



Workshop on Comparative Genomics

Colorado State University

Installing QIIME: Full install, virtual & cloud machines

Antonio González Peña
University of Colorado at Boulder
Knight Lab

Why we care?

- Microbes are ubiquitous

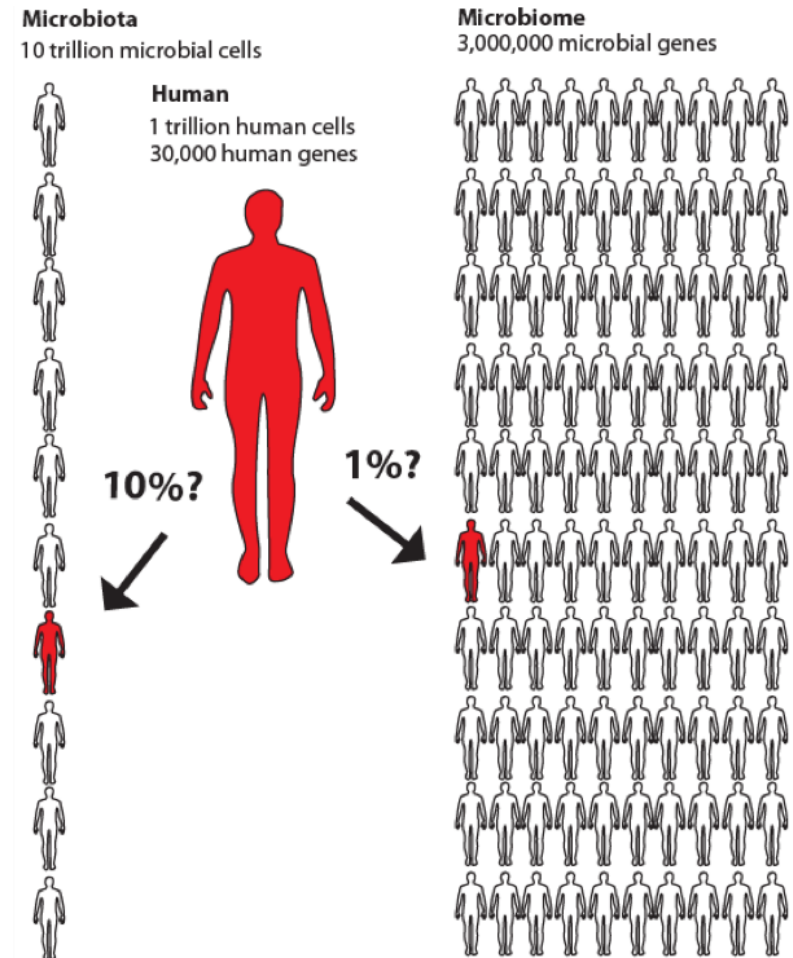
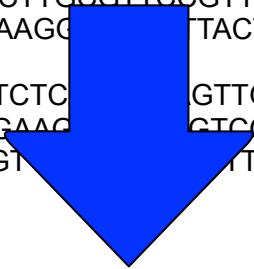


Image by Micah Hamady

Why we need QIIME?

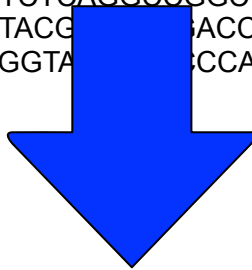
```
>Sa823_0 CAGATACACTTC 000448_3980_2031
CTGGGCCCGTGTCTCAGTCCCAGTGTGGCCGAACACCCTCTCAGGCCGGCTACCCATCATTACCTTGGTAGGCCGTTACCTACC
AACTAGCTAATAGGACGCAGGCTCATCCCCAAGCGCCTTACGGCTTTGACCCTGGAAGCTTATGCCCCAGGGTATCATGCGGT
ATTAGACCCCTTTTCGGGGGGCTTATTCTCACTCGGGGGGTAGATTGCCACGTGTTACTCACCCGTTCCGCACTCA
>Sa1261_1 CATACCAGTAGC 000784_3964_2774
CTGGGCCCGTGTCTCAGTCCCAATGTGGCCGATCAACCTCTCAGTTCGGCTACCGATCGTCGCCTTGGTGGGCCGTTACTCCGC
CAACTAGCTAATCGGACGCGGGGCCCATCTTTCACCGGATTGCTCCTTTGACCACTAAAGGGTGTCTTCTATGGTCTTATGCGG
TATTAGCACAAGTTTCCCTGTGTTATCCCCAGTAAAA
>Sa1232_2 CAGTACGATCTT 001023_3894_1089
CTGGGCCCGTGTCTCAGTCCCAGTGTGGCTGGTTCGTCCTCTCAGACCAGCTACAGATCGTCGGCTTGGTAAGCTTTTATCCCACC
AACTACCTAATCTGCCATCGGCCGCTCCAAACGCGCAAGGCCTTGCGGTCCCCTGCTTTTCATCCTTAGATCGTATGCGGTATTAG
CAAAGCTTTCGCCTCGTTATCCCCACGTCTGGGCACGTTCCGATGTATTACT
>Sa822R_3 CATCGTATCAAC 001047_3770_2500
CTCGCCCGTGTCTCAGTGCGAGTGTGGCTGGTTCGTCCTCTCAGACCAGCTACCCGTCATCGCCTTGGTGGTCCATTACACCGCC
AACTAGCTGATAGGCCGCAAGCTCATCCTCAAGCGGTTCTTGCGTTCCGTTTAAATGGCCAGGACTTCCGTCCCGGTACACCATC
CGGTATTAGCCAACCTTTTCGGCTGGTTGTGCCGGTCTTAAGCTTACTTACGTGTTA
>Sa1522_4 CATATACTCGCA 001249_3847_3836
CTGGGCCCGTGTCTCAGTCCCAGTGTGGCTGAACATCCTCTCTGTTGCGGATCATCGCCTTGGTAGACCATTACCCTACCA
ACTAGCTAATCGGACGCAAGCCCCTCCCCAAGCGCCCGAAGCTCCTTCCGCCTTACCGAAAGGACCACATACGGTATTA
GCTCCGATTTCTCAGAGTTATTCCCTACTTAGGGGCAGGTCTTACTCACCCATTCTG
```



Split libraries, pick OTU's, chimera-check,
align, phylogenetic tree, taxonomic classification, diversity
analyses, multivariate statistics, plots

Why we need QIIME?

```
>Sa823_0 CAGATACACTTC 000448_3980_2031
CTGGGCCGTGTCTCAGTCCCAGTGTGGCCGAACACCCTCTCAGGCCGGCTACCCATCATTACCTTGGTAGGCCGTTACCTACC
AACTAGCTAATAGGACGCAGGCTCATCCCCAAGCGCCTTACG ACCCTGGAAGCTTATGCCCCAGGGTATCATGCGGT
ATTAGACCCCCTTTTCGGGGGCTTATTCCTCACTCGGGGGGTA CCACGTGTTACTCACCCGTTTCGCCACTCA
```



Align, Chimera-check:

Upload to greengenes.lbl.gov server, 500 at a time

pick OTU's:

Dotur/Mothur through terminal on personal machine

phylogenetic tree:

ARB, upload to RAxML server (1 week for 2500 seqs)

taxonomic classification:

greengenes.lbl.gov server, 500 at a time or ARB (count by hand)

diversity analyses

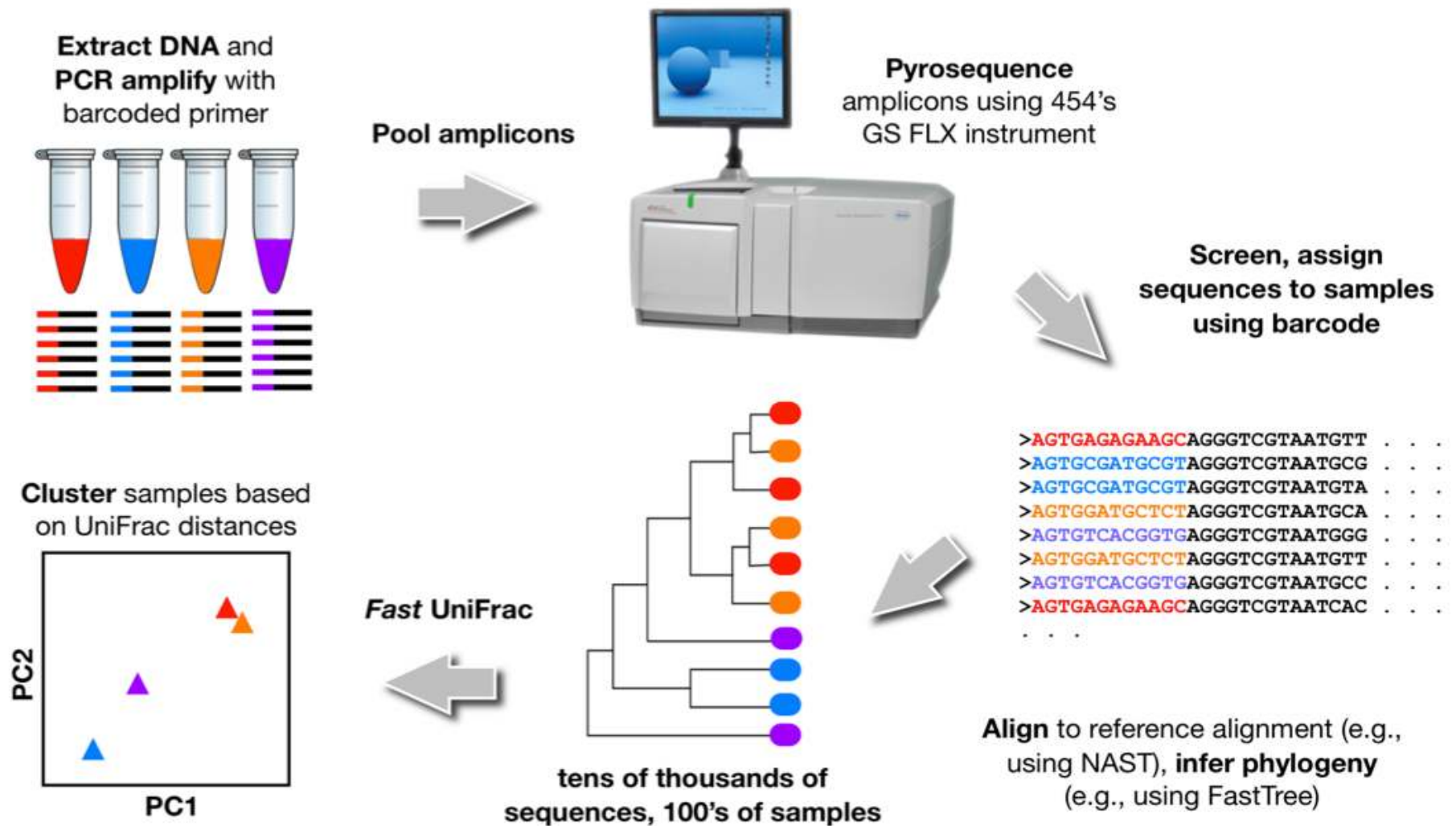
DOTUR/MOTHUR, upload to UniFrac server distance matrices

upload to UniFrac server, vegan package in R, PrimerE

multivariate statistics, plots

vegan and MASS packages in R, MATLAB, PrimerE

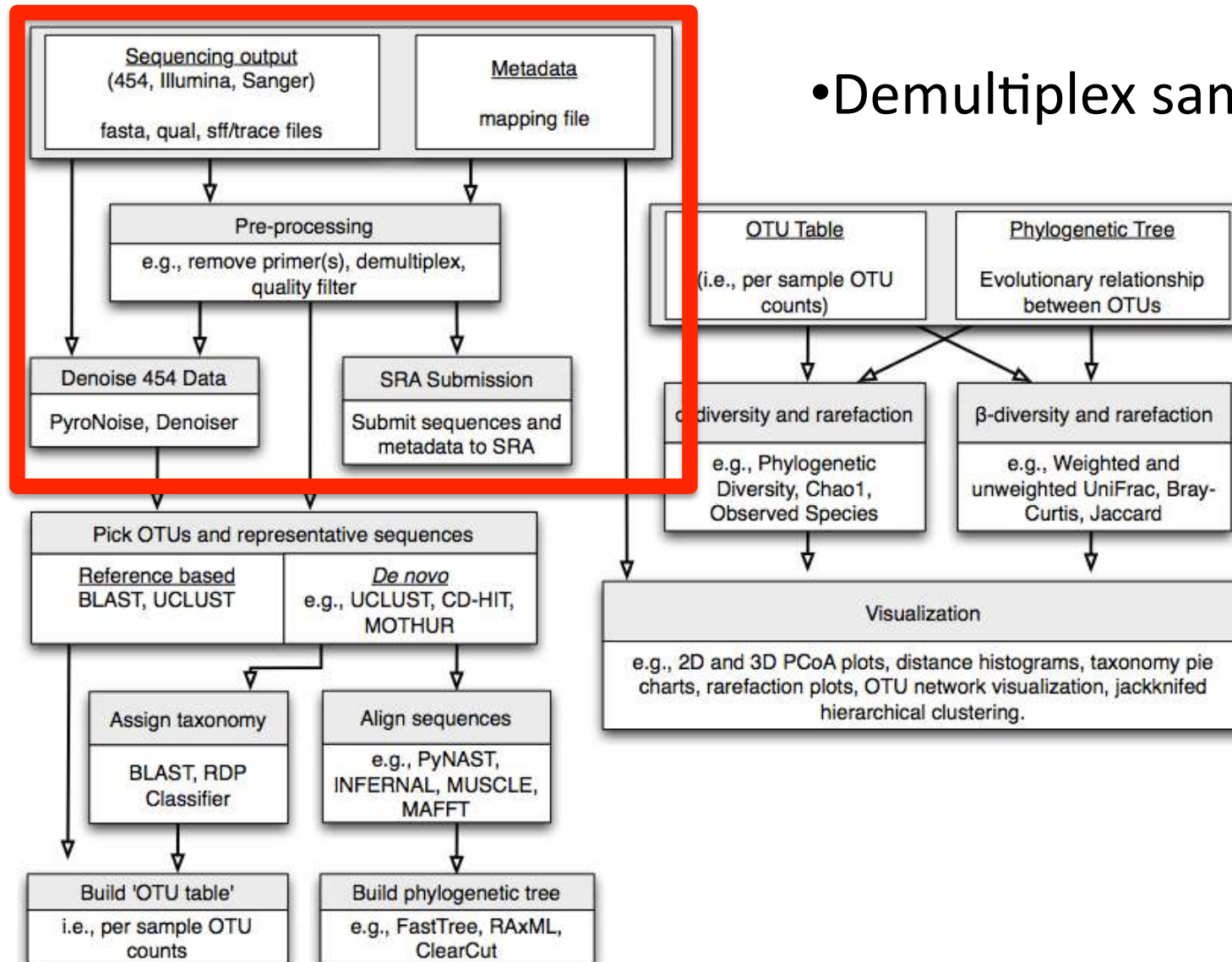
QIIME



Hamady et al. Error-correcting barcodes for pyrosequencing hundreds of samples in multiplex. *Nature Methods*, 2008.

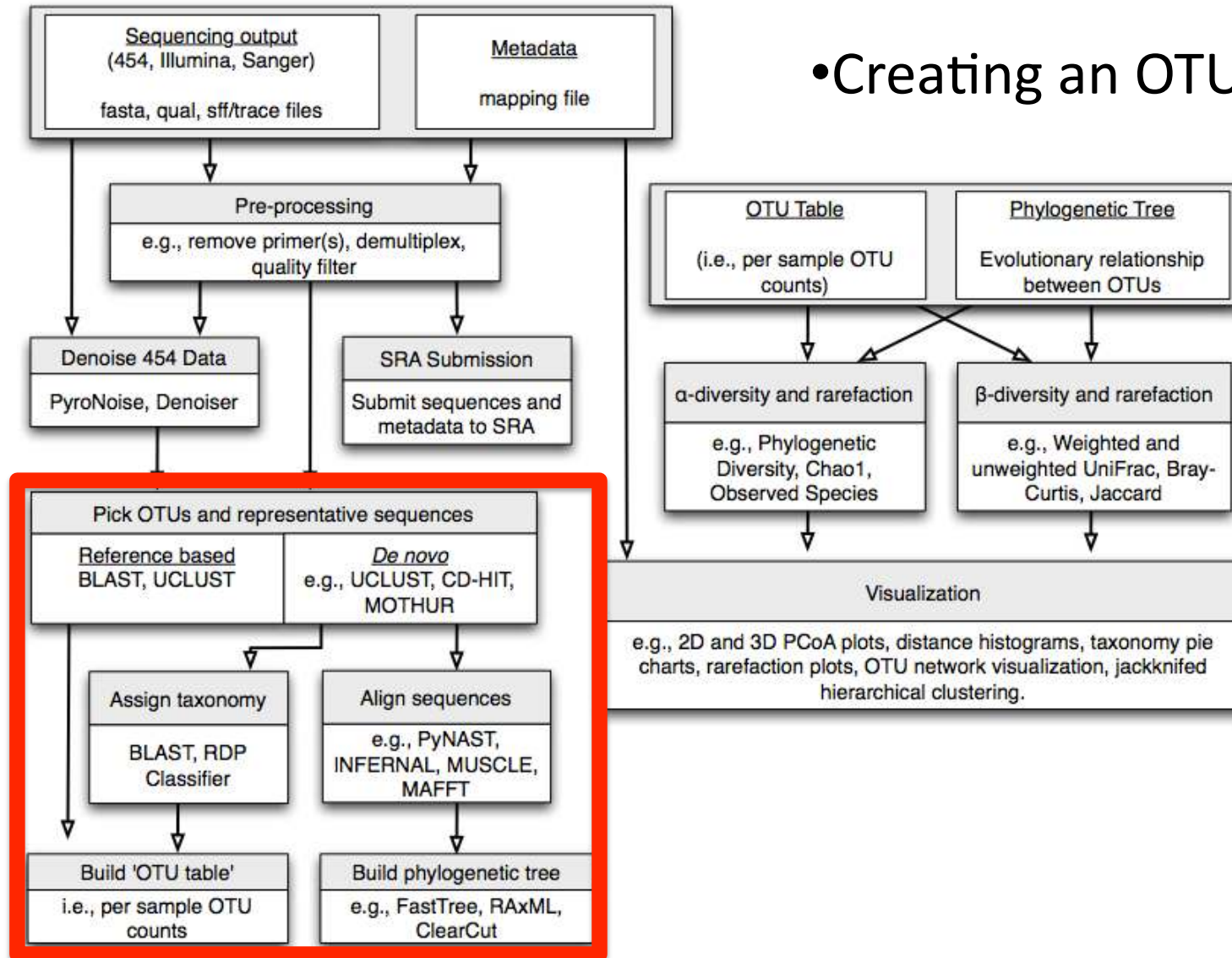
QIIME

- Demultiplex samples



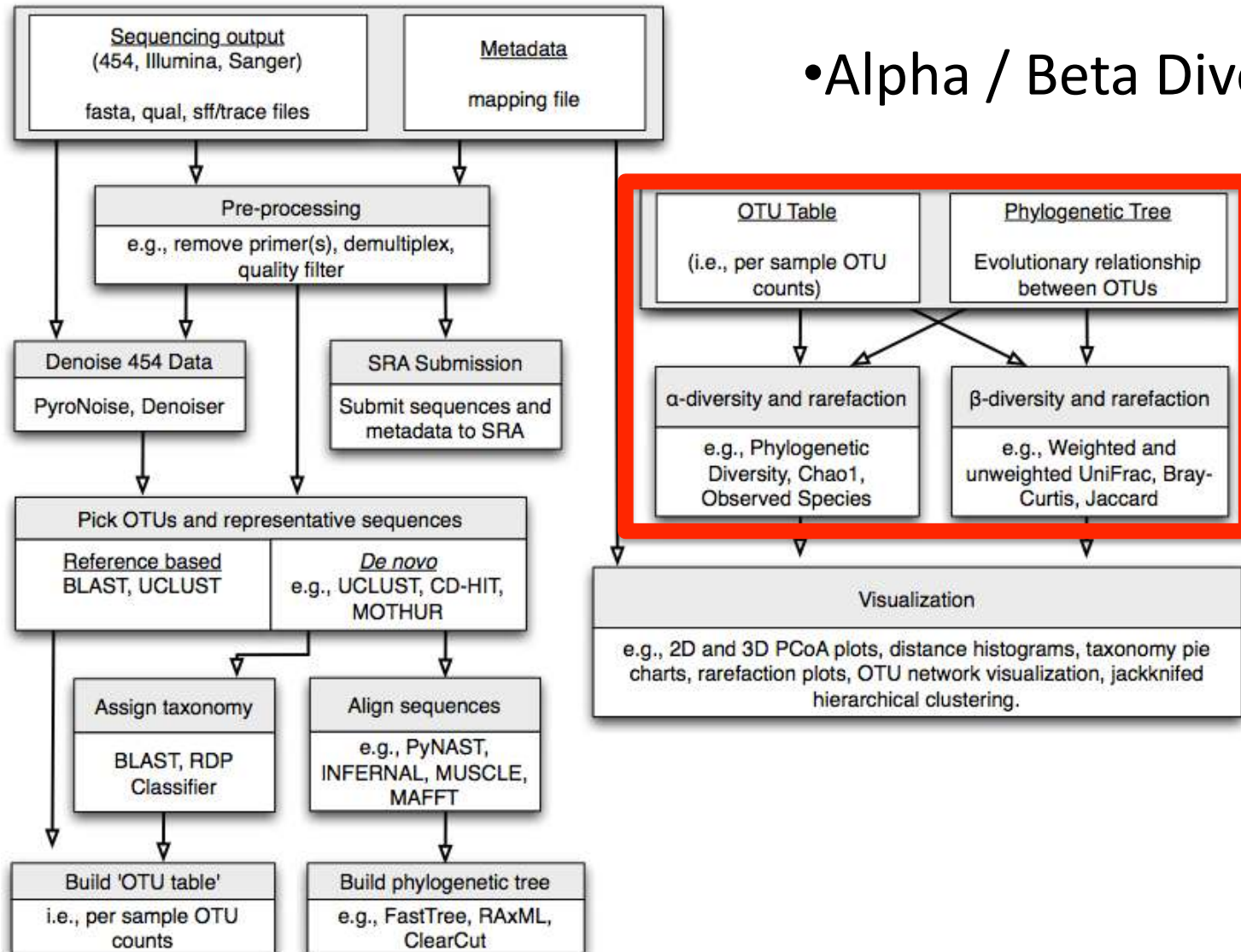
QIIME

•Creating an OTU table



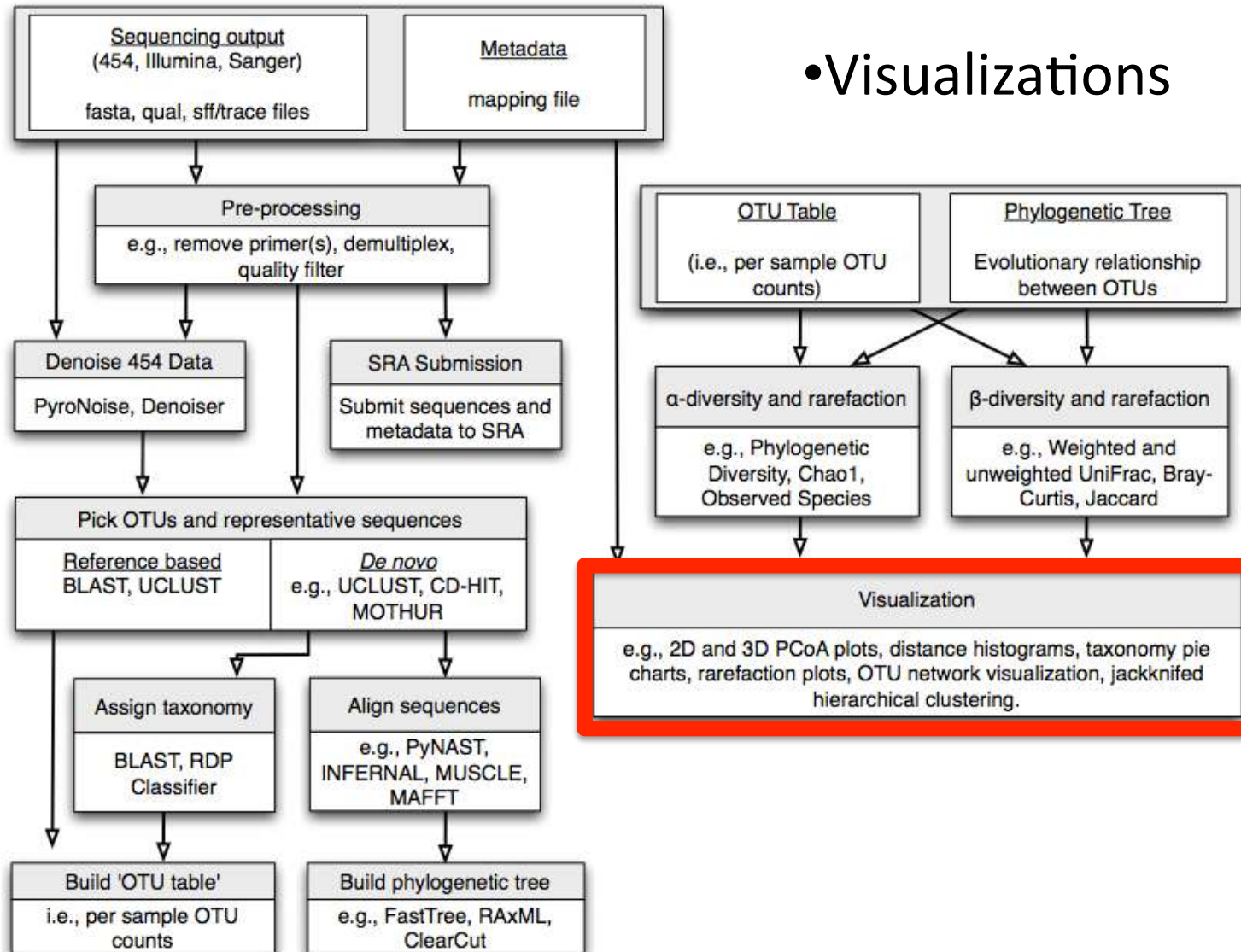
QIIME

•Alpha / Beta Diversity



QIIME

•Visualizations



What do I need to run QIIME?

1. Python-2.6.4
2. Qiime-1.3.0
3. setuptools-0.6c11
4. MySQL-python-1.2.3
5. SQLAlchemy-0.7.1
6. PyCogent-1.5.1
7. PyNAST-1.1
8. numpy-1.3.0
9. matplotlib-0.98.5.3
10. mpi4py-1.2.2
11. lxml-2.2.7
12. Sphinx-1.0.4
13. RAxML-7.0.3.tar.bz2
14. FastTree-2.1.3.c
15. cdbfasta
16. cd-hit-2007-0131
17. rdp_classifier_2.2
18. blast-2.2.22-x64-linux
19. muscle3.6_linux_ia32
20. infernal-1.0.2
21. cytoscapeSource-v2.7.0
22. Clearcut.source
23. Mothur.1.6.0
24. uclustq1.2.22_i86linux64
25. R-2.12.0
26. AmpliconNoiseV1.21
27. ViennaRNA-1.8.4
28. pprospector-1.0.1
29. microbiomeutil_2010-04-29

QIIME installs

- Install every single dependency by hand
- Use the app-deploy.py script
- Download the Virtual Machine QIIME hard drive
- Use the latest EC2 QIIME release

QIIME installs

-  Install every single dependency by hand
 - Use the app-deploy.py script
 - Download the Virtual Machine QIIME hard drive
 - Use the latest EC2 QIIME release

app-deploy.py

- Only works in Linux 64bit machines
- `python app-deploy.py [install dir] -f [conf file]`
- Multi-process install
 - `wget http://bmf.colorado.edu/QIIME/app-deploy-qiime-1.3.0.tgz`
 - `tar zxvf app-deploy-qiime-1.3.0.tgz`
 - `cd app-deploy-qiime-1.3.0`
 - `python app-deploy.py /software/ -f etc/qiime_1.3.0.conf`
 - `exit`

app-deploy.py



- Only works in Linux 64bit machines
- `python app-deploy.py [install dir] -f [conf file]`
- Multi-process install
 - `wget http://bmf.colorado.edu/QIIME/app-deploy-qiime-1.3.0.tgz`
 - `tar zxvf app-deploy-qiime-1.3.0.tgz`
 - `cd app-deploy-qiime-1.3.0`
 - `python app-deploy.py /software/ -f etc/qiime_1.3.0.conf`
 - `exit`

app-deploy.py

```
2011-06-30 17:19:21,297 - lib.application - INFO - Deploying app: qiime
2011-06-30 17:19:21,298 - lib.application - INFO - Making deploy directory: /var/lib/hudson/jobs/TestQIIME/workspace/software/qiime-1.3.0-release
2011-06-30 17:19:21,299 - lib.application - INFO - Downloading qiime
2011-06-30 17:19:21,300 - lib.util - INFO - Downloading http://sourceforge.net/projects/qiime/files/releases/Qiime-1.3.0.tar.gz
2011-06-30 17:19:41,171 - lib.util - INFO - Building qiime
2011-06-30 17:19:56,909 - lib.util - INFO - Installing qiime
2011-06-30 17:19:58,279 - lib.application - INFO - Successfully built qiime.
2011-06-30 17:19:58,843 - lib.deploy - INFO - Deploy of qiime succeeded.
```

```
INFO Launched initial deploy,
INFO Deploying app: qiime
INFO Making deploy directory:
INFO Downloading qiime
INFO Building qiime
INFO Installing qiime
INFO Successfully built qiime.
INFO Deploy of qiime
INFO Generating new ~/.qiime_config
INFO Backing up /var/lib/hudson/.qiime_config
INFO Generated new ~/.qiime_config
```

Virtual Machines



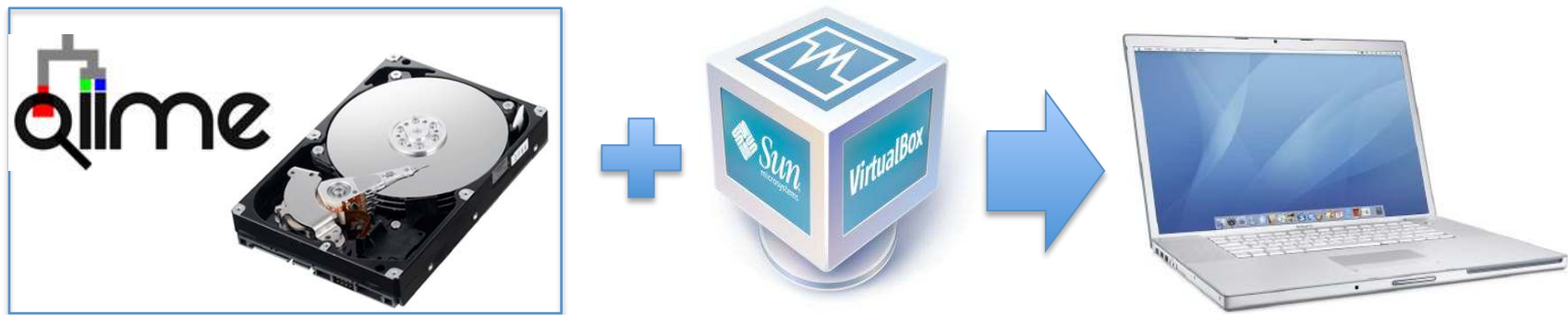
Virtualization Software Packages

- VirtualBox
- Windows Virtual PC
- Xen
- VMware
- Many, many more ...



QIIME Virtual Machine

- VirtualBox
- QIIME virtual hard drive
- http://qiime.org/install/virtual_box.html
- <http://videos.qiime.org/>



Cloud Computing

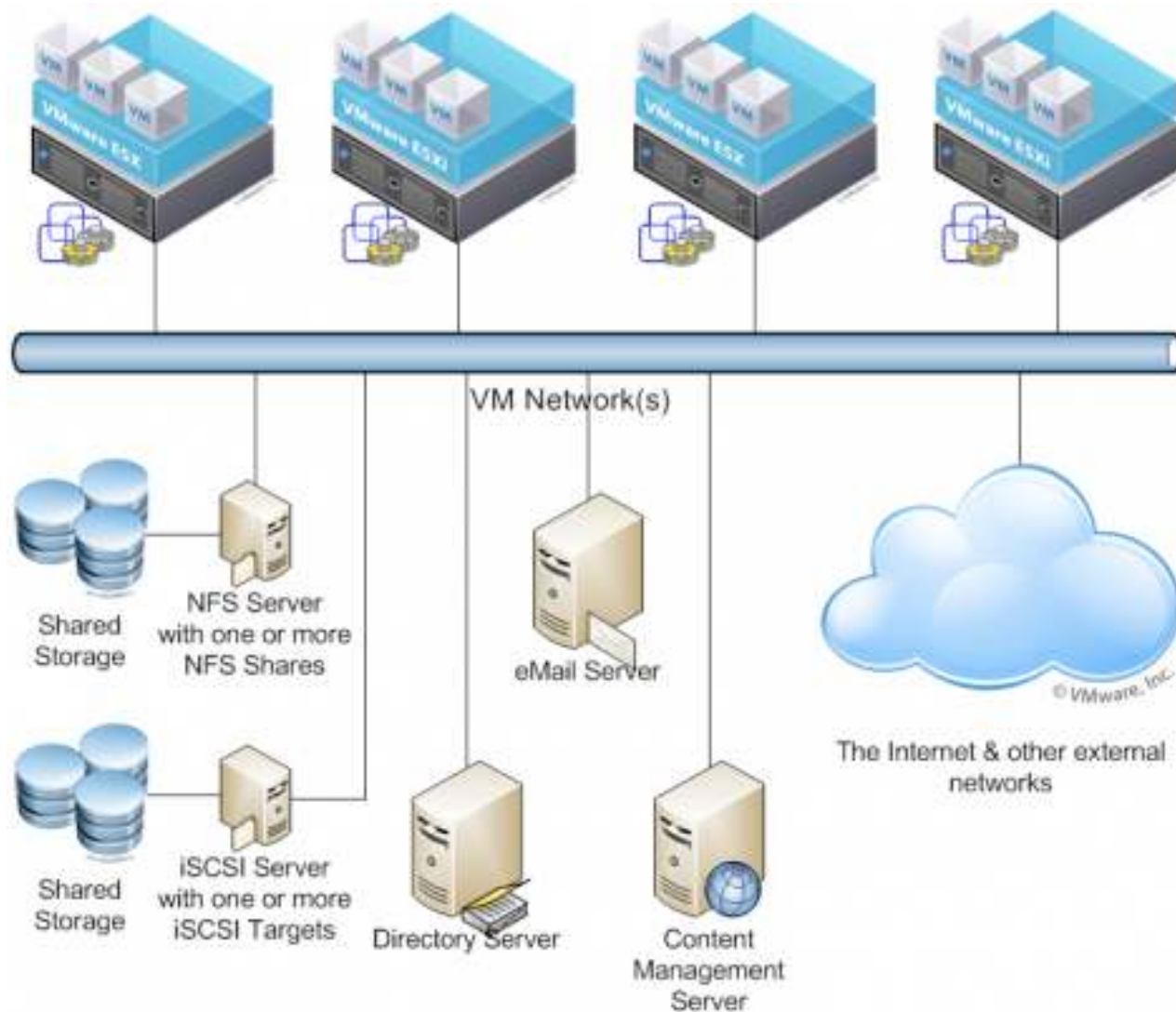


Image from: http://www.techwarelabs.com/articles/editorials/virtual_double/

Cloud computing options

- Amazon Elastic Compute Cloud (EC2)
- GoGrid
- Magellan – Argonne's DOE Cloud Computing
- Data Intensive Academic Grid (DIAG) –
Institute for Genome Sciences (IGS), University
of Maryland School of Medicine (UMSOM)

EC2 demo

<http://aws.amazon.com>

EC2 – Region \$\$

US East (Virginia)

US East (Virginia)

US West (Northern California)

EU (Ireland)

Asia Pacific (Singapore)

Asia Pacific (Tokyo)

Instance

Region: US East (Virginia)	
Linux/UNIX Usage	
Standard On-Demand Instances	
Small (Default)	\$0.085 per hour
Large	\$0.34 per hour
Extra Large	\$0.68 per hour

Instance

Region: Asia Pacific (Tokyo)	
Linux/UNIX Usage	
Standard On-Demand Instances	
Small (Default)	\$0.10 per hour
Large	\$0.40 per hour
Extra Large	\$0.80 per hour

EC2 – Region \$\$

US East (Virginia) 

- US East (Virginia)
- US West (Northern California)
- EU (Ireland)
- Asia Pacific (Singapore)
- Asia Pacific (Tokyo)

Instance

Region: US East (Virginia) 	
Linux/UNIX Usage	
Standard On-Demand Instances	
Small (Default)	\$0.085 per hour
Large	\$0.34 per hour
Extra Large	\$0.68 per hour

Spot Instance

Region: US East (Virginia) 	
Linux/UNIX Usage	
Standard Spot Instances	
Small (Default)	\$0.03 per hour
Large	\$0.12 per hour
Extra Large	\$0.24 per hour

EC2 – Region \$\$

	Linux/UNIX Usage	Windows Usage
Standard On-Demand Instances		
Small (Default)	\$0.085 per hour	\$0.12 per hour
Large	\$0.34 per hour	\$0.48 per hour
Extra Large	\$0.68 per hour	\$0.96 per hour
Micro On-Demand Instances		
Micro	\$0.02 per hour	\$0.03 per hour
Hi-Memory On-Demand Instances		
Extra Large	\$0.50 per hour	\$0.62 per hour
Double Extra Large	\$1.00 per hour	\$1.24 per hour
Quadruple Extra Large	\$2.00 per hour	\$2.48 per hour
Hi-CPU On-Demand Instances		
Medium	\$0.17 per hour	\$0.29 per hour
Extra Large	\$0.68 per hour	\$1.16 per hour
Cluster Compute Instances		
Quadruple Extra Large	\$1.60 per hour	N/A*
Cluster GPU Instances		
Quadruple Extra Large	\$2.10 per hour	N/A*

AWS in Education (Grants)

- <http://aws.amazon.com/education/>

QIIME

- The code is tested (properly)
- The documentation is updated constantly based on users suggestions
- The help in the QIIME-forum has a collaborative spirit (developers & users sharing their research experiences)



Thanks

<http://qiime.org>

<http://forum.qiime.org>

<http://videos.qiime.org>

Antonio Gonzalez Peña
antgonza@gmail.com

