

---

# METGENOX v4.0

## User Guide





Copyright (C) 2021, Team Metgeno

### Team MetgenoX, 2017-2021

Hasan Iqbal

Bilal Wajid

Safee Ullah Chaudhry

Faria Anwar

Imran Wajid

### Metgeno 3.0 User Guide

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License 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 General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see: <https://www.gnu.org/licenses/>

#### Disclaimer of Warranty:

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD

THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

#### Limitation of Liability:

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

#### Trademarks:

Sabz Qalam and the "feather" logo are registered trademarks of Sabz Qalam. All other marks are property of their respective owners.

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

REF MT-300

**Acknowledgment:**

This work has been supported by Sabz-Qalam via grant SQ-2021-Bioinfo-2.



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

<b>About MetgenoX</b>	<b>1</b>
Abstract . . . . .	1
Purpose . . . . .	1
Intended Audience . . . . .	1
Required Background . . . . .	1
How This Guide is Organized . . . . .	2
Contact . . . . .	2
<b>MetgenoX Package</b>	<b>3</b>
Software Selection Criteria . . . . .	3
Software Categorization Scheme . . . . .	9
MetgenoX Packaged Tools . . . . .	12
<b>Software Installation Process</b>	<b>17</b>
Installing MetgenoX software . . . . .	17
<b>Microsoft Windows Support</b>	<b>31</b>
Installing MetgenoX software on Microsoft Windows . . . . .	31
<b>macOS Support</b>	<b>41</b>
Installing MetgenoX software on macOS . . . . .	41



# About MetgenoX

---

## Abstract

MetgenoX is a free, easily distributable Ubuntu based software installation package for freeware tools related to Metagenomics.

MetgenoX is a software package that facilitate easy download and installation of tools related to Metegenomics software pipeline as per user requirement. It also downloads the dependencies prior to the tools for which they are required. MetgenoX provides an interactive graphical user interface which is easily understandable for users unaware of UNIX-shell language.

---

## Purpose

The basic purpose of this document is to provide the complete description about the installation of MetgenoX shell file. This document also explains the complete pipeline of Metagenomics; also shows how this pipeline is implemented in the installation package. Running MetgenoX on Microsoft Windows and macOS through Virtualization is also explained.

---

## Intended Audience

MetgenoX is intended for researchers, life scientists and users who wish to install the tools related to Metagenomics.

---

## Required Background

Team MetgenoX has made every attempt to make this a step by step guide. However, some familiarity with Linux operating system as well as software and hardware requirements of MetgenoX are assumed.

---

## How This Guide is Organized

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

- About This Guide (Chapter 1) describes this document's purpose and intended audience.
- MetgenoX Package (Chapter 2) has tables for Software Selection Criteria, Categorization Scheme and Included Tools
- Software Installation Process (Chapter 3) describes how to install MetgenoX.
- Microsoft Windows Support (Chapter 4) describes a step by step procedure to configure a virtual OS for MetgenoX on Windows.
- macOS Support (Chapter 5) describes a step by step procedure to configure a virtual OS for MetgenoX on macOS.

---

## Contact

For any further queries and suggestions, contact us at: [hasaniqbal777@gmail.com](mailto:hasaniqbal777@gmail.com)

# MetgenoX Package

## Software Selection Criteria

A specific criteria is used for enclosure of tools in MetgenoX. Table shows this selection criteria. Software which are included are due to these specific reasons:

- Publication Date (later then 2010)
- Freeware License
- Linux based
- Offline

Sr. no	Software Tool Name	Pub. Date (2012-)	Offline	Linux based	Free	Installed
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	STPipeline[17]		✓	✓	✓	
18	AlmostSignificant[18]		✓	✓	✓	
19	SOAPnuke[19]	✓	✓	✓	✓	✓
20	miARRna-Seq[20]		✓	✓	✓	
21	CoVaCS[21]	✓		✓		
22	NGSQCToolkit[22]		✓	✓	✓	
23	RNA-QC-Chain[23]		✓	✓	✓	
24	miRPursuit[24]		✓	✓	✓	
25	SolexaQA[25]	✓	✓	✓	✓	✓
26	mubimomics[26]	✓	✓		✓	
27	Oncorep[27]		✓	✓	✓	
28	S-MART[28]		✓	✓	✓	
29	ReQON[29]		✓	✓	✓	
30	TRAPLINE[30]	✓	✓		✓	
31	PyroTrimmer[31]		✓	✓	✓	
32	Rnnotator[32]		✓	✓	✓	
33	TRUFA[33]	✓		✓		

## MetgenoX v4.0

---

34	BIGpre[34]		✓	✓	✓	
35	QuaCRS[35]	✓		✓		
36	preAssemble[36]		✓	✓	✓	
37	PRINSEQ[37]	✓	✓	✓	✓	✓
38	Pyrocleaner[38]		✓	✓	✓	
39	PANDAseq[39]	✓	✓	✓	✓	✓
40	BWA	✓	✓	✓	✓	✓
41	ClustalOmega[40]	✓	✓	✓	✓	✓
42	Greengenes[41]	✓	✓	✓	✓	✓
43	Nonpareil[42]	✓	✓	✓	✓	✓
44	CorQ[43]		✓	✓	✓	
45	FQC[44]	✓	✓	✓	✓	✓
46	QA[45]	✓	✓	✓	✓	✓
47	FaQCs[46]	✓	✓	✓	✓	✓
48	AfterQC[47]		✓	✓	✓	
49	HTQC[48]	✓	✓	✓	✓	✓
50	SUGAR[49]	✓	✓	✓	✓	✓
51	KAT[50]	✓	✓	✓	✓	✓
52	FastQ_brew[51]		✓	✓	✓	
53	QUASR[52]	✓	✓	✓	✓	✓
54	NGS-QCGenerator[53]		✓	✓	✓	
55	NGS-eval[54]	✓		✓		
56	mistagging[55]		✓	✓	✓	
57	ShortRead[56]	✓	✓	✓	✓	✓
58	sigQC[57]		✓	✓	✓	
59	StatsDB[58]		✓	✓	✓	
60	G-CNV[59]		✓	✓	✓	
61	fitGCP[60]	✓	✓	✓	✓	✓
62	Pheniqs[61]		✓	✓	✓	
63	QASDRA[62]		✓	✓	✓	
64	Fast-GBS[63]		✓	✓	✓	
65	ClinQC[64]		✓	✓	✓	
66	SeqAssist[65]		✓	✓	✓	
67	PathoQC[66]	✓	✓	✓	✓	✓
68	Zseq[67]		✓	✓	✓	
69	MuffinInfo[68]		✓	✓	✓	
70	A-Game[69]	✓		✓		
71	MT-Toolbox[70]		✓	✓	✓	
72	StreamingTrim[71]		✓	✓	✓	
73	FASTQC[72]	✓	✓	✓	✓	✓
74	ACDC[73]	✓	✓	✓	✓	✓
75	SqClean[74]		✓	✓	✓	
76	Vecuum[75]		✓	✓	✓	
77	PhylOligo[76]	✓	✓	✓	✓	✓
78	AFS[77]		✓	✓	✓	
79	MCSC[78]	✓	✓	✓	✓	✓
80	DeconSeq[79]		✓	✓	✓	
81	Kraken[80]	✓	✓	✓	✓	✓
82	Conpair[81]	✓	✓	✓	✓	✓
83	ContEst[82]		✓	✓	✓	
84	decontam[83]		✓	✓	✓	
85	SIDR[84]	✓	✓	✓	✓	✓
86	MIDAS[85]	✓	✓	✓	✓	✓
87	Coockiecutter[86]	✓			✓	
88	Blobology[87]	✓	✓	✓	✓	✓
89	PhagePhisher[88]	✓	✓	✓	✓	✓
90	Taxoblast[89]	✓	✓	✓	✓	✓
91	HYSYS[90]	✓	✓	✓	✓	✓
92	Orione[91]	✓		✓		
93	MetAMOS[92]	✓	✓	✓	✓	✓
94	dDocent[93]	✓	✓	✓	✓	✓
95	RayMeta[94]	✓	✓	✓	✓	✓

96	TaxMan[95]	✓		✓		
97	DRAP[96]		✓	✓	✓	
98	MetaVelvet[97]	✓	✓	✓	✓	✓
99	IDBA-UD[98]	✓	✓	✓	✓	✓
100	MetaSort[99]	✓	✓	✓	✓	✓
101	Snowball[100]	✓	✓	✓	✓	✓
102	Meta-IDBA[101]		✓	✓	✓	
103	PRICE[102]	✓	✓	✓	✓	✓
104	Phrap[103]		✓	✓	✓	
105	BBAP[104]	✓	✓	✓	✓	✓
106	BBMAP[105]	✓	✓	✓	✓	✓
107	MEGAHIT[106]	✓	✓	✓	✓	✓
108	Xander[107]		✓	✓	✓	
109	InteMAP[108]		✓	✓	✓	
110	Genovo[109]		✓	✓	✓	
111	MetaStorm[110]	✓		✓		
112	Faucet[111]	✓	✓	✓	✓	✓
113	Athena[112]	✓	✓	✓	✓	✓
114	MetaFinisherSC[113]	✓	✓	✓	✓	✓
115	MetaVelvet-SL[114]		✓	✓	✓	
116	Mapsembler[115]		✓	✓	✓	
117	Spherical[116]	✓	✓	✓	✓	✓
118	BinSanity[117]		✓	✓	✓	
119	Xgenovo[118]		✓	✓	✓	
120	SAT-Assembler[119]		✓	✓	✓	
121	DIME[120]		✓	✓	✓	
122	VirtualMetagenome[121]	✓		✓		
123	RAMBL[122]	✓	✓	✓	✓	✓
124	SePLA[123]	✓	✓	✓	✓	✓
125	EMIRGE[124]	✓	✓	✓	✓	✓
126	MATAM[125]	✓	✓	✓	✓	✓
127	REAGO[126]	✓	✓	✓	✓	✓
128	TRFMA[127]			✓		
129	GeneStitch[128]	✓	✓	✓	✓	✓
130	BUSCO[129]		✓	✓	✓	
131	CONCOCT[130]	✓	✓	✓	✓	✓
132	MaxBin[131]	✓	✓	✓	✓	✓
133	MetaCompass[132]		✓	✓	✓	
134	MetaCluster-TA[133]		✓	✓	✓	
135	CAMI[134, 135]		✓	✓	✓	
136	ALE[136]		✓	✓	✓	
137	MegaGTA[137]		✓	✓	✓	
138	MetaBAT[138]	✓	✓	✓	✓	✓
139	AbundanceBin[139]	✓	✓	✓	✓	✓
140	CheckM[140]		✓	✓	✓	
141	MetaQuast[41]		✓	✓	✓	
142	GroopM[142]	✓	✓	✓	✓	✓
143	MetaPhlAn[143]		✓	✓	✓	
144	MEGAN[144]	✓	✓	✓	✓	✓
145	Pyrotools[145]		✓	✓	✓	
146	Taxonomer[146]	✓		✓		
147	Centrifuge[147]	✓	✓	✓	✓	✓
148	mOTU[148]	✓	✓	✓	✓	✓
149	CARMA[149]		✓	✓	✓	
150	Treephyler[150]		✓	✓	✓	
151	SATIVA[151]	✓	✓	✓	✓	✓
152	MGmapper[152]		✓	✓	✓	
153	CSSSCL[153]	✓	✓	✓	✓	✓
154	TaxSOM[154]		✓	✓	✓	
155	Kaiju[155]	✓	✓	✓	✓	✓
156	WebMGA[156]		✓	✓	✓	
157	MetaPhyler[157]		✓	✓	✓	

158	k-SLAM[158]	✓	✓	✓	✓	✓
159	LMAT[159]	✓	✓	✓	✓	✓
160	Binning_refiner[160]		✓	✓	✓	
161	MLTreeMap[161]				✓	
162	TACOA[162]		✓	✓	✓	
163	deSPI[163]		✓	✓	✓	
164	GOTCHA[164]	✓	✓	✓	✓	✓
165	MBMC[165]		✓	✓	✓	
166	LCAStar[166]		✓	✓	✓	
167	PROTAX[167]		✓	✓	✓	
168	DUDes[168]	✓	✓	✓	✓	✓
169	HTTMM[169]		✓	✓	✓	
170	Bracken[170]		✓	✓	✓	
171	OneCodex[171]	✓		✓		
172	MTR[172]		✓	✓	✓	
173	SPANNER[173]		✓	✓	✓	
174	TAEC[174]	✓	✓	✓	✓	✓
175	PhyloSift[175]	✓	✓	✓	✓	✓
176	MetaBinG[176]		✓	✓	✓	
177	TAMER[177]		✓	✓	✓	
178	TAC-ELM[178]		✓	✓	✓	
179	Kraken[179]		✓	✓	✓	
180	LikelyBin[180]		✓	✓	✓	
181	Genometa[181]		✓	✓	✓	
182	MyTaxa[182]		✓	✓	✓	
183	WEVOTE[183]	✓	✓	✓	✓	✓
184	MetaAmp[184]	✓		✓		
185	MIPE[185]		✓	✓	✓	
186	ClaMS[186]		✓	✓	✓	
187	MetaCV[187]		✓	✓	✓	
188	CoMeta[188]		✓	✓	✓	
189	Sequedex[189]		✓	✓	✓	
190	RIEMS[190]		✓	✓	✓	
191	NBC[191]	✓	✓	✓	✓	✓
192	RITA[192]		✓	✓	✓	
193	WGSQuikr[193]		✓	✓	✓	
194	SHERA[194]		✓	✓	✓	
195	PPS[195]	✓	✓	✓	✓	✓
196	d2SBin[196]	✓	✓	✓	✓	✓
197	cuCLARK[197]	✓	✓	✓	✓	✓
198	FROGS[198]	✓	✓	✓	✓	✓
199	TOSS[199]	✓	✓	✓	✓	✓
200	MetaFast[200]	✓	✓	✓	✓	✓
201	COCACOLA[201]		✓	✓	✓	
202	PHYSCIMM[202]		✓	✓	✓	
203	VizBin[203]	✓	✓	✓	✓	✓
204	MetaProb[204]	✓	✓	✓	✓	✓
205	MetaWatt[205]	✓	✓	✓	✓	✓
206	AMBER[206]	✓	✓	✓	✓	✓
207	MrGBP[207]	✓	✓	✓	✓	✓
208	MyCC[208]	✓	✓	✓	✓	✓
209	DectlCOI[209]		✓	✓	✓	
210	MBBC[210]		✓	✓	✓	
211	Simka[211]	✓	✓	✓	✓	✓
212	BiMeta[212]		✓	✓	✓	
213	crAss[213]	✓		✓		
214	CompostBin[214]	✓	✓	✓	✓	✓
215	PhymmBL[215]	✓	✓	✓	✓	✓
216	Glimmer[216]	✓	✓	✓	✓	✓
217	FragGeneScan[217]	✓	✓	✓	✓	✓
218	GeneMark[218]		✓	✓	✓	
219	MICRA[219]	✓	✓	✓	✓	✓

220	MGA[220]		✓	✓	✓	
221	Plastid[221]	✓	✓	✓	✓	✓
222	Orphelia[222]	✓	✓	✓	✓	✓
223	RAP[223]	✓	✓	✓	✓	✓
224	Toplogical-Pressure[224]	✓	✓		✓	
225	Proteingenerator[225]	✓	✓	✓	✓	✓
226	MetaGeneTach[226]		✓	✓	✓	
227	TransGeneScan[227]	✓	✓		✓	
228	MOCAT[228]	✓	✓	✓	✓	✓
229	Prodigal[229]	✓	✓	✓	✓	✓
230	meta_rna[230]	✓	✓	✓	✓	✓
231	Barnap[231]	✓	✓	✓	✓	✓
232	RNAammer[232]			✓		
233	Pplacer[233]	✓	✓	✓	✓	✓
234	HMMER[234]	✓	✓	✓	✓	✓
235	InterProScan[235]	✓	✓	✓	✓	✓
236	biomart[236]	✓	✓	✓	✓	✓
237	GhostKOALA[237]	✓		✓		
238	SALT[238]	✓	✓	✓	✓	✓
239	Diamond-blastx[239]	✓	✓	✓	✓	✓
240	ROCKer[240]	✓	✓	✓	✓	✓
241	UProC[241]	✓	✓	✓	✓	✓
242	Woods[242]		✓	✓	✓	
243	RTMg[243]		✓	✓	✓	
244	MAGpy[244]	✓	✓	✓	✓	✓
245	metaSAMS[245]	✓		✓		
246	MEBS[246]	✓	✓	✓	✓	✓
247	phraSED-ML[247]	✓	✓	✓	✓	✓
248	Kamneva[248]	✓	✓	✓	✓	✓
249	ScrumPY[249]	✓	✓	✓	✓	✓
250	MetaboSignal[250]	✓	✓	✓	✓	✓
251	JMassBalance[251]	✓	✓	✓	✓	✓
252	Pyabolism[252]	✓	✓	✓	✓	✓
253	MappingTool[253]	✓	✓	✓	✓	✓
254	MWASTools[254, 255]	✓	✓	✓	✓	✓
255	Pybrn[256]	✓	✓	✓	✓	✓
256	Metabox[257]	✓	✓	✓	✓	✓
257	zeroSum[258]	✓	✓	✓	✓	✓
258	Papi[259]	✓	✓	✓	✓	✓
259	MetaboliteIDConverter[260]	✓	✓	✓	✓	✓
260	cPath[261]	✓	✓	✓	✓	✓
261	BURRITO[262]	✓		✓		
262	Metacoder[263]	✓	✓	✓	✓	✓
263	Explicit[264]	✓	✓	✓	✓	✓
264	GrammR[265]		✓	✓	✓	
265	Krona[266]	✓	✓	✓	✓	✓
266	VAMPS[267]	✓		✓		
267	GraPhlAn[268]	✓	✓	✓	✓	✓
268	ASAR[269]	✓	✓	✓	✓	✓
269	Metaviz[270]	✓		✓		
270	SynTView[271]		✓	✓	✓	
271	AmphoraVizu[272]	✓		✓		
272	FuncTree[273]	✓		✓		
273	MetaGenSense[274]	✓	✓	✓	✓	✓
274	Elviz[275]	✓		✓		
275	metaWRAP[276]	✓	✓	✓	✓	✓
276	Strainer[277]	✓	✓	✓	✓	✓

**NOTE** *The Authors have relaxed the selection criteria. 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*

## Software Categorization Scheme

Following table shows a categorization scheme which shows the category in which a software lie. This table mainly depicts the need of a package like MetgenoX which handles the setting up of the whole Metagenomics pipeline. The software is placed in the lowest category it is present.

Sr. No.	Software	Pre-processing	Quality Control	Seq. Contamination Det.	Assembly	16S rRNA Reconstruction	Binning	Homology-based Taxonomy	Prediction-based Taxonomy	Gene Prediction	Ribosomal Gene Prediction	Functional Annotation	Metabolic Profiling	Data Visualization
1	MultiQC[1]	✓												
2	HTSeq[2]	✓												
3	MapSplice[4]	✓												
4	NextClip[8]	✓												
5	Eouslan[9]	✓												
6	SOAPnuke[19]	✓												
7	SolexaQA[25]	✓												
8	PRINSEQ[37]	✓												
9	PANDAseq[39]	✓												
10	BWA	✓												
11	ClustalOmega[40]		✓											
12	Greengenes[41]		✓											
13	Nonpareil[42]		✓											
14	FQC[44]		✓											
15	QA[45]		✓											
16	FaQCs[46]		✓											
17	HTQC[48]		✓											
18	SUGAR[49]		✓											
19	KAT[50]		✓											
20	QUASR[52]		✓											
21	ShortRead[56]		✓											
22	fitGCP[60]		✓											
23	PathoQC[66]		✓											
24	FASTQC[72]		✓											
25	ACDC[73]		✓											
26	PhyloOligo[76]			✓										
27	MCSC[78]			✓										
28	Kraken[80]			✓										
29	Conpair[81]			✓										
30	SIDR[84]			✓										
31	MIDAS[85]			✓										
32	Blobology[87]			✓										
33	PhagePhisher[88]			✓										
34	Taxoblast[89]			✓										
35	HYSYS[90]			✓										
36	MetAMOS[92]				✓									
37	dDocent[93]				✓									
38	RayMeta[94]				✓									
39	MetaVelvet[97]				✓									
40	IDBA-UD[98]				✓									
41	MetaSort[99]				✓									
42	Snowball[100]				✓									
43	PRICE[102]				✓									
44	BBAP[104]				✓									

45	BBMAP[105]				✓										
46	MEGAHIT[106]				✓										
47	Faucet[111]				✓										
48	Athena[112]				✓										
49	MetaFinisherSC[113]				✓										
50	Spherical[116]				✓										
51	RAMBL[122]					✓									
52	SePIA[123]					✓									
53	EMIRGE[124]					✓									
54	MATAM[125]					✓									
55	REAGO[126]					✓									
56	GeneStitch[128]					✓									
57	CONCOCT[130]						✓								
58	MaxBin[131]						✓								
59	MetaBAT[138]						✓								
60	AbundanceBin[139]						✓								
61	GroopM[142]						✓								
62	MEGAN[144]							✓							
63	Centrifuge[147]							✓							
64	mOTU[148]							✓							
65	SATIVA[151]							✓							
66	CSSSCL[153]							✓							
67	Kaiju[155]							✓							
68	R-SLAM[158]							✓							
69	LMAT[159]							✓							
70	GOTTCHA[164]							✓							
71	DUDes[168]							✓							
72	TAEC[174]							✓							
73	PhyloSift[175]							✓							
74	WEVOTE[183]							✓							
75	NBC[191]							✓							
76	PPS[195]								✓						
77	d2SBin[196]								✓						
78	cuCLARK[197]								✓						
79	FROGS[198]								✓						
80	TOSS[199]								✓						
81	MetaFast[200]								✓						
82	VizBin[203]								✓						
83	MetaProb[204]									✓					
84	MetaWatt[205]									✓					
85	AMBER[206]									✓					
86	MrGBT[207]									✓					
87	MyCC[208]									✓					
88	Simka[211]									✓					
89	CompostBin[214]									✓					
90	PhymmBL[215]									✓					
91	Glimmer[216]										✓				
92	FragGeneScan[217]										✓				
93	MICRA[219]										✓				
94	Plastid[221]										✓				
95	Orphelia[222]										✓				
96	RAP[223]										✓				
97	Proteingenerator[225]										✓				
98	MOCAT[228]										✓				
99	Prodigal[229]											✓			
100	meta_rna[230]											✓			
101	Barrnap[231]											✓			
102	Pplacer[233]											✓			
103	HMMER[234]											✓			
104	InterProScan[235]												✓		
105	biomart[236]												✓		
106	SALT[238]													✓	

---

I07	Diamond-blastx[239]										✓		
I08	ROCKer[240]										✓		
I09	UProC[241]										✓		
I10	MAGpy[244]										✓		
I11	MEBS[246]										✓		
I12	phraSED-ML[247]										✓		
I13	Kamneva[248]										✓		
I14	ScrumPY[249]										✓		
I15	MetaboSignal[250]										✓		
I16	JMassBalance[251]										✓		
I17	Pyabolism[252]										✓		
I18	MappingTool[253]										✓		
I19	MWASTools[254, 255]										✓		
I20	Pybrn[256]										✓		
I21	Metabox[257]										✓		
I22	zeroSum[258]										✓		
I23	Papi[259]										✓		
I24	MetaboliteIDConverter[260]										✓		
I25	cPath[261]										✓		
I26	Metacoder[263]											✓	
I27	Explicit[264]											✓	
I28	Krona[266]											✓	
I29	GraPhlAn[268]											✓	
I30	ASAR[269]											✓	
I31	MetaGenSense[274]											✓	
I32	metaWRAP[276]											✓	
I33	Strainer[277]												✓

## MetgenoX Packaged Tools

List of software tools and plugins included in MetgenoX package are shown in the following table. Software which are recommended are also mentioned in the table. Download size, version and Latest update of each software is mentioned for user convenience.

Sr. No.	Software Name	Size (MB)	Version	Latest Update
<b>Pre-processing</b>				
1	MultiQC[1]	4.0	1.1.0	2021
2	HTSeq[2]	12.0	0.13.5	2020
3	MapSplice[4]	40.0	2.2.1	2020
4	NextClip[8]	10.0	1.3.1	2013
5	Eoulsan[9]	15.0	2.5.0	2021
6	SOAPnuke[19]	0.1	2.1.6	2021
7	SolexaQA[25]	10.0	3.1.7	2021
8	PRINSEQ[37]	0.5	0.20.4	2013
9	PANDAseq[39]	5.0	2.11.0	2016
10	ClustalOmega[40]	7.0	1.2.4	2016
11	BWA	1.0	0.7.17	2017
<b>Quality Control</b>				
12	Greengenes[41]	45.0	1.0.0	2020
13	Nonpareil[42]	57.0	3.3.4	2020
14	FQC[44]	10.0	1.5.7	2017
15	QA[45]	5.0	1.0.0	2020
16	FaQCs[46]	10.0	2.10.0	2020
17	HTQC[48]	1.0	1.92.3	2016
18	SUGAR[49]	93.0	1.0.0	2020
19	KAT[50]	40.0	2.4.1	2017
20	QUASR[52]	2.0	1.32.0	2020
21	ShortRead[56]	40.0	1.50.1	2020
22	fitGCP[60]	29.0	1.0.0	2015
23	PathoQC[66]	24.0	1.0.0	2020

24	FASTQC[72]	37.0	0.11.9	2019
25	ACDC[73]	17.0	1.0.0	2019
<b>Sequence Contamination Detection</b>				
26	PhyloOligo[76]	26.0	0.1.0	2019
27	MCSC[78]	1600.0	1.0.0	2020
28	Kraken[80]	1.5	2.0.0	2019
29	Conpair[81]	13.0	0.2.0	2019
30	SIDR[84]	0.1	0.0.2	2018
31	MIDAS[85]	26.0	1.3.2	2017
32	Blobology[87]	1.0	1.0.0	2014
33	PhagePhisher[88]	0.1	1.2.0	2015
34	Taxoblast[89]	30.0	1.0.0	2020
35	HYSYS[90]	0.1	1.0.0	2021
<b>Assembly</b>				
36	MetAMOS[92]	402.0	1.5.0	2014
37	dDocent[93]	0.4	2.8.13	2021
38	RayMeta[94]	0.4	1.0.0	2021
39	MetaVelvet[97]	0.2	1.2.01	2012
40	IDBA-UD[98]	2	1.1.6	2016
41	MetaSort[99]	33.0	1.2.0	2016
42	Snowball[100]	1.0	1.2.0	2017
43	PRICE[102]	9.0	1.2.0	2014
44	BBAP[104]	1.4	1.0.0	2020
45	BBMAP[105]	26.0	38.94.0	2021
46	MEGAHIT[106]	0.5	1.2.9	2019
47	Faucet[111]	3.0	0.5.0	2017
48	Athena[112]	0.1	1.1.0	2016
49	MetaFinisherSC[113]	36.0	5.1.0	2015
50	Spherical[116]	34.0	1.0.0	2018
<b>16S rRNA Reconstruction</b>				
51	RAMBL[122]	0.4	1.0.0	2017

52	SePIA[123]	20.0	1.0.0	2020
53	EMIRGE[124]	2.0	0.61.0	2016
54	MATAM[125]	6.0	1.6.0	2019
55	REAGO[126]	0.3	1.1.0	2015
56	GeneStitch[128]	5.4	1.2.1	2014
<b>Binning</b>				
57	CONCOCT[130]	7.0	1.1.0	2019
58	MaxBin[131]	4.0	2.2.7	2020
59	MetaBAT[138]	13.0	2.12.1	2017
60	AbundanceBin[139]	7.0	1.01.0	2012
61	GroopM[142]	30.0	1.0.0	2016
<b>Taxonomic Classification</b>				
Homology-based Taxonomy				
62	MEGAN[144]	102.0	6.21.6	2020
63	Centrifuge[147]	0.8	1.0.4	2018
64	MOTU[148]	26.0	3.0.1	2021
65	SATIVA[151]	0.6	1.0.0	2020
66	CSSSCL[153]	0.1	1.0.0	2015
67	Kaiju[155]	0.6	1.8.2	2021
68	k-SLAM[158]	0.1	1.0.0	2018
69	LMAT[159]	2.0	1.2.4	2018
70	GOTTCHA[164]	13.0	1.0.0	2017
71	DUDes[168]	35.0	0.08.0	2017
72	TAEC[174]	7.0	3.0.2	2015
73	PhyloSift[175]	33.0	1.0.0	2015
74	WEVOTE[183]	5.0	1.0.0	2018
75	NBC[191]	2.0	1.0.0	2016
Prediction-based Taxonomy				
76	PPS[195]	12.0	1.0.0	2020
77	d2SBin[196]	23.0	1.0.0	2018
78	cuCLARK[197]	0.1	1.0.0	2016

79	FROGS[198]	77.0	1.0.0	2020
80	TOSS[199]	0.1	1.0.0	2016
81	MetaFast[200]	25.0	1.0.0	2021
82	VizBin[203]	103.0	1.0.0	2020
83	MetaProb[204]	9.0	2.0.0	2018
84	MetaWatt[205]	12.0	3.5.3	2016
85	AMBER[206]	0.5	2.0.3	2021
86	MrGBP[207]	0.6	1.0.0	2020
87	MyCC[208]	121.0	1.0.0	2017
88	Simka[211]	0.3	1.5.3	2021
89	CompostBin[214]	158.0	0.0.1	2020
90	PhymmBL[215]	5.0	4.0.0	2014
<b>Gene Prediction</b>				
91	Glimmer[216]	45.0	3.0.2	2021
92	FragGeneScan[217]	36.0	1.31.0	2018
93	MICRA[219]	30.0	1.0.0	2017
94	Plastid[221]	4.0	0.5.1	2020
95	Orphelia[222]	68.0	1.0.0	2021
96	RAP[223]	1.9	1.0.0	2020
97	Proteingenerator[225]	32.0	1.0.0	2020
98	MOCAT[228]	400.0	2.0.0	2021
99	Prodigal[229]	26.0	1.20.0	2021
<b>Ribosomal Gene Prediction</b>				
100	meta_rna[230]	6.0	1.0.0	2020
101	Barrnap[231]	12.0	0.9.0	2018
102	Pplacer[233]	36.0	1.1.0	2021
103	HMMER[234]	18.6	3.3.2	2021
<b>Functional Annotation</b>				
104	InterProScan[235]	16000	5.52.0	2020
105	biomartr[236]	0.3	0.9.1	2019
106	SALT[238]	0.1	1.0.0	2020

107	Diamond-blastx[ <a href="#">239</a> ]	46.0	2.0.13	2021
108	ROCKER[ <a href="#">240</a> ]	0.1	1.0.0	2020
109	UProC[ <a href="#">241</a> ]	0.2	2.0.0	2015
110	MAGpy[ <a href="#">244</a> ]	260.0	1.0.0	2021
<b>Metabolic Profiling</b>				
111	MEBS[ <a href="#">246</a> ]	73.2	1.0.0	2017
112	phraSED-ML[?]	1.8	1.0.3	2018
113	Kamneva 2016[ <a href="#">248</a> ]	101.8	1.0.0	2016
114	ScrumPy[ <a href="#">249</a> ]	1093.9	1.0.0	2018
115	MetaboSignal[ <a href="#">250</a> ]	169.1	1.14.0	2019
116	JMassBalance[ <a href="#">251</a> ]	3.7	1.0.0	2013
117	Pyabolism[?]	0.4	1.0.0	2017
118	MappingTool[ <a href="#">253</a> ]	1.9	1.0.0	2020
119	MWASTools[ <a href="#">254</a> , <a href="#">255</a> ]	1.9	1.0.0	2020
120	Pybrn[ <a href="#">256</a> ]	0.6	0.4.3	2016
121	Metabox[ <a href="#">257</a> ]	53.0	1.2.0	2016
122	zeroSum[ <a href="#">258</a> ]	1.0	2.0.0	2019
123	PAPi[?]	0.6	1.24.0	2019
124	M.IDConvertor[?]	0.2	1.0.0	2010
125	cPath[ <a href="#">261</a> ]	1.9	1.0.0	2020
<b>Data Visualization</b>				
126	Metacoder[ <a href="#">263</a> ]	2.0	0.3.5	2021
127	Explicit[ <a href="#">264</a> ]	13.0	2.10.5	2021
128	Krona[ <a href="#">266</a> ]	1.2	1.0.0	2020
129	GraPhlAn[ <a href="#">268</a> ]	4.0	1.1.3	2021
130	ASAR[ <a href="#">269</a> ]	31.0	1.0.0	2018
131	MetaGenSense[ <a href="#">274</a> ]	4.0	1.0.0	2016
132	metaWRAP[ <a href="#">276</a> ]	2.0	1.3.0	2021
133	Strainer[ <a href="#">277</a> ]	4.0	1.0.0	2020

# Software Installation Process

## Installing MetgenoX software

For downloading of MetgenoX, visit its website:

<https://github.com/hasaniqbal777/MetgenoX-bin>

1. MetgenoX can also be downloaded from the following command through git:

```
$ git clone https://github.com/hasaniqbal777/MetgenoX-bin
```

2. Now run the following command on terminal:

```
$ chmod +x MetgenoX_setup_enUS  
$ sudo ./MetgenoX_setup_enUS
```

3. 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).

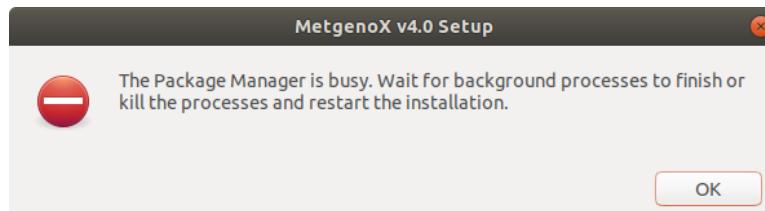


Figure 1: Password Screen

4. Enter your superuser credentials to start MetgenoX Installation.



Figure 2: Password Screen

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

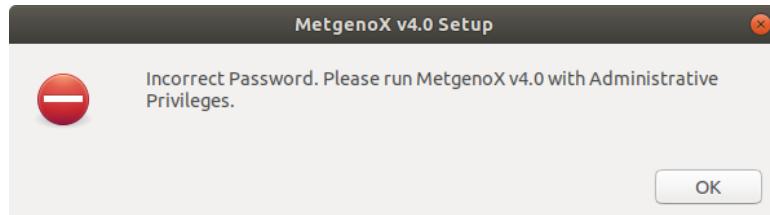


Figure 3: Password Screen

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



Figure 4: Internet Connection Successful Screen

**NOTE** Installation of MetgenoX require a proper internet connection to proceed, otherwise the installation terminates.

7. Installation is **terminated** if there is no internet connection.



Figure 5: Internet Connection failed Screen

8. Installation wizard of MetgenoX will start.



Figure 6: Splash Screen

9. Select you want to install, uninstall or exit MetgenoX.

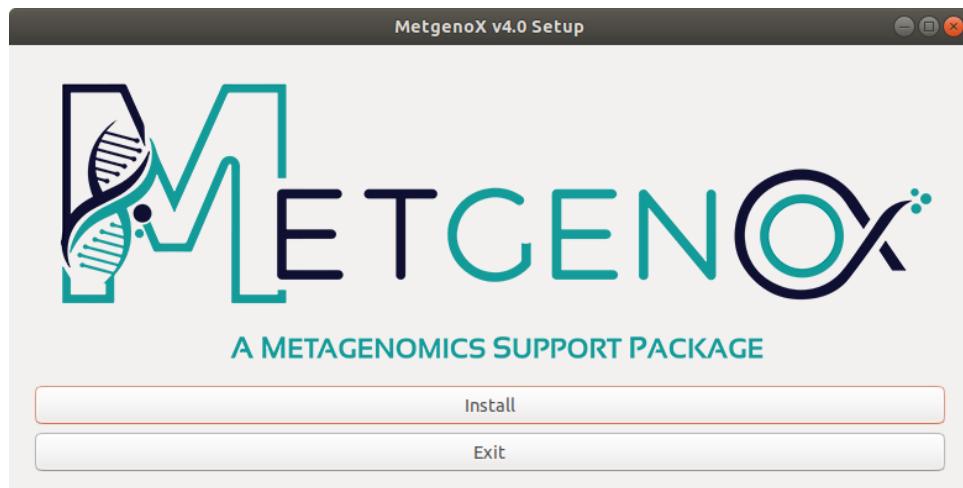


Figure 7: Installer Selection Screen

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

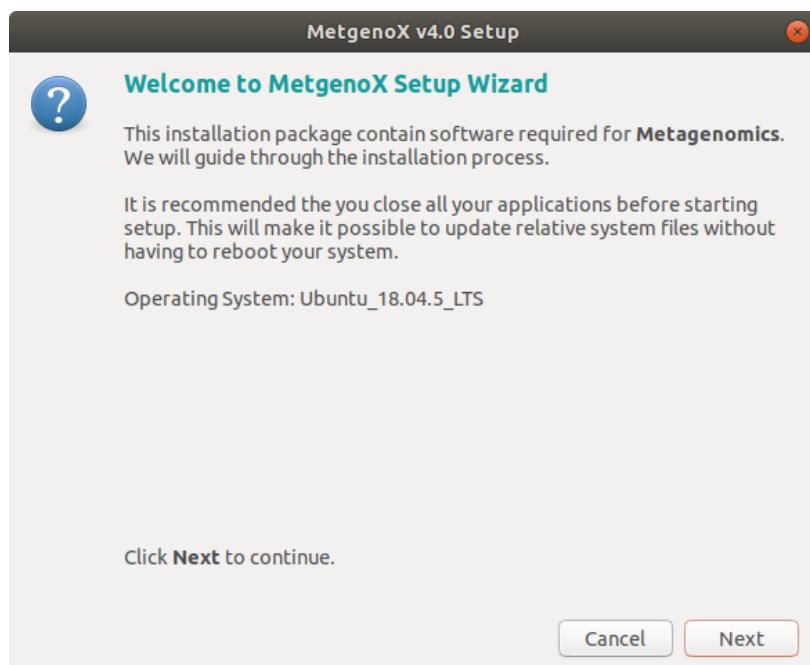


Figure 8: Installation Welcome Screen

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

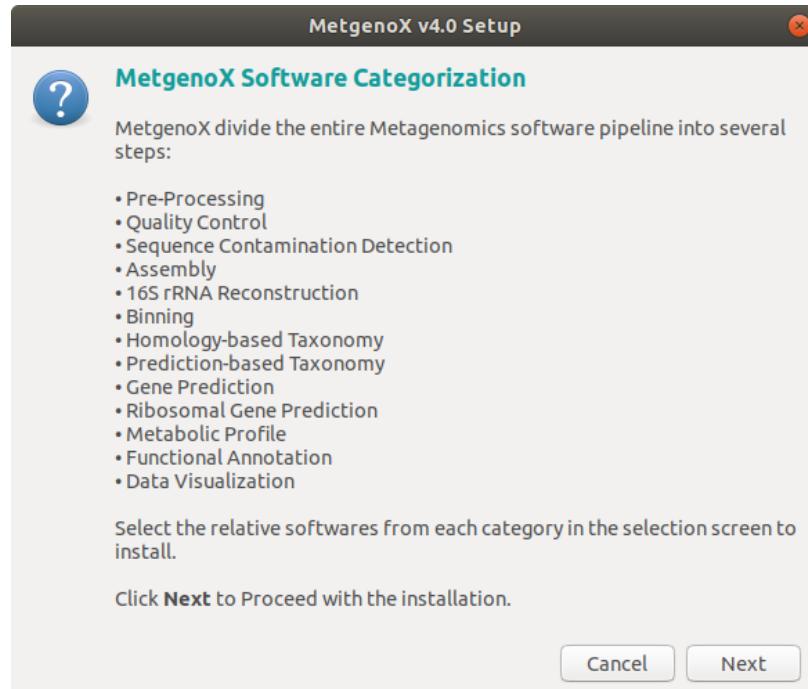


Figure 9: Pipeline information screen

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

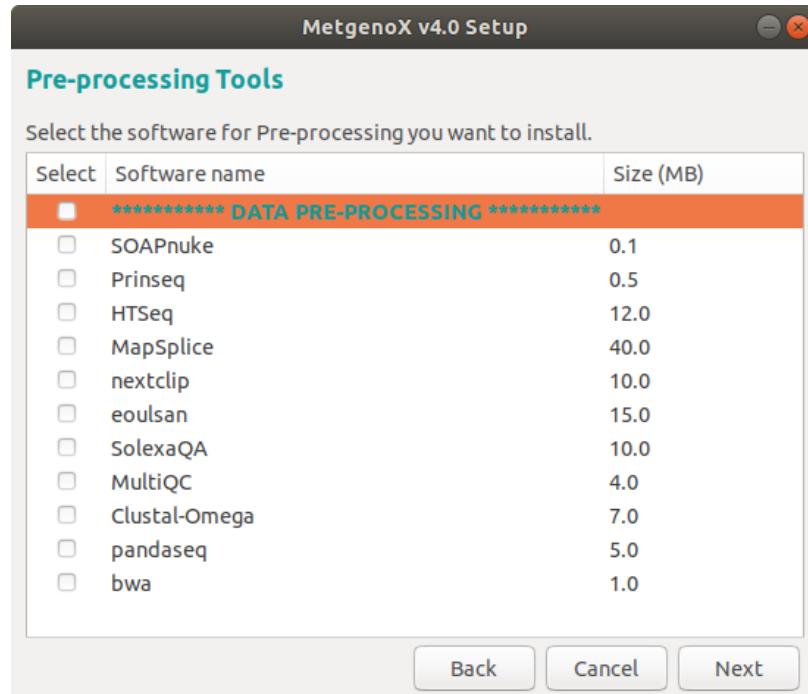


Figure 10: Pre-processing tools selection screen

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

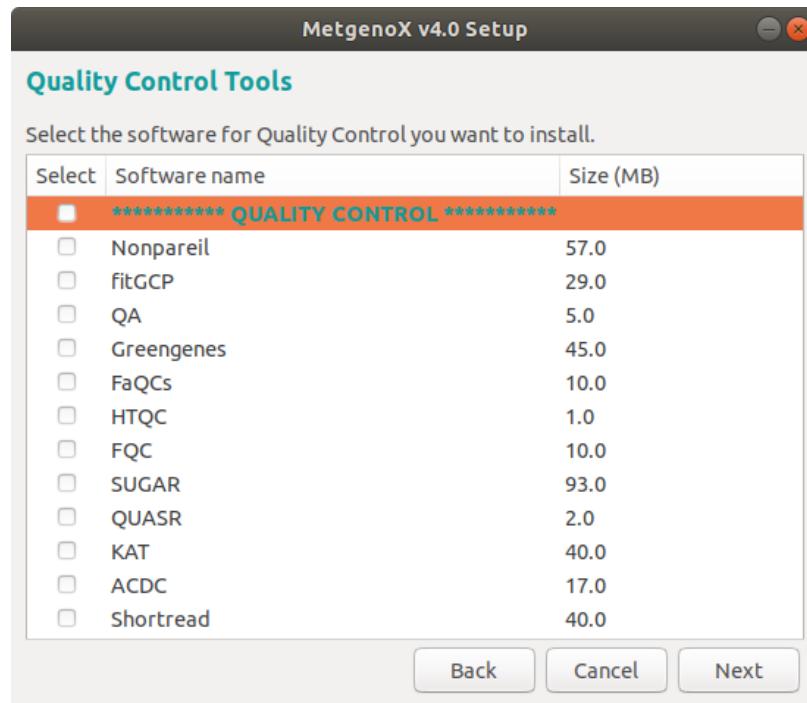


Figure 11: Quality Control tools selection screen

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

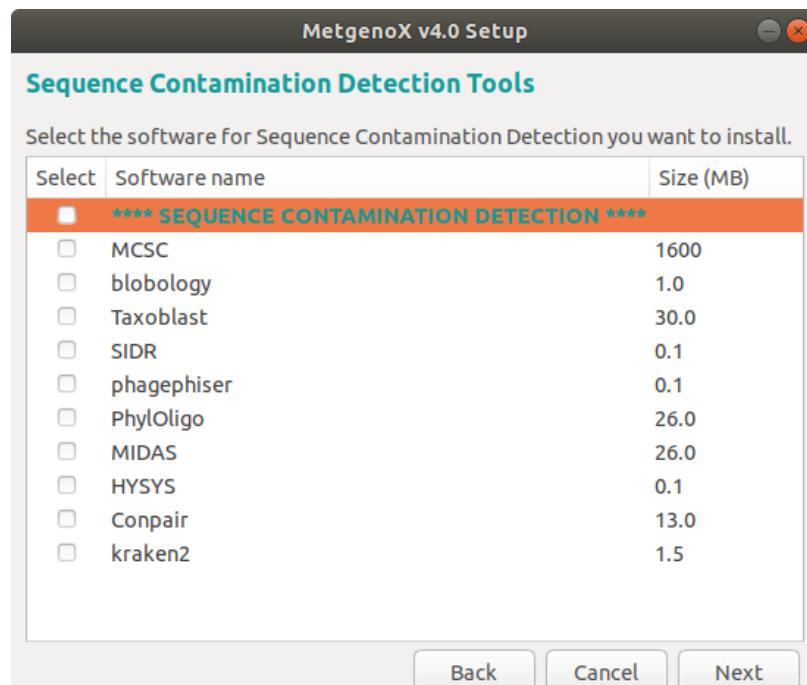


Figure 12: Sequence Contamination Detection tools selection screen

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

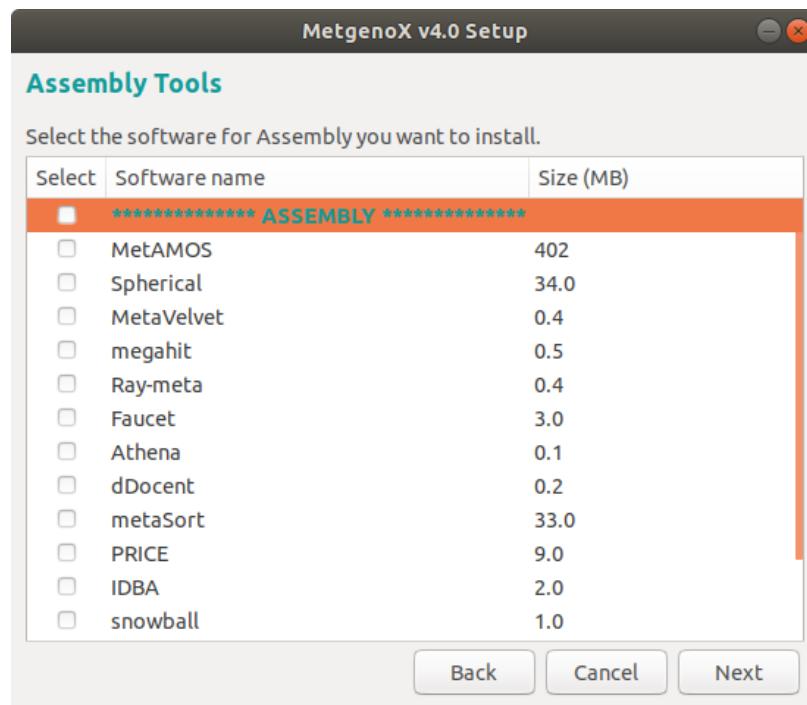


Figure 13: Assembly tools selection screen

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

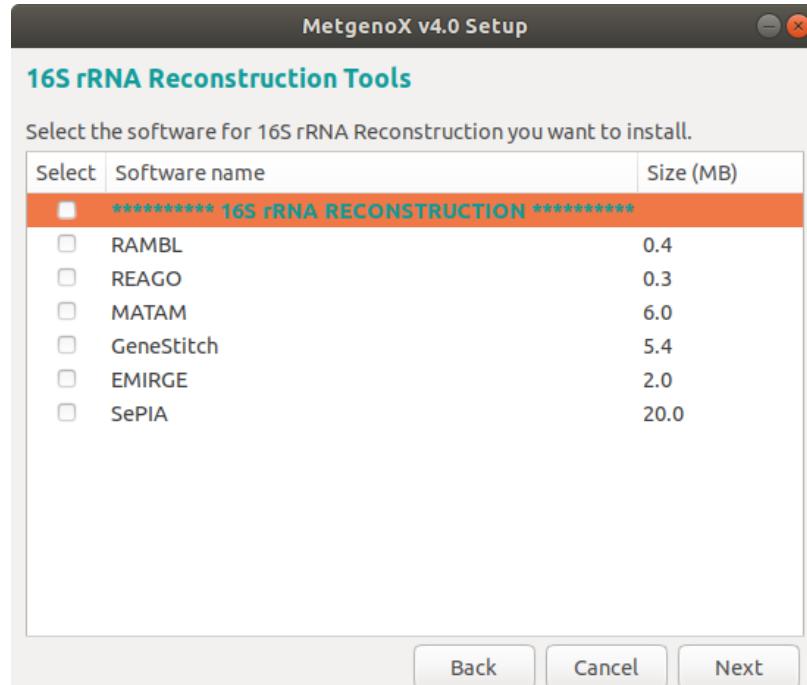


Figure 14: 16S rRNA Reconstruction tools selection screen

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

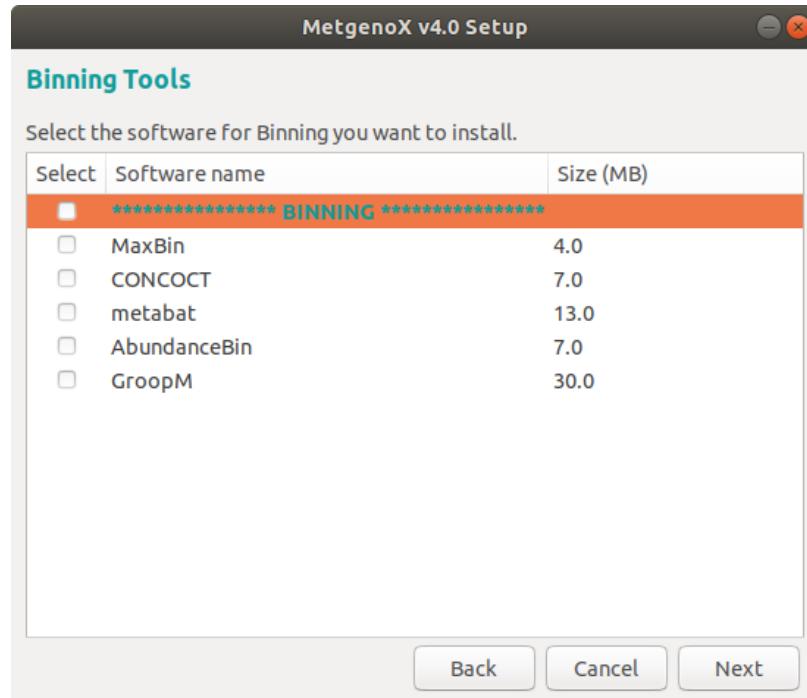


Figure 15: Binning tools selection screen

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

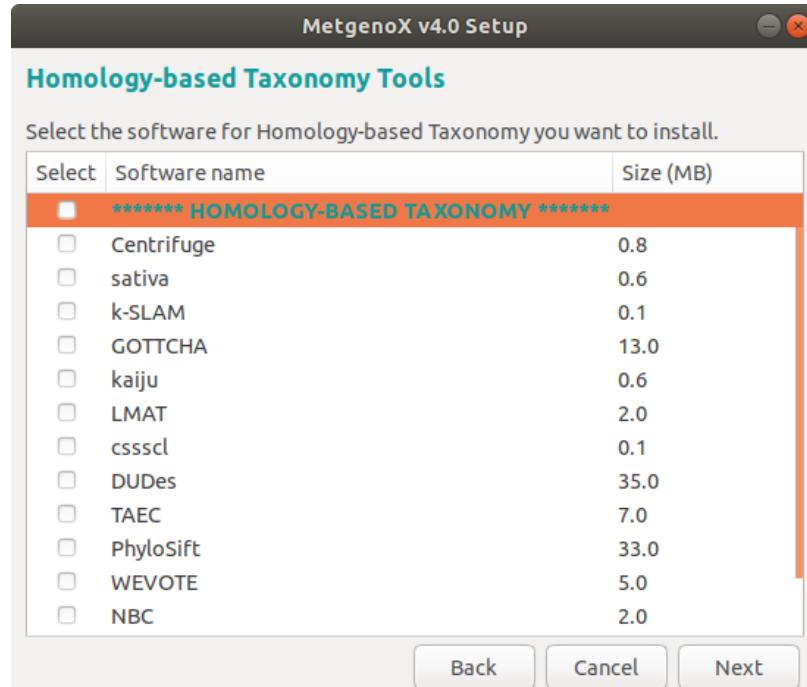


Figure 16: Homology-based Taxonomy tools selection screen

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

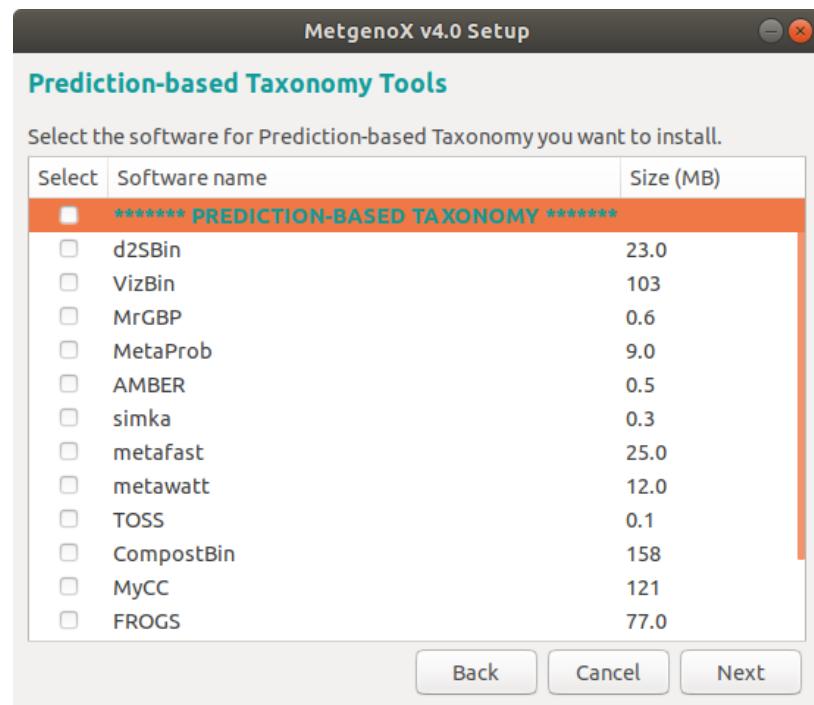


Figure 17: Prediction-based Taxonomy tools selection screen

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

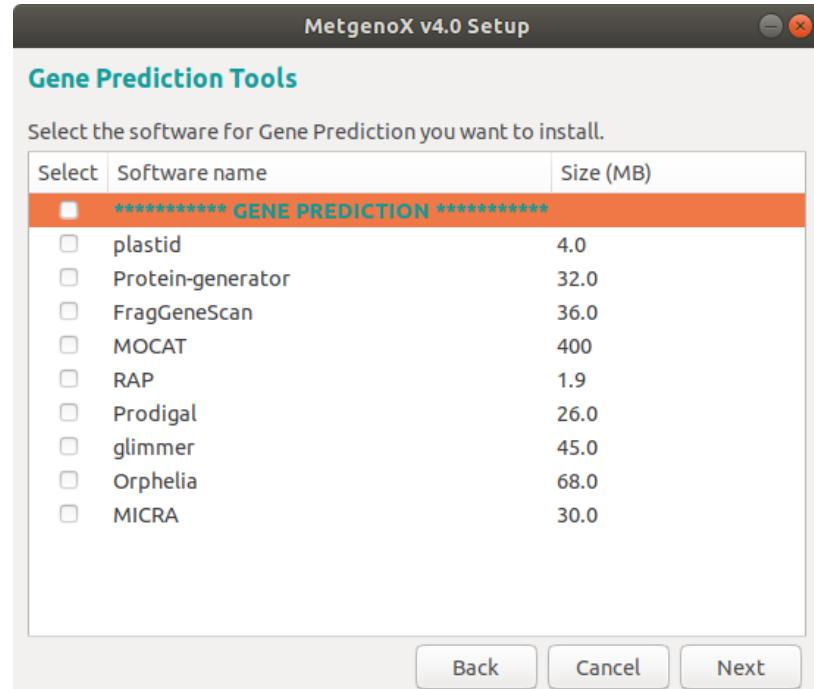


Figure 18: Gene Prediction tools selection screen

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

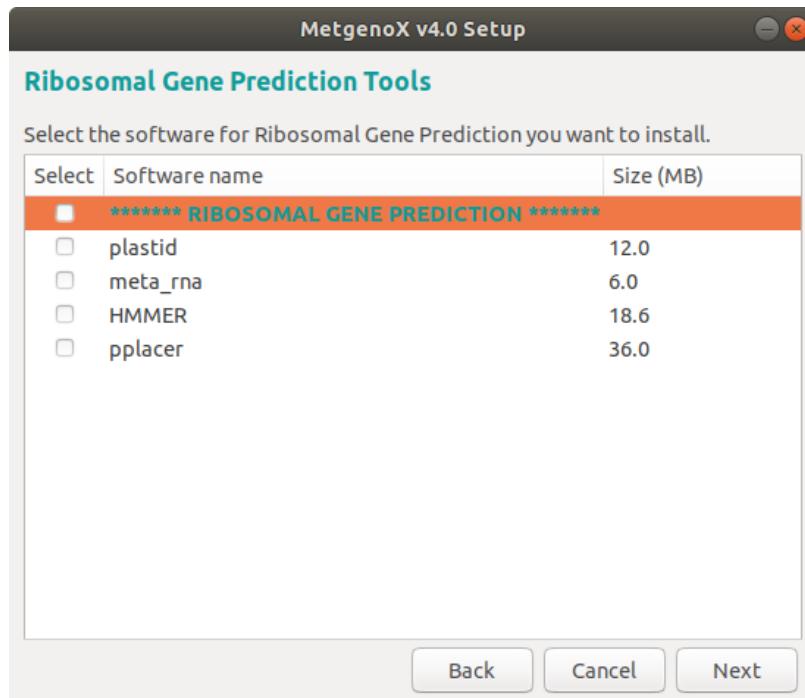


Figure 19: Ribosomal Gene Prediction tools selection screen

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

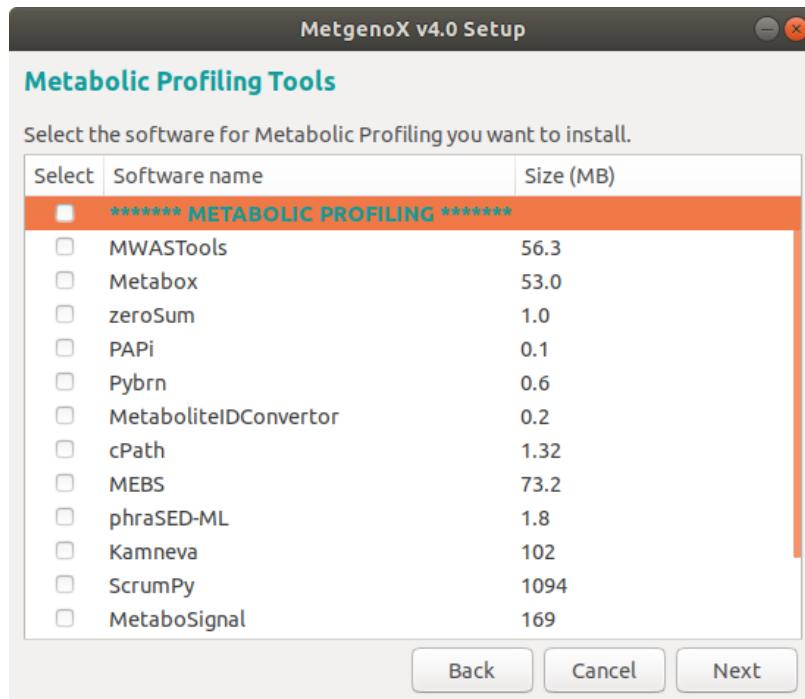


Figure 20: Metabolic Profiling tools selection screen

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

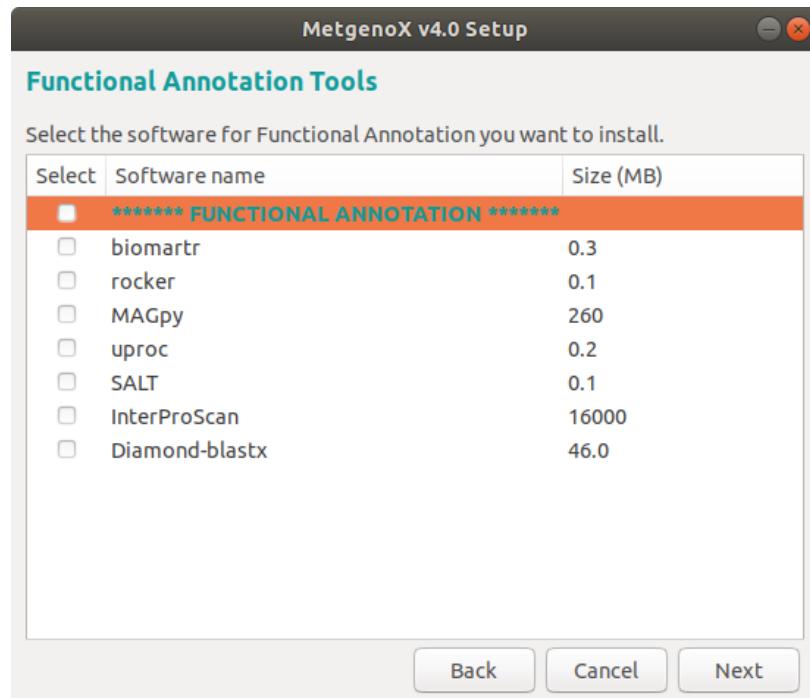


Figure 21: Functional Annotation tools selection screen

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

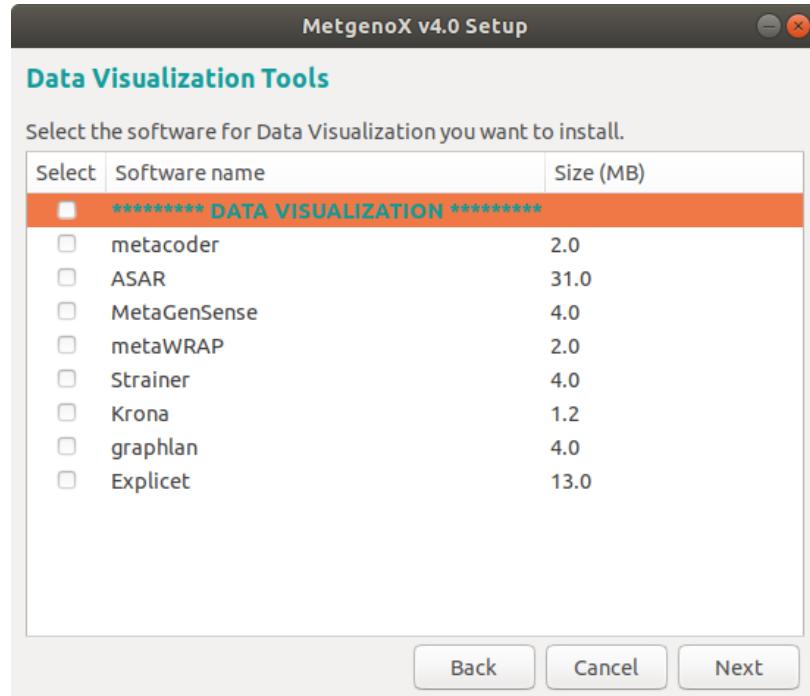
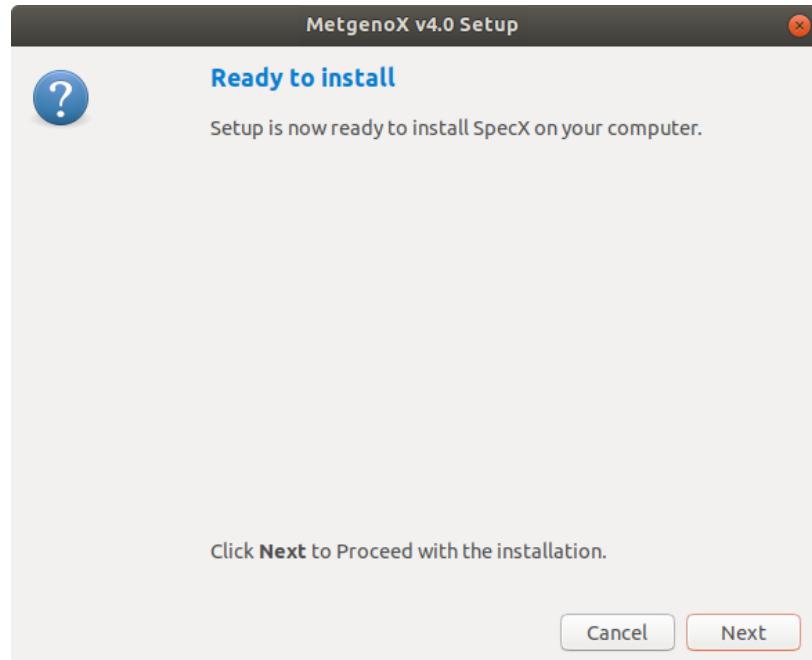


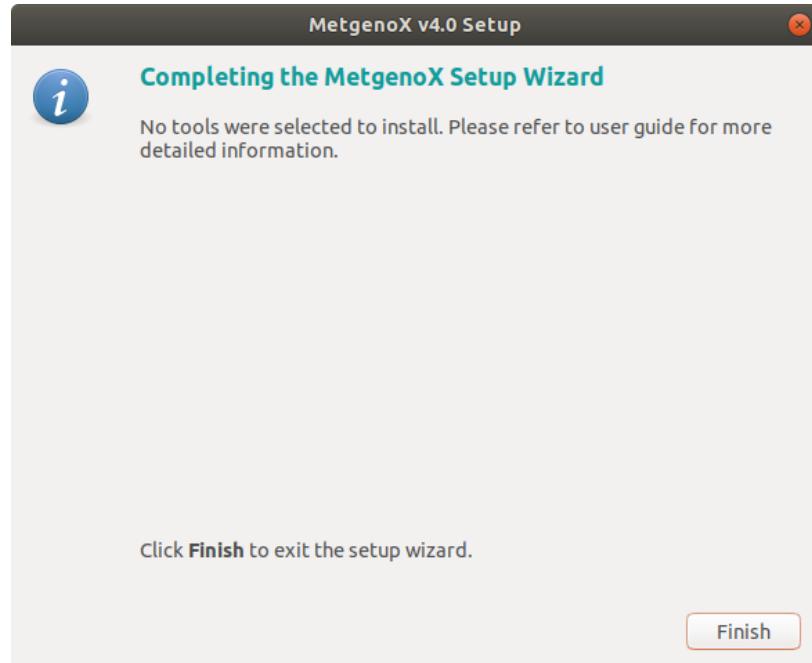
Figure 22: Data Visualization tools selection screen

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



*Figure 23: Ready to Install Screen*

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



*Figure 24: Ready to Install Screen*

27. **Dependencies** related to software will start installing.

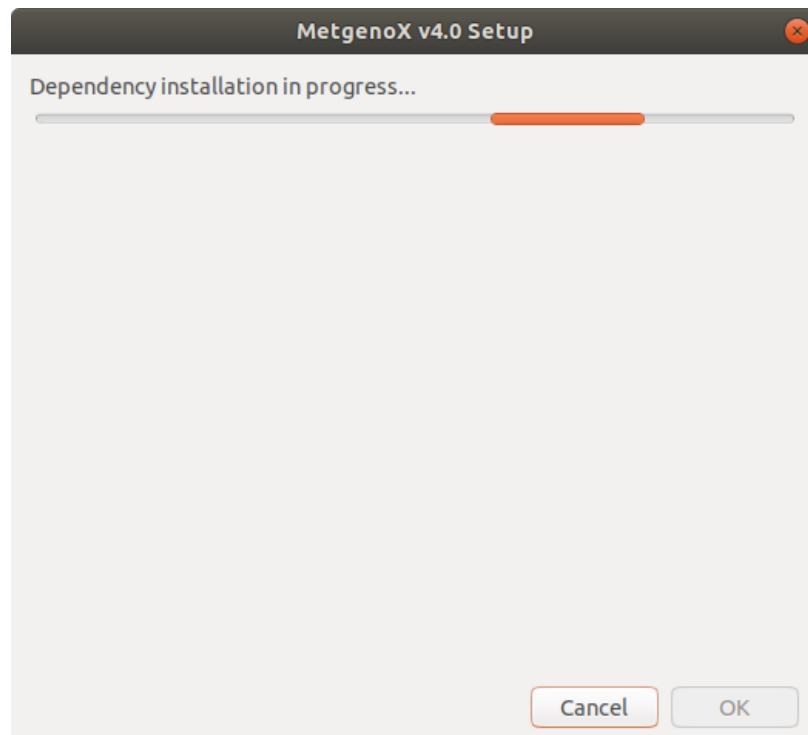


Figure 25: Installation Screen

28. All the dependencies which are installed are displayed at the end of the installation. Click **Ok**.

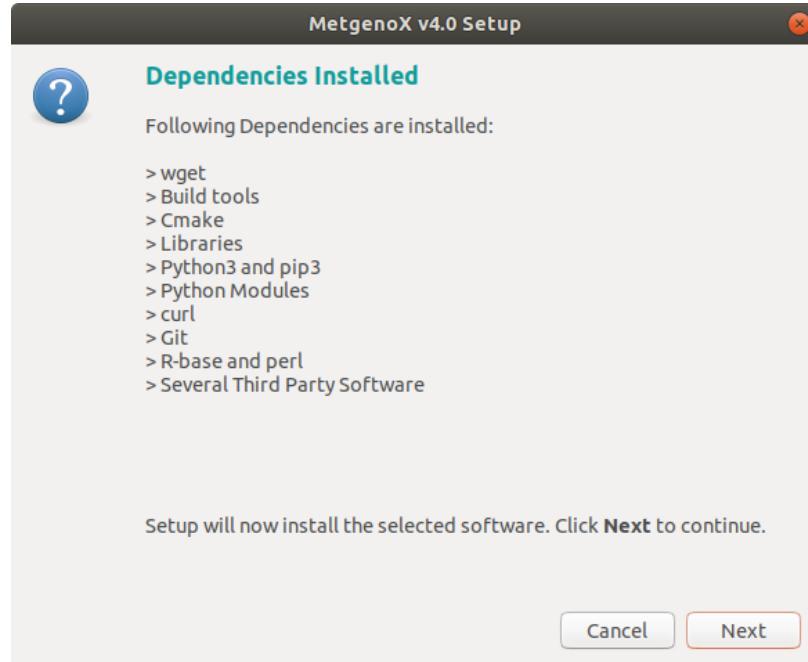


Figure 26: Dependencies installation finished screen

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

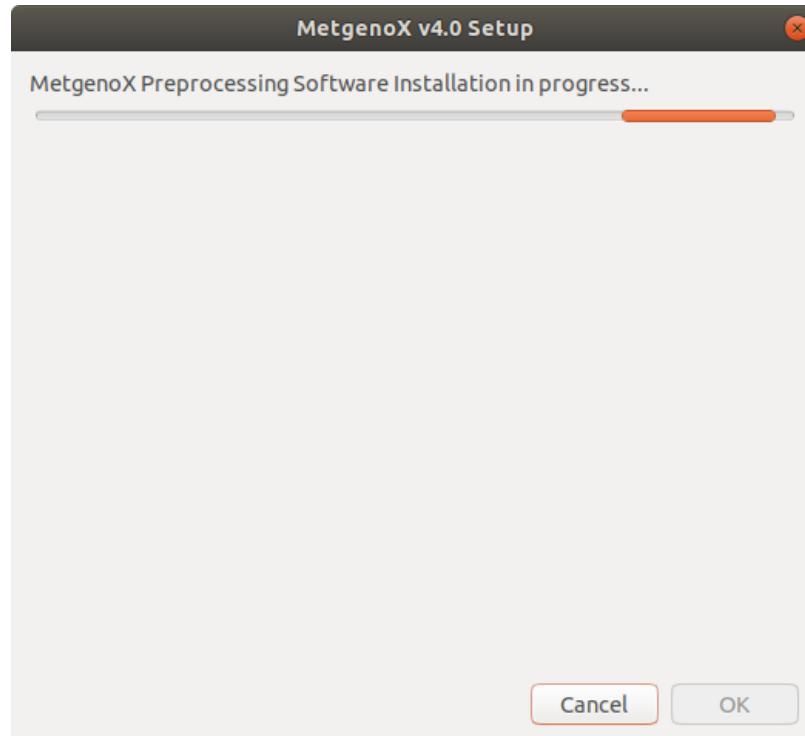


Figure 27: Downloading Screen

30. Installation of MetgenoX is now finished. Click **Finish** to use the tools.

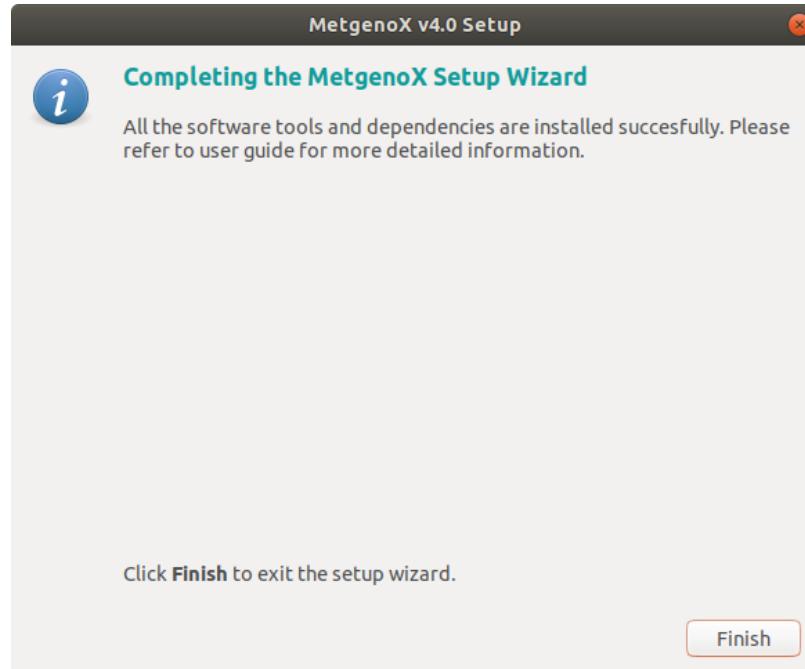


Figure 28: Finishing Installation Screen



# Microsoft Windows Support

## Installing MetgenoX software on Microsoft Windows

MetgenoX is software package for Linux, specifically Ubuntu users, but Windows users can use MetgenoX using a Virtual Machine. Follow these steps to initialize a Virtual Machine on your Windows host.

**NOTE** *MetgenoX support a 64bit Windows host. Microsoft Windows XP and Vista support is discontinued.*

1. First you have to **download** the image file of the required OS (Ubuntu) from its website or use the following link:

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

2. **Download** the virtualization software (Oracle VM VirtualBox) from the following link:

<http://download.virtualbox.org/virtualbox/6.0.10/VirtualBox-6.0.10-132072-Win.exe>

3. **Install** this software in Windows OS host.

4. **Start** Virtual Box, and click on the **New** symbol.

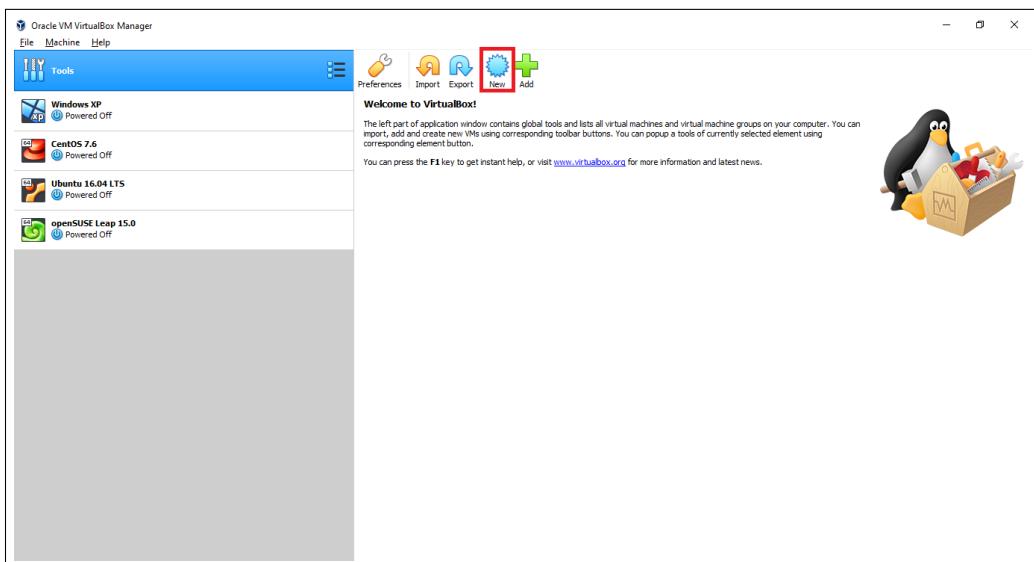


Figure 1: New Virtual OS

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

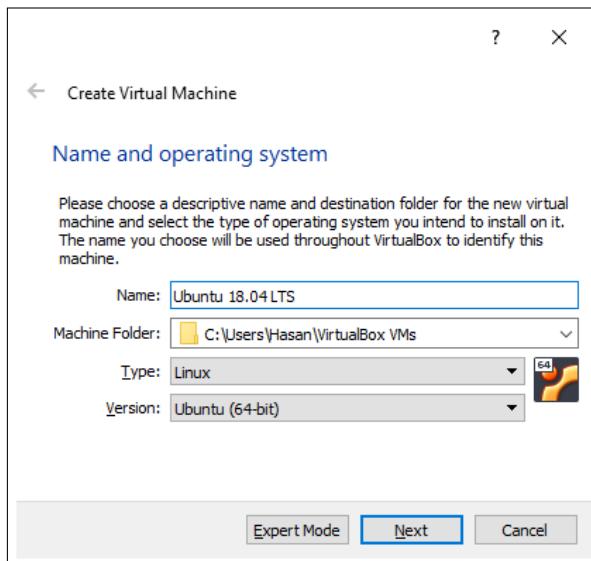


Figure 2: Assigning Name

6. Allocate RAM to the virtual OS. Following system has 8GB of RAM so 2GB of RAM is allocated. You can use more RAM if your system has enough extra RAM.

**NOTE** Allocate about half of the RAM to the virtual OS. Click Next.

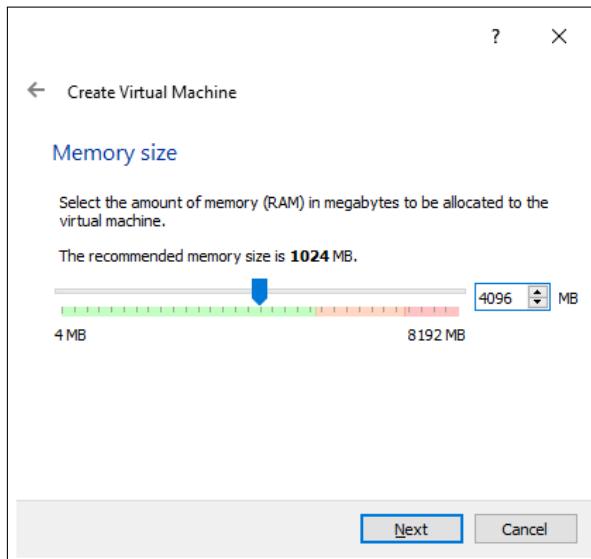


Figure 3: Memory Allocation

7. **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**.

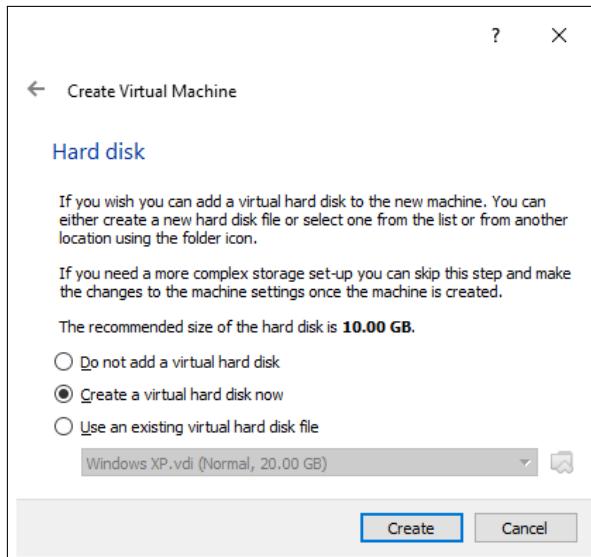


Figure 4: Creating a virtual hard disk

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

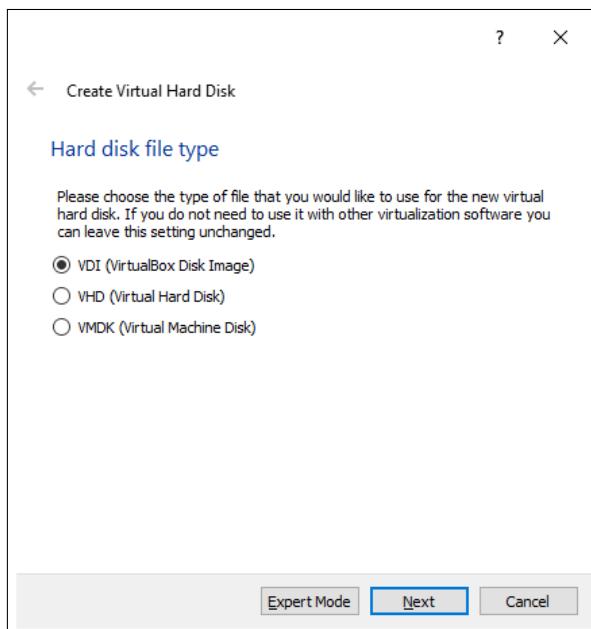


Figure 5: Hard disk file type

9. 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.*

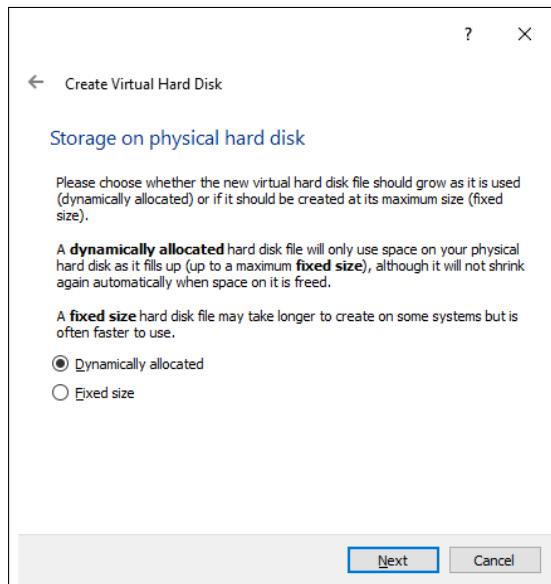


Figure 6: Storage type

10. Select **VDI file type** here (recommended) and Select the **Hard Disk size**. (recommended size: 100 GB). Click **Create**.

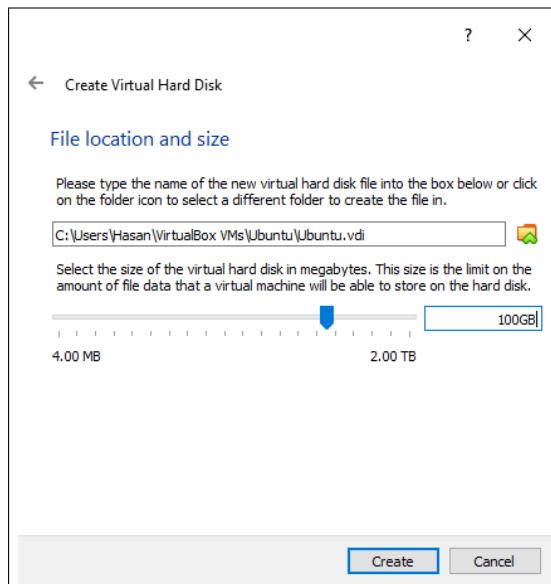


Figure 7: Hard disk size

11. Click **Next**. Now, Select **Settings** to assign the image file of respective OS to VB.

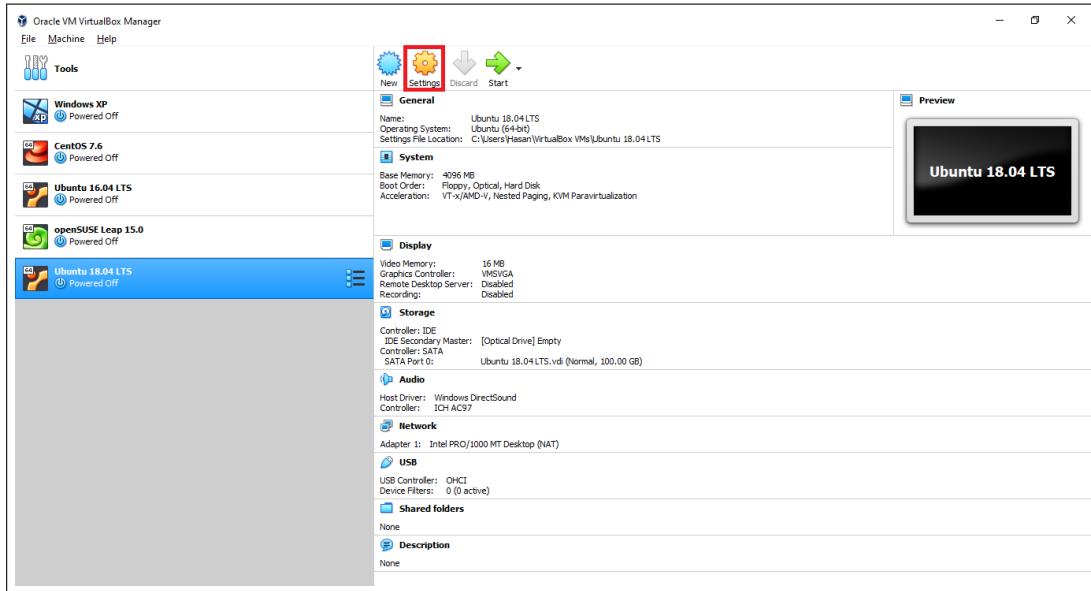


Figure 8: Select Settings

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

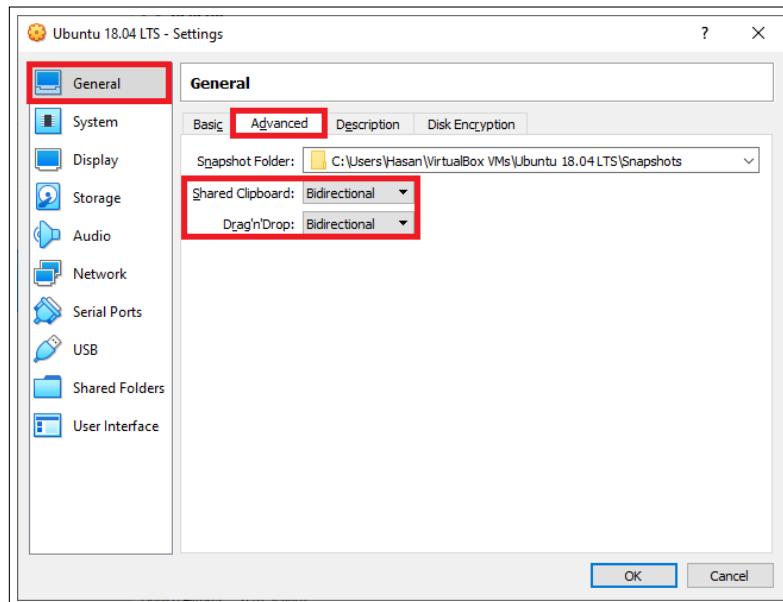


Figure 9: Advanced Settings

13. 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**.

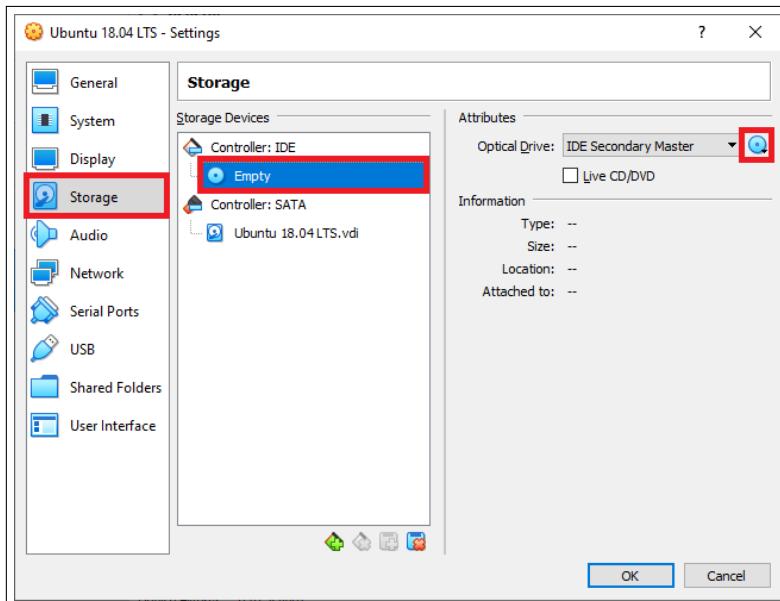


Figure 10: Providing Image file

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

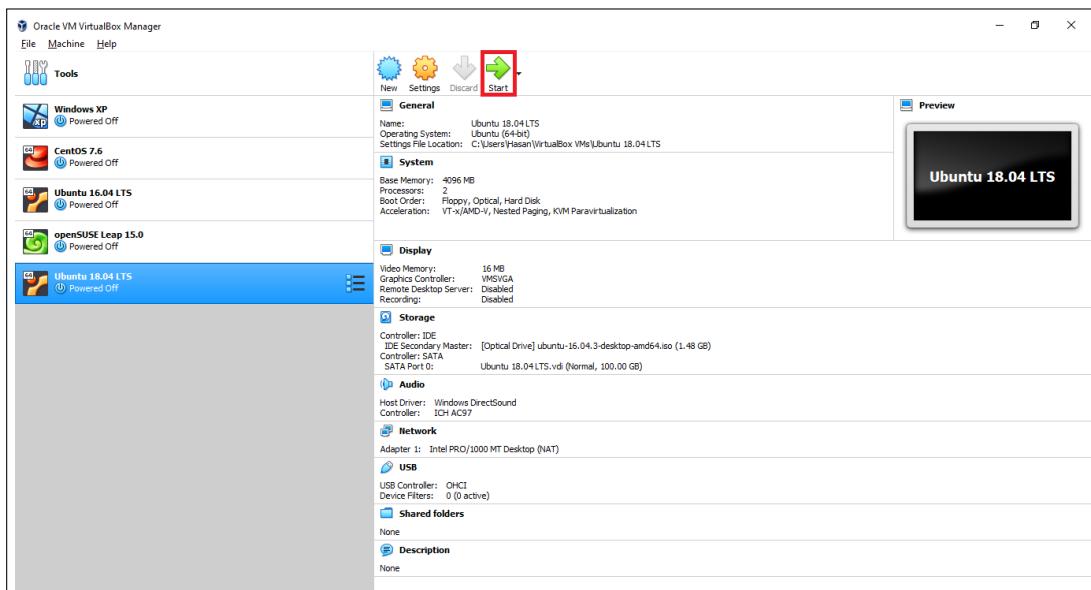


Figure 11: Starting Virtual OS

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

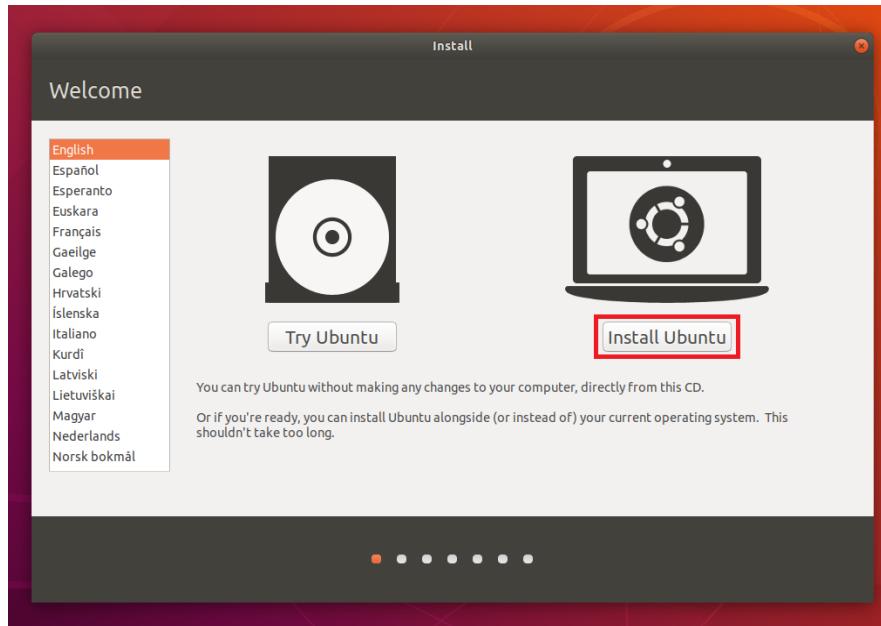


Figure 12: Installing Ubuntu

16. Continue with Normal Installation.

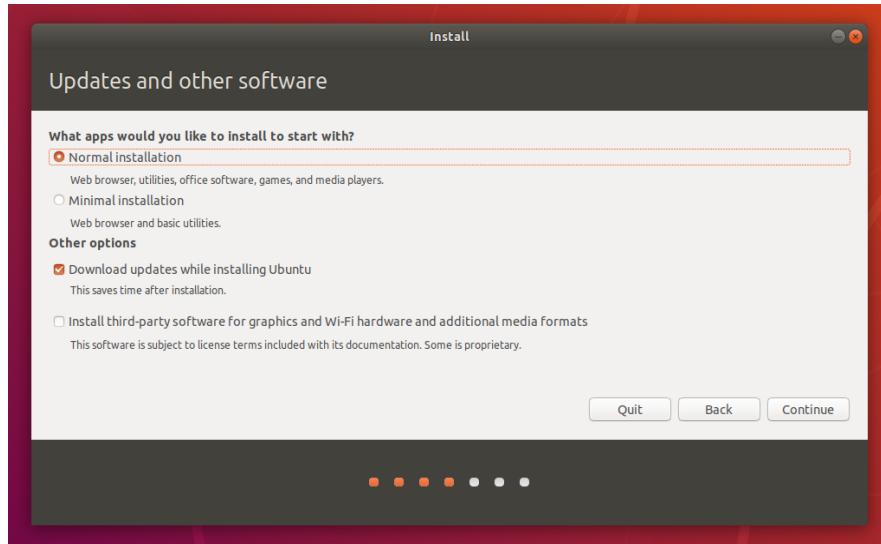


Figure 13: Update Screen

17. In Installation type screen, select **Erase disk** and **Install Now** option.

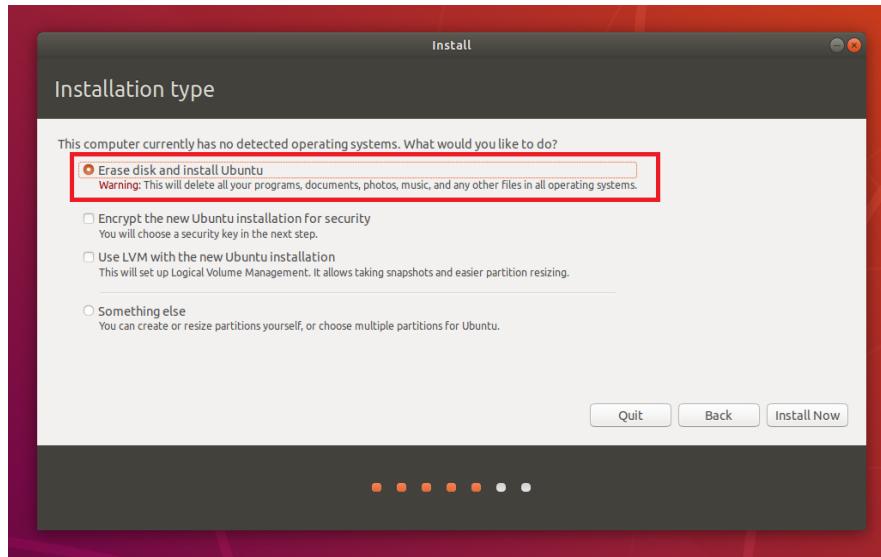


Figure 14: Installation type Screen

18. Select Continue.

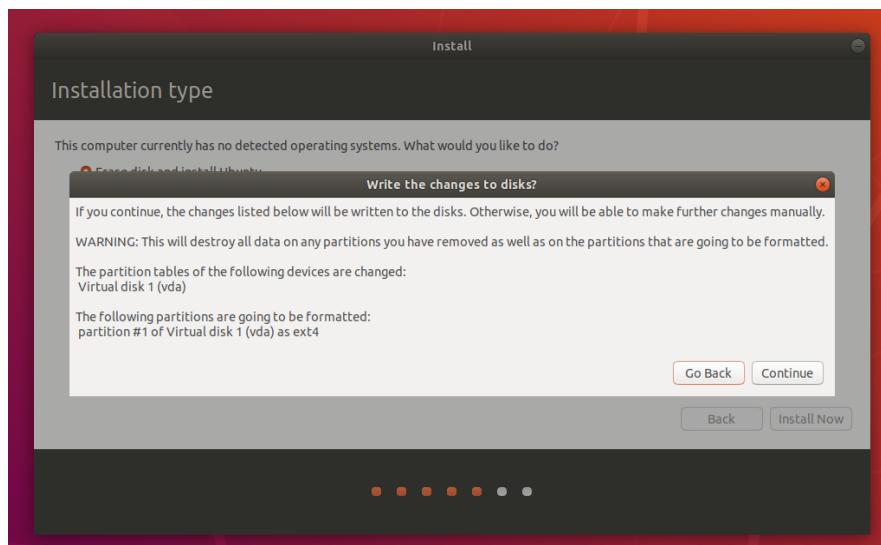


Figure 15: Confirmation Screen

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

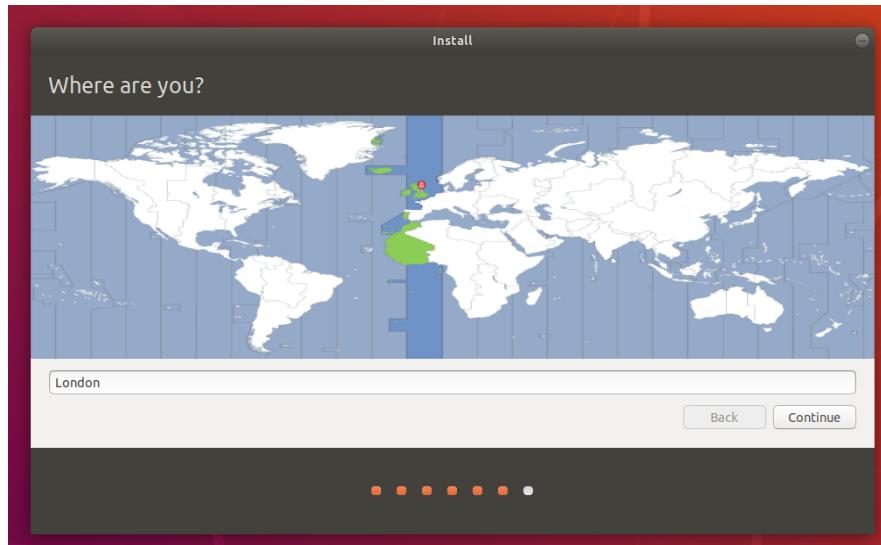


Figure 16: Location selection screen

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

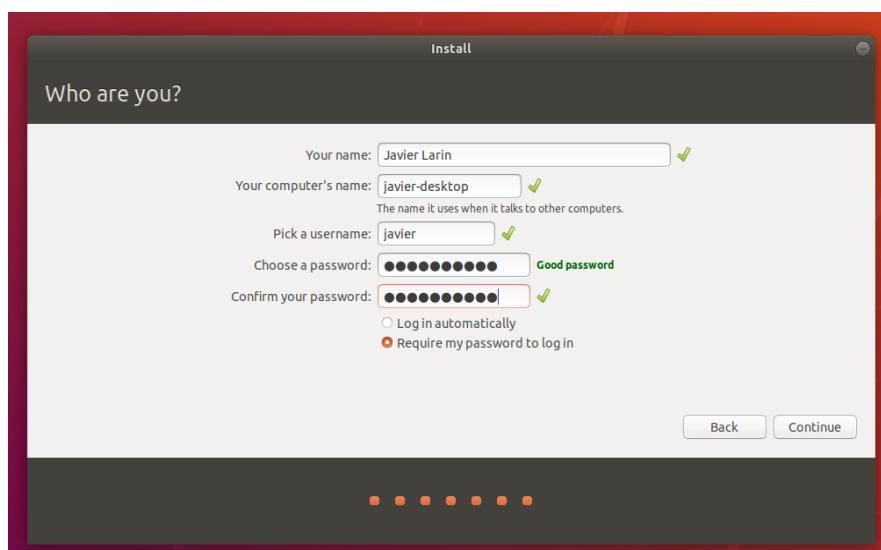
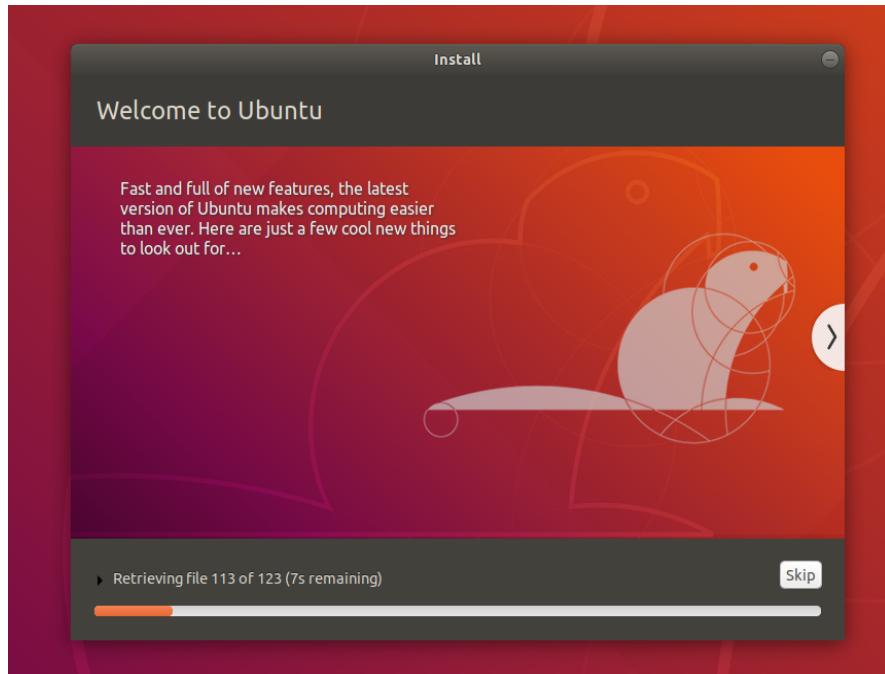


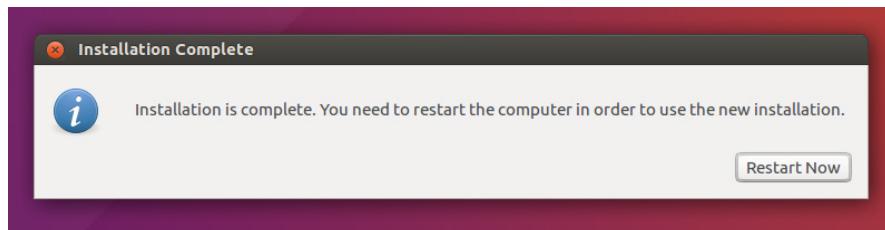
Figure 17: Intro Screen

21. Installation will Continue.



*Figure 18: Installation Screen*

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



*Figure 19: Complete Installation Screen*

MetgenoX can be installed in the Virtual Ubuntu OS normally as described in the previous chapter.

# macOS Support

## Installing MetgenoX software on macOS

MetgenoX is software package for Linux, but macOS users can use MetgenoX using a Virtual Machine. Follow these steps to initialize a Virtual Machine on your OS X host.

**NOTE** *MetgenoX doesn't support 10.10.x (Yosemite) or later running on Intel hardware (PowerPC hardware is not supported nor is building an X11 variant). For installation you need minimum of 2 GB RAM that can be dedicated and 10 GB of storage.*

1. First you have to **download** the image file of the required OS (For Example Ubuntu) from its website or use the following link:

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

2. **Download** the virtualization software (Oracle VM VirtualBox) from the following link:

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

3. **Click** the downloaded dmg file to install the software in macOS host.

4. **Start** Virtual Box, and click on the **New** symbol.



Figure 1: New Virtual OS

5. 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**.

**NOTE** *Allocate about half of the RAM to the virtual OS. Click Next.*

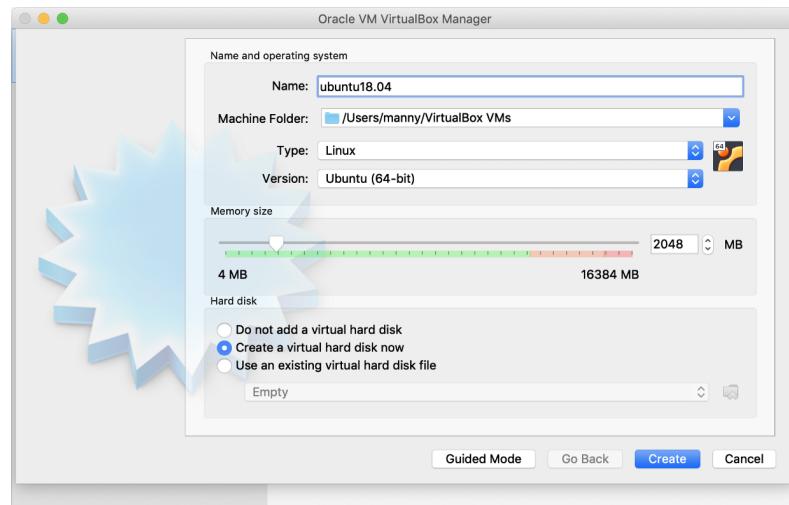


Figure 2: Assigning information

6. 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.*

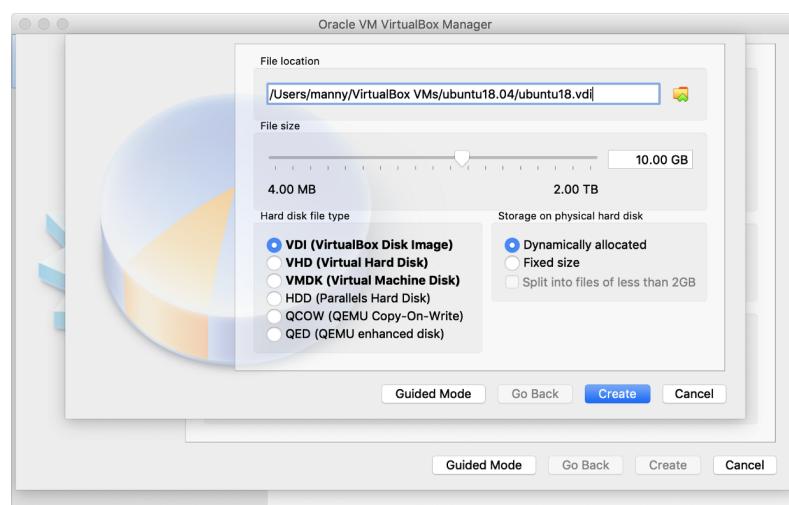


Figure 3: Hard disk file type

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

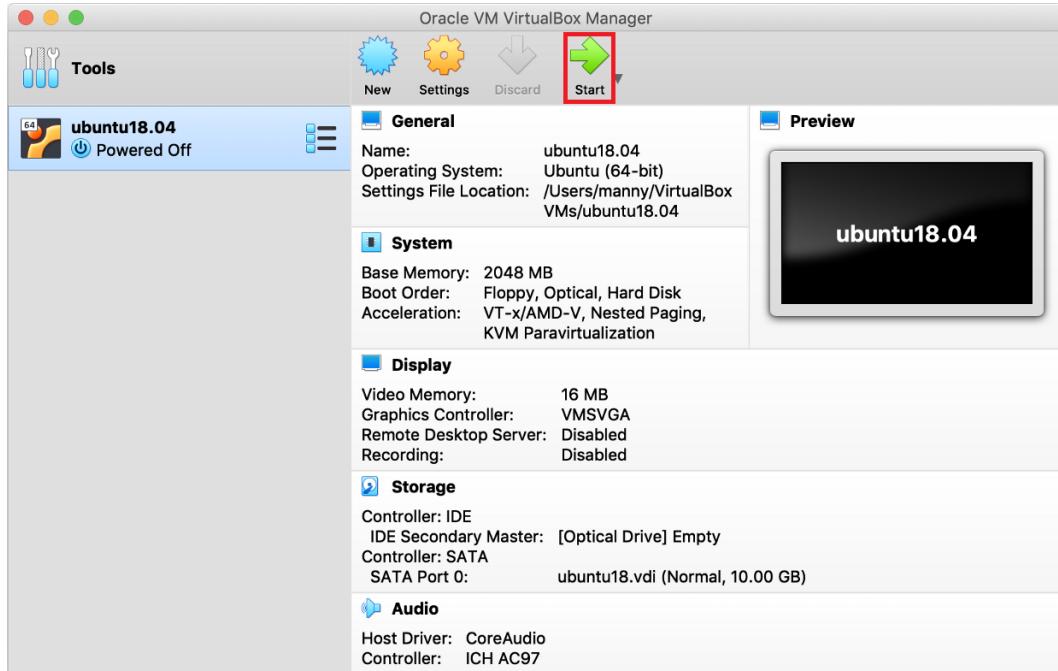


Figure 4: Splash Screen

8. 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 5: Assigning image file

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

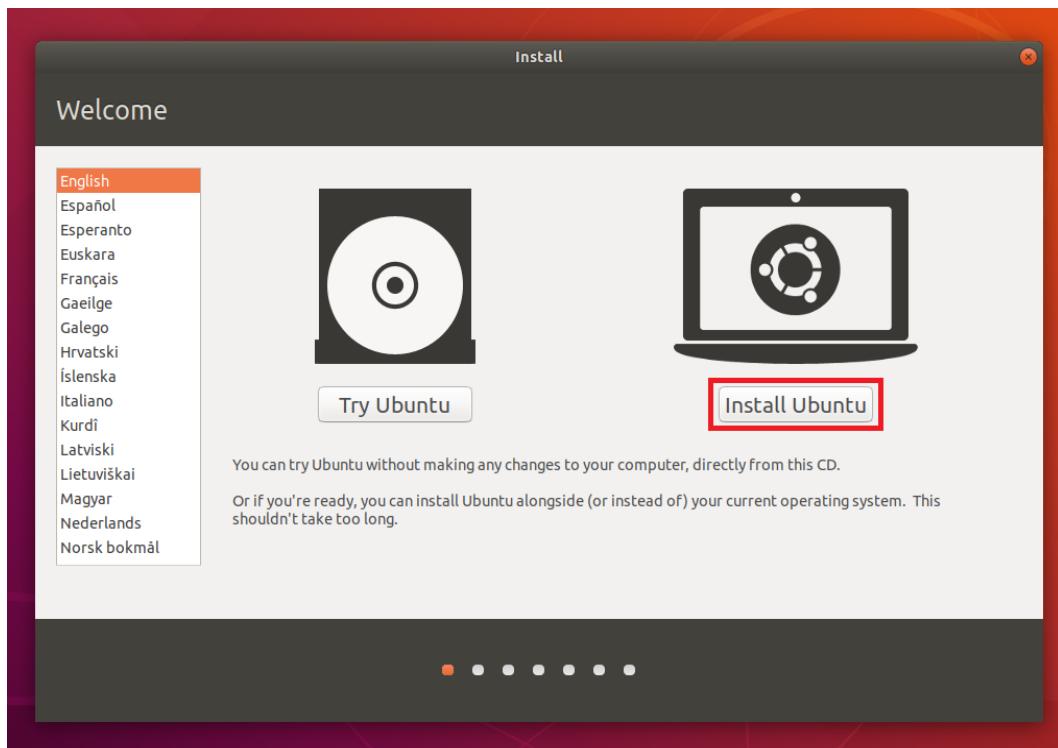


Figure 6: Installing Ubuntu

10. Continue with Normal Installation.

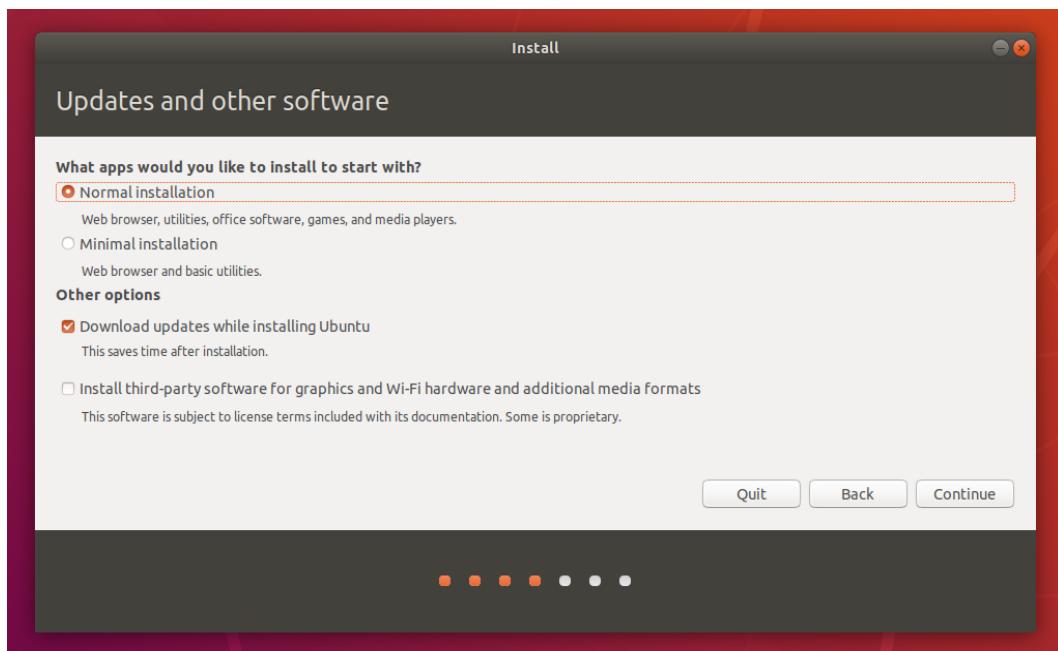


Figure 7: Update Screen

11. In Installation type screen, select **Erase disk** and **Install Now** option.

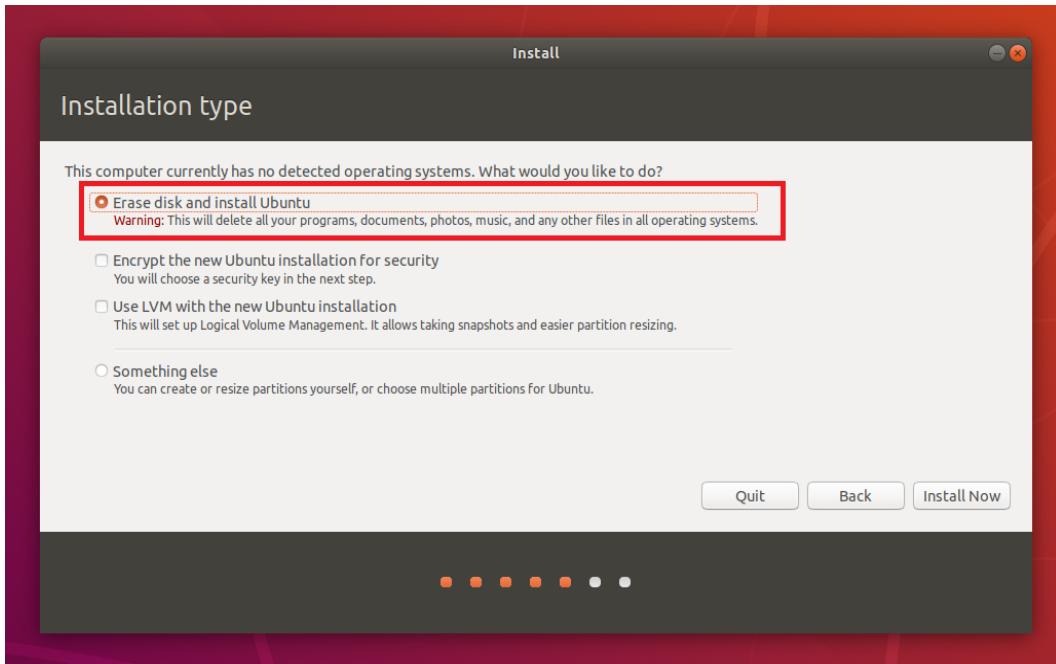


Figure 8: Installation type Screen

12. Select Continue.

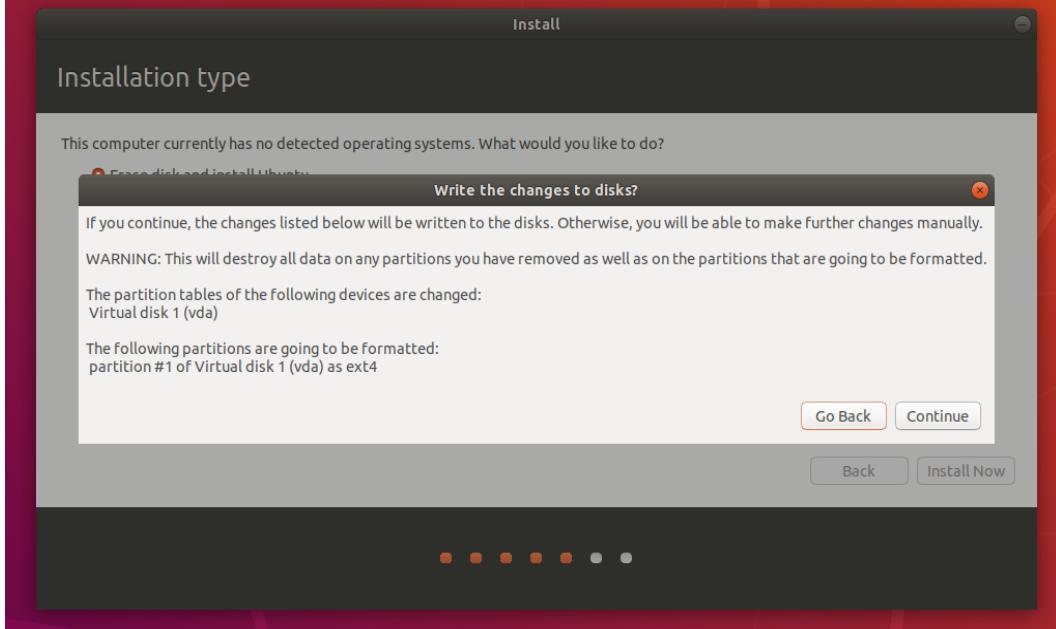


Figure 9: Confirmation Screen

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

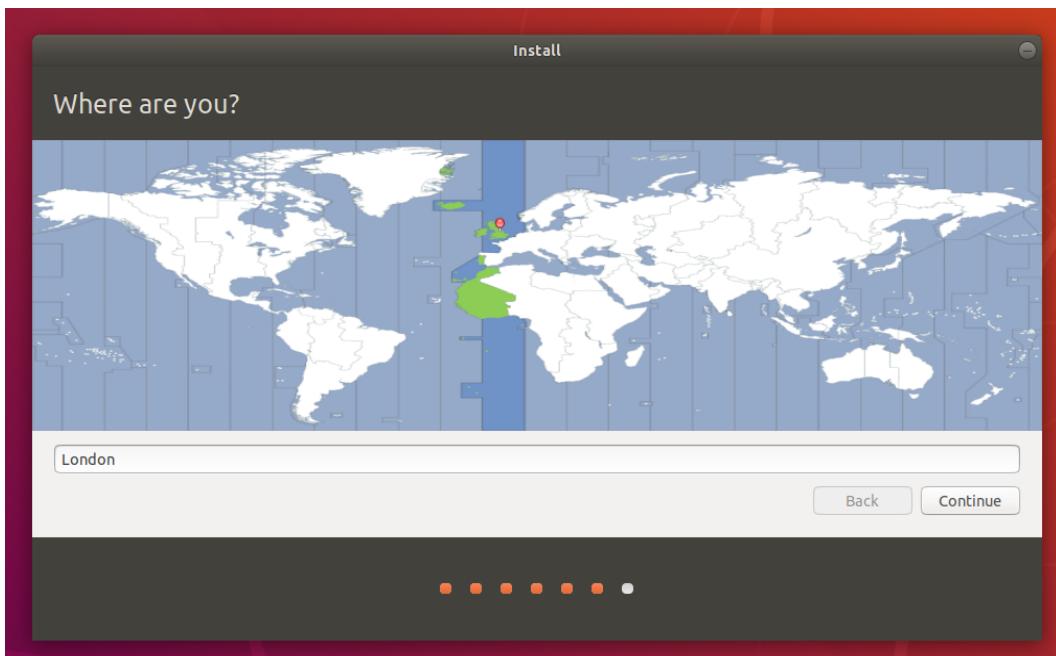


Figure 10: Location selection screen

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

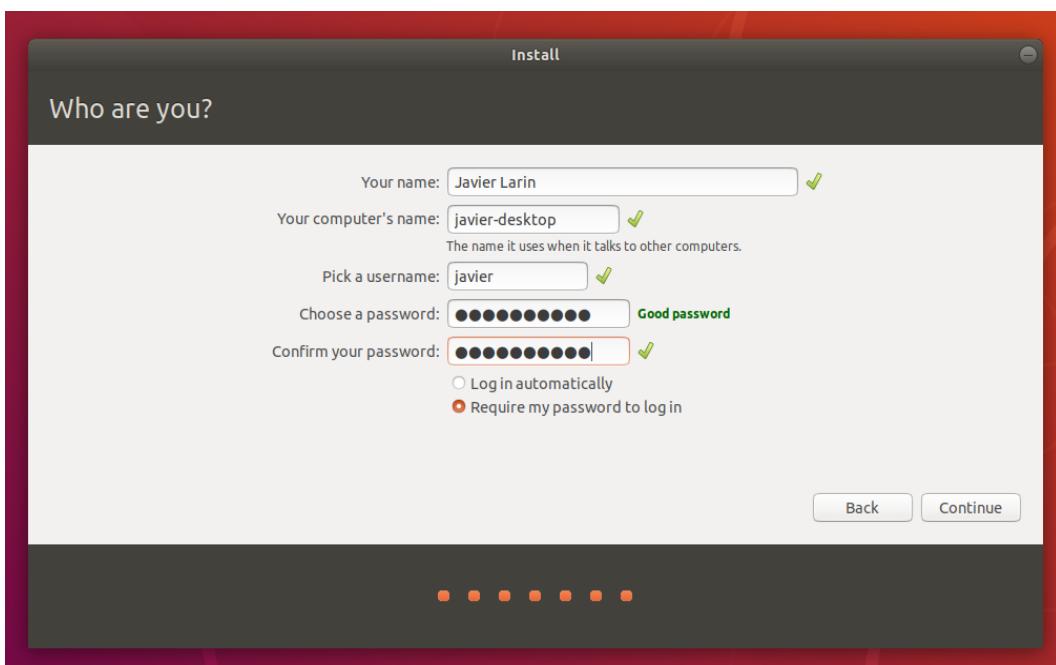


Figure 11: Intro Screen

15. Installation will Continue.

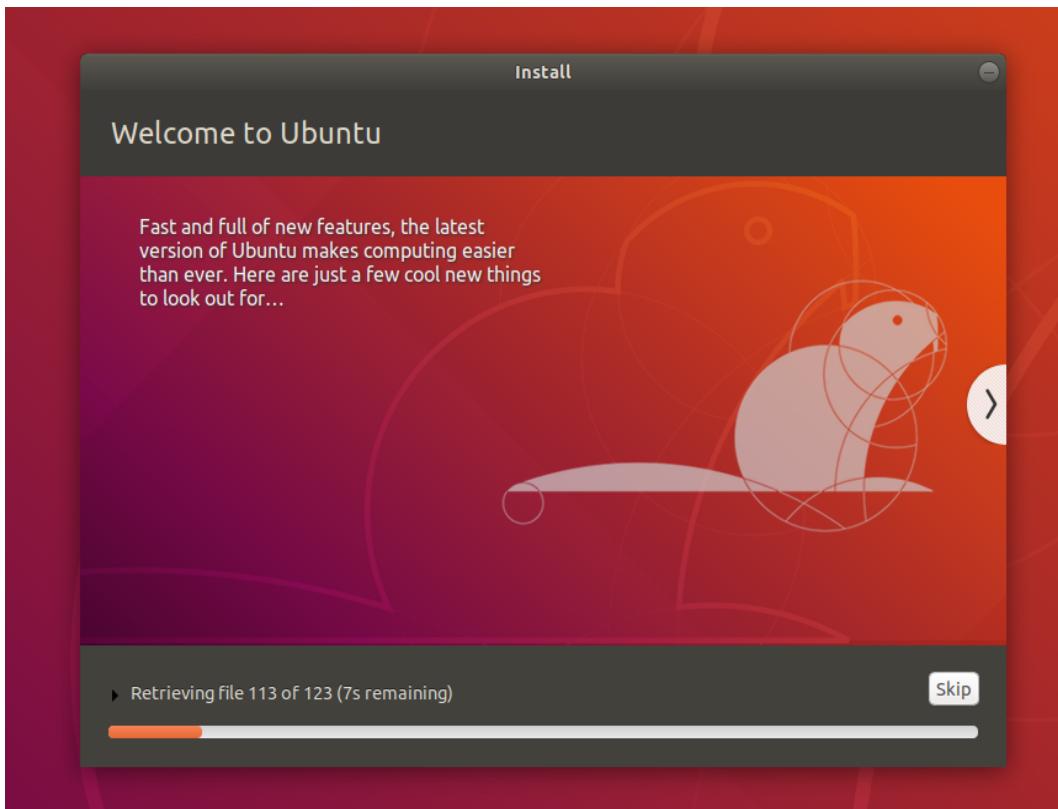


Figure 12: Installation Screen

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

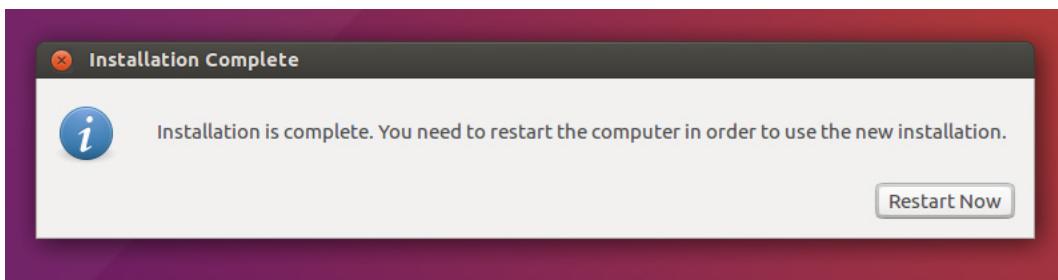


Figure 13: Complete Installation Screen

MetgenoX can be installed in the Virtual Linux OS normally as described in the previous chapter.



# Bibliography

- [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, pp. 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, pp. 166–169, 2015.
- [3] M. Lohse, A. M. Bolger, A. Nagel, A. R. Fernie, J. E. Lunn, M. Stitt, and B. Usadel, "R obi na: A user-friendly, integrated software solution for rna-seq-based transcriptomics," *Nucleic acids research*, vol. 40, no. W1, pp. 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, pp. 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 cancer genome analysis using high-performance computing," *Human mutation*, vol. 38, no. 10, pp. 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, pp. 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, pp. 566–568, 2013.
- [9] L. Jourdren, 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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. D'Antonio, 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, pp. 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. Rimmbach, 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, pp. 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, pp. EBO-523 873, 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, pp. 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, pp. CIN-SI4 022, 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, pp. 863–864, 2013.
  - [38] J. Mariette, C. Noirot, and C. Klopp, "Assessment of replicate bias in 454 pyrosequencing and a multi-purpose read-filtering tool," *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, "Pandaseq: 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, pp. 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, pp. 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, pp. 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. Narai, 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, pp. 574–576, 2016.
- [51] D. M. O'Halloran, "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*. Springer, 2016, pp. 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, pp. 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, pp. 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, p. 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, pp. 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, "Seqasssist: a novel toolkit for preliminary analysis of next-generation sequencing data," vol. 15, no. 11. BioMed Central, 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, pp. 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, pp. 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, pp. 426–434, 2014.
- [72] R. Schmieder and R. Edwards, "Quality control and preprocessing of metagenomic datasets," *Bioinformatics*, vol. 27, no. 6, pp. 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, pp. 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, pp. 3283–3285, 2017.
- [77] Y. Liu, F. Ripp, R. Koeppel, H. Schmidt, S. L. Hellmann, M. Weber, C. F. Krombholz, B. Schmidt, and T. Hankeln, "Afs: identification and quantification of species composition by metagenomic sequencing," *Bioinformatics*, vol. 33, no. 9, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 865–880, 2013.
- [103] D. A. Filatov, "Processing and population genetic analysis of multigenic datasets with proseq3 software," *Bioinformatics*, vol. 25, no. 23, pp. 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, pp. 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. Jovic, and D. Koller, "Genovo: de novo assembly for metagenomes," *Journal of Computational Biology*, vol. 18, no. 3, pp. 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, pp. 147–154, 2017.
  - [112] E. Moss, A. Bishara, E. Tkachenko, J. B. Kang, T. M. Andermann, C. Wood, C. Handy, H. Ji, S. Batzoglou, 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, pp. 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, pp. 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, pp. 159–177, 2015.
  - [121] S. Okuda, Y. Tsuchiya, C. Kiriyma, 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, pp. 1447–1456, 2017.
  - [123] K. Icay, 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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," vol. 15, no. 1. BioMed Central, 2014, p. S12.
  - [134] A. Bremges and A. C. McHardy, "Critical assessment of metagenome interpretation enters the second round," *MSystems*, vol. 3, no. 4, 2018.
  - [135] 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, pp. 1063–1071, 2017.
  - [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, pp. 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, pp. 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, pp. 1043–1055, 2015.
  - [141] A. Mikheenko, V. Saveliev, and A. Gurevich, "Metaquast: evaluation of metagenome assemblies," *Bioinformatics*, vol. 32, no. 7, pp. 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," in *Methods in enzymology*. Elsevier, 2013, vol. 531, pp. 465–485.
  - [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, pp. 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. Rynearson *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, pp. 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, pp. 2230–2239, 2008.
  - [150] F. Schreiber, P. Gumrich, R. Daniel, and P. Meinicke, "Treephyler: fast taxonomic profiling of metagenomes," *Bioinformatics*, vol. 26, no. 7, pp. 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, pp. 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, "Csscl: a python package that uses combined sequence similarity scores for accurate taxonomic classification of long and short sequence reads," *Bioinformatics*, vol. 32, no. 3, pp. 453–455, 2015.
- [154] M. Weber, H. Teeling, S. Huang, J. Waldmann, M. Kassabgy, B. M. Fuchs, A. Klindworth, C. Klockow, A. Wichels, G. Gerdts *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. Raczy, 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, pp. 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, pp. 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, pp. 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.

- 
- [I66] 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, pp. 3535–3542, 2016.
  - [I67] P. Somervuo, S. Koskela, J. Pennanen, R. Henrik Nilsson, and O. Ovaskainen, "Unbiased probabilistic taxonomic classification for dna barcoding," *Bioinformatics*, vol. 32, no. 19, pp. 2920–2927, 2016.
  - [I68] V. C. Piro, M. S. Lindner, and B. Y. Renard, "Dudes: a top-down taxonomic profiler for metagenomics," *Bioinformatics*, vol. 32, no. 15, pp. 2272–2280, 2016.
  - [I69] 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, pp. 1112–1122, 2015.
  - [I70] 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.
  - [I71] 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.
  - [I72] 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, pp. 196–203, 2010.
  - [I73] M. S. Porter and R. G. Beiko, "Spanner: taxonomic assignment of sequences using pyramid matching of similarity profiles," *Bioinformatics*, vol. 29, no. 15, pp. 1858–1864, 2013.
  - [I74] 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.
  - [I75] 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.
  - [I76] 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.
  - [I77] 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.
  - [I78] Z. Rasheed and H. Rangwala, "Metagenomic taxonomic classification using extreme learning machines," *Journal of bioinformatics and computational biology*, vol. 10, no. 05, p. 1250015, 2012.
  - [I79] D. E. Wood and S. L. Salzberg, "Kraken: ultrafast metagenomic sequence classification using exact alignments," *Genome biology*, vol. 15, no. 3, p. R46, 2014.
  - [I80] 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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 c<sub>30</sub> fullerene by pyrrole-n-carboxylic acid (n= 2, 3): Density functional theory studies," *Zeitschrift für Naturforschung A*, vol. 73, no. 1, pp. 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. 17, 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, pp. 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*. Springer, 2008, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 1063–1071, 2017.
- [225] M. Moniruzzaman, G. R. LeCleir, 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, pp. 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, pp. 114–116, 2012.
- [227] W. M. Ismail, Y. Ye, and H. Tang, "Gene finding in metatranscriptomic sequences," vol. 15, no. 9. BioMed Central, 2014, p. S8.
- [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, pp. 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, pp. 2223–2230, 2012.
- [230] Y. Huang, P. Gilna, and W. Li, "Identification of ribosomal rna genes in metagenomic fragments," *Bioinformatics*, vol. 25, no. 10, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, pp. 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, "Rocker: 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, pp. e14–e14, 2016.
- [241] P. Meinicke, "Uproc: tools for ultra-fast protein domain classification," *Bioinformatics*, vol. 31, no. 9, pp. 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, pp. 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, pp. 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, pp. 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. Konig, 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," *IET Proceedings-Systems Biology*, vol. 153, no. 5, pp. 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, pp. 773–775, 2017.
- [251] G. Basler and Z. Nikoloski, "Jmassbalance: Mass-balanced randomization and analysis of metabolic networks," *Bioinformatics*, vol. 27, no. 19, pp. 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 <sup>1</sup>h nmr spectra,” *Bioinformatics*, vol. 35, no. 11, pp. 1916–1922, 2019.
- [255] A. Rodriguez-Martinez, J. M. Posma, R. Ayala, A. L. Neves, M. Anwar, E. Petretto, C. Emanueli, D. Gauguier, J. K. Nicholson, and M.-E. Dumas, “Mwastools: an r/biocconductor package for metabolome-wide association studies,” *Bioinformatics*, vol. 34, no. 5, pp. 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, pp. 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, pp. 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, pp. 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, “Explicit: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data,” *Bioinformatics*, vol. 29, no. 23, pp. 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, pp. 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, p. 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, "Syntview—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, pp. 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. Vandebogaert, and V. Caro, "MetagenseSense: 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.