓	✓
18	Binning_refiner [160]	✓	✓	✓	✓
19	MLTreeMap [161]				✓
20	TACOA [162]		✓	✓	✓
21	deSPI [163]	✓	✓	✓	✓
22	GOTTCHA [164]	✓	✓	✓	✓
23	MBMC [165]	✓	✓	✓	✓
24	LCASstar [166]	✓	✓	✓	✓
25	PROTAX [167]	✓	✓	✓	✓
26	DUDes [168]	✓	✓	✓	✓
27	HTTMM [169]	✓	✓	✓	✓
28	Bracken [170]	✓	✓	✓	✓
29	One Codex [171]	✓	✓		
30	MTR [172]		✓	✓	✓
31	SPANNER [173]	✓	✓	✓	✓
32	TAEC [174]	✓	✓	✓	✓
33	PhyloSift [175]	✓	✓	✓	✓
34	MetaBinG [176]		✓	✓	✓
35	TAMER [177]	✓	✓	✓	✓
36	TAC-ELM [178]	✓	✓	✓	✓
37	Kraken [179]	✓	✓	✓	✓
38	LikelyBin [180]		✓	✓	✓
39	Genometa [181]	✓	✓	✓	✓
40	MyTaxa [182]	✓	✓	✓	✓
41	WEVOTE [183]	✓	✓	✓	✓
42	MetaAmp [184]	✓	✓		
43	MIPE [185]	✓	✓	✓	✓
44	ClaMS [186]		✓	✓	✓
45	MetaCV [187]	✓	✓	✓	✓
46	CoMeta [188]	✓	✓	✓	✓
47	Sequedex [189]	✓	✓	✓	✓
48	RIEMS [190]	✓	✓	✓	✓
49	NBC [191]		✓	✓	✓
50	RITA [192]	✓	✓	✓	✓
51	WGSQuikr [193]	✓	✓	✓	✓
Prediction-based Taxonomy					
1	SHERA [194]		✓	✓	✓

2	PPS [195]	✓	✓	✓	✓
3	d2SBin [196]	✓	✓	✓	✓
4	cuCLARK [197]	✓	✓	✓	✓
5	FROGS [198]	✓	✓	✓	✓
6	TOSS [199]	✓	✓	✓	✓
7	MetaFAST [200]	✓	✓	✓	✓
8	COCACOLA [201]	✓	✓	✓	✓
9	PHYSCIMM [202]		✓	✓	✓
10	VizBin [203]	✓	✓	✓	✓
11	MetaProb [204]	✓	✓	✓	✓
12	MetaWatt [205]	✓	✓	✓	✓
13	AMBER [206]	✓	✓	✓	✓
14	MrGBP [207]	✓	✓	✓	✓
15	MyCC [208]	✓	✓	✓	✓
16	DectICO [209]	✓	✓	✓	✓
17	MBBC [210]	✓	✓	✓	✓
18	Simka [211]	✓	✓	✓	✓
19	BiMeta [212]	✓	✓	✓	✓
20	crAss [213]	✓	✓		
21	CompostBin [214]		✓	✓	✓
22	PhymmBL [215]		✓	✓	✓
Gene Prediction					
1	Glimmer [216]		✓		
2	FragGeneScan [217]	✓	✓	✓	✓
3	GeneMark [218]		✓	✓	✓
4	MICRA [219]	✓	✓	✓	✓
5	MGA [220]		✓	✓	✓
6	Plastid [221]	✓	✓	✓	✓
7	Orphelia [222]		✓	✓	✓
8	RAP [223]	✓			✓
9	Topological-Pressure [224]	✓		✓	✓
10	Protein generator [225]	✓	✓	✓	✓
11	MetaGeneTack [226]	✓	✓	✓	✓
12	TransGeneScan [227]	✓		✓	✓
13	MOCAT [228]	✓	✓	✓	✓
14	Prodigal [229]	✓	✓	✓	✓
Ribosomal Gene Prediction					
1	meta_rna [230]	✓	✓	✓	✓
2	Barrnap [231]	✓	✓	✓	✓
3	RNAmmmer [232]		✓		
4	Pplacer [233]		✓	✓	✓
5	HMMER [234]	✓	✓	✓	✓

Functional Annotation					
1	InterProScan [235]	✓	✓	✓	✓
2	biomartr [236]	✓	✓	✓	✓
3	GhostKOALA [237]	✓	✓		
4	SALT [238]	✓	✓	✓	✓
5	Diamond-blastx [239]	✓	✓	✓	✓
6	ROCKeR [240]	✓	✓	✓	✓
7	UProC [241]	✓	✓	✓	✓
8	Woods [242]	✓	✓	✓	✓
9	RTMg [243]	✓	✓	✓	✓
10	MAGpy [244]	✓	✓	✓	✓
11	metaSAMS [245]	✓	✓		
Metabolic Profiling					
1	MEBS [246]	✓	✓	✓	✓
2	phraSED-ML [247]	✓	✓	✓	✓
3	Kamneva [248]	✓	✓	✓	✓
4	ScrumPY [249]		✓	✓	✓
5	MetaboSignal [250]	✓	✓	✓	✓
6	JMassBalance [251]		✓	✓	✓
7	Pyabolism [252]	✓	✓	✓	✓
8	Mapping Tool [253]		✓	✓	✓
9	MWASTools [254], [255]	✓	✓	✓	✓
10	Pybrn [256]	✓	✓	✓	✓
11	Metabox [257]	✓	✓	✓	✓
12	zeroSum [258]	✓	✓	✓	✓
13	Papi [259]		✓	✓	✓
14	MetaboliteIDConveter [260]	✓	✓	✓	✓
15	cPath [261]		✓	✓	✓
Data Visualization					
1	BURRITO [262]	✓	✓		
2	Metacoder [263]	✓	✓	✓	✓
3	Explicet [264]	✓	✓	✓	✓
4	GrammR [265]	✓	✓	✓	✓
5	Krona [266]		✓	✓	✓
6	VAMPS [267]	✓	✓		
7	GraPhlAn [268]	✓	✓		
8	ASAR [269]	✓	✓	✓	✓
9	MetaViz [270]	✓	✓		
10	SynTVView [271]	✓	✓	✓	✓
11	AmphoraVizu [272]	✓	✓		
12	FuncTree [273]	✓	✓		
13	MetaGenSense [274]	✓	✓	✓	✓

14	Elviz [275]	✓	✓		
15	metaWRAP [276]	✓	✓	✓	✓
16	Strainer [277]	✓	✓	✓	✓

Chapter-5: Packaged Tools

Note: The selection criterion (already mentioned in Table I) filtered most of the available software retaining only about **133** software tools which are now packaged in Metgeno and mentioned in **Table II** with the download size for user convenience. The remaining tools either does not lie on the basic selection criteria or have following notable issues:




-  Software is free but its dependencies are paid
-  Obsolete or not available
-  Credentials or Subscription needed to avail the software

TABLE II. Metgeno (Packaged Tools)

<i>Sr. No.</i>	<i>Software Tool Name</i>	<i>Size (MB)</i>
Pre-processing		
1	MultiQC [1]	4
2	HTSeq [2]	12
3	MapSplice [4]	40
4	NextClip [8]	10
5	Eoulsan [9]	15
6	SOAPnuke [19]	0.1
7	SolexaQA [25]	10
8	PRINSEQ [37]	0.5
9	PANDAsseq [39]	5
10	Clustal Omega [40]	7
11	BWA	1
Quality Control		
12	Greengenes [41]	45
13	Nonpareil [42]	57
14	FQC [44]	10
15	QA [45]	5
16	FaQCs [46]	10
17	HTQC [48]	1
18	SUGAR [49]	93
19	KAT [50]	40
20	QUASR [52]	2
21	ShortRead [56]	40
22	fitGCP [60]	29
23	PathoQC [66]	24
24	FASTQC [72]	37

25	ACDC [73]	17
Sequence Contamination Detection		
26	PhylOligo [76]	26
27	MCSC [78]	16 GB
28	Kraken [80]	1.5
29	Conpair [81]	13
30	SIDR [84]	0.1
31	MIDAS [85]	26
32	Blobology [87]	1
33	PhagePhisher [88]	0.1
34	Taxoblast [89]	30
35	HYSYS [90]	0.1
Assembly		
36	MetAMOS [92]	402
37	dDocent [93]	0.4
38	Ray Meta [94]	0.4
39	MetaVelvet [97]	0.2
40	IDBA-UD [98]	2
41	MetaSort [99]	33
42	Snowball [100]	1
43	PRICE [102]	9
44	BBAP [104]	1.4
45	BBMAP [105]	26
46	MEGAHIT [106]	0.5
47	Faucet [111]	3
48	Athena [112]	0.1
49	MetaFinisherSC [113]	36
50	Spherical [116]	34
16S rRNA Reconstruction		
51	RAMBL [122]	0.4
52	SePIA [123]	20
53	EMIRGE [124]	2
54	MATAM [125]	6
55	REAGO [126]	0.3
56	GeneStitch [128]	5.4
Binning		
57	CONCOCT [130]	7
58	MaxBin [131]	4
59	MetaBAT [138]	13
60	AbundanceBin [139]	7
61	GroopM [142]	30

Homology-based Taxonomy		
62	MEGAN [144]	102
63	Centrifuge [147]	0.8
64	mOTU [148]	26
65	SATIVA [151]	0.6
66	CSSSCL [153]	0.1
67	Kaiju [155]	0.6
68	k-SLAM [158]	0.1
69	LMAT [159]	2
70	GOTTCHA [164]	13
71	DUDes [168]	35
72	TAEC [174]	7
73	PhyloSift [175]	33
74	WEVOTE [183]	5
75	NBC [191]	2
Prediction-based Taxonomy		
76	PPS [195]	12
77	d2SBin [196]	23
78	cuCLARK [197]	0.1
79	FROGS [198]	77
80	TOSS [199]	0.1
81	MetaFAST [200]	25
82	VizBin [203]	103
83	MetaProb [204]	9
84	MetaWatt [205]	12
85	AMBER [206]	0.5
86	MrGBP [207]	0.6
87	MyCC [208]	121
88	Simka [211]	0.3
89	CompostBin [214]	158
90	PhymmBL [215]	5
Gene Prediction		
91	Glimmer [216]	45
92	FragGeneScan [217]	36
93	MICRA [219]	30
94	Plastid [221]	4
95	Orphelia [222]	68
96	RAP [223]	1.9
97	Protein generator [225]	32
98	MOCAT [228]	400
99	Prodigal [229]	26

Ribosomal Gene Prediction		
100	meta_rna [230]	6
101	Barrnap [231]	12
102	Pplacer [233]	36
103	HMMER [234]	18.6
Functional Annotation		
104	InterProScan [235]	16 GB
105	biomartr [236]	0.3
106	SALT [238]	0.1
107	Diamond-blastx [239]	46
108	ROCK [240]	0.1
109	UProC [241]	0.2
110	MAGpy [244]	260
Metabolic Profiling		
111	MEBS [246]	73.2
112	phraSED-ML [247]	1.8
113	Kamneva [248]	102
114	ScrumPY [249]	1.07 GB
115	MetaboSignal [250]	169
116	JMassBalance [251]	4
117	Pyabolism [252]	0.4
118	Mapping Tool [253]	15
119	MWASTools [254], [255]	56.3
120	Pybrn [256]	0.6
121	Metabox [257]	53
122	zeroSum [258]	1
123	Papi [259]	0.1
124	MetaboliteIDConveter [260]	0.2
125	cPath [261]	1.32
Data Visualization		
126	Metacoder [263]	2
127	Explicet [264]	13
128	Krona [266]	1.2
129	GraPhlAn [268]	4
130	ASAR [269]	31
131	MetaGenSense [274]	4
132	metaWRAP [276]	2
133	Strainer [277]	4

Chapter-6: Software Pipeline

Note: The First figure shows the pipeline with steps used for Metagenomical Analysis. The Second Figure shows the pipeline with steps including all the software packaged in **Metgeno**.

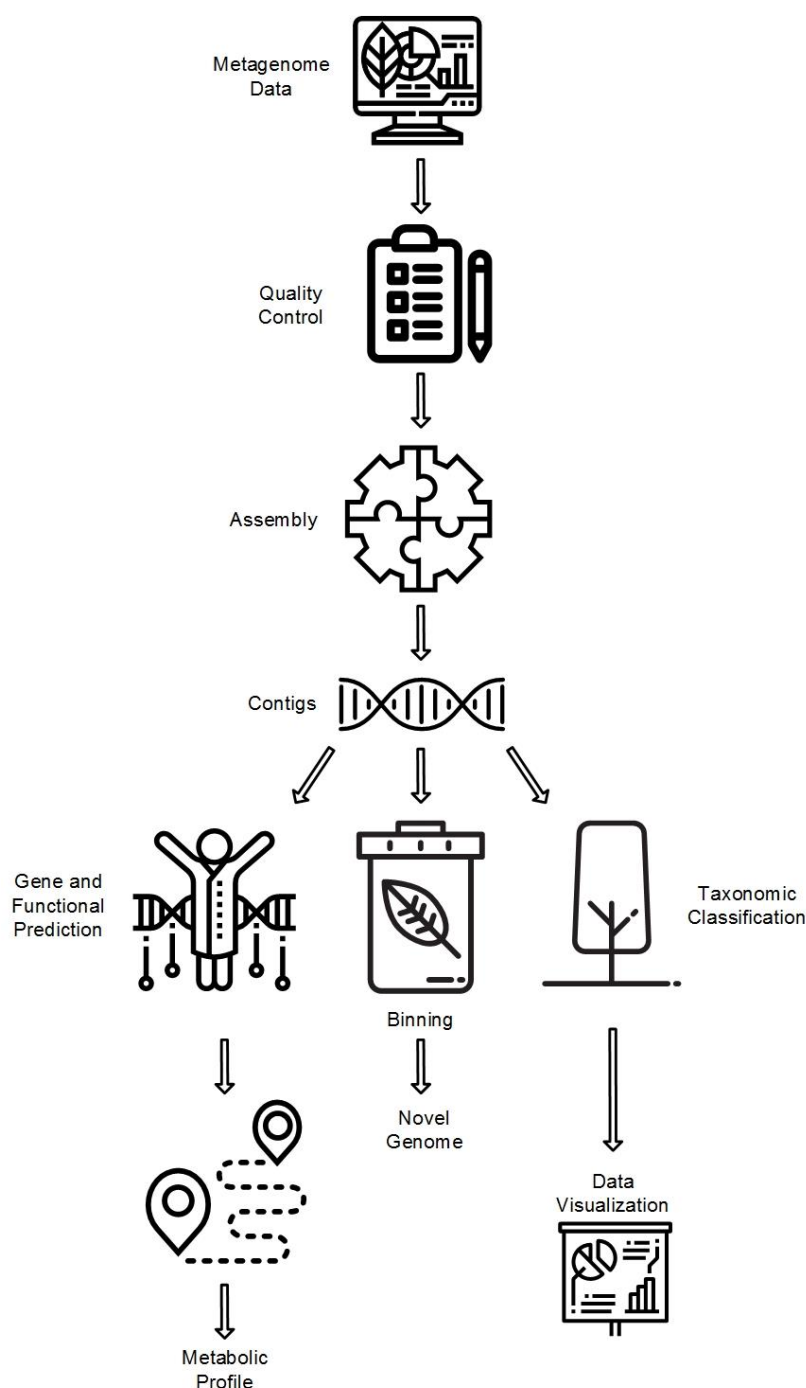
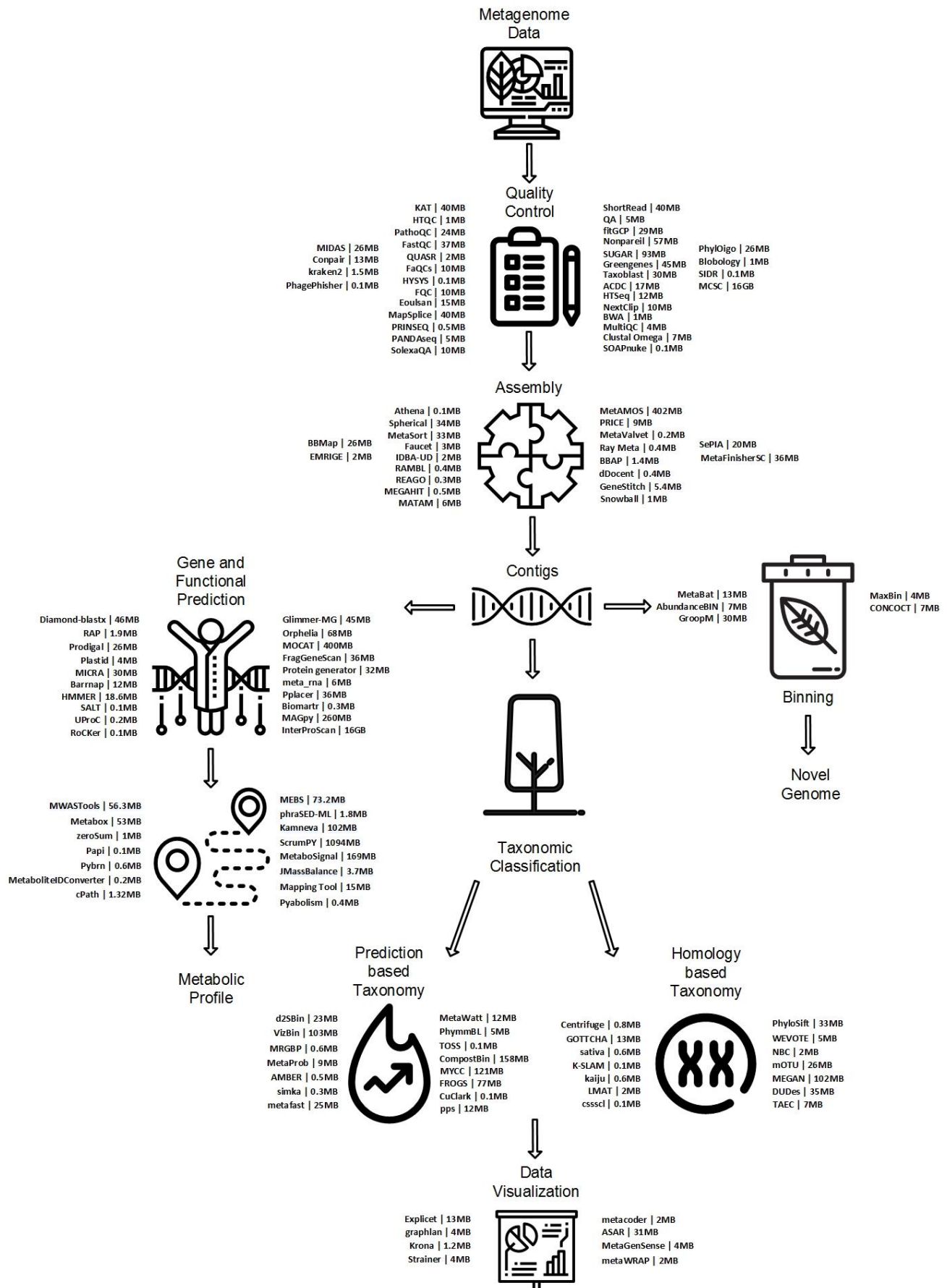


Figure 5.1: Genomes of multiple microorganisms (metagenomics) are obtained as reads through Next Generation Sequencing of environmental samples. These reads are filtered and trimmed to obtain the best quality reads. These reads are assembled to form contigs. Contigs belonging to the same organism are either binned together to obtain final assemblies, or undergo taxonomic classification to understand their standing within the tree-of-life. Alternatively, the same contigs may be employed to expound upon their functional roles.



References

- [1] P. Ewels, M. Magnusson, S. Lundin, and M. Käller, "MultiQC: summarize analysis results for multiple tools and samples in a single report," *Bioinformatics*, vol. 32, no. 19, p. 3047–3048, 2016.
 - [2] S. Anders, P. T. Pyl, and W. Huber, "HTSeq—a Python framework to work with high-throughput sequencing data," *Bioinformatics*, vol. 31, no. 2, p. 166–169, 2015.
 - [3] M. Lohse, A. M. Bolger, A. Nagel, A. R. Fernie, J. E. Lunn, M. Stitt, and B. Usadel, "R obina: A user-friendly, integrated software solution for RNA-Seq-based transcriptomics," *Nucleic acids research*, vol. 40, no. W1, p. W622–W627, 2012.
 - [4] K. Wang, D. Singh, Z. Zeng, S. J. Coleman, Y. Huang, G. L. Savich, X. He, P. Mieczkowski, S. A. Grimm, C. M. Perou *et al.*, "MapSplice: accurate mapping of RNA-seq reads for splice junction discovery," *Nucleic acids research*, vol. 38, no. 18, p. e178–e178, 2010.
 - [5] C. Firtina, Z. Bar-Joseph, C. Alkan, and A. E. Cicek, "Hercules: a profile HMM-based hybrid error correction algorithm for long reads," *bioRxiv*, p. 233080, 2017.
 - [6] G. Crispatzu, P. Kulkarni, M. R. Toliat, P. Nürnberg, M. Herling, C. D. Herling, and P. Frommolt, "Semi-automated cncr genome analysis using high-performance computing," *Human mutation*, vol. 38, no. 10, p. 1325–1335, 2017.
 - [7] T. Lassmann, Y. Hayashizaki, and C. O. Daub, "SAMStat: monitoring biases in next generation sequencing data," *Bioinformatics*, vol. 27, no. 1, p. 130–131, 2010.
 - [8] R. M. Leggett, B. J. Clavijo, L. Clissold, M. D. Clark, and M. Caccamo, "NextClip: an analysis and read preparation tool for Nextera Long Mate Pair libraries," *Bioinformatics*, vol. 30, no. 4, p. 566–568, 2013.
 - [9] L. Jourden, M. Bernard, M.-A. Dillies, and S. Le Crom, "Eoulsan: a cloud computing-based framework facilitating high throughput sequencing analyses," *Bioinformatics*, vol. 28, no. 11, p. 1542–1543, 2012.
 - [10] W. Zhao, W. Liu, D. Tian, B. Tang, Y. Wang, C. Yu, R. Li, Y. Ling, J. Wu, S. Song *et al.*, "wapRNA: a web-based application for the processing of RNA sequences," *Bioinformatics*, vol. 27, no. 21, p. 3076–3077, 2011.
 - [11] R. Sanges, F. Cordero, and R. A. Calogero, "oneChannelGUI: a graphical interface to Bioconductor tools, designed for life scientists who are not familiar with R language," *Bioinformatics*, vol. 23, no. 24, p. 3406–3408, 2007.
 - [12] J. Chu, S. Sadeghi, A. Raymond, S. D. Jackman, K. M. Nip, R. Mar, H. Mohamadi, Y. S. Butterfield, A. G. Robertson, and I. Birol, "BioBloom tools: fast, accurate and memory-efficient host species sequence screening using bloom filters," *Bioinformatics*, vol. 30, no. 23, p. 3402–3404, 2014.
 - [13] Y. Wang, G. Mehta, R. Mayani, J. Lu, T. Souaiaia, Y. Chen, A. Clark, H. J. Yoon, L. Wan, O. V. Evgrafov *et al.*, "RseqFlow: workflows for RNA-Seq data analysis," *Bioinformatics*, vol. 27, no. 18, p. 2598–2600, 2011.
 - [14] A. Alonso, B. N. Lasseigne, K. Williams, J. Nielsen, R. C. Ramaker, A. A. Hardigan, B. Johnston, B. S. Roberts, S. J. Cooper, S. Marsal *et al.*, "aRNApipe: a balanced, efficient and distributed pipeline for processing RNA-seq data in high-performance computing environments," *Bioinformatics*, vol. 33, no. 11, p. 1727–1729, 2017.
-

- [15] C. M. Schroeder, F. J. Hilke, M. W. Löffler, M. Bitzer, F. Lenz, and M. Sturm, "A comprehensive quality control workflow for paired tumor-normal NGS experiments," *Bioinformatics*, vol. 33, no. 11, p. 1721–1722, 2017.
- [16] J. Singer, H.-J. Ruscheweyh, A. L. Hofmann, T. Thurnherr, F. Singer, N. C. Toussaint, C. K. Ng, S. Piscuoglio, C. Beisel, G. Christofori *et al.*, "NGS-pipe: a flexible, easily extendable and highly configurable framework for NGS analysis," *Bioinformatics*, vol. 34, no. 1, p. 107–108, 2017.
- [17] J. F. Navarro, J. Sjöstrand, F. Salmén, J. Lundeberg, and P. L. Ståhl, "ST Pipeline: an automated pipeline for spatial mapping of unique transcripts," *Bioinformatics*, vol. 33, no. 16, p. 2591–2593, 2017.
- [18] J. Ward, C. Cole, M. Febrer, and G. J. Barton, "AlmostSignificant: simplifying quality control of high-throughput sequencing data," *Bioinformatics*, vol. 32, no. 24, p. 3850–3851, 2016.
- [19] Y. Chen, Y. Chen, C. Shi, Z. Huang, Y. Zhang, S. Li, Y. Li, J. Ye, C. Yu, Z. Li *et al.*, "SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data," *GigaScience*, vol. 7, no. 1, p. gix120, 2017.
- [20] E. Andrés-León, R. Núñez-Torres, and A. M. Rojas, "miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis," *Scientific reports*, vol. 6, p. 25749, 2016.
- [21] M. Chiara, S. Gioiosa, G. Chillemi, M. DAntonio, T. Flati, E. Picardi, F. Zambelli, D. S. Horner, G. Pesole, and T. Castrignanò, "CoVaCS: a consensus variant calling system," *BMC genomics*, vol. 19, no. 1, p. 120, 2018.
- [22] R. K. Patel and M. Jain, "NGS QC Toolkit: a toolkit for quality control of next generation sequencing data," *PloS one*, vol. 7, no. 2, p. e30619, 2012.
- [23] Q. Zhou, X. Su, G. Jing, S. Chen, and K. Ning, "RNA-QC-chain: comprehensive and fast quality control for RNA-Seq data," *BMC genomics*, vol. 19, no. 1, p. 144, 2018.
- [24] I. Chaves, B. V. Costa, A. S. Rodrigues, A. Bohn, and C. M. Miguel, "mi RP ursuit—a pipeline for automated analyses of small RNA s in model and nonmodel plants," *FEBS letters*, vol. 591, no. 15, p. 2261–2268, 2017.
- [25] M. P. Cox, D. A. Peterson, and P. J. Biggs, "SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data," *BMC bioinformatics*, vol. 11, no. 1, p. 485, 2010.
- [26] B. C. Smith, T. McAndrew, Z. Chen, A. Harari, D. M. Barris, S. Viswanathan, A. C. Rodriguez, P. Castle, R. Herrero, M. Schiffman *et al.*, "The cervical microbiome over 7 years and a comparison of methodologies for its characterization," *PloS one*, vol. 7, no. 7, p. e40425, 2012.
- [27] T. Meißner, K. M. Fisch, L. Gioia, and A. I. Su, "OncoRep: an n-of-1 reporting tool to support genome-guided treatment for breast cancer patients using RNA-sequencing," *BMC medical genomics*, vol. 8, no. 1, p. 24, 2015.
- [28] M. Zytnicki and H. Quesneville, "S-MART, a software toolbox to aid RNA-Seq data analysis," *PloS one*, vol. 6, no. 10, p. e25988, 2011.
- [29] C. R. Cabanski, K. Cavin, C. Bizon, M. D. Wilkerson, J. S. Parker, K. C. Wilhelmsen, C. M. Perou, J. Marron, and D. N. Hayes, "ReQON: a Bioconductor package for recalibrating quality scores from next-generation sequencing data," *BMC bioinformatics*, vol. 13, no. 1, p. 221, 2012.

-
- [30] M. Wolfien, C. Rimbach, U. Schmitz, J. J. Jung, S. Krebs, G. Steinhoff, R. David, and O. Wolkenhauer, "TRAPLINE: a standardized and automated pipeline for RNA sequencing data analysis, evaluation and annotation," *BMC bioinformatics*, vol. 17, no. 1, p. 21, 2016.
- [31] J. Oh, B. K. Kim, W.-S. Cho, S. G. Hong, and K. M. Kim, "PyroTrimmer: a software with GUI for pre-processing 454 amplicon sequences," *Journal of Microbiology*, vol. 50, no. 5, p. 766–769, 2012.
- [32] J. Martin, V. M. Bruno, Z. Fang, X. Meng, M. Blow, T. Zhang, G. Sherlock, M. Snyder, and Z. Wang, "Rnnotator: an automated de novo transcriptome assembly pipeline from stranded RNA-Seq reads," *BMC genomics*, vol. 11, no. 1, p. 663, 2010.
- [33] E. Kornobis, L. Cabellos, F. Aguilar, C. Frías-López, J. Rozas, J. Marco, and R. Zardoya, "TRUFA: a user-friendly web server for de novo RNA-seq analysis using cluster computing," *Evolutionary Bioinformatics*, vol. 11, p. EBO–S23873, 2015.
- [34] T. Zhang, Y. Luo, K. Liu, L. Pan, B. Zhang, J. Yu, and S. Hu, "BIGpre: a quality assessment package for next-generation sequencing data," *Genomics, proteomics & bioinformatics*, vol. 9, no. 6, p. 238–244, 2011.
- [35] K. W. Kroll, N. E. Mokaram, A. R. Pelletier, D. E. Frankhouser, M. S. Westphal, P. A. Stump, C. L. Stump, R. Bundschuh, J. S. Blachly, and P. Yan, "Quality Control for RNA-Seq (QuaCRS): an integrated quality control pipeline," *Cancer informatics*, vol. 13, p. CIN–S14022, 2014.
- [36] A. A. Adzhubei, J. K. Laerdahl, and A. V. Vlasova, "preAssemble: a tool for automatic sequencer trace data processing," *BMC bioinformatics*, vol. 7, no. 1, p. 22, 2006.
- [37] R. Schmieder and R. Edwards, "Quality control and preprocessing of metagenomic datasets," *Bioinformatics*, vol. 27, no. 6, p. 863–864, 2013.
- [38] J. Mariette, C. Noirot, and C. Klopp, "Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering too," *BMC research notes*, vol. 4, no. 1, p. 149, 2011.
- [39] A. P. Masella, A. K. Bartram, J. M. Truszkowski, D. G. Brown, and J. D. Neufeld, "PANDAsq: paired-end assembler for illumina sequences," *BMC bioinformatics*, vol. 13, no. 1, p. 31, 2012.
- [40] B. Haider, T.-H. Ahn, B. Bushnell, J. Chai, A. Copeland, and C. Pan, "Omega: an Overlap-graph de novo Assembler for Metagenomics," *Bioinformatics*, vol. 30, no. 19, p. 2717–2722, 2014.
- [41] T. DeSantis, P. Hugenholtz, K. Keller, E. Brodie, N. Larsen, Y. Piceno, R. Phan, and G. L. Andersen, "NASt: a multiple sequence alignment server for comparative analysis of 16S rRNA genes," *Nucleic acids research*, vol. 34, no. suppl_2, p. W394–W399, 2006.
- [42] L. M. Rodriguez-r and K. T. Konstantinidis, "Estimating coverage in metagenomic data sets and why it matters," *The ISME journal*, vol. 8, no. 11, p. 2349, 2014.
- [43] S. Iyer, H. Bouzek, W. Deng, B. Larsen, E. Casey, and J. I. Mullins, "Quality score based identification and correction of pyrosequencing errors," *PLoS One*, vol. 8, no. 9, p. e73015, 2013.
- [44] J. Brown, M. Pirrung, and L. A. McCue, "FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool," *Bioinformatics*, vol. 33, no. 19, p. 3137–3139, 2017.
-

- [45] R. T. Ramos, A. R. Carneiro, J. Baumbach, V. Azevedo, M. P. Schneider, and A. Silva, "Analysis of quality raw data of second generation sequencers with Quality Assessment Software," *BMC research notes*, vol. 4, no. 1, p. 130, 2011.
- [46] C.-C. Lo and P. S. Chain, "Rapid evaluation and quality control of next generation sequencing data with FaQCs," *BMC bioinformatics*, vol. 15, no. 1, p. 366, 2014.
- [47] S. Chen, T. Huang, Y. Zhou, Y. Han, M. Xu, and J. Gu, "AfterQC: automatic filtering, trimming, error removing and quality control for fastq data," *BMC bioinformatics*, vol. 18, no. 3, p. 80, 2017.
- [48] X. Yang, D. Liu, F. Liu, J. Wu, J. Zou, X. Xiao, F. Zhao, and B. Zhu, "HTQC: a fast quality control toolkit for Illumina sequencing data," *BMC bioinformatics*, vol. 14, no. 1, p. 33, 2013.
- [49] Y. Sato, K. Kojima, N. Nariai, Y. Yamaguchi-Kabata, Y. Kawai, M. Takahashi, T. Mimori, and M. Nagasaki, "SUGAR: graphical user interface-based data refiner for high-throughput DNA sequencing," *BMC genomics*, vol. 15, no. 1, p. 664, 2014.
- [50] D. Mapleson, G. Garcia Accinelli, G. Kettleborough, J. Wright, and B. J. Clavijo, "KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies," *Bioinformatics*, vol. 33, no. 4, p. 574–576, 2016.
- [51] D. M. OHalloran, "fastQ_brew: module for analysis, preprocessing, and reformatting of FASTQ sequence data," *BMC research notes*, vol. 10, no. 1, p. 275, 2017.
- [52] S. J. Watson, M. R. Welkers, D. P. Depledge, E. Coulter, J. M. Breuer, M. D. de Jong, and P. Kellam, "Viral population analysis and minority-variant detection using short read next-generation sequencing," *Phil. Trans. R. Soc. B*, vol. 368, no. 1614, p. 20120205, 2013.
- [53] M. A. Mendoza-Parra, M.-A. M. Saleem, M. Blum, P.-E. Cholley, and H. Gronemeyer, "NGS-QC generator: a quality control system for ChIP-Seq and related deep sequencing-generated datasets," in *Statistical Genomics*, 2016, p. 243–265.
- [54] A. May, S. Abeln, M. J. Buijs, J. Heringa, W. Crielaard, and B. W. Brandt, "NGS-eval: NGS Error analysis and novel sequence VArIant detection tooL," *Nucleic acids research*, vol. 43, no. W1, p. W301–W305, 2015.
- [55] P. Esling, F. Lejzerowicz, and J. Pawlowski, "Accurate multiplexing and filtering for high-throughput amplicon-sequencing," *Nucleic acids research*, vol. 43, no. 5, p. 2513–2524, 2015.
- [56] M. Morgan, S. Anders, M. Lawrence, P. Aboyoun, H. Pages, and R. Gentleman, "ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data," *Bioinformatics*, vol. 25, no. 19, p. 2607, 2009.
- [57] A. Dhawan, A. Barberis, W.-C. Cheng, E. Domingo, C. West, T. Maughan, J. Scott, A. L. Harris, and F. M. Buffa, "sigQC: A procedural approach for standardising the evaluation of gene signatures," *bioRxiv*, p. 203729, 2017.
- [58] R. H. Ramirez-Gonzalez, R. M. Leggett, D. Waite, A. Thanki, N. Drou, M. Caccamo, and R. Davey, "StatsDB: platform-agnostic storage and understanding of next generation sequencing run metrics," *F1000Research*, vol. 2, 2013.
- [59] A. Manconi, E. Manca, M. Moscatelli, M. Gnocchi, A. Orro, G. Armano, and L. Milanesi, "G-CNV: a GPU-based tool for preparing data to detect CNVs with read-depth methods," *Frontiers in bioengineering and biotechnology*, vol. 3, no. 28, 2015.

-
- [60] M. S. Lindner, M. Kollock, F. Zickmann, and B. Y. Renard, "Analyzing genome coverage profiles with applications to quality control in metagenomics," *Bioinformatics*, vol. 29, no. 10, p. 1260–1267, 2013.
 - [61] L. Galanti, D. Shasha, and K. Gunsalus, "Pheniqs: Fast and flexible quality-aware sequence demultiplexing," *bioRxiv*, p. 128512, 2017.
 - [62] M. O. Külekci, A. Fotouhi, and M. Majidi, "Quality Assessment of High-throughput DNA Sequencing Data via Range analysis," *bioRxiv*, p. 101469, 2017.
 - [63] D. Torkamaneh, J. Laroche, M. Bastien, A. Abed, and F. Belzile, "Fast-GBS: a new pipeline for the efficient and highly accurate calling of SNPs from genotyping-by-sequencing data," *BMC bioinformatics*, vol. 18, no. 1, p. 5, 2017.
 - [64] R. V. Pandey, S. Pabinger, A. Kriegner, and A. Weinhäusel, "ClinQC: a tool for quality control and cleaning of Sanger and NGS data in clinical research," *BMC bioinformatics*, vol. 17, no. 1, p. 56, 2016.
 - [65] Y. Peng, A. S. Maxwell, N. D. Barker, J. G. Laird, A. J. Kennedy, N. Wang, C. Zhang, and P. Gong, "SeqAssist: a novel toolkit for preliminary analysis of next-generation sequencing data," in *BMC bioinformatics*, vol. 15, 2014, p. S10.
 - [66] C. Hong, S. Manimaran, Y. Shen, J. F. Perez-Rogers, A. L. Byrd, E. Castro-Nallar, K. A. Crandall, and W. E. Johnson, "PathoScope 2.0: a complete computational framework for strain identification in environmental or clinical sequencing samples," *Microbiome*, vol. 2, no. 1, p. 33, 2014.
 - [67] A. Alkhateeb and L. Rueda, "Zseq: an approach for preprocessing next-generation sequencing data," *Journal of Computational Biology*, vol. 24, no. 8, p. 746–755, 2017.
 - [68] A. S. Alic and I. Blanquer, "MuffinInfo: HTML5-Based Statistics Extractor from Next-Generation Sequencing Data," *Journal of Computational Biology*, vol. 23, no. 9, p. 750–755, 2016.
 - [69] M. Chiara, A. Placido, E. Picardi, L. R. Ceci, D. S. Horner, and G. Pesole, "A-GAME: improving the assembly of pooled functional metagenomics sequence data," *BMC genomics*, vol. 19, no. 1, p. 44, 2018.
 - [70] S. M. Yourstone, D. S. Lundberg, J. L. Dangl, and C. D. Jones, "MT-Toolbox: improved amplicon sequencing using molecule tags," *BMC bioinformatics*, vol. 15, no. 1, p. 284, 2014.
 - [71] G. Bacci, M. Bazzicalupo, A. Benedetti, and A. Mengoni, "StreamingTrim 1.0: a Java software for dynamic trimming of 16S rRNA sequence data from metagenetic studies," *Molecular ecology resources*, vol. 14, no. 2, p. 426–434, 2014.
 - [72] R. Schmieder and R. Edwards, "Quality control and preprocessing of metagenomic datasets," *Bioinformatics*, vol. 27, no. 6, p. 863–864, 2013.
 - [73] M. Lux, J. Krüger, C. Rinke, I. Maus, A. Schlüter, T. Woyke, A. Sczyrba, and B. Hammer, "acdc—Automated Contamination Detection and Confidence estimation for single-cell genome data," *BMC bioinformatics*, vol. 17, no. 1, p. 543, 2016.
 - [74] W. Xie, F. Wang, L. Guo, Z. Chen, S. M. Sievert, J. Meng, G. Huang, Y. Li, Q. Yan, S. Wu *et al.*, "Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries," *The ISME journal*, vol. 5, no. 3, p. 414, 2011.
 - [75] E.-Y. Lee, D.-Y. Choi, D.-K. Kim, J.-W. Kim, J. O. Park, S. Kim, S.-H. Kim, D. M. Desiderio, Y.-K. Kim, K.-P. Kim *et al.*, "Gram-positive bacteria produce membrane vesicles:
-

- Proteomics-based characterization of Staphylococcus aureus-derived membrane vesicles," *Proteomics*, vol. 9, no. 24, p. 5425–5436, 2009.
- [76] L. Mallet, T. Bitard-Feildel, F. Cerutti, and H. Chiapello, "PhylOligo: a package to identify contaminant or untargeted organism sequences in genome assemblies," *Bioinformatics*, vol. 33, no. 20, p. 3283–3285, 2017.
- [77] Y. Liu, F. Ripp, R. Koepfel, H. Schmidt, S. L. Hellmann, M. Weber, C. F. Kromholz, B. Schmidt, and T. Hankeln, "Afs: identification and quantification of species composition by metagenomic sequencing," *Bioinformatics*, vol. 33, no. 9, p. 1396–1398, 2017.
- [78] J. Lafond-Lapalme, M.-O. Duceppe, S. Wang, P. Moffett, and B. Mimee, "A new method for decontamination of de novo transcriptomes using a hierarchical clustering algorithm," *Bioinformatics*, vol. 33, no. 9, p. 1293–1300, 2017.
- [79] R. Schmieder and R. Edwards, "Fast identification and removal of sequence contamination from genomic and metagenomic datasets," *PloS one*, vol. 6, no. 3, p. e17288, 2011.
- [80] M. P. Davis, S. van Dongen, C. Abreu-Goodger, N. Bartonicek, and A. J. Enright, "Kraken: a set of tools for quality control and analysis of high-throughput sequence data," *Methods*, vol. 63, no. 1, p. 41–49, 2013.
- [81] E. A. Bergmann, B.-J. Chen, K. Arora, V. Vacic, and M. C. Zody, "Conpair: concordance and contamination estimator for matched tumor--normal pairs," *Bioinformatics*, vol. 32, no. 20, p. 3196–3198, 2016.
- [82] K. Cibulskis, A. McKenna, T. Fennell, E. Banks, M. DePristo, and G. Getz, "ContEst: estimating cross-contamination of human samples in next-generation sequencing data," *Bioinformatics*, vol. 27, no. 18, p. 2601–2602, 2011.
- [83] N. M. Davis, D. Proctor, S. P. Holmes, D. A. Relman, and B. J. Callahan, "Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data," *bioRxiv*, p. 221499, 2017.
- [84] J. L. Fierst and D. A. Murdock, "Decontaminating eukaryotic genome assemblies with machine learning," *BMC bioinformatics*, vol. 18, no. 1, p. 533, 2017.
- [85] S. Nayfach, B. Rodriguez-Mueller, N. Garud, and K. S. Pollard, "An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography," *Genome research*, vol. 26, no. 11, p. 1612–1625, 2016.
- [86] E. Starostina, G. Tamazian, P. Dobrynin, S. O'Brien, and A. Komissarov, "Cookiecutter: a tool for kmer-based read filtering and extraction," *bioRxiv*, p. 024679, 2015.
- [87] S. Kumar, M. Jones, G. Koutsovoulos, M. Clarke, and M. Blaxter, "Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots," *Frontiers in genetics*, vol. 4, p. 237, 2013.
- [88] T. Hatzopoulos, S. C. Watkins, and C. Putonti, "PhagePhisher: a pipeline for the discovery of covert viral sequences in complex genomic datasets," *Microbial Genomics*, vol. 2, no. 3, 2016.
- [89] S. M. Dittami and E. Corre, "Detection of bacterial contaminants and hybrid sequences in the genome of the kelp *Saccharina japonica* using Taxoblast," *PeerJ*, vol. 5, p. e4073, 2017.
- [90] J. Schröder, V. Corbin, and A. T. Papenfuss, "HYSYS: have you swapped your samples?," *Bioinformatics*, vol. 33, no. 4, p. 596–598, 2016.

-
- [91] G. Cuccuru, M. Orsini, A. Pinna, A. Sbardellati, N. Soranzo, A. Travaglione, P. Uva, G. Zanetti, and G. Fotia, "Orione, a web-based framework for NGS analysis in microbiology," *Bioinformatics*, vol. 30, no. 13, p. 1928–1929, 2014.
- [92] T. J. Treangen, S. Koren, D. D. Sommer, B. Liu, I. Astrovskaia, B. Ondov, A. E. Darling, A. M. Phillippy, and M. Pop, "MetAMOS: a modular and open source metagenomic assembly and analysis pipeline," *Genome biology*, vol. 14, no. 1, p. R2, 2013.
- [93] J. B. Puritz, C. M. Hollenbeck, and J. R. Gold, "dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms," *PeerJ*, vol. 2, p. e431, 2014.
- [94] S. Boisvert, F. Raymond, É. Godzaridis, F. Laviolette, and J. Corbeil, "Ray Meta: scalable de novo metagenome assembly and profiling," *Genome biology*, vol. 13, no. 12, p. R122, 2012.
- [95] B. W. Brandt, M. J. Bonder, S. M. Huse, and E. Zaura, "TaxMan: a server to trim rRNA reference databases and inspect taxonomic coverage," *Nucleic acids research*, vol. 40, no. W1, p. W82–W87, 2012.
- [96] C. Cabau, F. Escudié, A. Djari, Y. Guiguen, J. Bobe, and C. Klopp, "Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies," *PeerJ*, vol. 5, p. e2988, 2017.
- [97] T. Namiki, T. Hachiya, H. Tanaka, and Y. Sakakibara, "MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads," *Nucleic acids research*, vol. 40, no. 20, p. e155–e155, 2012.
- [98] Y. Peng, H. C. Leung, S.-M. Yiu, and F. Y. Chin, "IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth," *Bioinformatics*, vol. 28, no. 11, p. 1420–1428, 2012.
- [99] P. Ji, Y. Zhang, J. Wang, and F. Zhao, "MetaSort untangles metagenome assembly by reducing microbial community complexity," *Nature communications*, vol. 8, p. 14306, 2017.
- [100] I. Gregor, A. Schönhuth, and A. C. McHardy, "Snowball: strain aware gene assembly of metagenomes," *Bioinformatics*, vol. 32, no. 17, p. i649–i657, 2016.
- [101] Y. Peng, H. C. Leung, S.-M. Yiu, and F. Y. Chin, "Meta-IDBA: a de Novo assembler for metagenomic data," *Bioinformatics*, vol. 27, no. 13, p. i94–i101, 2011.
- [102] J. G. Ruby, P. Bellare, and J. L. DeRisi, "PRICE: software for the targeted assembly of components of (Meta) genomic sequence data," *G3: Genes, Genomes, Genetics*, vol. 3, no. 5, p. 865–880, 2013.
- [103] D. A. Filatov, "Processing and population genetic analysis of multigenic datasets with ProSeq3 software," *Bioinformatics*, vol. 25, no. 23, p. 3189–3190, 2009.
- [104] Y.-Y. Lin, C.-H. Hsieh, J.-H. Chen, X. Lu, J.-H. Kao, P.-J. Chen, D.-S. Chen, and H.-Y. Wang, "De novo assembly of highly polymorphic metagenomic data using in situ generated reference sequences and a novel BLAST-based assembly pipeline," *BMC bioinformatics*, vol. 18, no. 1, p. 223, 2017.
- [105] B. Bushnell, "BBMap: a fast, accurate, splice-aware aligner," Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), Tech. Rep., 2014.
- [106] D. Li, C.-M. Liu, R. Luo, K. Sadakane, and T.-W. Lam, "MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph," *Bioinformatics*, vol. 31, no. 10, p. 1674–1676, 2015.
- [107] Q. Wang, J. A. Fish, M. Gilman, Y. Sun, C. T. Brown, J. M. Tiedje, and J. R. Cole, "Xander: employing a novel method for efficient gene-targeted metagenomic assembly," *Microbiome*, vol. 3, no. 1, p. 32, 2015.
-

- [108] B. Lai, F. Wang, X. Wang, L. Duan, and H. Zhu, "InteMAP: Integrated metagenomic assembly pipeline for NGS short reads," *BMC bioinformatics*, vol. 16, no. 1, p. 244, 2015.
- [109] J. Laserson, V. Jojic, and D. Koller, "Genovo: de novo assembly for metagenomes," *Journal of Computational Biology*, vol. 18, no. 3, p. 429–443, 2011.
- [110] G. Arango-Argoty, G. Singh, L. S. Heath, A. Pruden, W. Xiao, and L. Zhang, "MetaStorm: A Public Resource for Customizable Metagenomics Annotation," *PloS one*, vol. 11, no. 9, p. e0162442, 2016.
- [111] R. Rozov, G. Goldshlager, E. Halperin, and R. Shamir, "Faucet: streaming de novo assembly graph construction," *Bioinformatics*, vol. 34, no. 1, p. 147–154, 2017.
- [112] E. Moss, A. Bishara, E. Tkachenko, J. B. Kang, T. M. Andermann, C. Wood, C. Handy, H. Ji, S. Batzoglu, and A. S. Bhatt, "De novo assembly of microbial genomes from human gut metagenomes using barcoded short read sequences," *bioRxiv*, p. 125211, 2017.
- [113] K.-K. Lam, R. Hall, A. Clum, and S. Rao, "BIGMAC: Breaking Inaccurate Genomes and Merging Assembled Contigs for long read metagenomic assembly," *BMC bioinformatics*, vol. 17, no. 1, p. 435, 2016.
- [114] K. Afiahayati and Y. Sato, "Sakakibara,(2015)'MetaVelvet-SL: An Extension of the Velvet Assembler to a De Novo Metagenomic Assembler Utilizing Supervised Learning'," *DNA Research*, vol. 22, no. 1, p. 69–77, 2015.
- [115] P. Peterlongo and R. Chikhi, "Mapsembler, targeted and micro assembly of large NGS datasets on a desktop computer," *BMC bioinformatics*, vol. 13, no. 1, p. 48, 2012.
- [116] T. C. Hitch, B. J. Thomas, J. C. Friedersdorff, H. Ougham, and C. J. Creevey, "Deep sequence analysis reveals the ovine rumen as a reservoir of antibiotic resistance genes," *Environmental Pollution*, vol. 235, p. 571–575, 2018.
- [117] E. D. Graham, J. F. Heidelberg, and B. J. Tully, "BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation," *PeerJ*, vol. 5, p. e3035, 2017.
- [118] K. Sato, Y. Sakakibara *et al.*, "An extended genovo metagenomic assembler by incorporating paired-end information," *PeerJ*, vol. 1, p. e196, 2013.
- [119] Y. Zhang, Y. Sun, and J. R. Cole, "A scalable and accurate targeted gene assembly tool (SAT-Assembler) for next-generation sequencing data," *PLoS computational biology*, vol. 10, no. 8, p. e1003737, 2014.
- [120] X. Guo, N. Yu, X. Ding, J. Wang, and Y. Pan, "Dime: A novel framework for de novo metagenomic sequence assembly," *Journal of Computational Biology*, vol. 22, no. 2, p. 159–177, 2015.
- [121] S. Okuda, Y. Tsuchiya, C. Kiriya, M. Itoh, and H. Morisaki, "Virtual metagenome reconstruction from 16S rRNA gene sequences," *Nature communications*, vol. 3, p. 1203, 2012.
- [122] F. Zeng, Z. Wang, Y. Wang, J. Zhou, and T. Chen, "Large-scale 16S gene assembly using metagenomics shotgun sequences," *Bioinformatics*, vol. 33, no. 10, p. 1447–1456, 2017.
- [123] K. Içay, P. Chen, A. Cervera, V. Rantanen, R. Lehtonen, and S. Hautaniemi, "SePIA: RNA and small RNA sequence processing, integration, and analysis," *BioData mining*, vol. 9, no. 1, p. 20, 2016.

-
- [124] C. S. Miller, B. J. Baker, B. C. Thomas, S. W. Singer, and J. F. Banfield, "EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data," *Genome biology*, vol. 12, no. 5, p. R44, 2011.
- [125] P. Pericard, Y. Dufresne, L. Couderc, S. Blanquart, and H. Touzet, "MATAM: reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes," *Bioinformatics*, 2017.
- [126] C. Yuan, J. Lei, J. Cole, and Y. Sun, "Reconstructing 16S rRNA genes in metagenomic data," *Bioinformatics*, vol. 31, no. 12, p. i35–i43, 2015.
- [127] Y. Nakano, T. Takeshita, M. Yasui, and Y. Yamashita, "Prediction of plausible bacterial composition based on terminal restriction fragment length polymorphisms using a Monte Carlo method," *Microbial ecology*, vol. 60, no. 2, p. 364–372, 2010.
- [128] Y.-W. Wu, M. Rho, T. G. Doak, and Y. Ye, "Stitching gene fragments with a network matching algorithm improves gene assembly for metagenomics," *Bioinformatics*, vol. 28, no. 18, p. i363–i369, 2012.
- [129] F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs," *Bioinformatics*, vol. 31, no. 19, p. 3210–3212, 2015.
- [130] J. Alneberg, B. S. Bjarnason, I. de Bruijn, M. Schirmer, J. Quick, U. Z. Ijaz, N. J. Loman, A. F. Andersson, and C. Quince, "CONCOCT: clustering contigs on coverage and composition," *arXiv preprint arXiv:1312.4038*, 2013.
- [131] Y.-W. Wu, B. A. Simmons, and S. W. Singer, "MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets," *Bioinformatics*, vol. 32, no. 4, p. 605–607, 2015.
- [132] V. Cepeda, B. Liu, M. Almeida, C. M. Hill, S. Koren, T. J. Treangen, and M. Pop, "Metacompass: Reference-guided assembly of metagenomes," *BioRxiv*, p. 212506, 2017.
- [133] Y. Wang, H. C. M. Leung, S. M. Yiu, and F. Y. L. Chin, "MetaCluster-TA: taxonomic annotation for metagenomic data based on assembly-assisted binning," *BMC genomics*, vol. 15, no. 1, p. S12, 2014.
- [134] A. Sczyrba, P. Hofmann, P. Belmann, D. Koslicki, S. Janssen, J. Dröge, I. Gregor, S. Majda, J. Fiedler, E. Dahms *et al.*, "Critical assessment of metagenome interpretation—a benchmark of metagenomics software," *Nature methods*, vol. 14, no. 11, p. 1063–1071, 2017.
- [135] A. Bremges and A. C. McHardy, "Critical assessment of metagenome interpretation enters the second round," *MSystems*, vol. 3, no. 4, 2018.
- [136] S. C. Clark, R. Egan, P. I. Frazier, and Z. Wang, "ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies," *Bioinformatics*, vol. 29, no. 4, p. 435–443, 2013.
- [137] D. Li, Y. Huang, C.-M. Leung, R. Luo, H.-F. Ting, and T.-W. Lam, "MegaGTA: a sensitive and accurate metagenomic gene-targeted assembler using iterative de Bruijn graphs," *BMC bioinformatics*, vol. 18, no. 12, p. 408, 2017.
- [138] D. D. Kang, J. Froula, R. Egan, and Z. Wang, "MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities," *PeerJ*, vol. 3, p. e1165, 2015.
-

- [139] Y.-W. Wu and Y. Ye, "A novel abundance-based algorithm for binning metagenomic sequences using l-tuples," *Journal of Computational Biology*, vol. 18, no. 3, p. 523–534, 2011.
- [140] D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, and G. W. Tyson, "CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes," *Genome research*, vol. 25, no. 7, p. 1043–1055, 2015.
- [141] A. Mikheenko, V. Saveliev, and A. Gurevich, "MetaQUAST: evaluation of metagenome assemblies," *Bioinformatics*, vol. 32, no. 7, p. 1088–1090, 2016.
- [142] M. Imelfort, D. Parks, B. J. Woodcroft, P. Dennis, P. Hugenholtz, and G. W. Tyson, "GroopM: an automated tool for the recovery of population genomes from related metagenomes," *PeerJ*, vol. 2, p. e603, 2014.
- [143] N. Segata, L. Waldron, A. Ballarini, V. Narasimhan, O. Jousson, and C. Huttenhower, "Metagenomic microbial community profiling using unique clade-specific marker genes," *Nature methods*, vol. 9, no. 8, p. 811, 2012.
- [144] D. H. Huson and N. Weber, "Microbial community analysis using MEGAN," *Methods in enzymology*, vol. 531, p. 465–485, 2013.
- [145] B. Singh, T. L. Crippen, L. Zheng, A. T. Fields, Z. Yu, Q. Ma, T. K. Wood, S. E. Dowd, M. Flores, J. K. Tomberlin *et al.*, "A metagenomic assessment of the bacteria associated with *Lucilia sericata* and *Lucilia cuprina* (Diptera: Calliphoridae)," *Applied microbiology and biotechnology*, vol. 99, no. 2, p. 869–883, 2015.
- [146] S. Flygare, K. Simmon, C. Miller, Y. Qiao, B. Kennedy, T. Di Sera, E. H. Graf, K. D. Tardif, A. Kapusta, S. Ryneerson *et al.*, "Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling," *Genome biology*, vol. 17, no. 1, p. 111, 2016.
- [147] D. Kim, L. Song, F. P. Breitwieser, and S. L. Salzberg, "Centrifuge: rapid and sensitive classification of metagenomic sequences," *Genome research*, vol. 26, no. 12, p. 1721–1729, 2016.
- [148] S. Sunagawa, D. R. Mende, G. Zeller, F. Izquierdo-Carrasco, S. A. Berger, J. R. Kultima, L. P. Coelho, M. Arumugam, J. Tap, H. B. Nielsen *et al.*, "Metagenomic species profiling using universal phylogenetic marker genes," *Nature methods*, vol. 10, no. 12, p. 1196, 2013.
- [149] L. Krause, N. N. Diaz, A. Goesmann, S. Kelley, T. W. Nattkemper, F. Rohwer, R. A. Edwards, and J. Stoye, "Phylogenetic classification of short environmental DNA fragments," *Nucleic acids research*, vol. 36, no. 7, p. 2230–2239, 2008.
- [150] F. Schreiber, P. Gumrich, R. Daniel, and P. Meinicke, "TreePhyler: fast taxonomic profiling of metagenomes," *Bioinformatics*, vol. 26, no. 7, p. 960–961, 2010.
- [151] A. M. Kozlov, J. Zhang, P. Yilmaz, F. O. Glöckner, and A. Stamatakis, "Phylogeny-aware identification and correction of taxonomically mislabeled sequences," *Nucleic acids research*, vol. 44, no. 11, p. 5022–5033, 2016.
- [152] T. N. Petersen, O. Lukjancenko, M. C. F. Thomsen, M. M. Sperotto, O. Lund, F. M. Aarestrup, and T. Sicheritz-Pontén, "MGMapper: Reference based mapping and taxonomy annotation of metagenomics sequence reads," *PloS one*, vol. 12, no. 5, p. e0176469, 2017.
- [153] I. Borozan and V. Ferretti, "CSSSCL: a python package that uses combined sequence similarity scores for accurate taxonomic classification of long and short sequence reads," *Bioinformatics*, vol. 32, no. 3, p. 453–455, 2015.

-
- [154] M. Weber, H. Teeling, S. Huang, J. Waldmann, M. Kassabgy, B. M. Fuchs, A. Klindworth, C. Klockow, A. Wichels, G. Gerdt *et al.*, "Practical application of self-organizing maps to interrelate biodiversity and functional data in NGS-based metagenomics," *The ISME journal*, vol. 5, no. 5, p. 918, 2011.
- [155] P. Menzel, K. L. Ng, and A. Krogh, "Fast and sensitive taxonomic classification for metagenomics with Kaiju," *Nature communications*, vol. 7, p. 11257, 2016.
- [156] S. Wu, Z. Zhu, L. Fu, B. Niu, and W. Li, "WebMGA: a customizable web server for fast metagenomic sequence analysis," *BMC genomics*, vol. 12, no. 1, p. 444, 2011.
- [157] B. Liu, T. Gibbons, M. Ghodsi, T. Treangen, and M. Pop, "Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences," *Genome biology*, vol. 12, no. 1, p. P11, 2011.
- [158] D. Ainsworth, M. J. Sternberg, C. Racz, and S. A. Butcher, "k-SLAM: accurate and ultra-fast taxonomic classification and gene identification for large metagenomic data sets," *Nucleic acids research*, vol. 45, no. 4, p. 1649–1656, 2017.
- [159] S. K. Ames, D. A. Hysom, S. N. Gardner, G. S. Lloyd, M. B. Gokhale, and J. E. Allen, "Scalable metagenomic taxonomy classification using a reference genome database," *Bioinformatics*, vol. 29, no. 18, p. 2253–2260, 2013.
- [160] W.-Z. Song and T. Thomas, "Binning_refiner: improving genome bins through the combination of different binning programs," *Bioinformatics*, vol. 33, no. 12, p. 1873–1875, 2017.
- [161] M. Stark, S. A. Berger, A. Stamatakis, and C. von Mering, "MLTreeMap-accurate Maximum Likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies," *BMC genomics*, vol. 11, no. 1, p. 461, 2010.
- [162] N. N. Diaz, L. Krause, A. Goesmann, K. Niehaus, and T. W. Nattkemper, "TACOA--Taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach," *BMC bioinformatics*, vol. 10, no. 1, p. 56, 2009.
- [163] D. Guan, B. Liu, and Y. Wang, "deSPI: efficient classification of metagenomic reads with lightweight de Bruijn graph-based reference indexing," *bioRxiv*, p. 080200, 2016.
- [164] R. Cao, C. Freitas, L. Chan, M. Sun, H. Jiang, and Z. Chen, "ProLanGO: Protein Function Prediction Using Neural Machine Translation Based on a Recurrent Neural Network," *Molecules*, vol. 22, no. 10, p. 1732, 2017.
- [165] C. Wang, D. Dong, H. Wang, K. Müller, Y. Qin, H. Wang, and W. Wu, "Metagenomic analysis of microbial consortia enriched from compost: new insights into the role of Actinobacteria in lignocellulose decomposition," *Biotechnology for biofuels*, vol. 9, no. 1, p. 22, 2016.
- [166] N. W. Hanson, K. M. Konwar, and S. J. Hallam, "LCA*: an entropy-based measure for taxonomic assignment within assembled metagenomes," *Bioinformatics*, vol. 32, no. 23, p. 3535–3542, 2016.
- [167] P. Somervuo, S. Koskela, J. Pennanen, R. Henrik Nilsson, and O. Ovaskainen, "Unbiased probabilistic taxonomic classification for DNA barcoding," *Bioinformatics*, vol. 32, no. 19, p. 2920–2927, 2016.
- [168] V. C. Piro, M. S. Lindner, and B. Y. Renard, "DUDes: a top-down taxonomic profiler for metagenomics," *Bioinformatics*, vol. 32, no. 15, p. 2272–2280, 2016.
-

- [169] Y.-Q. Qiu, X. Tian, and S. Zhang, "Infer Metagenomic Abundance and Reveal Homologous Genomes Based on the Structure of Taxonomy Tree," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 12, no. 5, p. 1112–1122, 2015.
- [170] J. Lu, F. P. Breitwieser, P. Thielen, and S. L. Salzberg, "Bracken: estimating species abundance in metagenomics data," *PeerJ Computer Science*, vol. 3, p. e104, 2017.
- [171] S. S. Minot, N. Krumm, and N. B. Greenfield, "One codex: a sensitive and accurate data platform for genomic microbial identification," *bioRxiv*, p. 027607, 2015.
- [172] F. Gori, G. Folino, M. S. Jetten, and E. Marchiori, "MTR: taxonomic annotation of short metagenomic reads using clustering at multiple taxonomic ranks," *Bioinformatics*, vol. 27, no. 2, p. 196–203, 2010.
- [173] M. S. Porter and R. G. Beiko, "SPANNER: taxonomic assignment of sequences using pyramid matching of similarity profiles," *Bioinformatics*, vol. 29, no. 15, p. 1858–1864, 2013.
- [174] M. B. Sohn, L. An, N. Pookhao, and Q. Li, "Accurate genome relative abundance estimation for closely related species in a metagenomic sample," *BMC bioinformatics*, vol. 15, no. 1, p. 242, 2014.
- [175] A. E. Darling, G. Jospin, E. Lowe, F. A. Matsen IV, H. M. Bik, and J. A. Eisen, "PhyloSift: phylogenetic analysis of genomes and metagenomes," *PeerJ*, vol. 2, p. e243, 2014.
- [176] P. Jia, L. Xuan, L. Liu, and C. Wei, "MetaBinG: Using GPUs to accelerate metagenomic sequence classification," *PloS one*, vol. 6, no. 11, p. e25353, 2011.
- [177] H. Jiang, L. An, S. M. Lin, G. Feng, and Y. Qiu, "A statistical framework for accurate taxonomic assignment of metagenomic sequencing reads," *PLoS One*, vol. 7, no. 10, p. e46450, 2012.
- [178] Z. Rasheed and H. Rangwala, "Metagenomic taxonomic classification using extreme learning machines," *Journal of bioinformatics and computational biology*, vol. 10, no. 5, p. 1250015, 2012.
- [179] D. E. Wood and S. L. Salzberg, "Kraken: ultrafast metagenomic sequence classification using exact alignments," *Genome biology*, vol. 15, no. 3, p. R46, 2014.
- [180] A. Kislyuk, S. Bhatnagar, J. Dushoff, and J. S. Weitz, "Unsupervised statistical clustering of environmental shotgun sequences," *BMC bioinformatics*, vol. 10, no. 1, p. 316, 2009.
- [181] C. F. Davenport, J. Neugebauer, N. Beckmann, B. Friedrich, B. Kameri, S. Kokott, M. Paetow, B. Siekmann, M. Wieding-Drewes, M. Wienhöfer *et al.*, "Genometa-a fast and accurate classifier for short metagenomic shotgun reads," *PloS one*, vol. 7, no. 8, p. e41224, 2012.
- [182] C. Luo, L. M. Rodriguez-r, and K. T. Konstantinidis, "MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences," *Nucleic acids research*, vol. 42, no. 8, p. e73–e73, 2014.
- [183] A. A. Metwally, Y. Dai, P. W. Finn, and D. L. Perkins, "WEVOTE: weighted voting taxonomic identification method of microbial sequences," *PloS one*, vol. 11, no. 9, p. e0163527, 2016.
- [184] X. Dong, M. Kleiner, C. E. Sharp, E. Thorson, C. Li, D. Liu, and M. Strous, "Fast and simple analysis of MiSeq amplicon sequencing data with MetaAmp," *Frontiers in microbiology*, vol. 8, p. 1461, 2017.
- [185] B. Zou, J. Li, Q. Zhou, and Z.-X. Quan, "MIPE: A metagenome-based community structure explorer and SSU primer evaluation tool," *PloS one*, vol. 12, no. 3, p. e0174609, 2017.

-
- [186] A. Pati, L. S. Heath, N. C. Kyrpides, and N. Ivanova, "ClaMS: a classifier for metagenomic sequences," *Standards in genomic sciences*, vol. 5, no. 2, p. 248, 2011.
- [187] J. Liu, H. Wang, H. Yang, Y. Zhang, J. Wang, F. Zhao, and J. Qi, "Composition-based classification of short metagenomic sequences elucidates the landscapes of taxonomic and functional enrichment of microorganisms," *Nucleic acids research*, vol. 41, no. 1, p. e3–e3, 2012.
- [188] J. Kawulok and S. Deorowicz, "CoMeta: classification of metagenomes using k-mers," *PloS one*, vol. 10, no. 4, p. e0121453, 2015.
- [189] J. Berendzen, W. J. Bruno, J. D. Cohn, N. W. Hengartner, C. R. Kuske, B. H. McMahon, M. A. Wolinsky, and G. Xie, "Rapid phylogenetic and functional classification of short genomic fragments with signature peptides," *BMC research notes*, vol. 5, no. 1, p. 460, 2012.
- [190] M. Scheuch, D. Höper, and M. Beer, "RIEMS: a software pipeline for sensitive and comprehensive taxonomic classification of reads from metagenomics datasets," *BMC bioinformatics*, vol. 16, no. 1, p. 69, 2015.
- [191] G. L. Rosen, E. R. Reichenberger, and A. M. Rosenfeld, "NBC: the Naive Bayes Classification tool webserver for taxonomic classification of metagenomic reads," *Bioinformatics*, vol. 27, no. 1, p. 127–129, 2010.
- [192] N. J. MacDonald, D. H. Parks, and R. G. Beiko, "Rapid identification of high-confidence taxonomic assignments for metagenomic data," *Nucleic acids research*, vol. 40, no. 14, p. e111–e111, 2012.
- [193] D. Koslicki, S. Foucart, and G. Rosen, "WGSQuikr: fast whole-genome shotgun metagenomic classification," *PloS one*, vol. 9, no. 3, p. e91784, 2014.
- [194] S. Rodrigue, A. C. Materna, S. C. Timberlake, M. C. Blackburn, R. R. Malmstrom, E. J. Alm, and S. W. Chisholm, "Unlocking short read sequencing for metagenomics," *PloS one*, vol. 5, no. 7, p. e11840, 2010.
- [195] K. R. Patil, L. Roune, and A. C. McHardy, "The PhyloPythiaS web server for taxonomic assignment of metagenome sequences," *PloS one*, vol. 7, no. 6, p. e38581, 2012.
- [196] Y. Wang, K. Wang, Y. Y. Lu, and F. Sun, "Improving contig binning of metagenomic data using oligonucleotide frequency dissimilarity," *BMC Bioinformatics*, vol. 18, no. 1, 2017.
- [197] R. Kobus, C. Hundt, A. Müller, and B. Schmidt, "Accelerating metagenomic read classification on CUDA-enabled GPUs," *BMC bioinformatics*, vol. 18, no. 1, p. 11, 2017.
- [198] F. de ric Escudie, L. Auer, M. Bernard, M. Mariadassou, L. Cauquil, K. Vidal, S. Maman, G. Hernandez-Raquet, S. Combes, and G. raldine Pascal, "FROGS: Find, Rapidly, OTUs with Galaxy Solution," *Bioinformatics*, vol. 1, p. 8, 2017.
- [199] O. Tanaseichuk, J. Borneman, and T. Jiang, "Separating metagenomic short reads into genomes via clustering," *Algorithms for Molecular Biology*, vol. 7, no. 1, p. 27, 2012.
- [200] V. I. Ulyantsev, S. V. Kazakov, V. B. Dubinkina, A. V. Tyakht, and D. G. Alexeev, "MetaFast: fast reference-free graph-based comparison of shotgun metagenomic data," *Bioinformatics*, vol. 32, no. 18, p. 2760–2767, 2016.
- [201] Y. Y. Lu, T. Chen, J. A. Fuhrman, and F. Sun, "COCACOLA: binning metagenomic contigs using sequence COMposition, read CoverAge, CO-alignment and paired-end read LinkAge," *Bioinformatics*, vol. 33, no. 6, p. 791–798, 2017.
- [202] D. R. Kelley and S. L. Salzberg, "Clustering metagenomic sequences with interpolated Markov models," *BMC bioinformatics*, vol. 11, no. 1, p. 544, 2010.
-

- [203] C. C. Laczny, T. Sternal, V. Plugaru, P. Gawron, A. Atashpendar, H. H. Margossian, S. Coronado, L. Van der Maaten, N. Vlassis, and P. Wilmes, "VizBin-an application for reference-independent visualization and human-augmented binning of metagenomic data," *Microbiome*, vol. 3, no. 1, p. 1, 2015.
- [204] S. Girotto, C. Pizzi, and M. Comin, "MetaProb: accurate metagenomic reads binning based on probabilistic sequence signatures," *Bioinformatics*, vol. 32, no. 17, p. i567–i575, 2016.
- [205] M. Strous, B. Kraft, R. Bisdorf, and H. Tegetmeyer, "The binning of metagenomic contigs for microbial physiology of mixed cultures," *Frontiers in microbiology*, vol. 3, p. 410, 2012.
- [206] F. Meyer, P. Hofmann, P. Belmann, R. Garrido-Oter, A. Fritz, A. Sczyrba, and A. C. McHardy, "AMBER: Assessment of Metagenome BinnERs," *bioRxiv*, p. 239582, 2017.
- [207] K. Harismah, M. Mirzaei, N. Ghasemi, and M. Nejati, "Non-Covalent Functionalisation of C30 Fullerene by Pyrrole-n-Carboxylic Acid (n= 2, 3): Density Functional Theory Studies," *Zeitschrift für Naturforschung A*, vol. 73, no. 1, p. 51–56, 2017.
- [208] H.-H. Lin and Y.-C. Liao, "Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes," *Scientific reports*, vol. 6, p. 24175, 2016.
- [209] X. Ding, F. Cheng, C. Cao, and X. Sun, "DectICO: an alignment-free supervised metagenomic classification method based on feature extraction and dynamic selection," *BMC bioinformatics*, vol. 16, no. 1, p. 323, 2015.
- [210] Y. Wang, H. Hu, and X. Li, "MBBC: an efficient approach for metagenomic binning based on clustering," *BMC bioinformatics*, vol. 16, no. 1, p. 36, 2015.
- [211] G. Benoit, P. Peterlongo, M. Mariadassou, E. Drezen, S. Schbath, D. Lavenier, and C. Lemaitre, "Multiple comparative metagenomics using multiset k-mer counting," *PeerJ Computer Science*, vol. 2, p. e94, 2016.
- [212] V. Van Le, L. Van Tran, and H. Van Tran, "A novel semi-supervised algorithm for the taxonomic assignment of metagenomic reads," *BMC bioinformatics*, vol. 22, no. 1, p. 22, 2016.
- [213] B. E. Dutilh, R. Schmieder, J. Nulton, B. Felts, P. Salamon, R. A. Edwards, and J. L. Mokili, "Reference-independent comparative metagenomics using cross-assembly: crAss," *Bioinformatics*, vol. 28, no. 24, p. 3225–3231, 2012.
- [214] S. Chatterji, I. Yamazaki, Z. Bai, and J. A. Eisen, "CompostBin: A DNA composition-based algorithm for binning environmental shotgun reads," in *Annual International Conference on Research in Computational Molecular Biology*, 2008, p. 17–28.
- [215] A. Brady and S. L. Salzberg, "Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models," *Nature methods*, vol. 6, no. 9, p. 673, 2009.
- [216] D. R. Kelley, B. Liu, A. L. Delcher, M. Pop, and S. L. Salzberg, "Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering," *Nucleic acids research*, vol. 40, no. 1, p. e9–e9, 2011.
- [217] M. Rho, H. Tang, and Y. Ye, "FragGeneScan: predicting genes in short and error-prone reads," *Nucleic acids research*, vol. 38, no. 20, p. e191–e191, 2013.
- [218] W. Zhu, A. Lomsadze, and M. Borodovsky, "Ab initio gene identification in metagenomic sequences," *Nucleic acids research*, vol. 38, no. 12, p. e132–e132, 2010.

-
- [219] S. Caboche, G. Even, A. Loywick, C. Audebert, and D. Hot, "MICRA: an automatic pipeline for fast characterization of microbial genomes from high-throughput sequencing data," *Genome biology*, vol. 18, no. 1, p. 233, 2017.
- [220] H. Noguchi, T. Taniguchi, and T. Itoh, "MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes," *DNA research*, vol. 15, no. 6, p. 387–396, 2008.
- [221] R. Ramanan, B.-H. Kim, D.-H. Cho, H.-M. Oh, and H.-S. Kim, "Algae–bacteria interactions: evolution, ecology and emerging applications," *Biotechnology advances*, vol. 34, no. 1, p. 14–29, 2016.
- [222] K. J. Hoff, T. Lingner, P. Meinicke, and M. Tech, "Orphelia: predicting genes in metagenomic sequencing reads," *Nucleic acids research*, vol. 37, no. suppl_2, p. W101–W105, 2009.
- [223] M. D'Antonio, P. D. De Meo, M. Pallocca, E. Picardi, A. M. D'Erchia, R. A. Calogero, T. Castrignanò, and G. Pesole, "RAP: RNA-Seq Analysis Pipeline, a new cloud-based NGS web application," *BMC genomics*, vol. 16, no. 6, p. S3, 2015.
- [224] D. Koslicki, "Education/Employment," *methods*, vol. 14, no. 11, p. 1063–1071, 2017.
- [225] M. Moniruzzaman, G. R. LeClerc, C. M. Brown, C. J. Gobler, K. D. Bidle, W. H. Wilson, and S. W. Wilhelm, "Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host–virus coevolution," *Virology*, vol. 466, p. 60–70, 2014.
- [226] S. Tang, I. Antonov, and M. Borodovsky, "MetaGeneTack: ab initio detection of frameshifts in metagenomic sequences," *Bioinformatics*, vol. 29, no. 1, p. 114–116, 2012.
- [227] W. M. Ismail, Y. Ye, and H. Tang, "Gene finding in metatranscriptomic sequences," *BMC bioinformatics*, vol. 15, no. 9, p. S8, 2014.
- [228] J. R. Kultima, L. P. Coelho, K. Forslund, J. Huerta-Cepas, S. S. Li, M. Driessen, A. Y. Voigt, G. Zeller, S. Sunagawa, and P. Bork, "MOCAT2: a metagenomic assembly, annotation and profiling framework," *Bioinformatics*, vol. 32, no. 16, p. 2520–2523, 2016.
- [229] D. Hyatt, P. F. LoCascio, L. J. Hauser, and E. C. Uberbacher, "Gene and translation initiation site prediction in metagenomic sequences," *Bioinformatics*, vol. 28, no. 17, p. 2223–2230, 2012.
- [230] Y. Huang, P. Gilna, and W. Li, "Identification of ribosomal RNA genes in metagenomic fragments," *Bioinformatics*, vol. 25, no. 10, p. 1338–1340, 2013.
- [231] T. Loman, "A novel method for predicting ribosomal rna genes in prokaryotic genomes," 2017, student Paper.
- [232] K. Lagesen, P. Hallin, E. A. Rødland, H.-H. Stærfeldt, T. Rognes, and D. W. Ussery, "RNAmmer: consistent and rapid annotation of ribosomal RNA genes," *Nucleic acids research*, vol. 35, no. 9, p. 3100–3108, 2007.
- [233] F. A. Matsen, R. B. Kodner, and E. V. Armbrust, "pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree," *BMC bioinformatics*, vol. 11, no. 1, p. 538, 2010.
- [234] S. C. Potter, A. Luciani, S. R. Eddy, Y. Park, R. Lopez, and R. D. Finn, "HMMER web server: 2018 update," *Nucleic acids research*, vol. 46, no. W1, p. W200–W204, 2018.
- [235] A. Mitchell, H.-Y. Chang, L. Daugherty, M. Fraser, S. Hunter, R. Lopez, C. McAnulla, C. McMenamin, G. Nuka, S. Pesseat *et al.*, "The InterPro protein families database: the
-

- classification resource after 15 years," *Nucleic acids research*, vol. 43, no. D1, p. D213–D221, 2014.
- [236] H.-G. Drost, A. Gabel, J. Liu, M. Quint, and I. Grosse, "myTAI: Evolutionary Transcriptomics with R," *Bioinformatics*, p. btx835, 2017.
- [237] M. Kanehisa, Y. Sato, and K. Morishima, "BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences," *Journal of molecular biology*, vol. 428, no. 4, p. 726–731, 2016.
- [238] Y. Zhang, Y. Sun, and J. R. Cole, "A Sensitive and Accurate protein domain cLassification Tool (SALT) for short reads," *Bioinformatics*, vol. 29, no. 17, p. 2103–2111, 2013.
- [239] B. Buchfink, C. Xie, and D. H. Huson, "Fast and sensitive protein alignment using DIAMOND," *Nature methods*, vol. 12, no. 1, p. 59, 2015.
- [240] L. H. Orellana, L. M. Rodriguez-R, and K. T. Konstantinidis, "ROCK: accurate detection and quantification of target genes in short-read metagenomic data sets by modeling sliding-window bitscores," *Nucleic acids research*, vol. 45, no. 3, p. e14–e14, 2016.
- [241] P. Meinicke, "UProC: tools for ultra-fast protein domain classification," *Bioinformatics*, vol. 31, no. 9, p. 1382–1388, 2014.
- [242] A. K. Sharma, A. Gupta, S. Kumar, D. B. Dhakan, and V. K. Sharma, "Woods: a fast and accurate functional annotator and classifier of genomic and metagenomic sequences," *Genomics*, vol. 106, no. 1, p. 1–6, 2015.
- [243] R. A. Edwards, R. Olson, T. Disz, G. D. Pusch, V. Vonstein, R. Stevens, and R. Overbeek, "Real time metagenomics: using k-mers to annotate metagenomes," *Bioinformatics*, vol. 28, no. 24, p. 3316–3317, 2012.
- [244] R. Stewart, M. Auffret, T. Snelling, R. Roehe, and M. Watson, "MAGpy: a reproducible pipeline for the downstream analysis of metagenome-assembled genomes (MAGs)," *bioRxiv*, p. 233544, 2017.
- [245] M. Zakrzewski, T. Bekel, C. Ander, A. Pühler, O. Rupp, J. Stoye, A. Schlüter, and A. Goesmann, "MetaSAMS—a novel software platform for taxonomic classification, functional annotation and comparative analysis of metagenome datasets," *Journal of biotechnology*, vol. 167, no. 2, p. 156–165, 2013.
- [246] V. De Anda, I. Zapata-Peñasco, A. C. Poot-Hernandez, L. E. Eguiarte, B. Contreras-Moreira, and V. Souza, "MEBS, a software platform to evaluate large (meta) genomic collections according to their metabolic machinery: unraveling the sulfur cycle," *GigaScience*, vol. 6, no. 11, p. gix096, 2017.
- [247] K. Choi, J. K. Medley, C. Cannistra, M. König, L. Smith, K. Stocking, and H. M. Sauro, "Tellurium: a python based modeling and reproducibility platform for systems biology," *BioRxiv*, p. 054601, 2016.
- [248] O. K. Kamneva, "Genome composition and phylogeny of microbes predict their co-occurrence in the environment," *PLoS computational biology*, vol. 13, no. 2, p. e1005366, 2017.
- [249] M. G. Poolman, "ScrumPy: metabolic modelling with Python," *IEE Proceedings-Systems Biology*, vol. 153, no. 5, p. 375–378, 2006.
- [250] A. Rodriguez-Martinez, R. Ayala, J. M. Posma, A. L. Neves, D. Gauguier, J. K. Nicholson, and M.-E. Dumas, "MetaboSignal: a network-based approach for topological analysis of

- metabotype regulation via metabolic and signaling pathways," *Bioinformatics*, vol. 33, no. 5, p. 773–775, 2017.
- [251] G. Basler and Z. Nikoloski, "JMassBalance: Mass-balanced randomization and analysis of metabolic networks," *Bioinformatics*, vol. 27, no. 19, p. 2761–2762, 2011.
- [252] J. J. Hamilton and J. L. Reed, "Identification of functional differences in metabolic networks using comparative genomics and constraint-based models," *PloS one*, vol. 7, no. 4, 2012.
- [253] L. B. Jorde, "Linkage disequilibrium as a gene-mapping tool.," *American journal of human genetics*, vol. 56, no. 1, p. 11, 1995.
- [254] A. Rodriguez-Martinez, R. Ayala, J. M. Posma, N. Harvey, B. Jiménez, K. Sonomura, T.-A. Sato, F. Matsuda, P. Zalloua, D. Gauguier *et al.*, "pJRES Binning Algorithm (JBA): a new method to facilitate the recovery of metabolic information from pJRES 1H NMR spectra," *Bioinformatics*, vol. 35, no. 11, p. 1916–1922, 2019.
- [255] A. Rodriguez-Martinez, J. M. Posma, R. Ayala, A. L. Neves, M. Anwar, E. Petretto, C. Emanuelli, D. Gauguier, J. K. Nicholson, and M.-E. Dumas, "MWASTools: an R/bioconductor package for metabolome-wide association studies," *Bioinformatics*, vol. 34, no. 5, p. 890–892, 2018.
- [256] S. Waldherr, "A guideline to model reduction by stoichiometric decomposition for biochemical network analysis," in *Proc. of the 21st International Symposium on Mathematical Theory of Networks and Systems*, 2014, p. 490–495.
- [257] K. Wanichthanarak, S. Fan, D. Grapov, D. K. Barupal, and O. Fiehn, "Metabox: A toolbox for metabolomic data analysis, interpretation and integrative exploration," *PloS one*, vol. 12, no. 1, 2017.
- [258] H. U. Zacharias, T. Rehberg, S. Mehrl, D. Richtmann, T. Wettig, P. J. Oefner, R. Spang, W. Gronwald, and M. Altenbuchinger, "Scale-invariant biomarker discovery in urine and plasma metabolite fingerprints," *Journal of proteome research*, vol. 16, no. 10, p. 3596–3605, 2017.
- [259] R. B. Aggio, K. Ruggiero, and S. G. Villas-Bôas, "Pathway Activity Profiling (PAPi): from the metabolite profile to the metabolic pathway activity," *Bioinformatics*, vol. 26, no. 23, p. 2969–2976, 2010.
- [260] S. C. Booth, A. M. Weljie, and R. J. Turner, "Computational tools for the secondary analysis of metabolomics experiments," *Computational and structural biotechnology journal*, vol. 4, no. 5, p. e201301003, 2013.
- [261] E. G. Cerami, G. D. Bader, B. E. Gross, and C. Sander, "cPath: open source software for collecting, storing, and querying biological pathways," *BMC bioinformatics*, vol. 7, no. 1, p. 497, 2006.
- [262] C. P. McNally, A. Eng, C. Noecker, W. C. Gagne-Maynard, and E. Borenstein, "BURRITO: An interactive multi-omic tool for visualizing taxa-function relationships in microbiome data," *bioRxiv*, p. 217315, 2017.
- [263] Z. S. Foster, T. J. Sharpton, and N. J. Grünwald, "Metacoder: An R package for visualization and manipulation of community taxonomic diversity data," *PLoS computational biology*, vol. 13, no. 2, p. e1005404, 2017.
- [264] C. E. Robertson, J. K. Harris, B. D. Wagner, D. Granger, K. Browne, B. Tatem, L. M. Feazel, K. Park, N. R. Pace, and D. N. Frank, "Explicitet: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data," *Bioinformatics*, vol. 29, no. 23, p. 3100–3101, 2013.

- [265] D. N. Ayyala and S. Lin, "GrammR: graphical representation and modeling of count data with application in metagenomics," *Bioinformatics*, vol. 31, no. 10, p. 1648–1654, 2015.
- [266] B. D. Ondov, N. H. Bergman, and A. M. Phillippy, "Interactive metagenomic visualization in a Web browser," *BMC bioinformatics*, vol. 12, no. 1, p. 385, 2011.
- [267] S. M. Huse, D. B. M. Welch, A. Voorhis, A. Shipunova, H. G. Morrison, A. M. Eren, and M. L. Sogin, "VAMPS: a website for visualization and analysis of microbial population structures," *BMC bioinformatics*, vol. 15, no. 1, p. 41, 2014.
- [268] F. Asnicar, G. Weingart, T. L. Tickle, C. Huttenhower, and N. Segata, "Compact graphical representation of phylogenetic data and metadata with GraPhlAn," *PeerJ*, vol. 3, p. e1029, 2015.
- [269] A. Orakov, N. Sakenova, A. Sorokin, and I. Goryanin, "ASAR: visual analysis of metagenomes in R," *Bioinformatics*, vol. 1, no. 2, 2017.
- [270] J. Wagner, F. Chelaru, J. Kancherla, J. N. Paulson, V. Felix, A. Mahurkar, and H. C. Bravo, "Metaviz: interactive statistical and visual analysis of metagenomic data," *bioRxiv*, p. 105205, 2017.
- [271] P. Lechat, E. Souche, and I. Moszer, "SynTVView—an interactive multi-view genome browser for next-generation comparative microorganism genomics," *BMC bioinformatics*, vol. 14, no. 1, p. 277, 2013.
- [272] C. Kerepesi, B. Szalkai, and V. Grolmusz, "Visual analysis of the quantitative composition of metagenomic communities: the AmphoraVizu webserver," *Microbial ecology*, vol. 69, no. 3, p. 695–697, 2015.
- [273] T. Uchiyama, M. Irie, H. Mori, K. Kurokawa, and T. Yamada, "FuncTree: functional analysis and visualization for large-scale omics data," *PloS one*, vol. 10, no. 5, p. e0126967, 2015.
- [274] D. Correia, O. Doppelt-Azeroual, J.-B. Denis, M. Vandenbogaert, and V. Caro, "MetaGenSense: A web-application for analysis and exploration of high throughput sequencing metagenomic data," *F1000Research*, vol. 4, 2015.
- [275] M. Cantor, H. Nordberg, T. Smirnova, M. Hess, S. Tringe, and I. Dubchak, "Elviz—exploration of metagenome assemblies with an interactive visualization tool," *BMC bioinformatics*, vol. 16, no. 1, p. 130, 2015.
- [276] G. V. Uritskiy, J. DiRuggiero, and J. Taylor, "MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis," *bioRxiv*, p. 277442, 2018.
- [277] J. M. Eppley, G. W. Tyson, W. M. Getz, and J. F. Banfield, "Strainer: software for analysis of population variation in community genomic datasets," *BMC bioinformatics*, vol. 8, no. 1, p. 398, 2013.