
Quantitative Insights Into Microbial Ecology
qiime
Mahidol University
QIIME Workshop
December 14th-15th, 2015
William (Tony) Walters
Jai Ram Rideout

Northern Arizona University

Agenda

http://bit.ly/1QfnzBb

Outline for today's session December 14th

08:00 - 08:30 Intro to microbial community analysis

08:30 - 09:00 Intro to QIIME: installation and usage

09:00 - 09:30 Preprocessing: metadata and demultiplexing

09:30 - 09:45 Break

09:45 - 11:15 Preprocessing: OTU picking methods

11:15 - 12:00 Diversity and statistical analysis Part I

12:00 - 13:30 Lunch break

13:30 - 15:30 Diversity and statistical analysis Part II

15:30 - 17:00 Advanced topics

First, some background about the microbiome and microbiome studies

A microbe dominated world

The *small subunit ribosomal RNA* gene is frequently used to "fingerprint" different microbial organisms.

Why this gene?

- It's ubiquitous.
- Contains regions conserved across organisms, and regions that are variable across organisms.

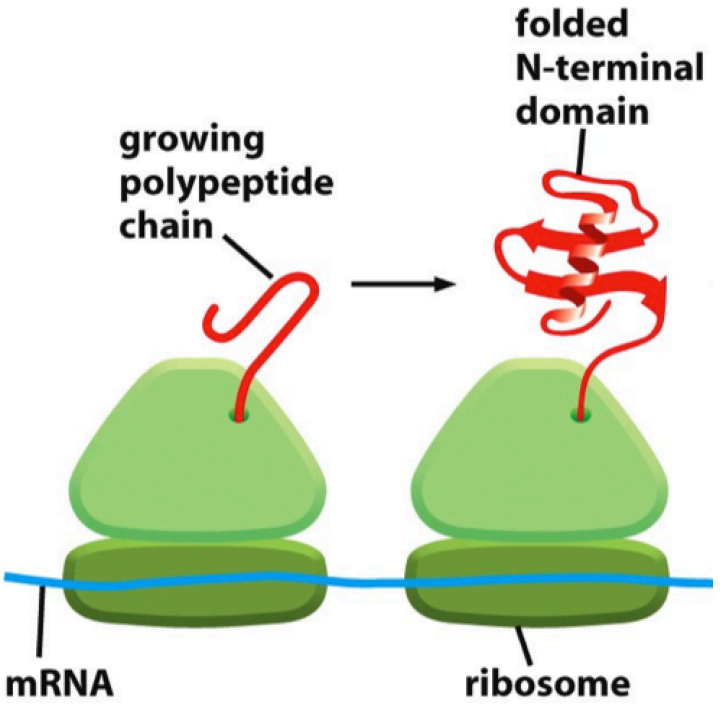
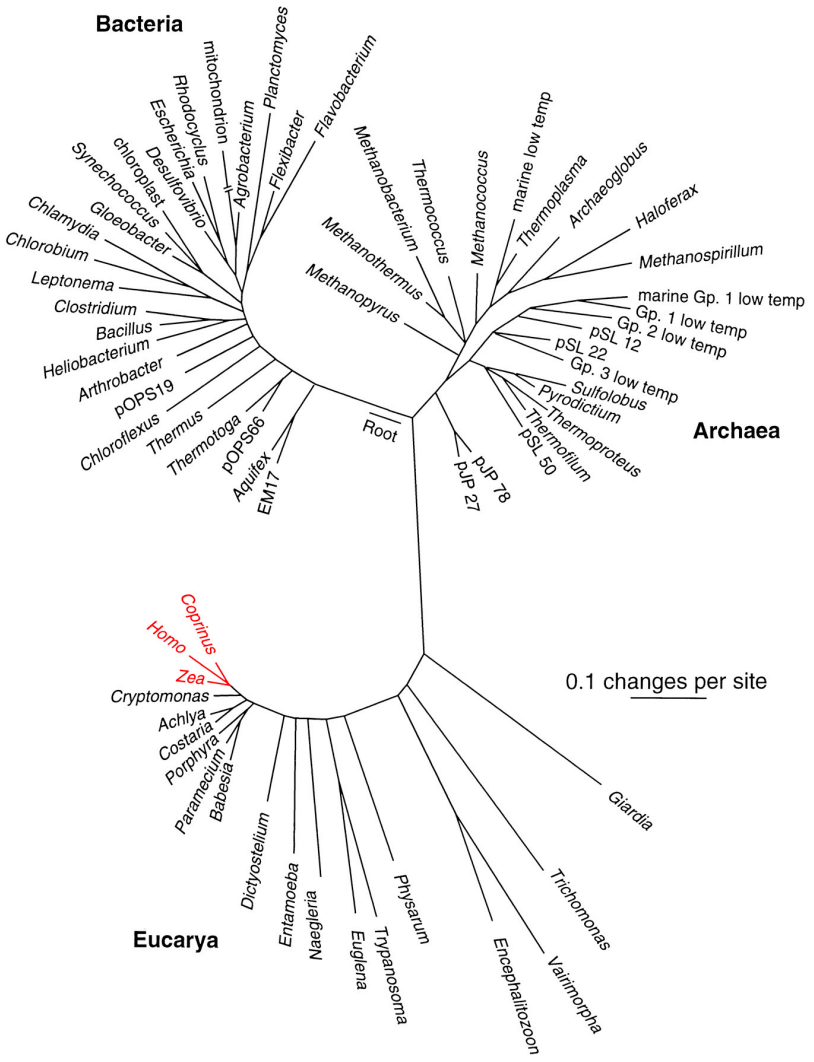


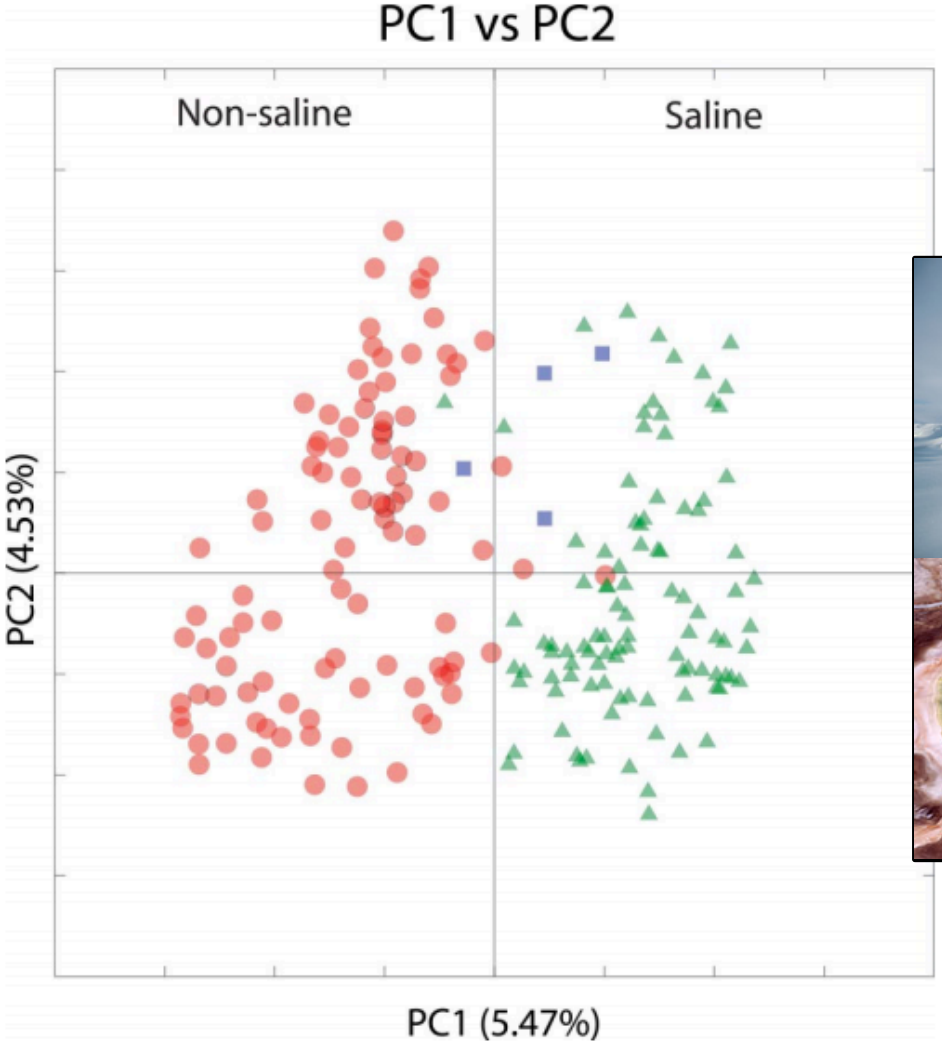
Figure 6-84 *Molecular Biology of the Cell* (© Garland Science 2008)

A microbe dominated world



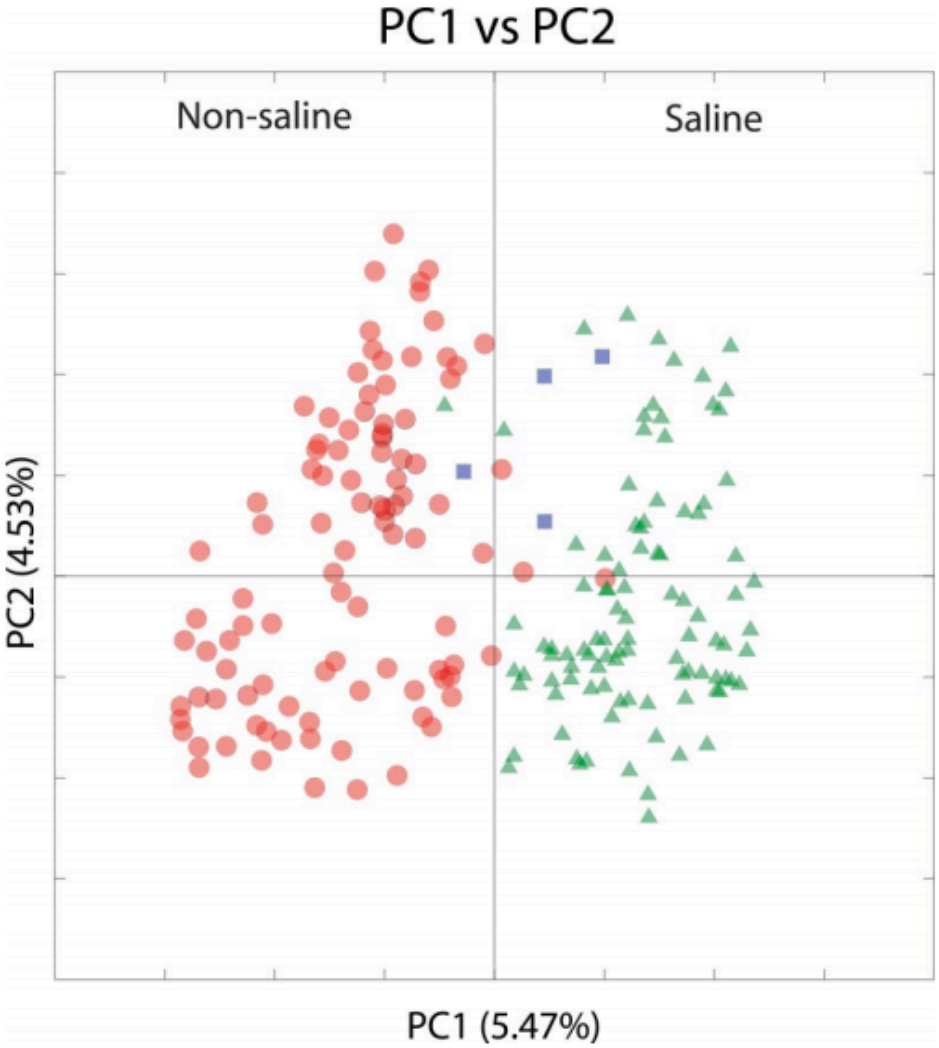
Pace NR. *PNAS*. 2001 Jan 30;98(3):805-8.

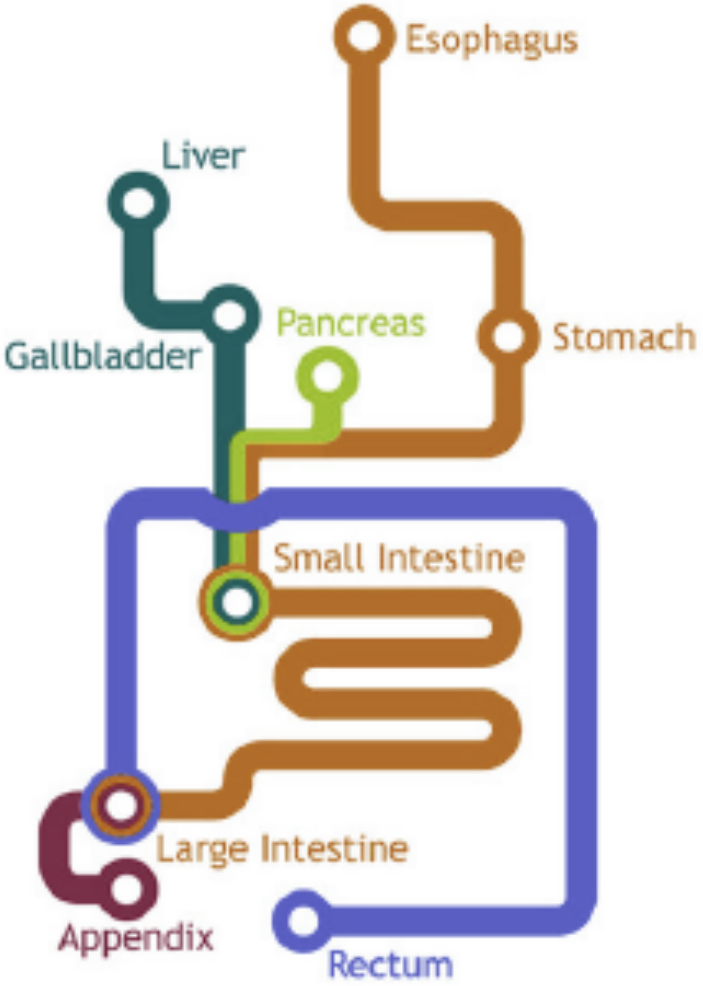
A microbe dominated world



Lozupone and Knight *PNAS*. 2007 Jul 3;104(27):11436-40

A microbe dominated world

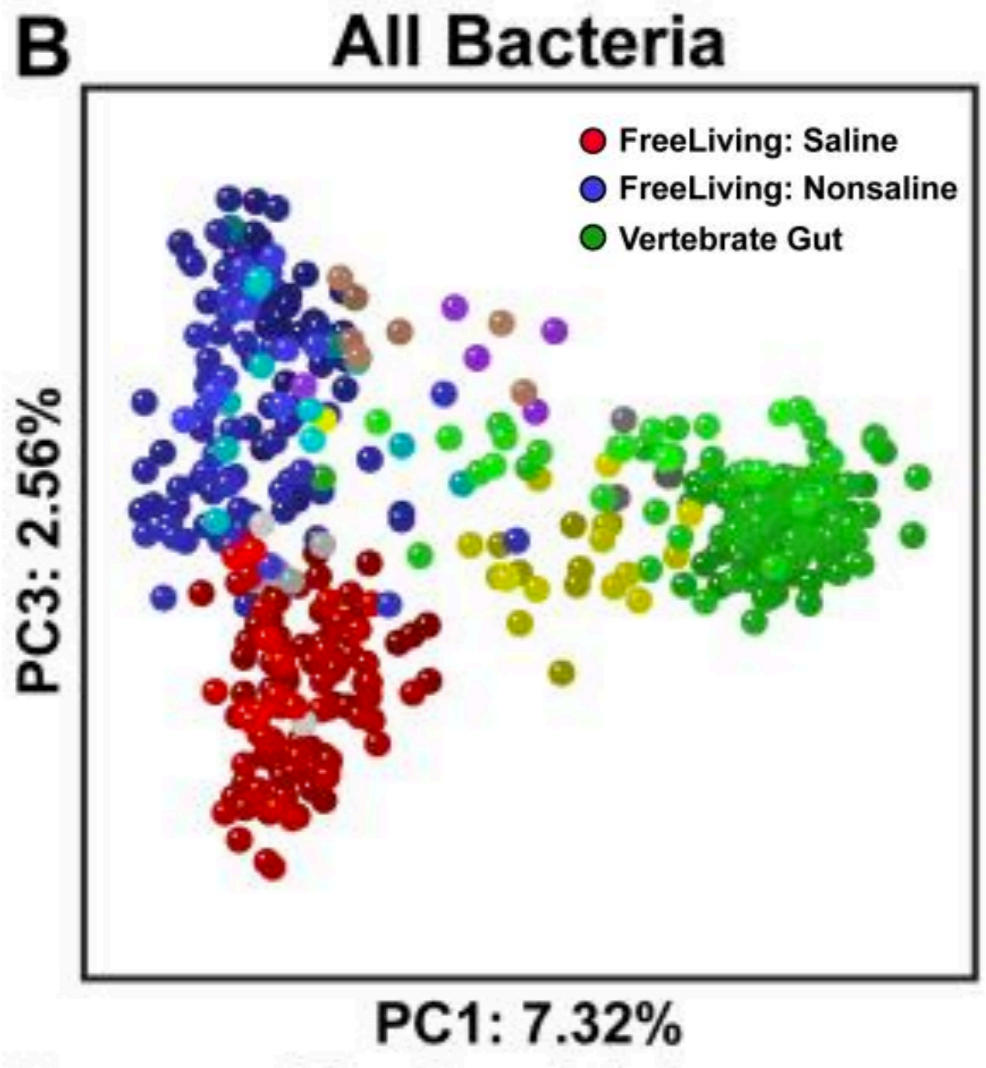


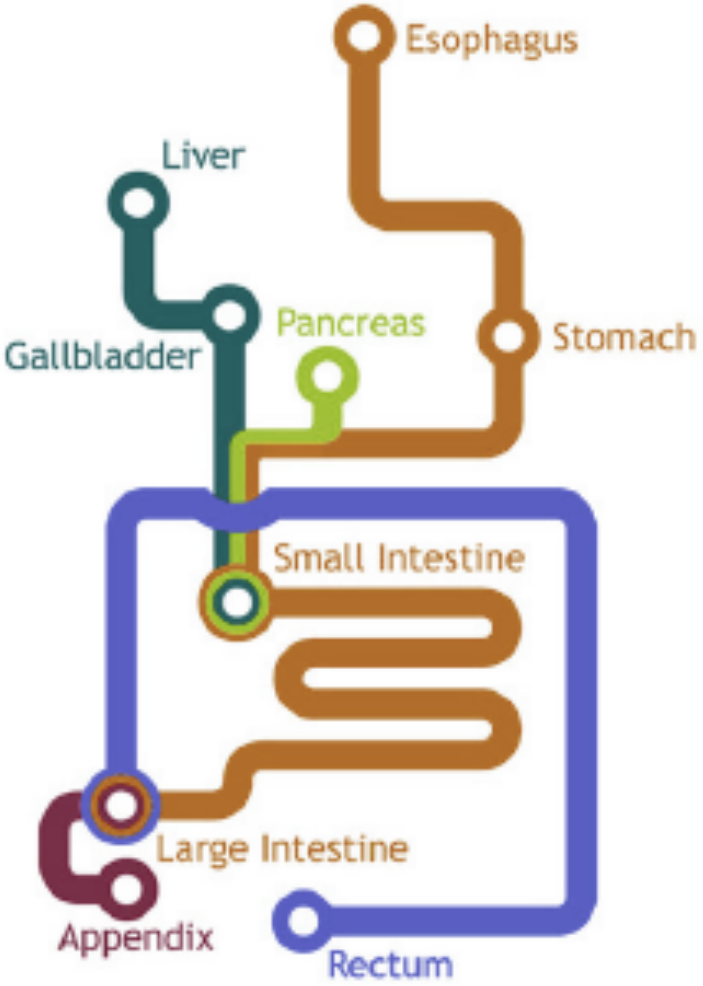


Lozupone and Knight PNAS. 2007 Jul 3;104(27):11436-40

Subway image adapted from http://laughingsquid.com/the-gastrointestinal-system-represented-as-a-subway-map/

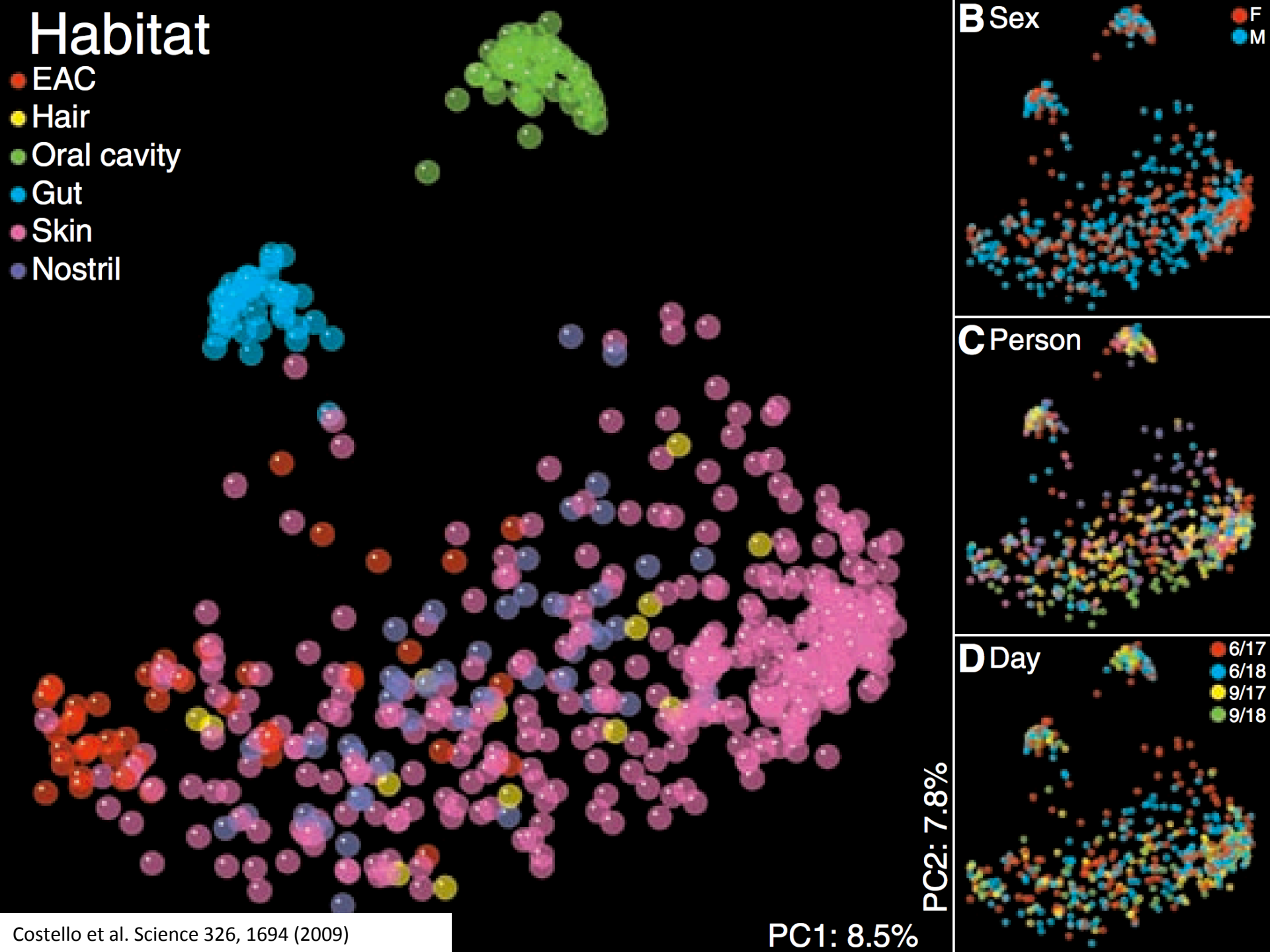
A microbe dominated world





Ley et al Nat Rev Microbiol. Oct 2008; 6(10): 776–788

Subway image adapted from http://laughingsquid.com/the-gastrointestinal-system-represented-as-a-subway-map/

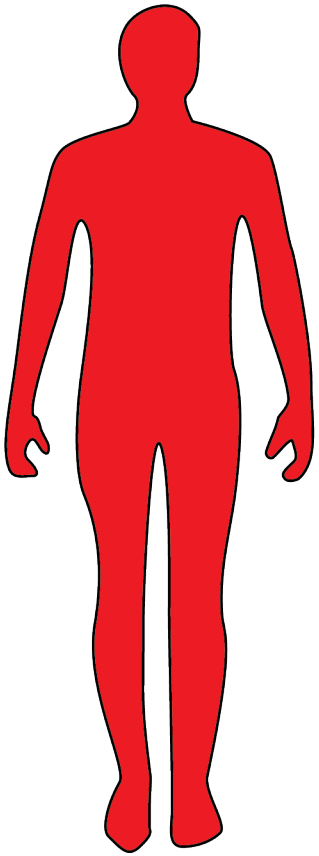


Costello et al. Science 326, 1694 (2009)

How “human” are we, really?

Human

10 trillion human cells
20,000 human genes



Hamady, 2009, PhD thesis

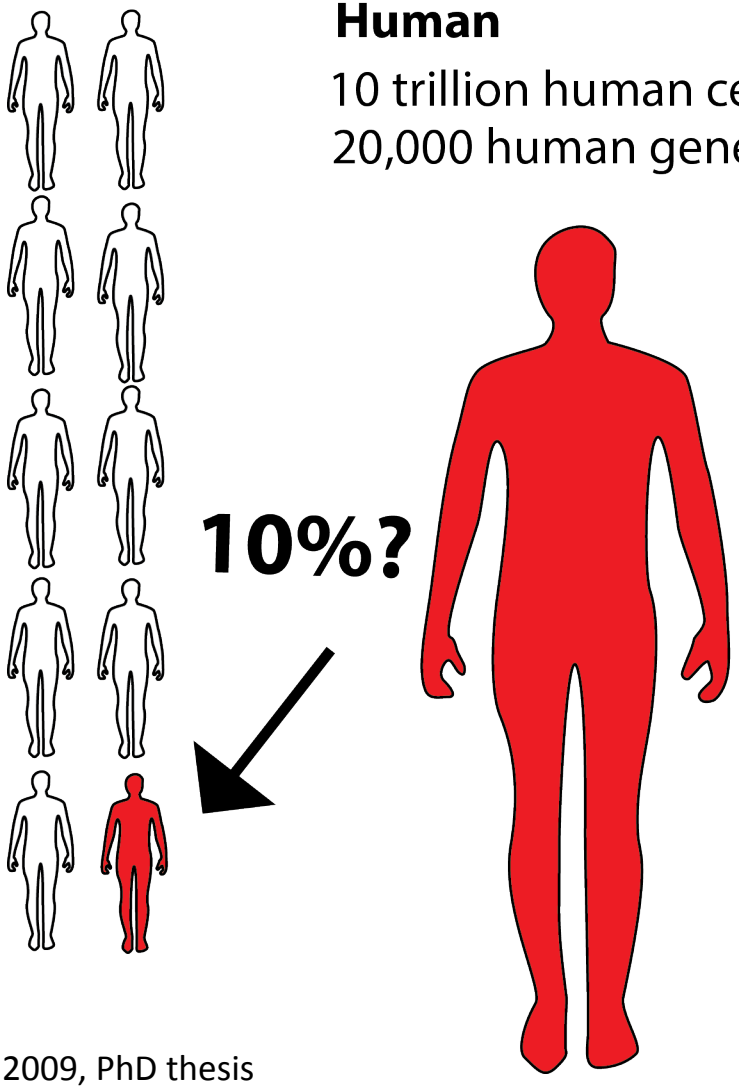
How “human” are we, really?

Microbiota

100 trillion microbial cells

Human

10 trillion human cells
20,000 human genes

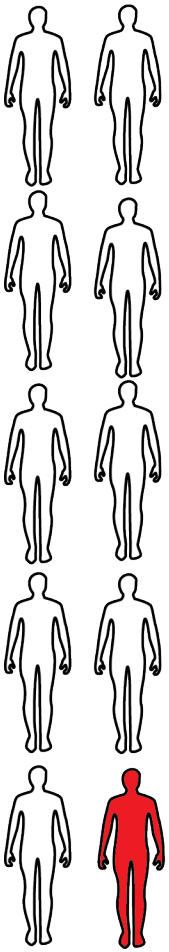


Hamady, 2009, PhD thesis

How “human” are we, really?

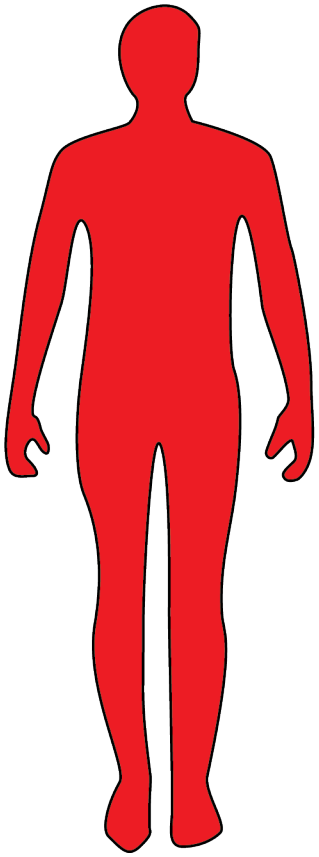
Microbiota

100 trillion microbial cells



Human

10 trillion human cells
20,000 human genes

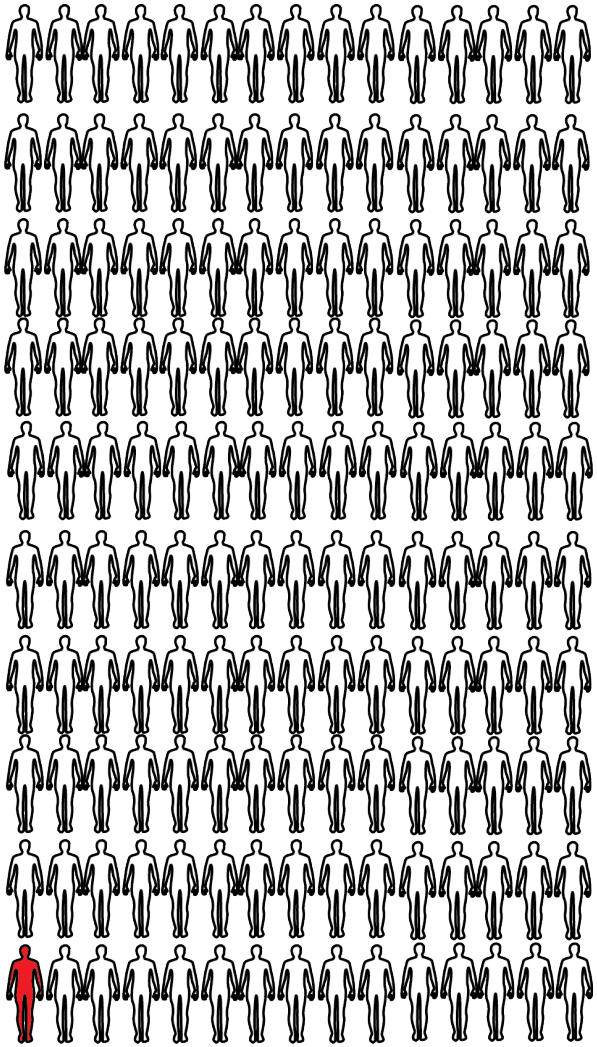


10%?

<1%?

Microbiome

3,000,000 microbial genes

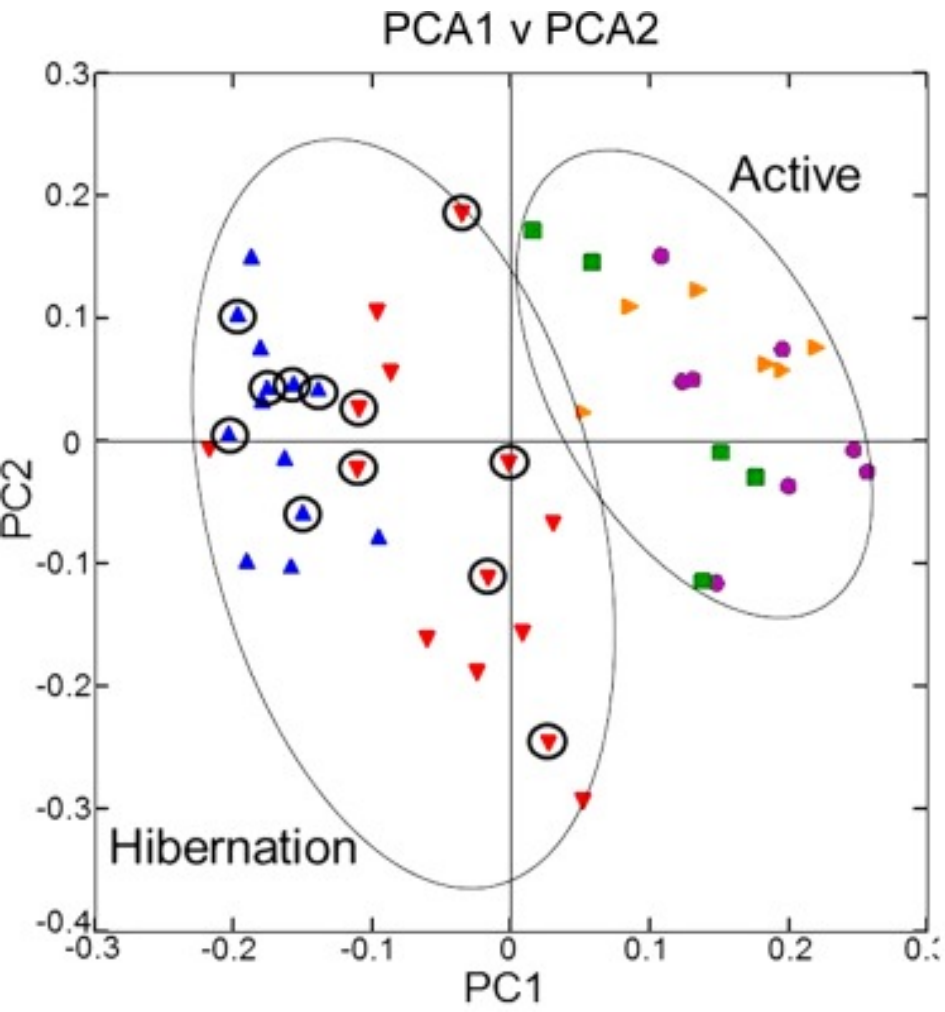


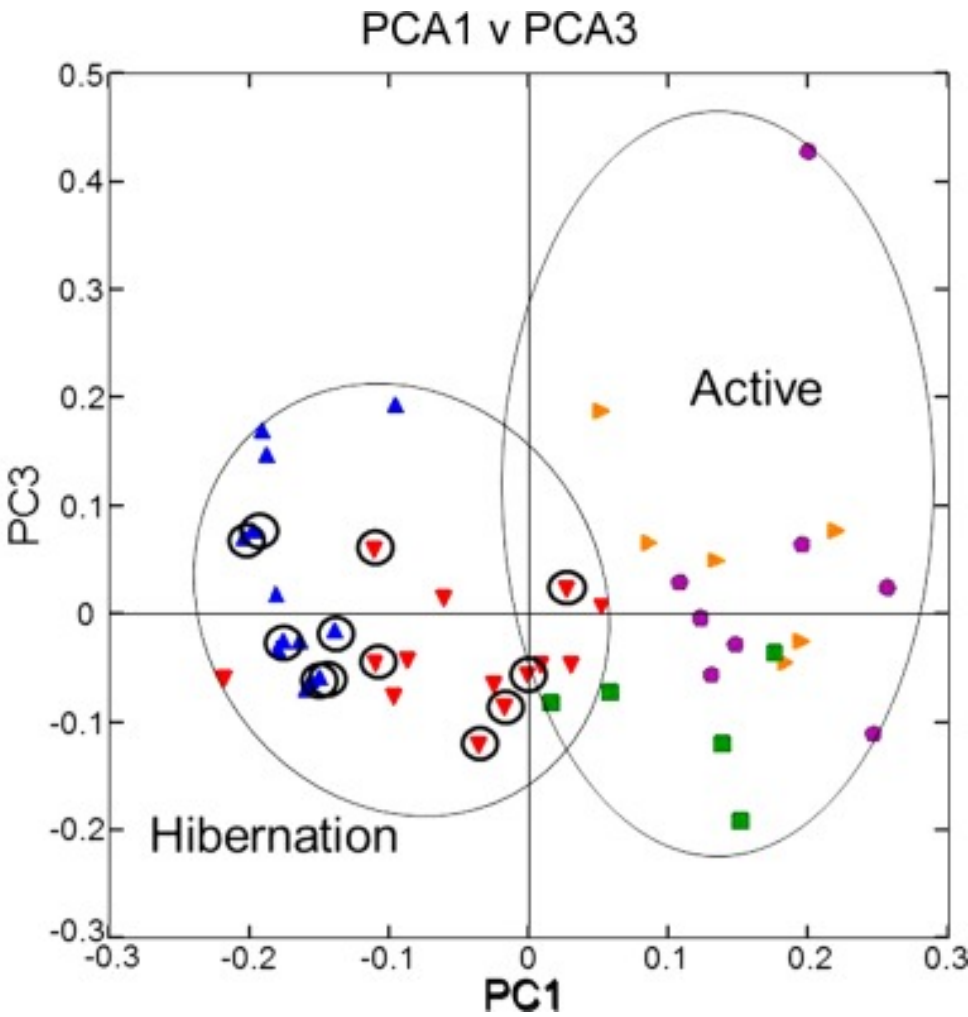
Hamady, 2009, PhD thesis

"Your gut is infallible, it is like the pope of your torso." –Stephen Colbert

Hibernating Ground Squirrels

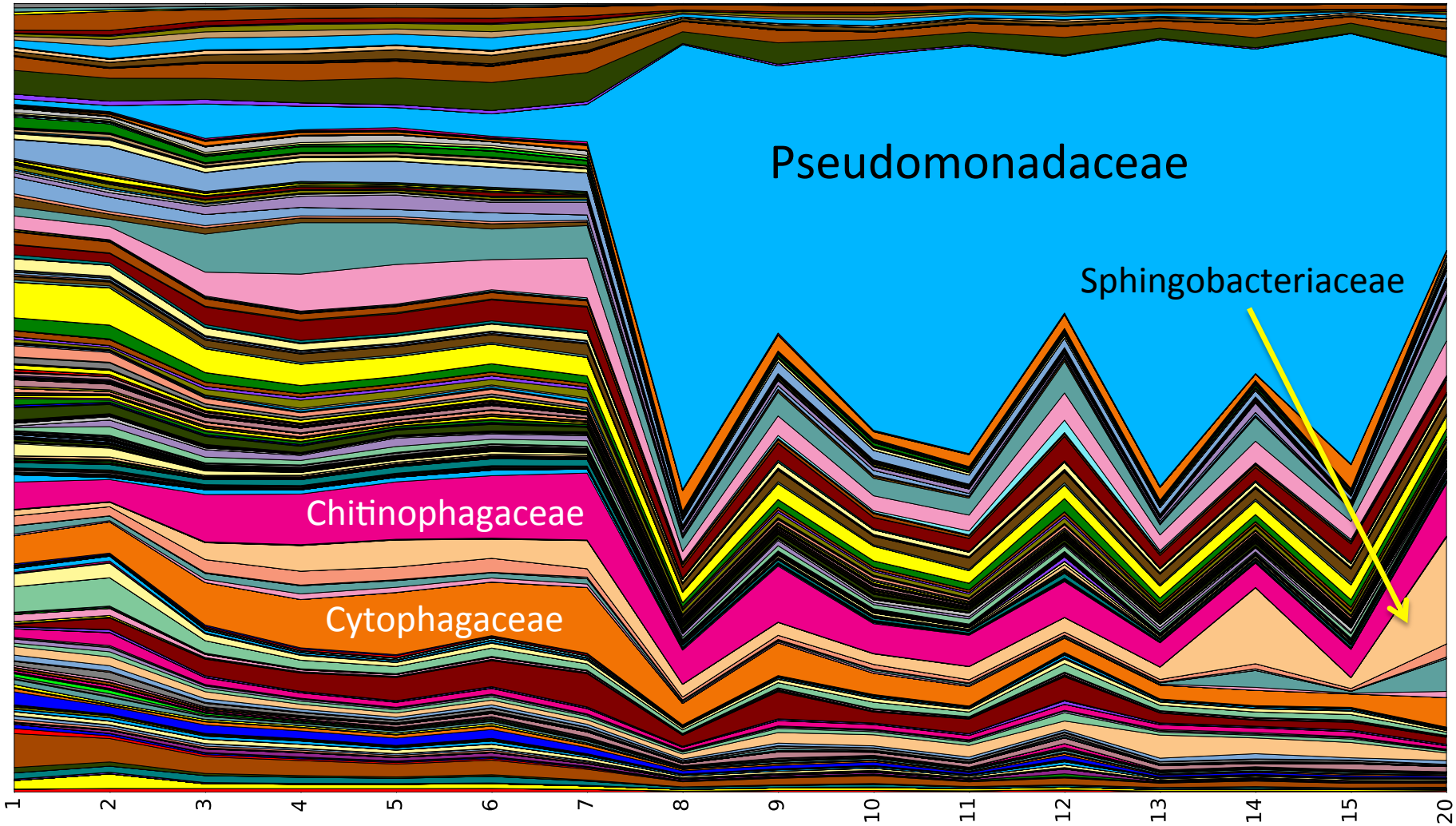
Results of ground squirrel cecal microbiota clustering from Hannah Carey et al, 2012.



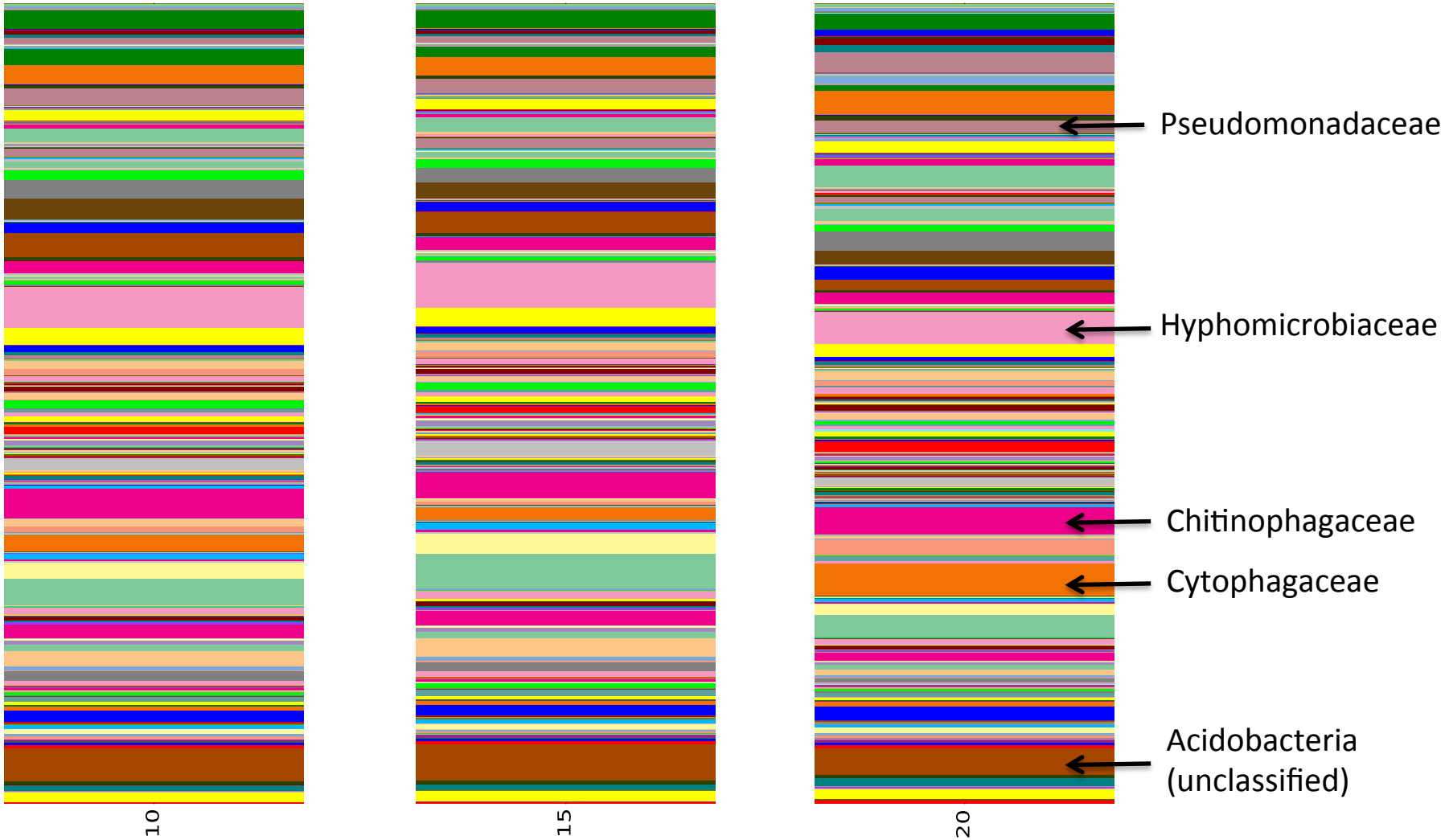


Mothers (yellow triangles) and their pups in Summer (purple circles), Early Winter (red triangles), Late Winter (blue triangles), and Spring (green squares)

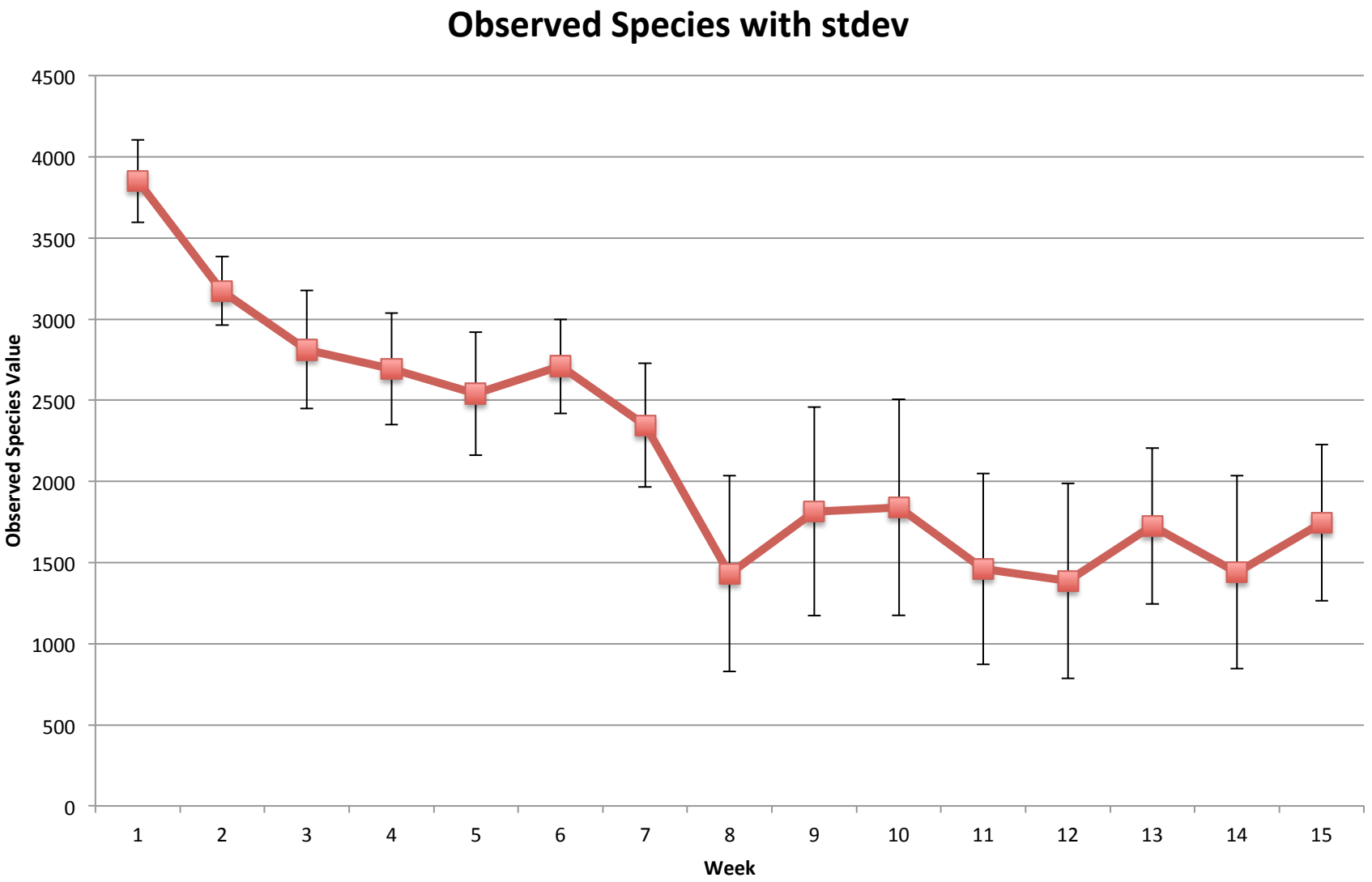
Maize rhizo(endo?)sphere taxa shifts over time



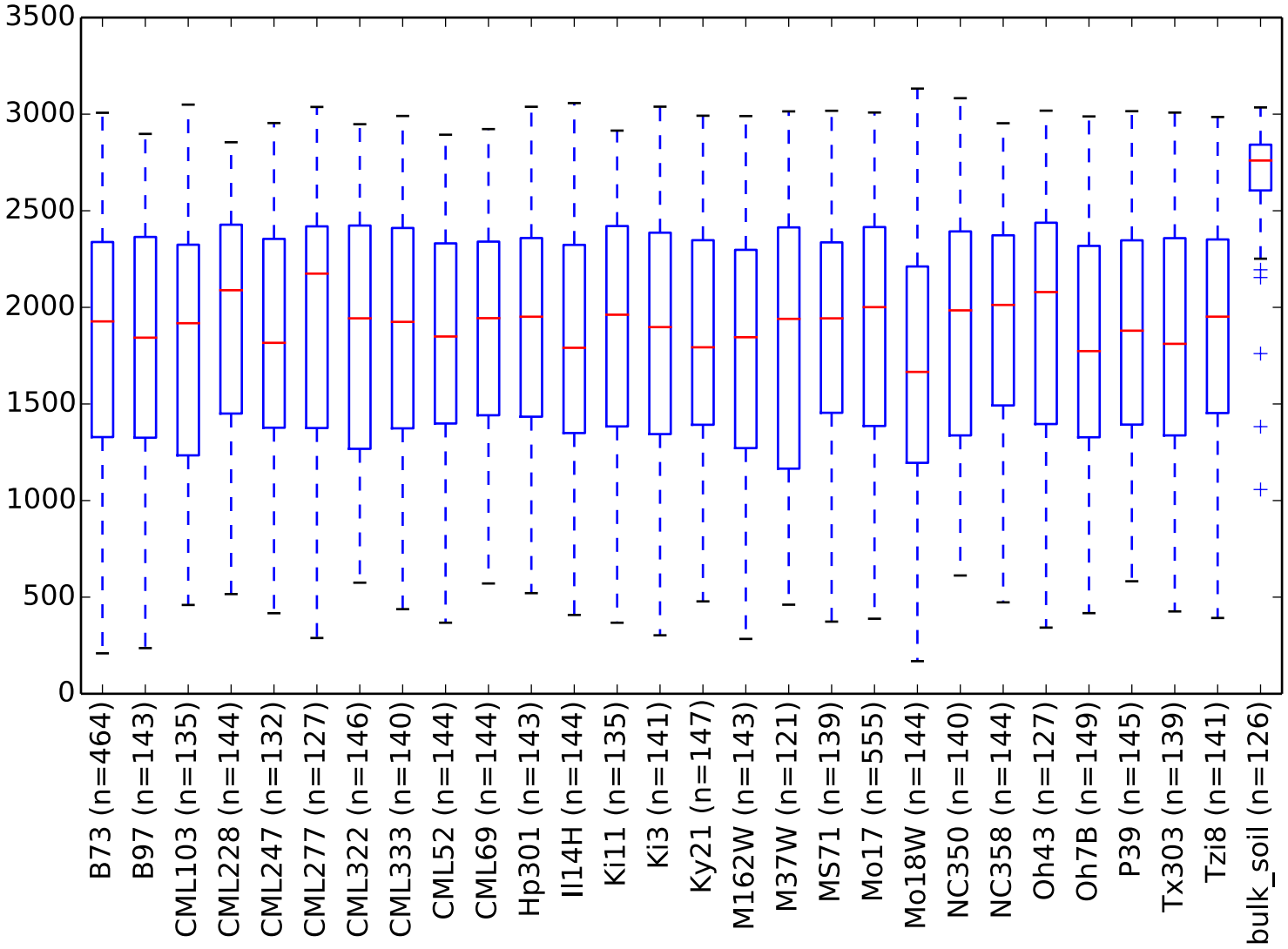
Taxonomy – Bulk Soils by Week



Alpha Diversity, all Maize



No significant difference between maize strains



Emperor Plots

What can we study with QIIME?

Metagenomics versus marker-gene surveys.

What can we study with QIIME?

Metagenomics versus marker-gene surveys.

What can we investigate?

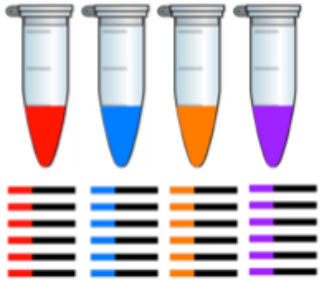
What can we study with QIIME?

Metagenomics versus marker-gene surveys.

What can we investigate?

What are the limitations?

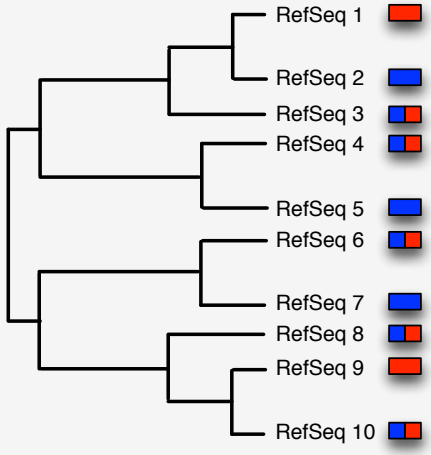
Intro to QIIME: installation and usage



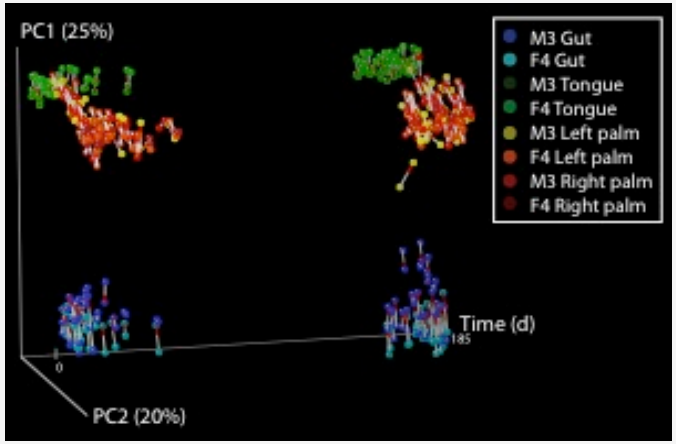
qiime
www.qiime.org

>GCACCTGAGGACAGGCATGAGGAA...
>GCACCTGAGGACAGGGGAGGAGGA...
>TCACATGAACCTAGGCAGGACGAA...
>CTACCGGAGGACAGGCATGAGGAT...
>TCACATGAACCTAGGCAGGAGGAA...
>GCACCTGAGGACACGCAGGACGAC...
>CTACCGGAGGACAGGCAGGAGGAA...
>CTACCGGAGGACACACAGGAGGAA...
>GAACCTTCACATAGGCAGGAGGAT...
>TCACATGAACCTAGGGGCAAGGAA...
>GCACCTGAGGACAGGCAGGAGGAA...

Assign reads to samples



Assign millions of sequences from thousands of samples to reference



Compare samples statistically and visually

A few of the >3600 articles citing QIIME (Google Scholar, 10 Dec 2015)

PERSPECTIVE

Is It Time for a Metagenomic Basis of Therapeutics?

Henry J. Haiser and Peter J. Turnbaugh*

The trillions of microbes associated with the human body are a key part of a comprehensive view of pharmacology. A mechanistic understanding of how the gut microbiota directly and indirectly affects drug metabolism is beginning to emerge.

Science (2012)

Gastrointestinal Microbiome Signatures of Pediatric Patients With Irritable Bowel Syndrome

Delphine M. Saulnier*, ‡, §, Kevin Riehle‖, ¶, Toni–Ann Mistretta*, ‡, Maria–Alejandra Diaz*, ‡, Debasmita Mandal#, Sabeen Raza*, ‡, Erica M. Weidler**, ‡‡, Xiang Qin§§, Cristian Coarfa‖, ¶, Aleksandar Milosavljevic‖, ¶, Joseph F. Petrosino§§, ‖‖, ¶¶, Sarah Highlander§§, ‖‖, Richard Gibbs§§, Susan V. Lynch#, Robert J. Shulman**, ‡‡, James Versalovic*, ‡, ‖, **, ‖‖,

Gastroenterology (2011)

Hydrocarbon-Degrading Bacteria and the Bacterial Community Response in Gulf of Mexico Beach Sands Impacted by the Deepwater Horizon Oil Spill

Joel E. Kostka[1,*], Om Prakash[1], Will A. Overholt[1], Stefan J. Green[1,3], Gina Freyer[1,2], Andy Canion[1], Jonathan Delgardio[1], Nikita Norton[1], Terry C. Hazen[4] and Markus Huettel[1]

Applied and Environmental Microbiology (2011)

Obesity (2012); **20** 4, 738–747. doi:10.1038/oby.2011.111

Responses of Gut Microbiota to Diet Composition and Weight Loss in Lean and Obese Mice

Yann Ravussin[1], Omry Koren[2], Ayme Spor[2], Charles LeDuc[1], Roee Gutman[1,5], Jesse Stombaugh[3], Rob Knight[3,4], Ruth E. Ley[2] and Rudolph L. Leibel[1]

The Microbiology of Malting and Brewing

Nicholas A. Bokulich[a,b] and Charles W. Bamforth[a]

Microbiology and Molecular Biology Reviews (2013)

Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice

Andrew L. Goodman[1], George Kallstrom, Jeremiah J. Faith, Alejandro Reyes, Aimee Moore, Gautam Dantas, and Jeffrey I. Gordon[2]

+ Author Affiliations

PNAS (2011)

Review

The Impact of the Gut Microbiota on Human Health: An Integrative View

Jose C. Clemente[1], Luke K. Ursell[1], Laura Wegener Parfrey[1], Rob Knight[1, 2],

Cell (2012)

Microbial diversity associated with four functional groups of benthic reef algae and the reef-building coral *Montastraea annularis*

Katie L. Barott[1,*], Beltran Rodriguez-Brito[1], Jan Janouškovec[2], Kristen L. Marhaver[3], Jennifer E. Smith[3], Patrick Keeling[2], Forest L. Rohwer[1]

Article first published online: 27 JAN 2011
DOI: 10.1111/j.1462-2920.2010.02419.x

Issue

Environmental Microbiology
Volume 13, Issue 5, pages 1192–1204, May 2011

PLoS one

Fecal Microbiota in Premature Infants Prior to Necrotizing Enterocolitis

Volker Mai[1*], Christopher Michael Young[2], Maria Ukhanova[1], Xiaoyu Wang[1], Yijun Sun[1], George Casella[4], Douglas Theriaque[5], Nan Li[2], Renu Sharma[3], Mark Hudak[3], Josef Neu[2]

Current Concepts of the Intestinal Microbiota and the Pathogenesis of Infection

Leslie H. Wardwell · Curtis Huttenhower · Wendy S. Garrett

Current Infectious Disease Reports (2011)

Composition of Human Skin Microbiota Affects Attractiveness to Malaria Mosquitoes

Niels O. Verhulst[1*], Yu Tong Qiu[1], Hans Beijleveld[1,2], Chris Maliepaard[3], Dan Knights[4], Stefan Schulz[5], Donna Berg-Lyons[6], Christian L. Lauber[6], Willem Verduijn[7], Geert W. Haasnoot[7], Roland Mumm[8,9], Harro J. Bouwmeester[10], Frans H. J. Claas[7], Marcel Dicke[1], Joop J. A. van Loon[1], Willem Takken[1], Rob Knight[4,11], Renate C. Smallegange[1]

PLoS One (2011)

qiime
Quantitative Insights Into Microbial Ecology

www.qiime.org

Native installation

OS X or Linux (laptops through 153,408-core compute cluster*)

Virtual machines

VirtualBox (local) or

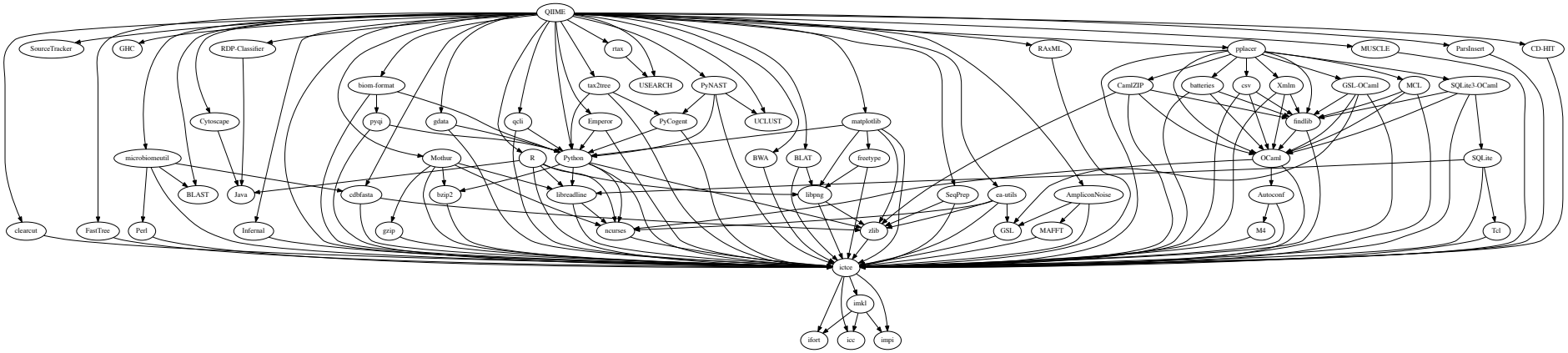
Cloud environments

(e.g., Amazon Web Services, iPlant)

NEW: Now available as a BaseSpace app!

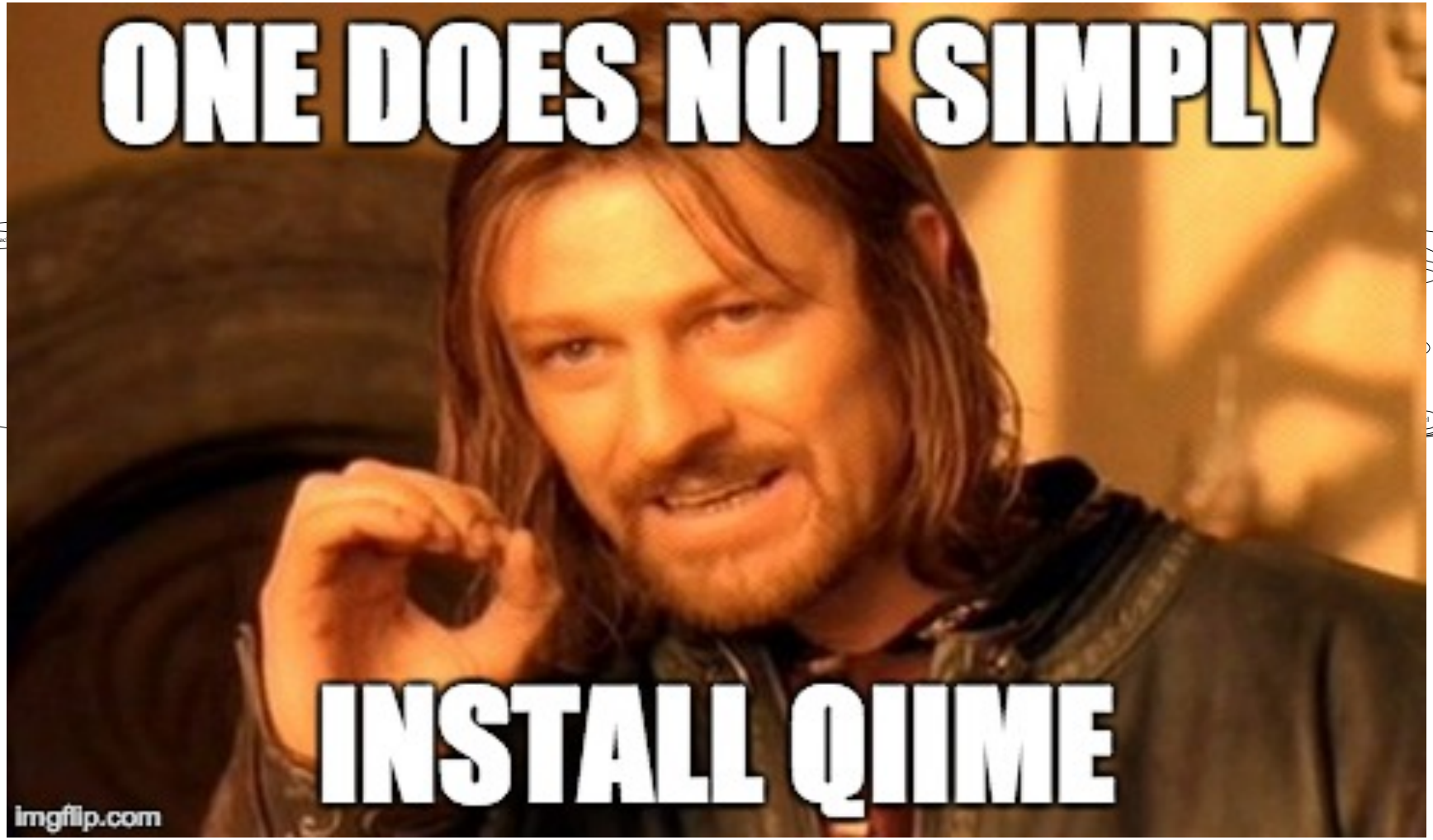
*Hopper (http://i.top500.org/system/176952)

QIIME software dependencies



http://users.ugent.be/~kehoste/QIIME.pdf

QIIME software dependencies



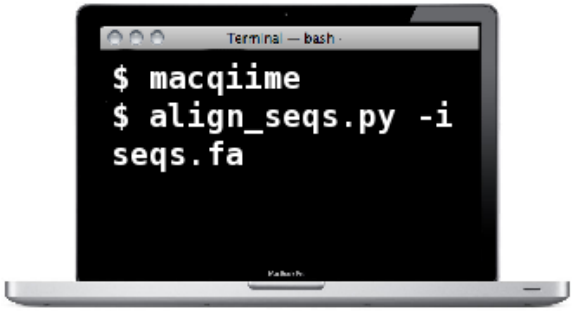
https://imgflip.com/memegenerator/One-Does-Not-Simply
http://users.ugent.be/~kehoste/QIIME.pdf

Native installation

- Installed directly on operating system / computer
- Pros
  - Best performance
  - Better use of resources (e.g., memory)
- Cons
  - Can be difficult to install

Native installation: MacQIIME

- Mac users: easiest way is MacQIIME
  - Excellent tool maintained by Jeff Werner's lab
  - Easiest way to get a (mostly) complete QIIME installation on a Mac
  - Can be difficult to customize/change the installation (e.g., updating specific pieces of MacQIIME)



More info: http://www.wernerlab.org/software/macqiime

Native installation: pip

- Mac/Linux users:

```
pip install qiime
```

- Basic QIIME installation
  - Supports basic upstream analyses and most downstream analyses
  - Install other dependencies as needed

Native installation: qiime-deploy

- Linux users (especially Ubuntu): easiest way to get a **full** QIIME installation
- Commonly used by system administrators to install QIIME in a cluster environment

More info: https://github.com/qiime/qiime-deploy

Virtual machines

- A “guest” operating system running within a “host” operating system
- A software implementation of a computer, that operates like a physical computer.
- A developer can create a virtual machine *image* which contains their tools pre-installed. Users can then *instantiate* that image to work with those tools.

Browse this page: http://en.wikipedia.org/wiki/Virtual_machine

Benefits that virtual machines offer bioinformatics

- Reproducibility: can publish protocols with a virtual machine instance id.
- Updates are burden of developer, not user.
- Coupled with cloud computing, it's the perfect model for users with sporadic compute needs.

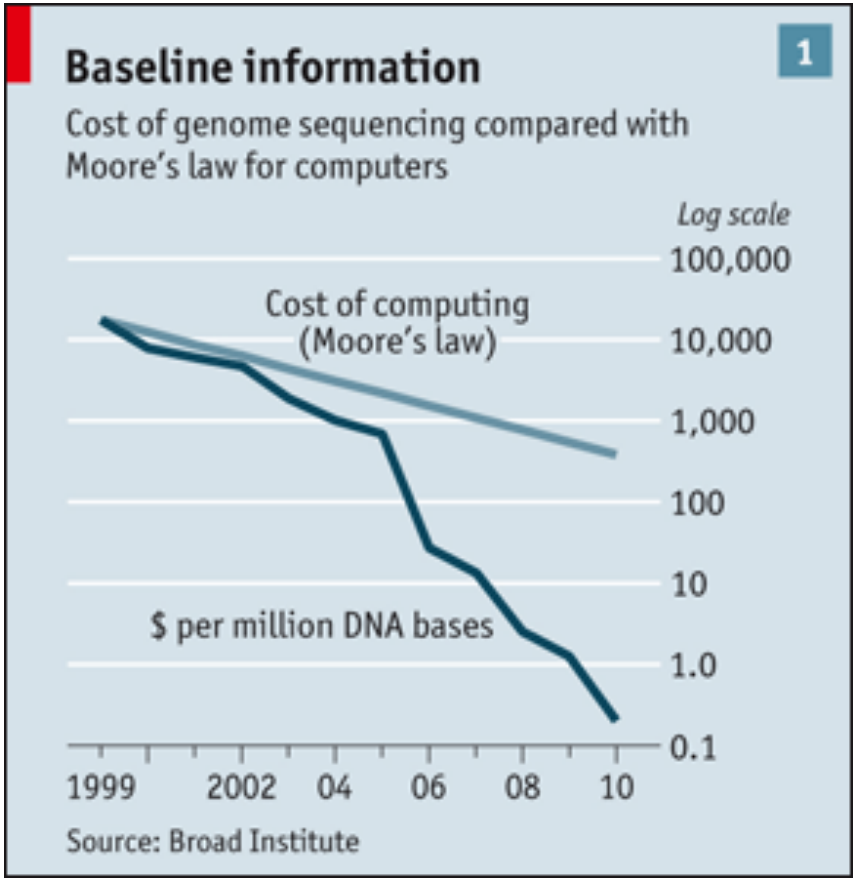
QIIME virtual machines

- VirtualBox (Ubuntu)
  - Easiest way to get a **full** QIIME installation on pretty much **any** platform
    - Especially for Windows users
    - Great for trying out QIIME
  - Slower than a native installation
- Cloud-based installations
  - Amazon Web Services (EC2)
  - iPlant Atmosphere
  - ANL Magellan

Isn’t my laptop powerful enough?

http://thinnestlaptopwallpaper.blogspot.com/2012/10/worlds-smallest-laptop.html

Why is parallel computing important in bioinformatics?



Why is parallel computing important in bioinformatics?

Platform	Sanger	454 (Titanium)	Illumina Genome Analyzer II	Illumina HiSeq2000	Illumina MiSeq
Read Length (bases)	~1000	~400	150 (single end)	100 (single end)	150 (single end); 250 soon
Number of reads	96 or 384	~1,000,000	~100,000,000	~1,600,000,000	~10,000,000
Maximum number of samples per run	n/a	1000	12,000 (barcode-limited)	24,000 (barcode-limited)	2500 (barcode-limited)
Sequences per $1 (sequencing costs only)	0.44	100	5000	200,000	12,500

Cluster computing

- Many computers connected to one another to serve as a larger compute resource.
- Compute-intensive jobs can be split over many systems and run in parallel.
- Similar to desktop compute hardware, but different casing, no (or only few) displays/ keyboards directly connected.
- Owned and maintained “in-house”.

Maintaining hardware is expensive

- Temperature (redundant cooling systems)
- Redundant network connections
- Hardware maintenance (e.g., replacing hard drives)
- Non-water fire suppression
- Backup power
- System administrator ($$)

Cloud computing (IaaS model)

- Implemented on a cluster (or grid), but compute power is rented as a service to support arbitrary applications.

Pay-as-you-go compute power

- Public clouds (e.g., Amazon) rent compute resources
- Log in, boot virtual machine image, run analyses, and terminate instance.
- Cheaper for many tasks than buying, maintaining, and supporting a compute cluster.

Interacting with the Amazon Cloud

- Boot virtual machine image via web interface (or a third-party tool like StarCluster).
- Log in and work via terminal (or via web interface with IPython Notebook)
- Move data back and forth via sftp/scp or a graphical sftp client (e.g., Cyberduck [free/cross-platform])

For information on costs, see http://www.ec2instances.info

Learning QIIME

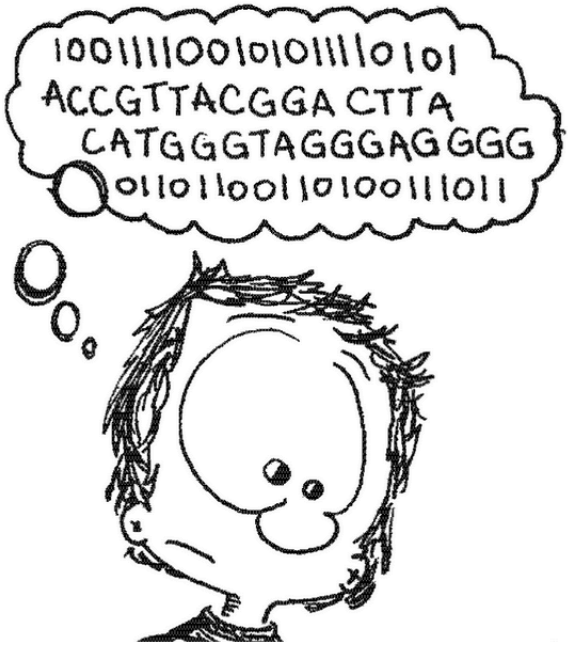
- Start with the tutorials
  - http://qiime.org/tutorials
- Call any script with –h to get help or see the script usage pages
  - http://scripts.qiime.org
- Ask questions on the QIIME Forum
  - http://help.qiime.org
  - http://forum.qiime.org
- Report bugs on the issue tracker
  - http://github.com/biocore/qiime/issues

An Introduction To Applied Bioinformatics

Interactive lessons in bioinformatics.

View the Project on GitHub
gregcaporaso/An-Introduction-To-Applied-Bioinformatics

Download ZIP File	Download TAR Ball	View On GitHub



Bioinformatics, as I see it, is the application of the tools of computer science (things like programming languages, algorithms, and databases) to address biological problems (for example, inferring the evolutionary relationship between a group of organisms based on fragments of their genomes, or understanding if or how the community of microorganisms that live in my gut changes if I modify my diet). Bioinformatics is a rapidly growing field, largely in response to the vast increase in the quantity of data that biologists now grapple with. Students from varied disciplines (e.g., biology, computer science, statistics, and biochemistry) and stages of their educational careers (undergraduate, graduate, or postdoctoral) are becoming interested in bioinformatics.

I teach bioinformatics at the undergraduate and graduate levels at Northern Arizona University. This repository contains some of the materials that I've developed in these courses, and represents an initial attempt to organize these materials in a standalone way. In some cases, I'm just linking out to other materials for now.

http://readIAB.org

ALFRED P. SLOAN FOUNDATION

coursera

Catalog | Search catalog

Institutions | Log In | Sign Up

Gut Check: Exploring Your Microbiome

by University of Colorado Boulder & University of Colorado System

Course Info

UNIVERSITY OF COLORADO BOULDER & UNIVERSITY OF COLORADO SYSTEM

Gut Check: Exploring Your Microbiome

About this Course

Imagine if there were an organ in your body that weighed as much as your brain, that affected your health, your weight, and even your behavior. Wouldn't you want to know more about it? There is such an organ — the collection of microbes in and on your body, your human microbiome.

Subtitles available in **English**

Log in to enroll in this course

Log In

Instructors

Professor Rob Knight
Professor
Howard Hughes Medical Institute, and Department of Chemistry & Biochemistry and Computer Science, and Biofrontiers Institute

Dr. Jessica Metcalf
Senior Research Associate
BioFrontiers Institute

Dr. Katherine Amato
Postdoctoral Research Associate
Department of Anthropology, BioFrontiers Institute

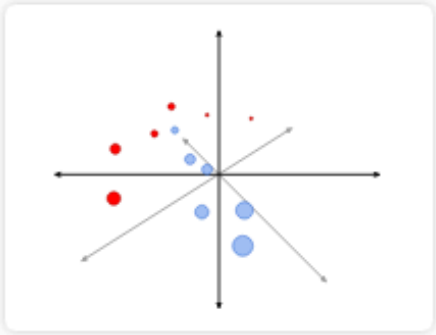
https://www.coursera.org/learn/microbiome

Welcome to the GUide to STasitical Analysis in Microbial Ecology (GUSTA ME)!

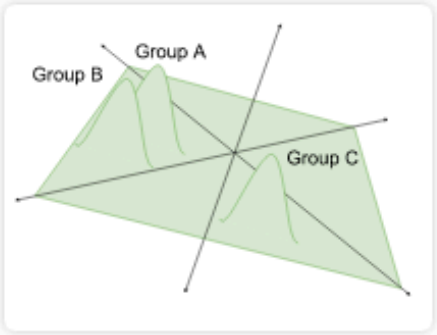
Where would you like to start?

You may start exploring the guide by browsing topics in the sidebar, using the search box at the top right of this page to find a particular method, or by clicking on one of the entry points below...

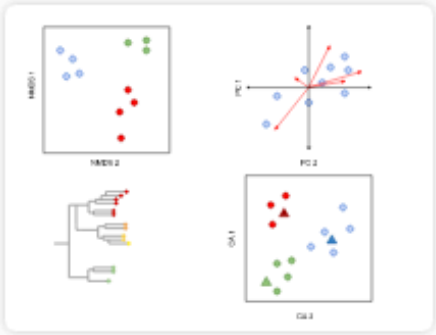
Explore data...



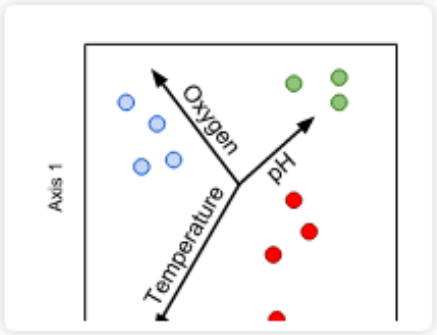
Test a hypothesis...



Browse visualisations...

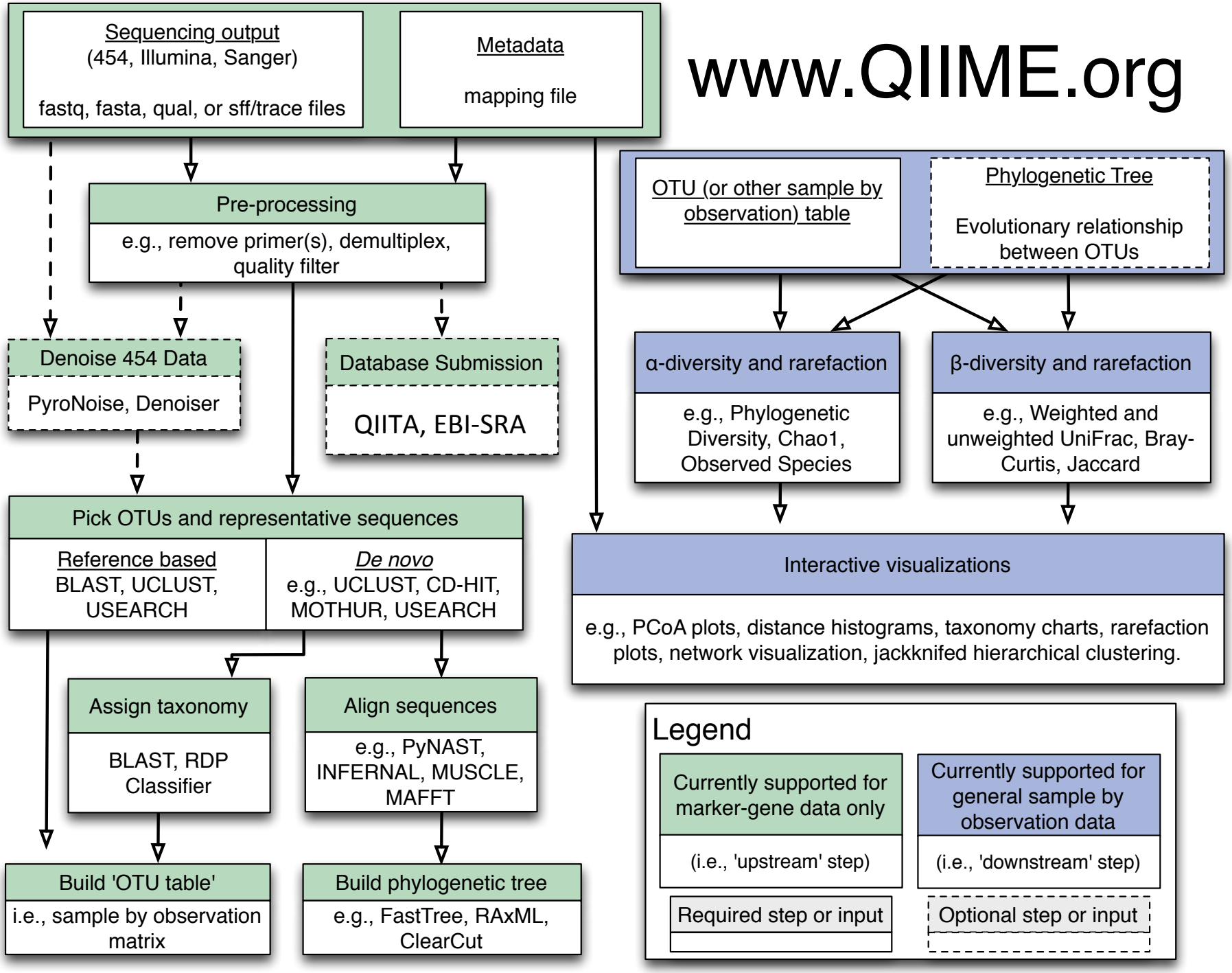


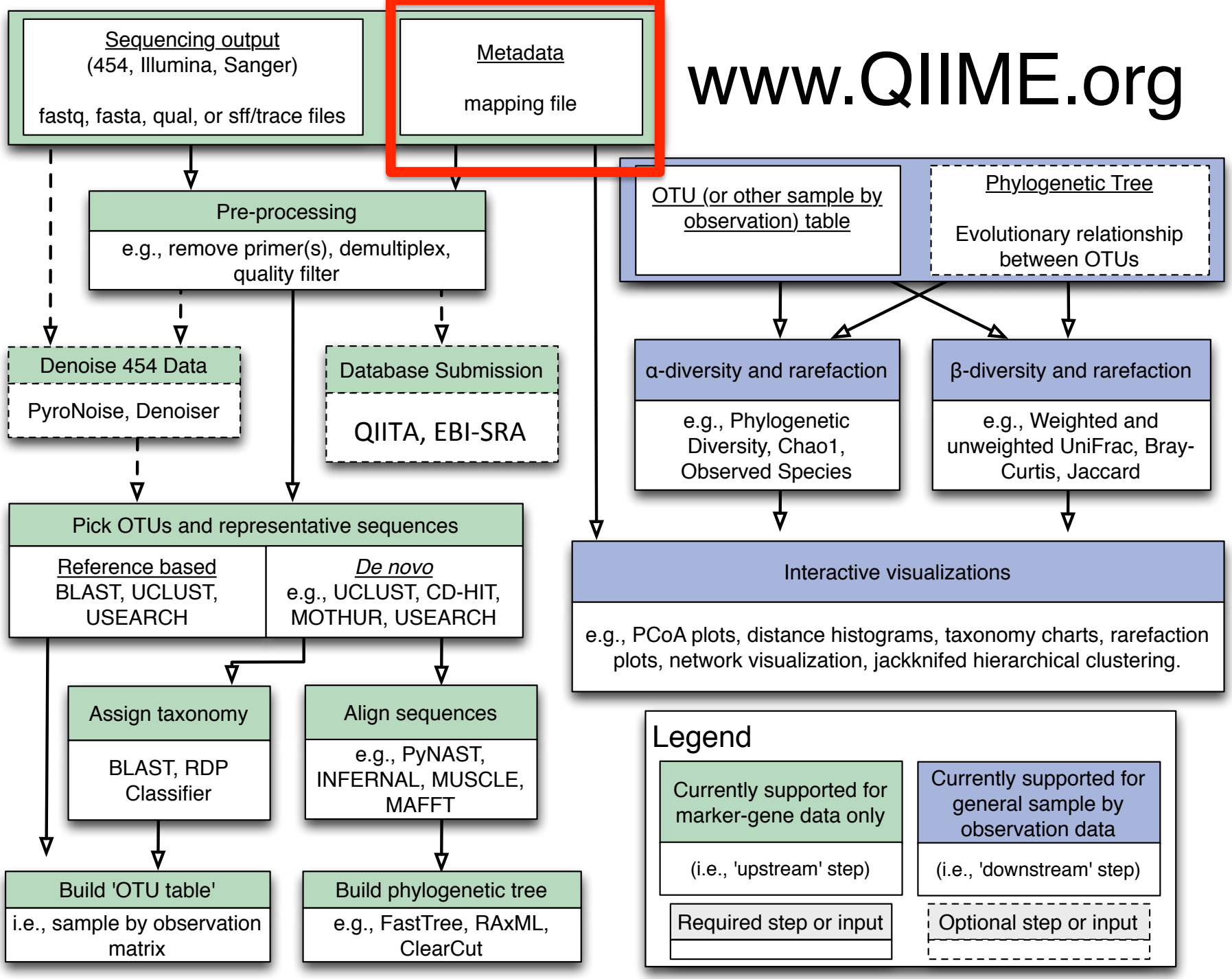
Explore environmental influence...



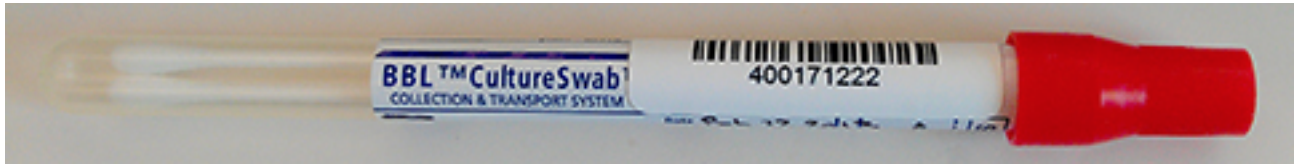
http://mb3is.megx.net/gustame

Preprocessing: metadata and demultiplexing


Sequencing output
(454, Illumina, Sanger)
fastq, fasta, qual, or sff/trace files
Metadata
mapping file
www.QIIME.org
Pre-processing
e.g., remove primer(s), demultiplex, quality filter
OTU (or other sample by observation) table
Phylogenetic Tree
Evolutionary relationship between OTUs
Denoise 454 Data
PyroNoise, Denoiser
Database Submission
QIITA, EBI-SRA
α-diversity and rarefaction
e.g., Phylogenetic Diversity, Chao1, Observed Species
β-diversity and rarefaction
e.g., Weighted and unweighted UniFrac, Bray-Curtis, Jaccard
Pick OTUs and representative sequences
Reference based
BLAST, UCLUST, USEARCH
De novo
e.g., UCLUST, CD-HIT, MOTHUR, USEARCH
Interactive visualizations
e.g., PCoA plots, distance histograms, taxonomy charts, rarefaction plots, network visualization, jackknifed hierarchical clustering.
Assign taxonomy
BLAST, RDP Classifier
Align sequences
e.g., PyNAST, INFERNAL, MUSCLE, MAFFT
Build 'OTU table'
i.e., sample by observation matrix
Build phylogenetic tree
e.g., FastTree, RAxML, ClearCut
Legend
Currently supported for marker-gene data only
(i.e., 'upstream' step)
Currently supported for general sample by observation data
(i.e., 'downstream' step)
Required step or input
Optional step or input


Sequencing output
(454, Illumina, Sanger)
fastq, fasta, qual, or sff/trace files
Metadata
mapping file
www.QIIME.org
Pre-processing
e.g., remove primer(s), demultiplex, quality filter
OTU (or other sample by observation) table
Phylogenetic Tree
Evolutionary relationship between OTUs
Denoise 454 Data
PyroNoise, Denoiser
Database Submission
QIITA, EBI-SRA
α-diversity and rarefaction
e.g., Phylogenetic Diversity, Chao1, Observed Species
β-diversity and rarefaction
e.g., Weighted and unweighted UniFrac, Bray-Curtis, Jaccard
Pick OTUs and representative sequences
Reference based
BLAST, UCLUST, USEARCH
De novo
e.g., UCLUST, CD-HIT, MOTHUR, USEARCH
Interactive visualizations
e.g., PCoA plots, distance histograms, taxonomy charts, rarefaction plots, network visualization, jackknifed hierarchical clustering.
Assign taxonomy
BLAST, RDP Classifier
Align sequences
e.g., PyNAST, INFERNAL, MUSCLE, MAFFT
Build 'OTU table'
i.e., sample by observation matrix
Build phylogenetic tree
e.g., FastTree, RAxML, ClearCut
Legend
Currently supported for marker-gene data only
(i.e., 'upstream' step)
Currently supported for general sample by observation data
(i.e., 'downstream' step)
Required step or input
Optional step or input

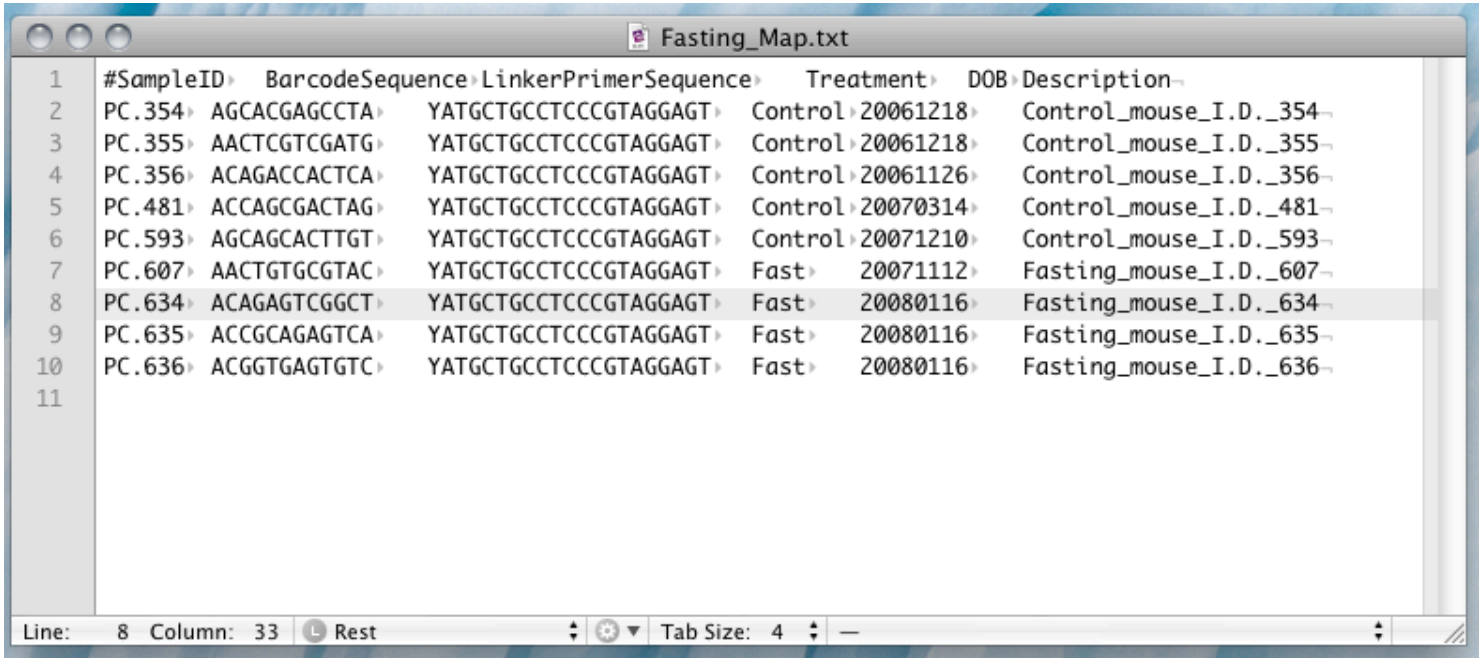
Metadata relates samples to variables



SampleID	Sex	SampleType	Plotting your doom?
400171222	Male	feces	Naturally

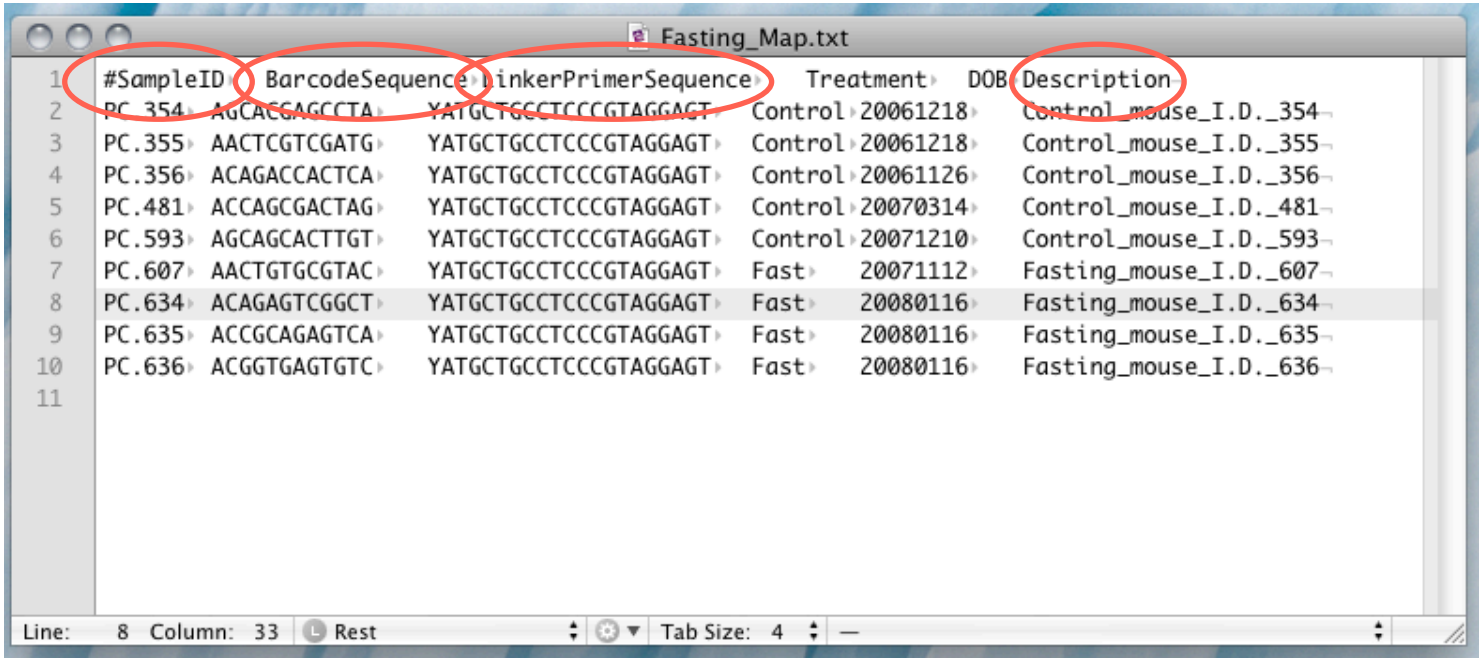
Cat image: http://www.heatherbrookes.org/hb-blog/?p=2341

Mapping file

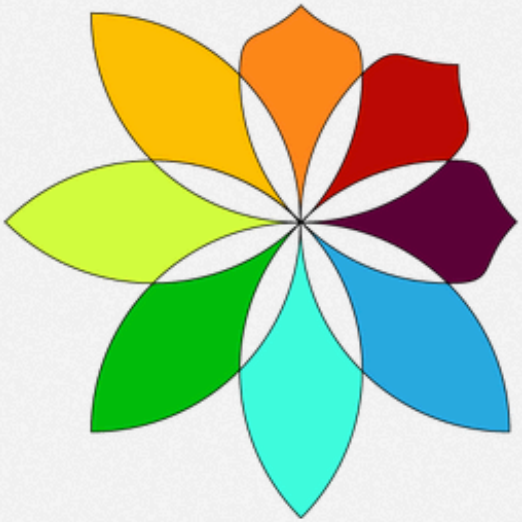


Mapping file: always run

`validate_mapping_file.py`



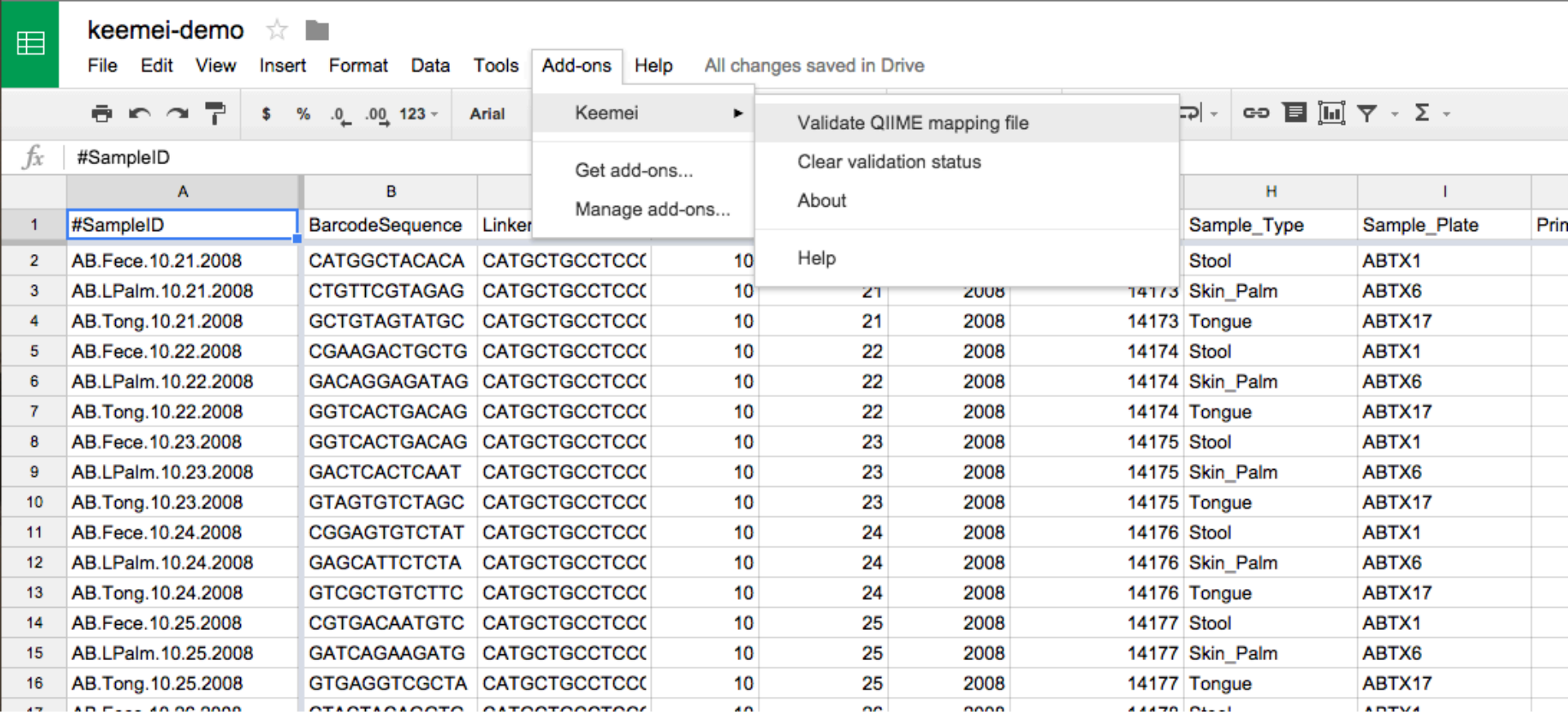
= required field



Keemei: Validate bioinformatics metadata in Google Sheets

Keemei (canonically pronounced *key may*) is an open source Google Sheets add-on for validating bioinformatics metadata, including QIIME mapping files.

keemei.qiime.org



keemei.qiime.org

keemei-demo

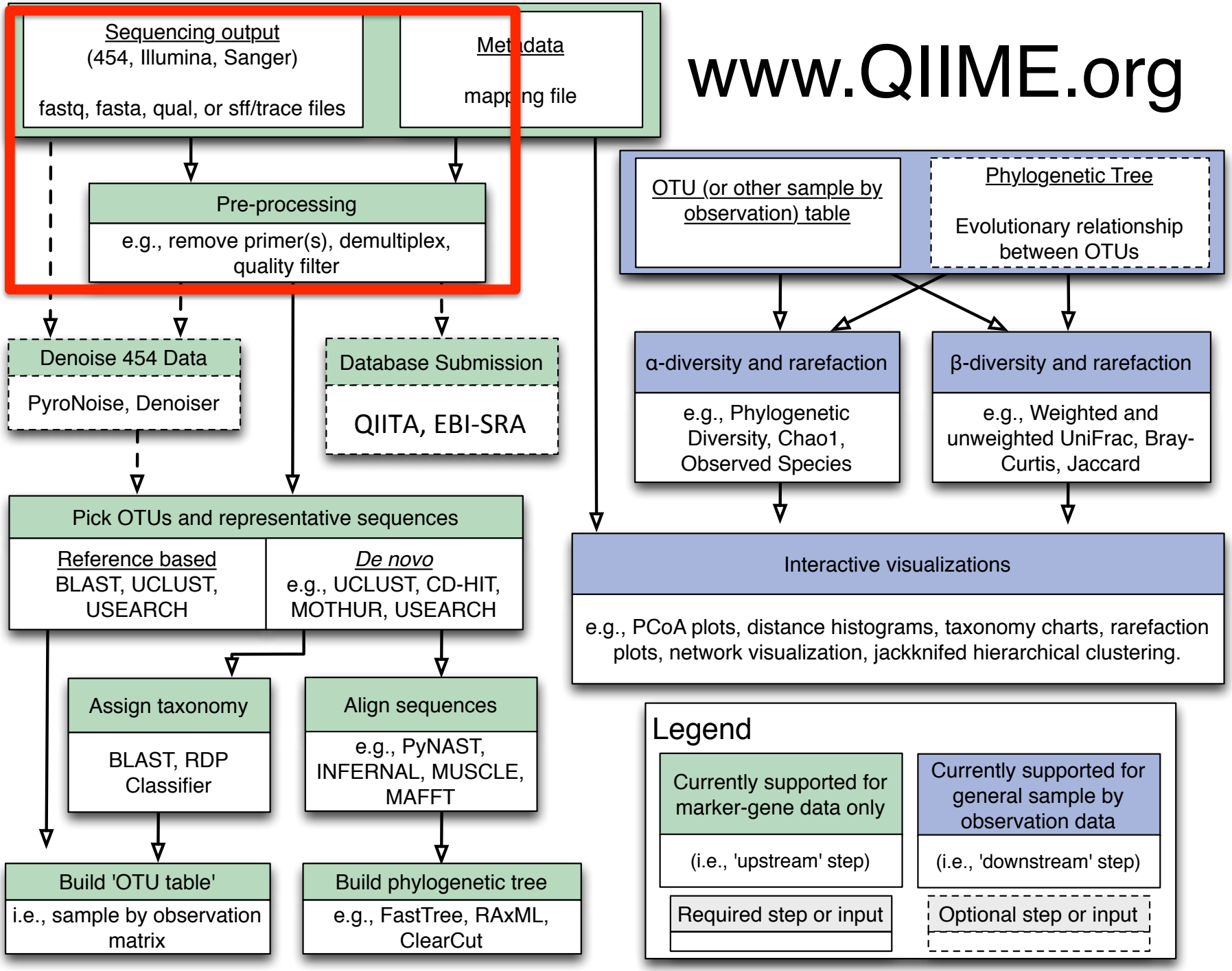
File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

#SampleID

	A	B	C	D	E	F	G	H	I
1	#SampleID	BarcodeSequence	LinkerPrimerSequer	Month	Day	Year	days_since_epoch	Sample_Type	Sample_Plate
2	AB.Fece.10.21.2008	CATGGCTACACA	CATGCTGCCTCCC	10	21	2008	14173	Stool	ABTX1
3	AB.LPalm.10.21.2008	CTGTTCGTAGAG	CATGCTGCCTCCC	10	21	2008	14173	Skin_Palm	ABTX6
4	AB.Tong.10.21.2008	GCTGTAGTATGC	CATGCTGCCTCCC	10	21	2008	14173	Tongue	ABTX17
5	AB.Fece.10.22.2008	CGAAGACTGCTG	CATGCTGCCTCCC	10	22	2008	14174	Stool	ABTX1
6	AB.LPalm.10.22.2008	GACAGGAGATAG	CATGCTGCCTCCC	10	22	2008	14174	Skin_Palm	ABTX6
7	AB.Tong.10.22.2008	GGTCACTGACAG			22	2008	14174	Tongue	ABTX17
8	AB.Fece.10.23.2008	GGTCACTGACAG			23	2008	14175	Stool	ABTX1
9	AB.LPalm.10.23.2008	GACTCACTCAAT			23	2008	14175	Skin_Palm	ABTX6
10	AB.Tong.10.23.2008	GTAGTGTCTAGC			23	2008	14175	Tongue	ABTX17
11	AB.Fece.10.24.2008	CGGAGTGTCTAT	CATGCTGCCTCCC	10	24	2008	14176	Stool	ABTX1
12	AB.LPalm.10.24.2008	GAGCATTCTCTA	CATGCTGCCTCCC	10	24	2008	14176	Skin_Palm	ABTX6
13	AB.Tong.10.24.2008	GTCGCTGTCTTC	CATGCTGCCTCCC	10	24	2008	14176	Tongue	ABTX17
14	AB.Fece.10.25.2008	CGTGACAATGTC	CATGCTGCCTCCC	10	25	2008	14177	Stool	ABTX1
15	AB.LPalm.10.25.2008	GATCAGAAGATG	CATGCTGCCTCCC	10	25	2008	14177	Skin_Palm	ABTX6

Error: Duplicate barcode sequence.
Duplicates in B7, B8

keemei.qiime.org


Sequencing output
(454, Illumina, Sanger)
fastq, fasta, qual, or sff/trace files
Metadata
mapping file
www.QIIME.org
Pre-processing
e.g., remove primer(s), demultiplex, quality filter
OTU (or other sample by observation) table
Phylogenetic Tree
Evolutionary relationship between OTUs
Denoise 454 Data
PyroNoise, Denoiser
Database Submission
QIITA, EBI-SRA
α-diversity and rarefaction
e.g., Phylogenetic Diversity, Chao1, Observed Species
β-diversity and rarefaction
e.g., Weighted and unweighted UniFrac, Bray-Curtis, Jaccard
Pick OTUs and representative sequences
Reference based
BLAST, UCLUST, USEARCH
De novo
e.g., UCLUST, CD-HIT, MOTHUR, USEARCH
Interactive visualizations
e.g., PCoA plots, distance histograms, taxonomy charts, rarefaction plots, network visualization, jackknifed hierarchical clustering.
Assign taxonomy
BLAST, RDP Classifier
Align sequences
e.g., PyNAST, INFERNAL, MUSCLE, MAFFT
Build 'OTU table'
i.e., sample by observation matrix
Build phylogenetic tree
e.g., FastTree, RAxML, ClearCut
Legend
Currently supported for marker-gene data only
(i.e., 'upstream' step)
Currently supported for general sample by observation data
(i.e., 'downstream' step)
Required step or input
Optional step or input

Preprocessing of reads

The data that we will be processing tomorrow was generated in the format that comes from Caporaso 2011, PNAS, where one has separate reads and barcodes fastq files.

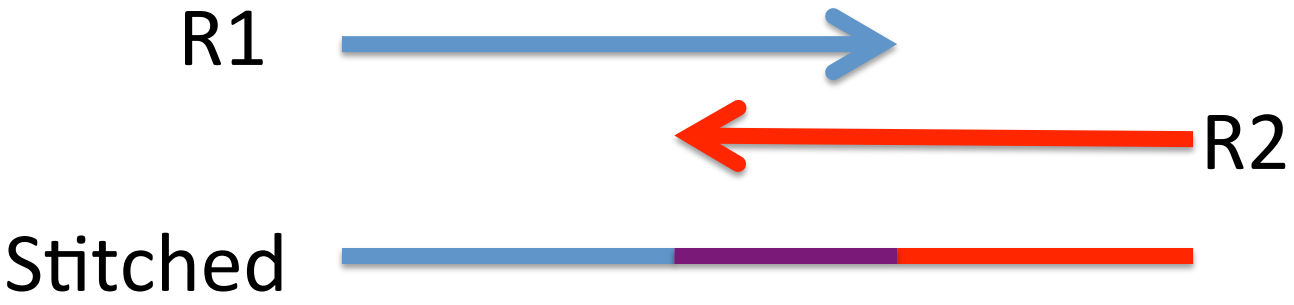
Preprocessing of reads

The data that we will be processing tomorrow was generated in the format that comes from Caporaso 2011, PNAS, where one has separate reads and barcodes fastq files.

For most cases, you will receive paired-end reads (with R1 and R2 in their filenames) for the read 1 and read 2 of Illumina data.

Preprocessing of reads

If one's reads overlap, then they can be stitched together. In QIIME, the join_paired_ends.py script can be used to stitch the reads.



Preprocessing of reads

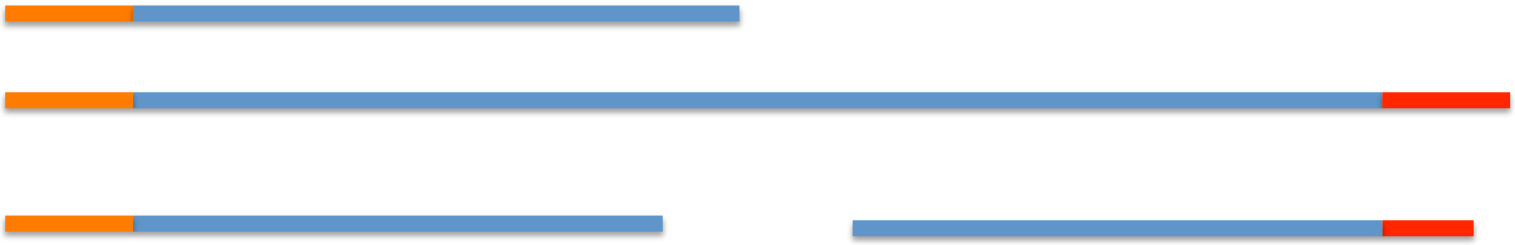
The output of join_paired_ends.py will be a set of "joined" reads, as well as unjoined R1 and R2 files. You want to use the joined reads for further processing and analysis.

If you have a separate barcode read, you want to pass this as a parameter when calling join_paired_ends.py, so the barcodes will be filtered to matched the joined reads.

One consideration: if there is low yields from the stitching process, it may be better to use a single end read, such as R1, than using a low number of stitched reads.

Alternative Barcoding

Sometimes (not in the tutorial data tomorrow), the barcodes will be at the end(s) of your reads.

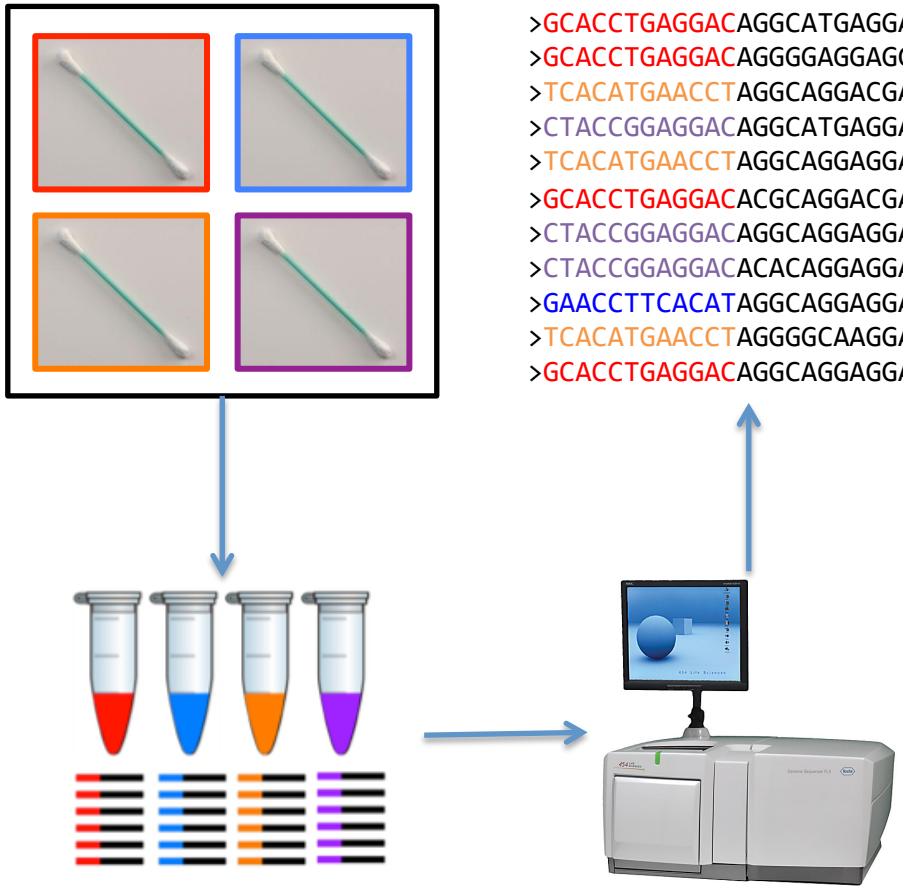


Orange/Red indicates barcodes

Alternative Barcoding

You can use the extract_barcodes.py script to generate separate barcodes and reads (with barcodes stripped) fastq files.

Error-correcting codes allow multiplex sequencing



>PC.634_1 FLP3FBN01ELBSX
CTGGGCCGTGTCTCAGTCCCAATGTGGCCGTTTACCCTCTCAGGCCGGCTAC
GCATCATCGCCTTGGTGGGCCGTTACCTCACCAACTAGCTAATGCGCCGCAG
GTCCATCCATGTTCACGCCTTGATGGGCGCTTTAATATACTGAGCATGCGCT
CTGTATACCTATCCGGTTTTAGCTACCGTTTCCAGCAGTTATCCCGGACACA
TGGGCTAGG

>PC.354_3 FLP3FBN01EEWKD
TTGGACCGTGTCTCAGTTCCAATGTGGGGGCCTTCCTCTCAGAACCCCTATC
CATCGAAGGCTTGGTGGGCCGTTACCCCGCCAACAACCTAATGGAACGCATC
CCCATCGATGACCGAAGTTCTTTAATAGTTCTACCATGCGGAAGAACTATGC
CATCGGGTATTAATCTTTCTTTCGAAAGGCTATCCCCGAGTCATCGGCAGGT
TGGATACGTGTTACTCACCCGTGCGCCGGT

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

Demultiplexing

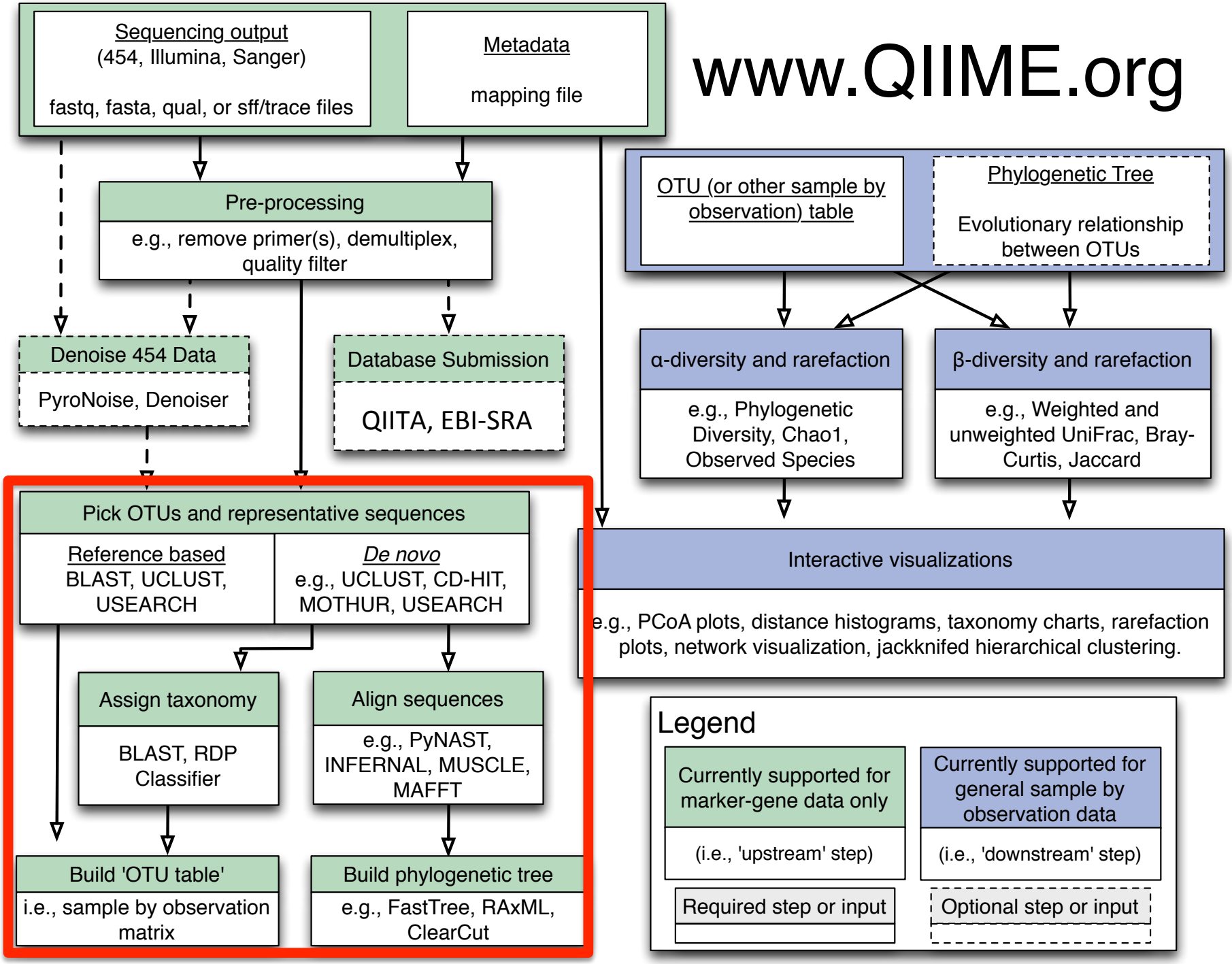
For Illumina data, the script is split_libraries_fastq.py.

For 454 and IonTorrent data, use split_libraries.py.

A common situation (not in this tutorial): If you get already-split up reads (separated into fastq file(s) by sample), use the script multiple_split_libraries_fastq.py

multiple_join_paired_ends.py and multiple_extract_barcodes.py also exist.

OTU picking methods

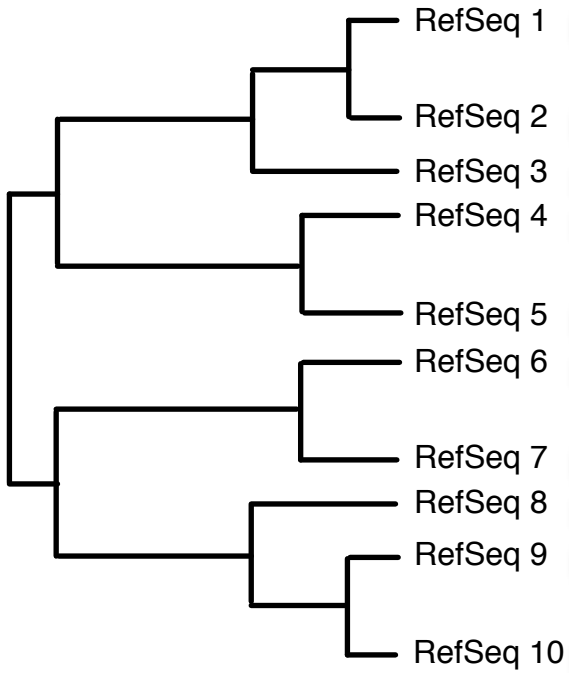

Sequencing output
(454, Illumina, Sanger)
fastq, fasta, qual, or sff/trace files
Metadata
mapping file
www.QIIME.org
Pre-processing
e.g., remove primer(s), demultiplex, quality filter
OTU (or other sample by observation) table
Phylogenetic Tree
Evolutionary relationship between OTUs
Denoise 454 Data
PyroNoise, Denoiser
Database Submission
QIITA, EBI-SRA
α-diversity and rarefaction
e.g., Phylogenetic Diversity, Chao1, Observed Species
β-diversity and rarefaction
e.g., Weighted and unweighted UniFrac, Bray-Curtis, Jaccard
Pick OTUs and representative sequences
Reference based
BLAST, UCLUST, USEARCH
De novo
e.g., UCLUST, CD-HIT, MOTHUR, USEARCH
Interactive visualizations
e.g., PCoA plots, distance histograms, taxonomy charts, rarefaction plots, network visualization, jackknifed hierarchical clustering.
Assign taxonomy
BLAST, RDP Classifier
Align sequences
e.g., PyNAST, INFERNAL, MUSCLE, MAFFT
Build 'OTU table'
i.e., sample by observation matrix
Build phylogenetic tree
e.g., FastTree, RAxML, ClearCut
Legend
Currently supported for marker-gene data only
(i.e., 'upstream' step)
Currently supported for general sample by observation data
(i.e., 'downstream' step)
Required step or input
Optional step or input

Which microbial organisms are represented by the rRNA gene sequences in each sample?

rRNA reference database (sequences are available for each 'tip' in the tree)

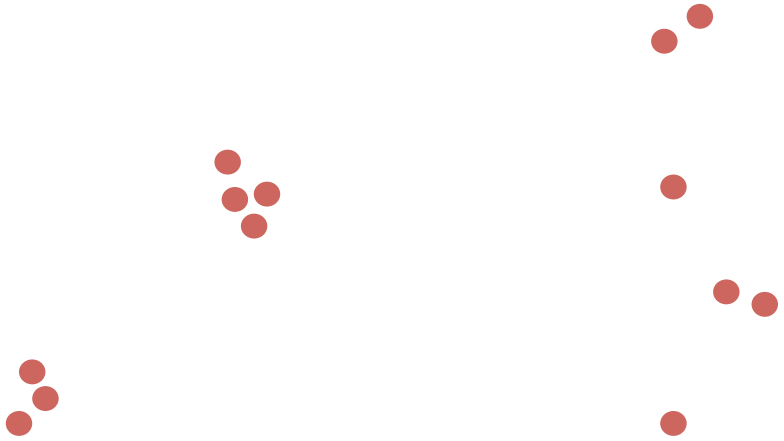
```
>PC.634_1 FLP3FBN01ELBSX
CTGGGCCGTGTCTCAGTCCCAATGTGGCCGTTTACCCTCTCAGGCCGG
CTACGCATCATCGCCTTGGTGGGCCGTTACCTCACCAACTAGCTAATG
CGCCGCAGGTCCATCCATGTTCACGCCTTGATGGGCGCTTTAATATAC
TGAGCATGCGCTCTGTATACCTATCCGGTTTTAGCTACCGTTTCCAGC
AGTTATCCCGGACACATGGGCTAGG
>PC.634_2 FLP3FBN01EG8AX
TTGGACCGTGTCTCAGTTCCAATGTGGGGGCCTTCCTCTCAGAACCCC
TATCCATCGAAGGCTTGGTGGGCCGTTACCCCGCCAACAACCTAATGG
AACGCATCCCCATCGATGACCGAAGTTCTTTAATAGTTCTACCATGCG
GAAGAACTATGCCATCGGGTATTAATCTTTCTTTCGAAAGGCTATCCC
CGAGTCATCGGCAGGTTGGATACGTGTTACTCACCCGTGCGCCGGT
>PC.354_3 FLP3FBN01EEWKD
TTGGGCCGTGTCTCAGTCCCAATGTGGCCGATCAGTCTCTTAACTCGG
CTATGCATCATTGCCTTGGTAAGCCGTTACCTTACCAACTAGCTAATG
CACCGCAGGTCCATCCAAGAGTGATAGCAGAACCATCTTTCAAACTCT
AGACATGCGTCTAGTGTTGTTATCCGGTATTAGCATCTGTTTCCAGGT
GTTATCCCAGTCTCTTGGG
```

Search against reference sequences

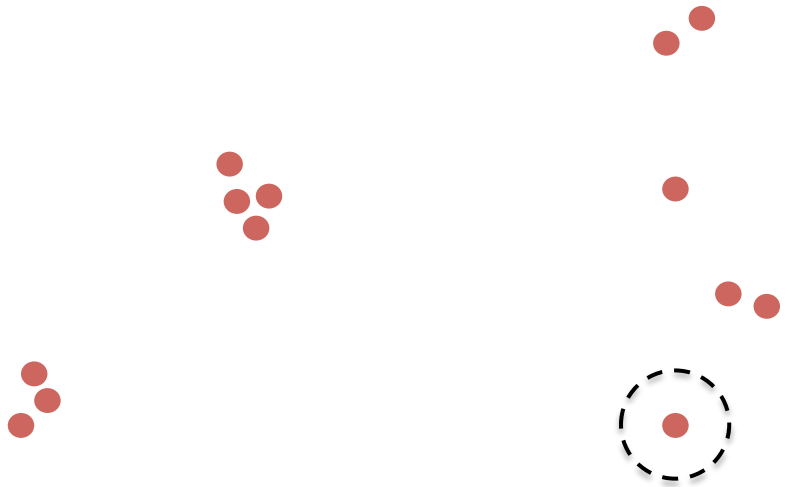


OTU table: counts of OTUs in each sample

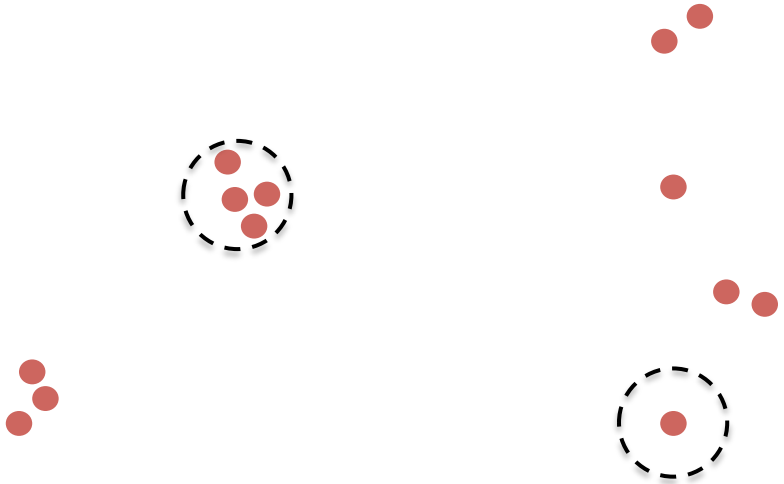
Understanding OTU picking



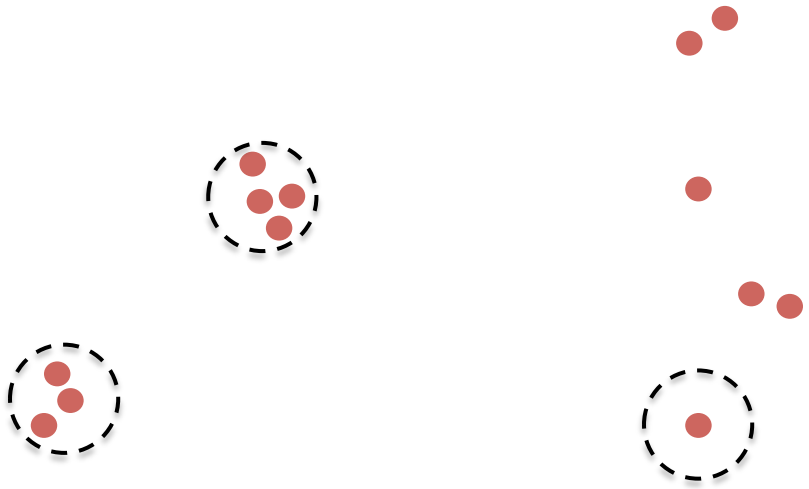
Understanding OTU picking



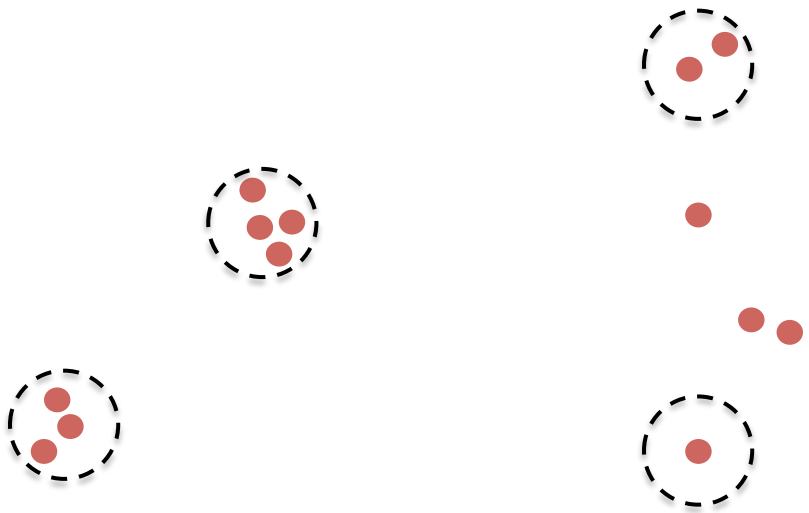
Understanding OTU picking



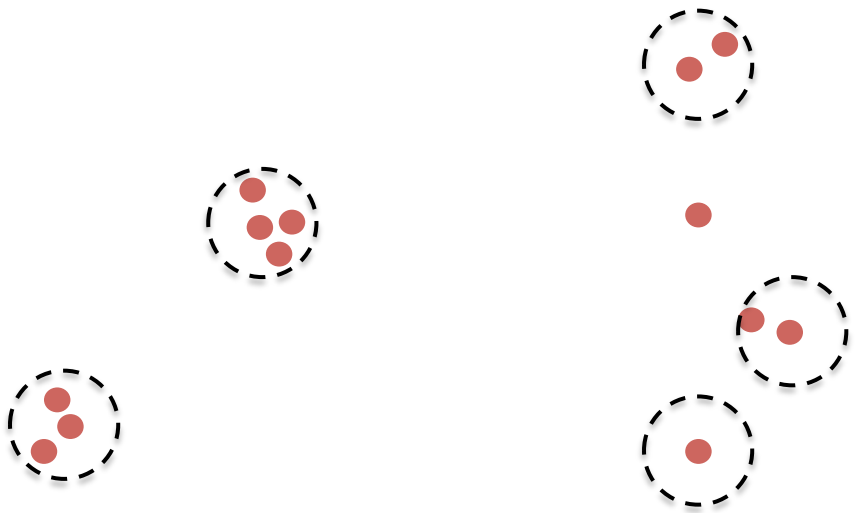
Understanding OTU picking



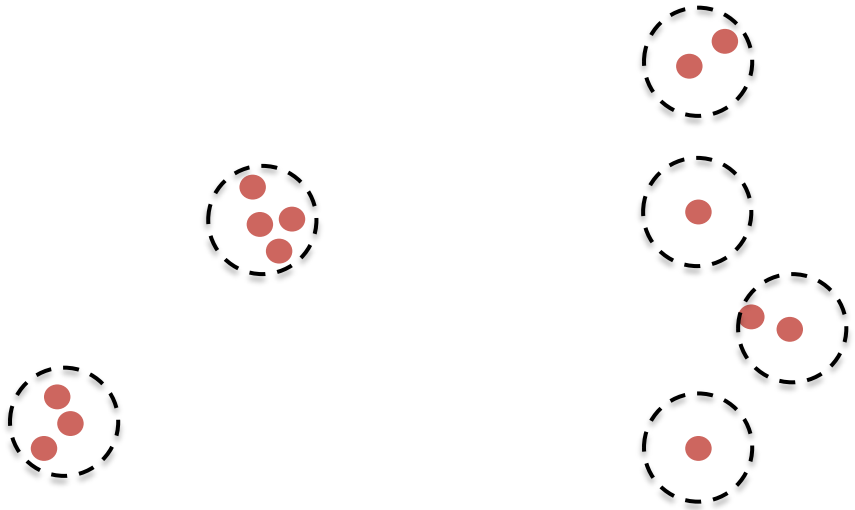
Understanding OTU picking



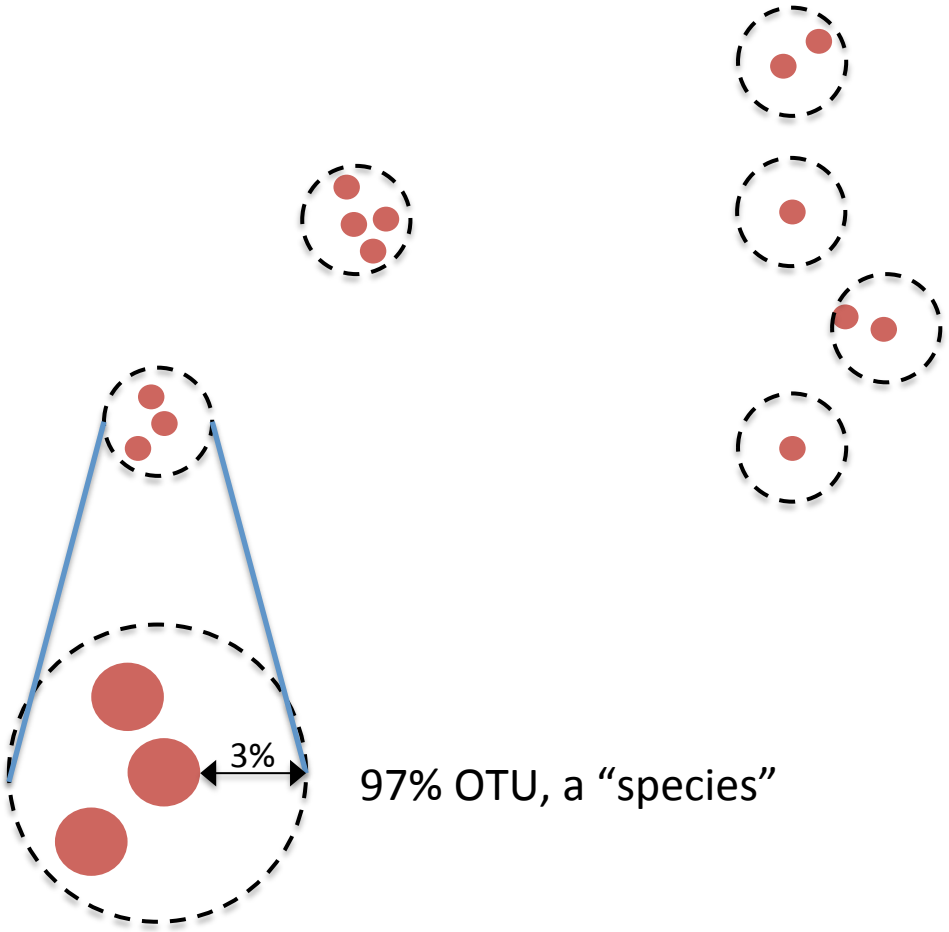
Understanding OTU picking



Understanding OTU picking

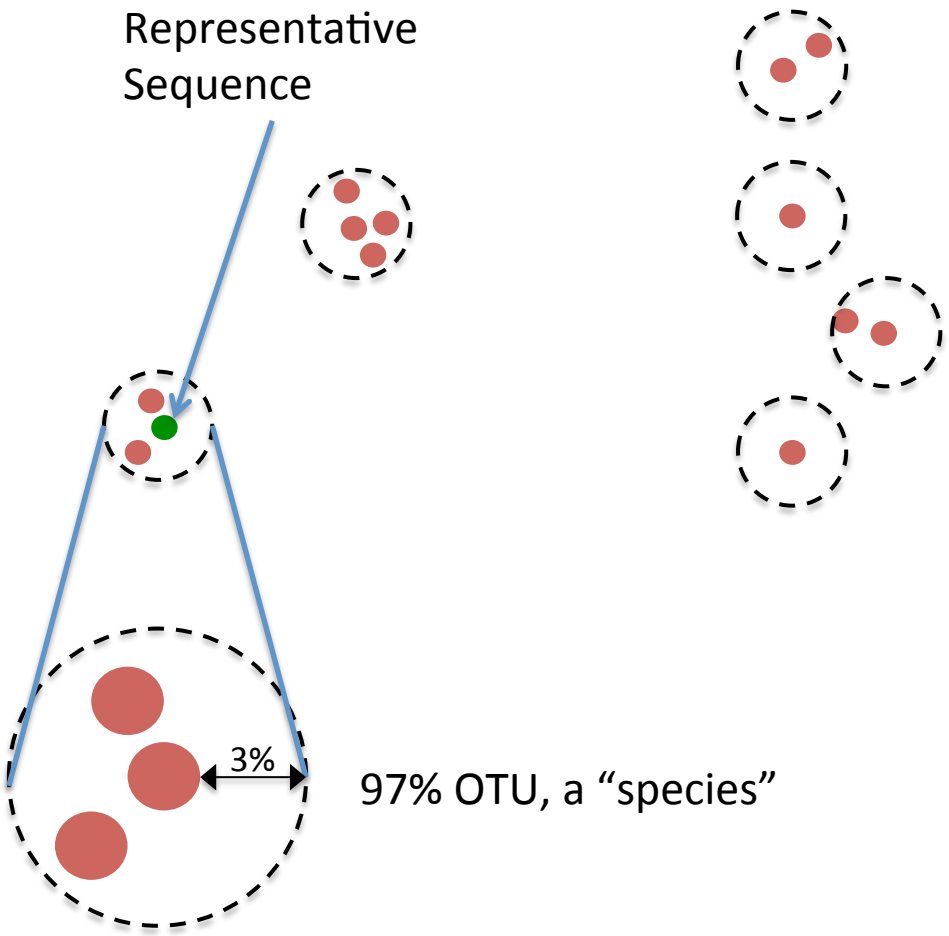


Understanding OTU picking

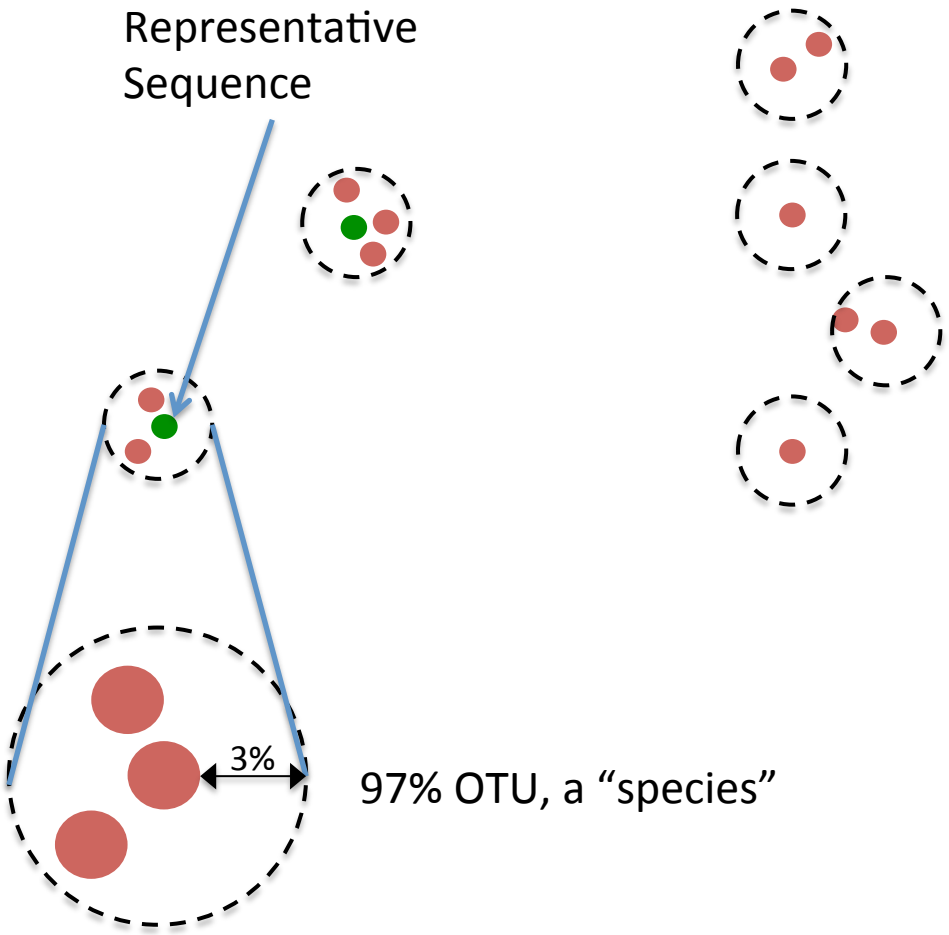


97% OTU, a “species”

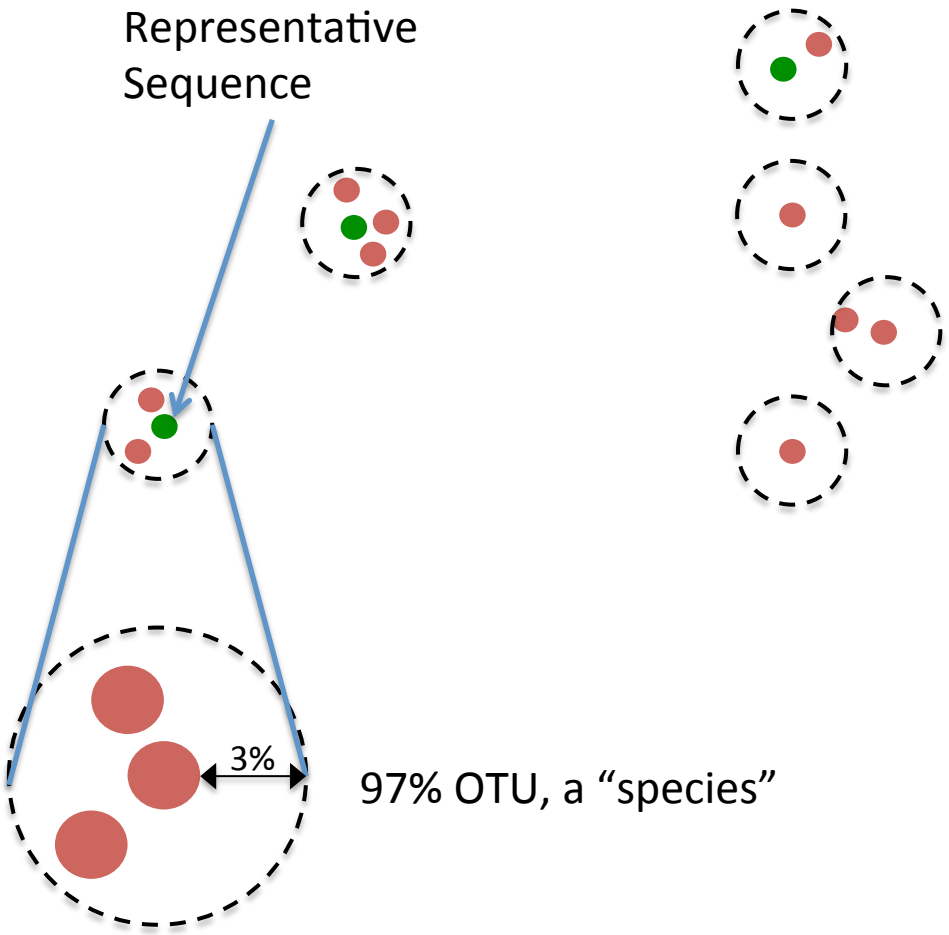
Understanding OTU picking



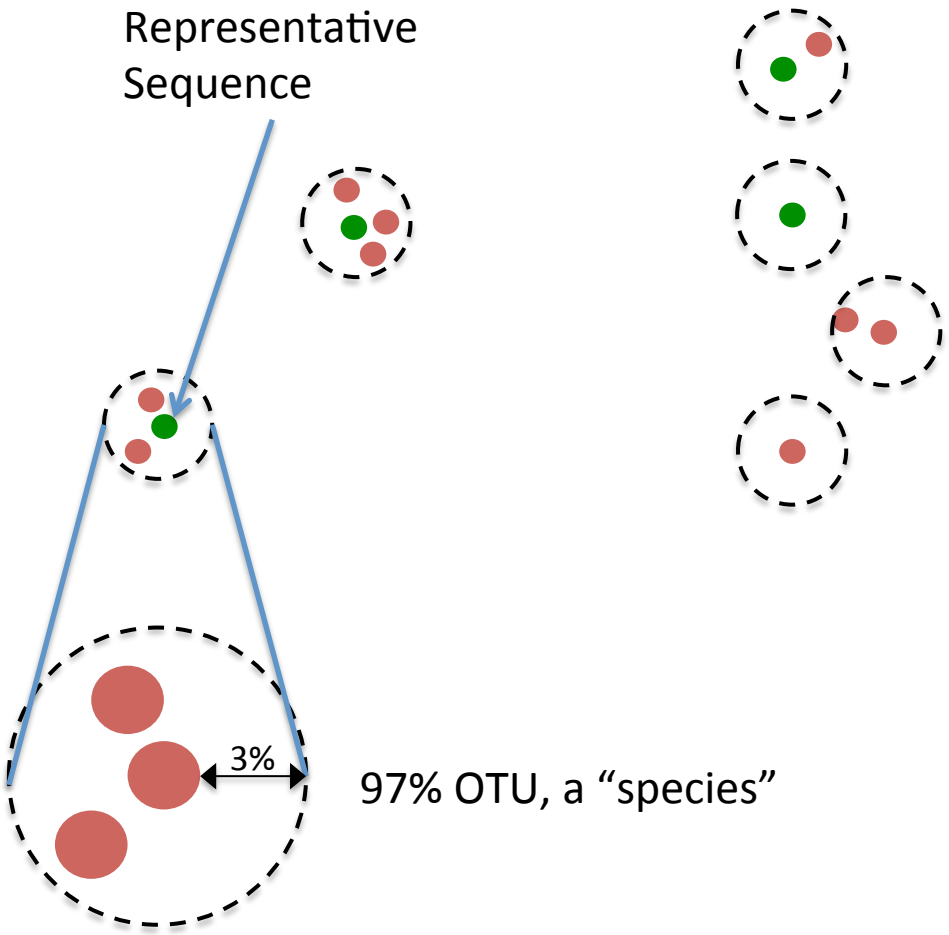
Understanding OTU picking



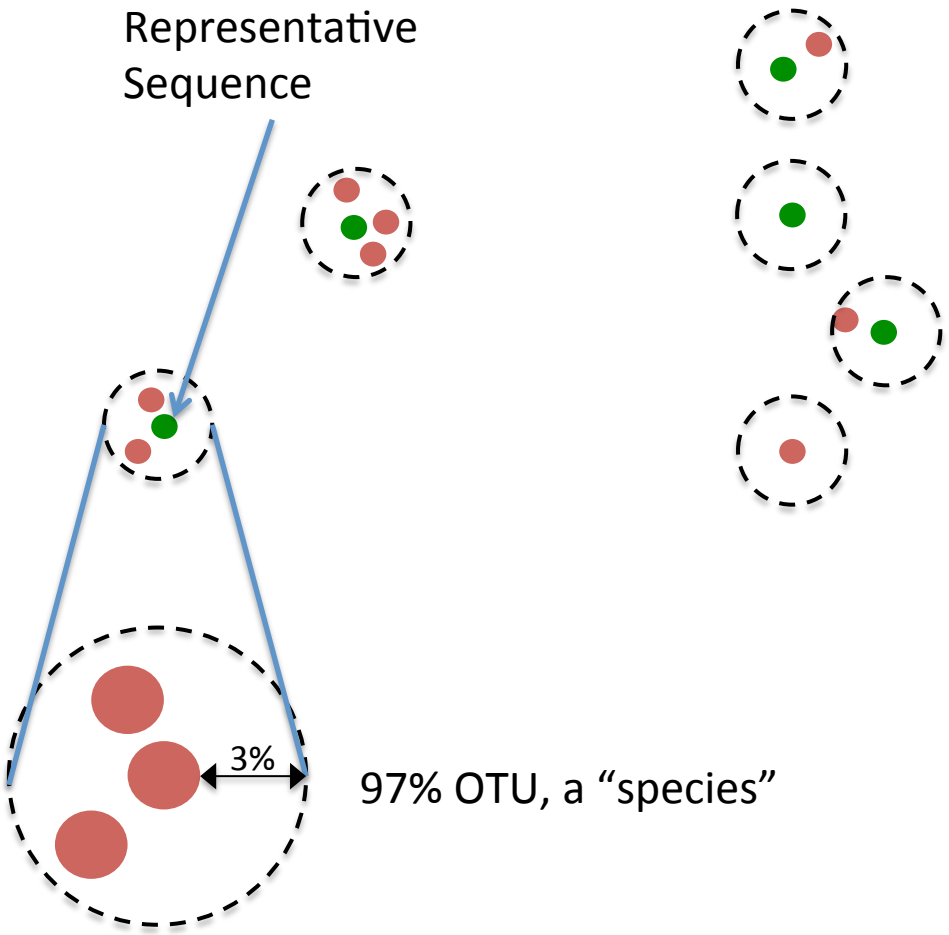
Understanding OTU picking



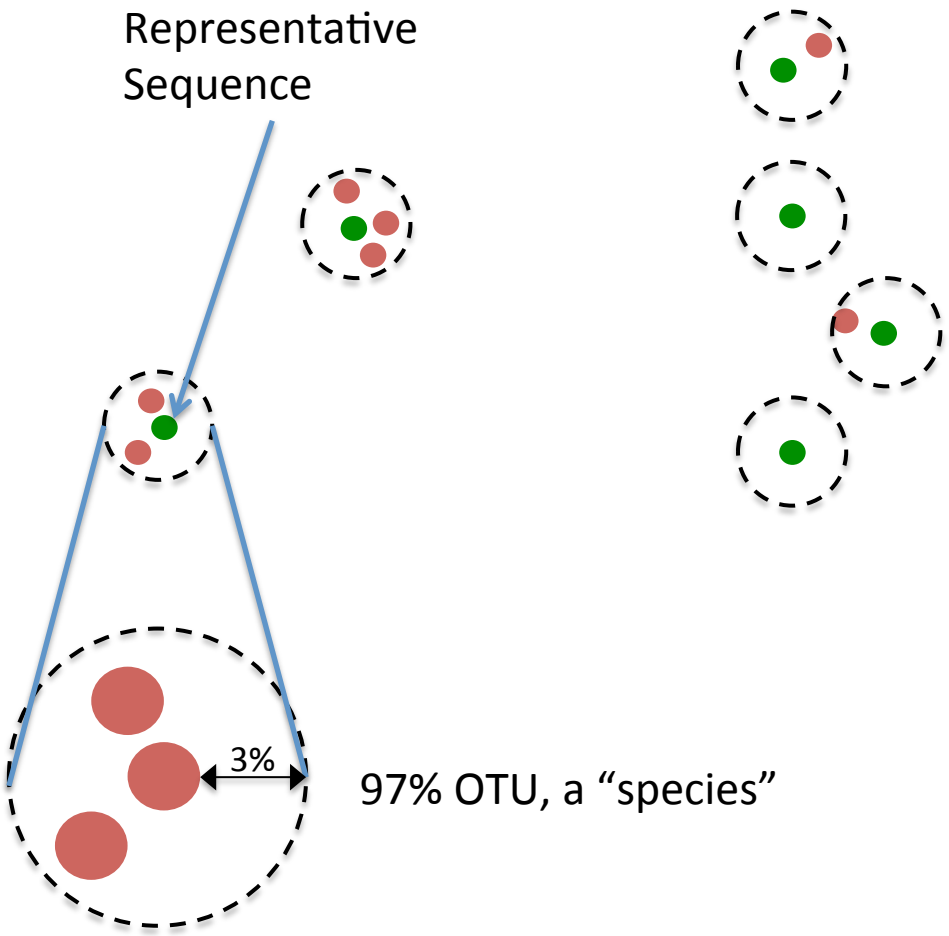
Understanding OTU picking



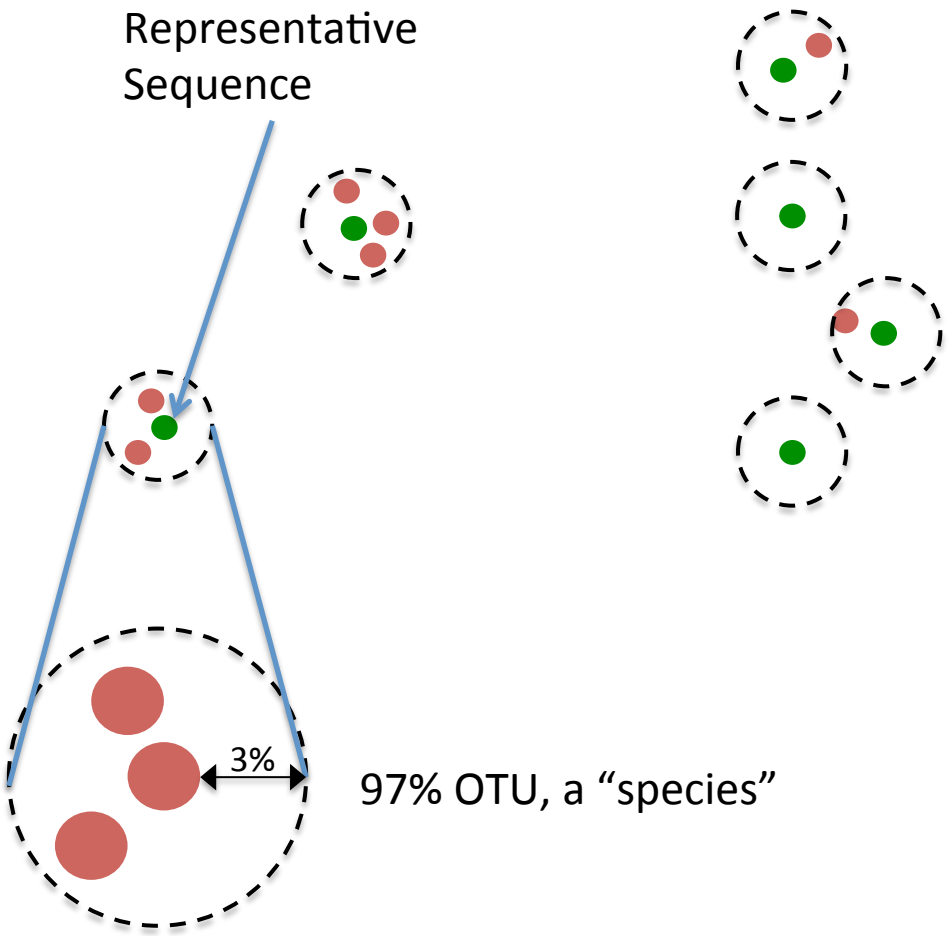
Understanding OTU picking



Understanding OTU picking



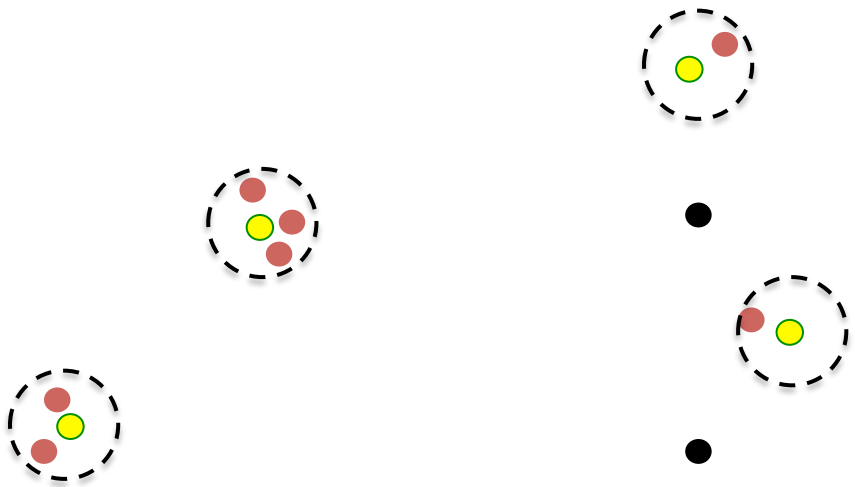
Understanding OTU picking



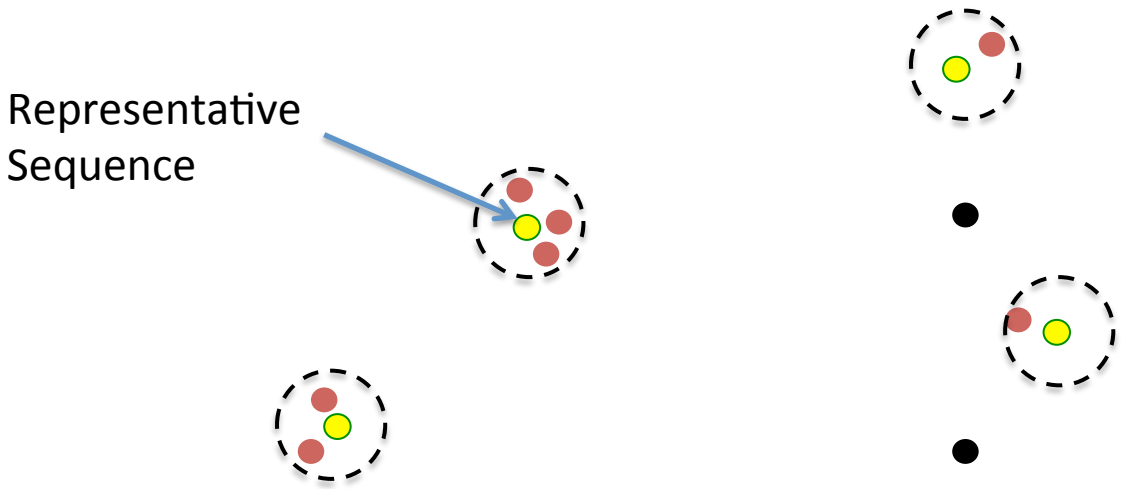
97% OTU, a “species”

De novo
OTU picking

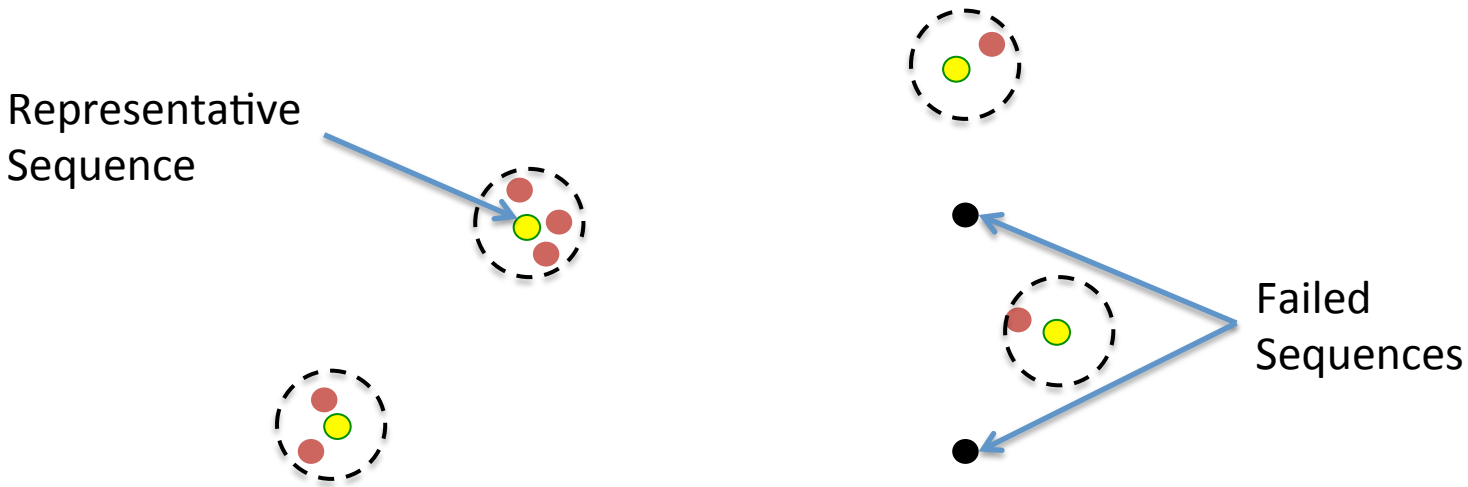
Understanding OTU picking



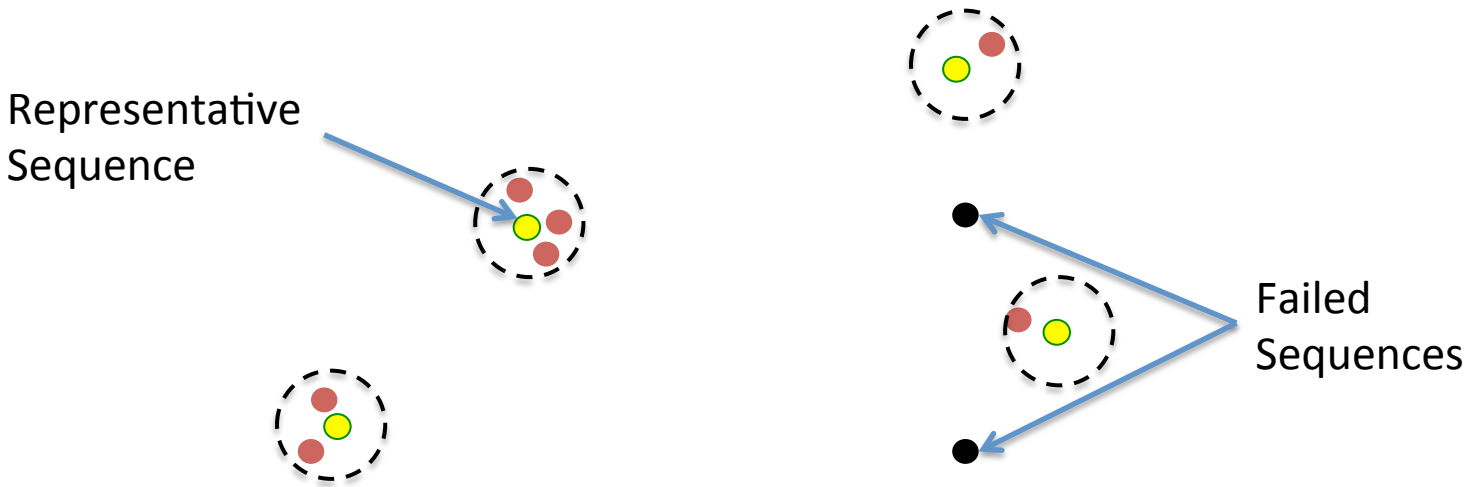
Understanding OTU picking



Understanding OTU picking

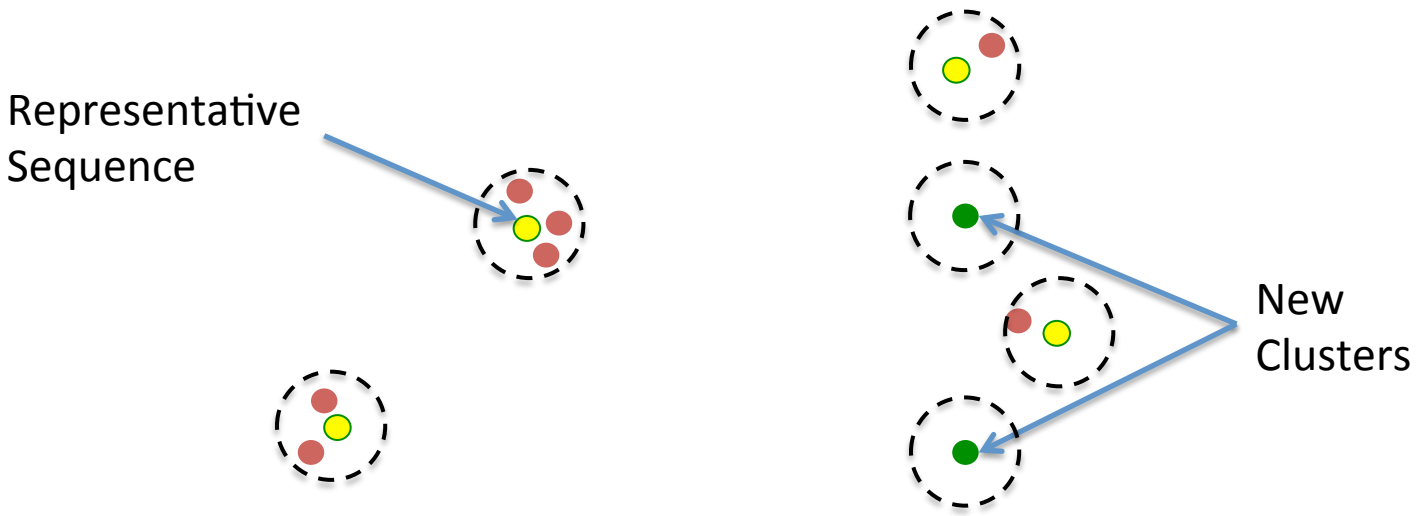


Understanding OTU picking

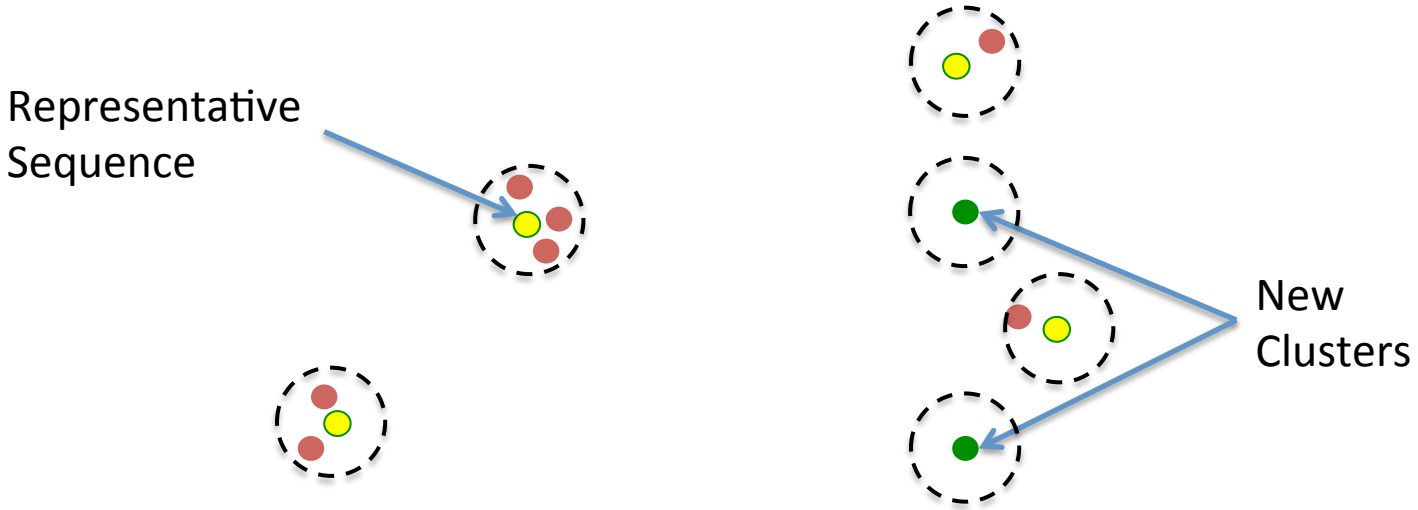


Closed reference
OTU picking

Understanding OTU picking



Understanding OTU picking



Open reference
OTU picking

De novo OTU picking

- Pros
  - All reads are clustered
- Cons
  - Not parallelizable
  - OTUs may be defined by erroneous reads

pick_de_novo_otus.py
http://qiime.org/tutorials/tutorial.html

De novo OTU picking

- You **must** use if:
  - You do not have a reference sequence collection to cluster against, for example because you're working with an infrequently used marker gene.
- You **cannot** use if:
  - You are comparing non-overlapping amplicons, such as the V2 and the V4 regions of the 16S rRNA.
  - You working with very large data sets, like a full HiSeq 2000 run. (Technically you can, but it will be *really* slow.)

pick_de_novo_otus.py
http://qiime.org/tutorials/tutorial.html

Closed-reference OTU picking

- Pros
  - Built-in quality filter
  - Easily parallelizable
  - OTUs are defined by high-quality, trusted sequences
- Cons
  - Reads that don't hit reference dataset are excluded, so you can never observe new OTUs

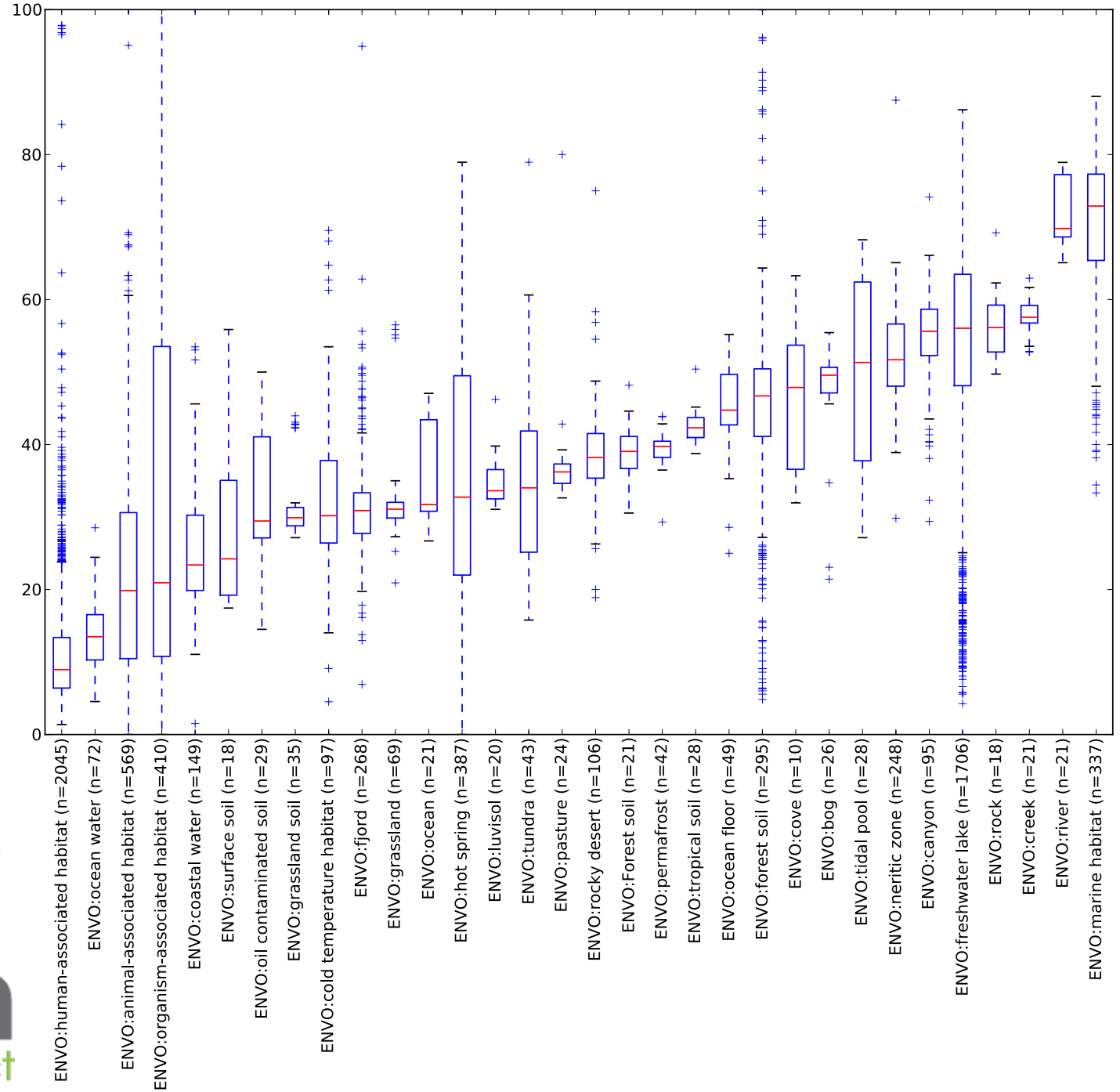
pick_closed_reference_otus.py

Closed-reference OTU picking

- You **must** use if:
  - You are comparing non-overlapping amplicons, such as the V2 and the V4 regions of the 16S rRNA. Your reference sequences must span both of the regions being sequenced.
- You **cannot** use if:
  - You do not have a reference sequence collection to cluster against, for example because you're working with an infrequently used marker gene.

pick_closed_reference_otus.py

Percentage of reads that do not hit the reference collection, by environment type.



earth microbiome project

Open-reference OTU picking

- Pros
  - All reads are clustered
  - Partially parallelizable
- Cons
  - Only *partially* parallelizable
  - Mix of high quality sequences defining OTUs (i.e., the database sequences) and possible low quality sequences defining OTUs (i.e., the sequencing reads)

pick_open_reference_otus.py
http://qiime.org/tutorials/illumina_overview_tutorial.html
http://qiime.org/tutorials/open_reference_illumina_processing.html
http://qiime.org/tutorials/fungal_its_analysis.html

Open-reference OTU picking

- You **cannot** use if:
  - You are comparing non-overlapping amplicons, such as the V2 and the V4 regions of the 16S rRNA.
  - You do not have a reference sequence collection to cluster against, for example because you're working with an infrequently used marker gene.

pick_open_reference_otus.py
http://qiime.org/tutorials/illumina_overview_tutorial.html
http://qiime.org/tutorials/open_reference_illumina_processing.html
http://qiime.org/tutorials/fungal_its_analysis.html

Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences

Jai Ram Rideout[1,2], Yan He[3], Jose A. Navas-Molina[4], William A. Walters[5], Luke K. Ursell[6], Sean M. Gibbons[7,10], John Chase[8], Daniel McDonald[4,9], Antonio Gonzalez[9], Adam Robbins-Pianka[4,9], Jose C. Clemente[2], Jack A. Gilbert[10,11], Susan M. Huse[12], Hong-Wei Zhou[3], Rob Knight[9,13], J. Gregory Caporaso[1,8]

Published August 21, 2014

PubMed 25177538

https://peerj.com/articles/545/

OTU picking methods vs algorithms

- QIIME provides 3 high-level **methods**
- QIIME wraps OTU clustering tools
  - Performs sequence clustering **algorithm**
  - uclust, usearch, SortMeRNA, SumaClust, swarm, …
  - Be sure to cite these tools too!

OTU table
OTU x sample matrix

Count table

	S1	S2	S3
OTU1	100	0	0
OTU2	100	40	600
OTU3	0	10	0

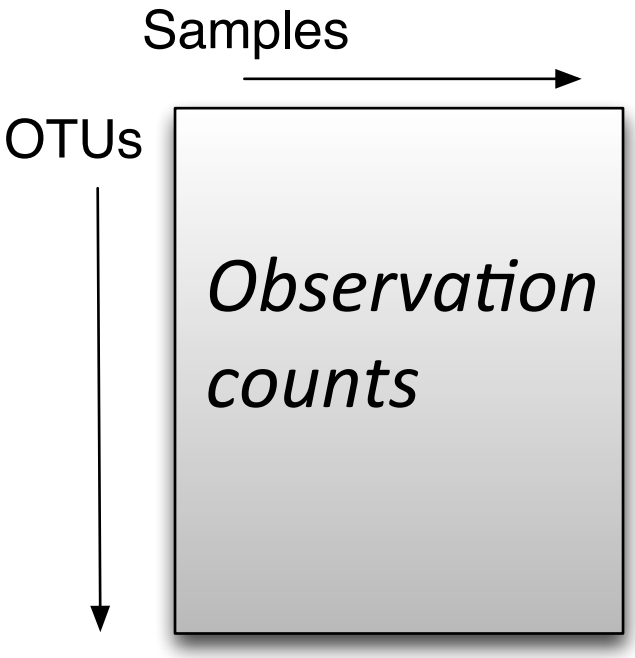
Relative abundance table

	S1	S2	S3
OTU1	.5	0	0
OTU2	.5	.8	1.0
OTU3	0	.2	0

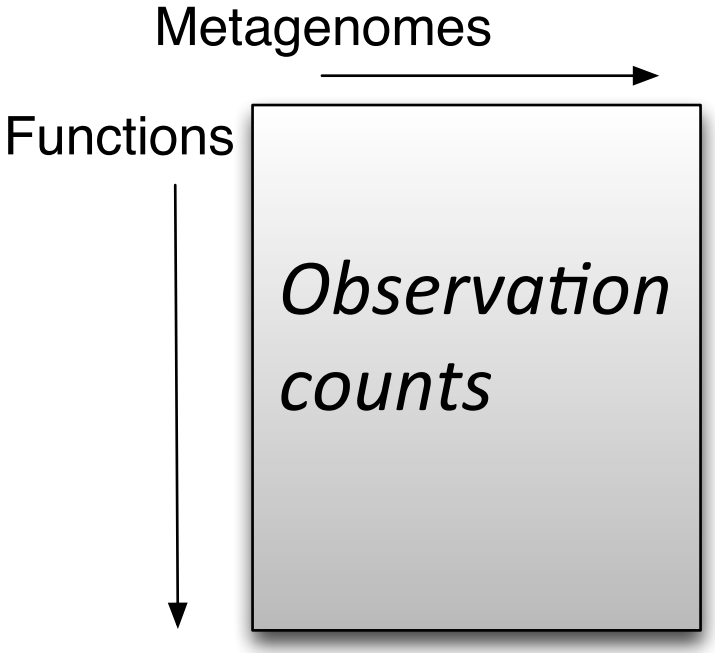
OTU tables are in Biological Observation Matrix (*.biom*) format

http://biom-format.org

sample x observation contingency matrix

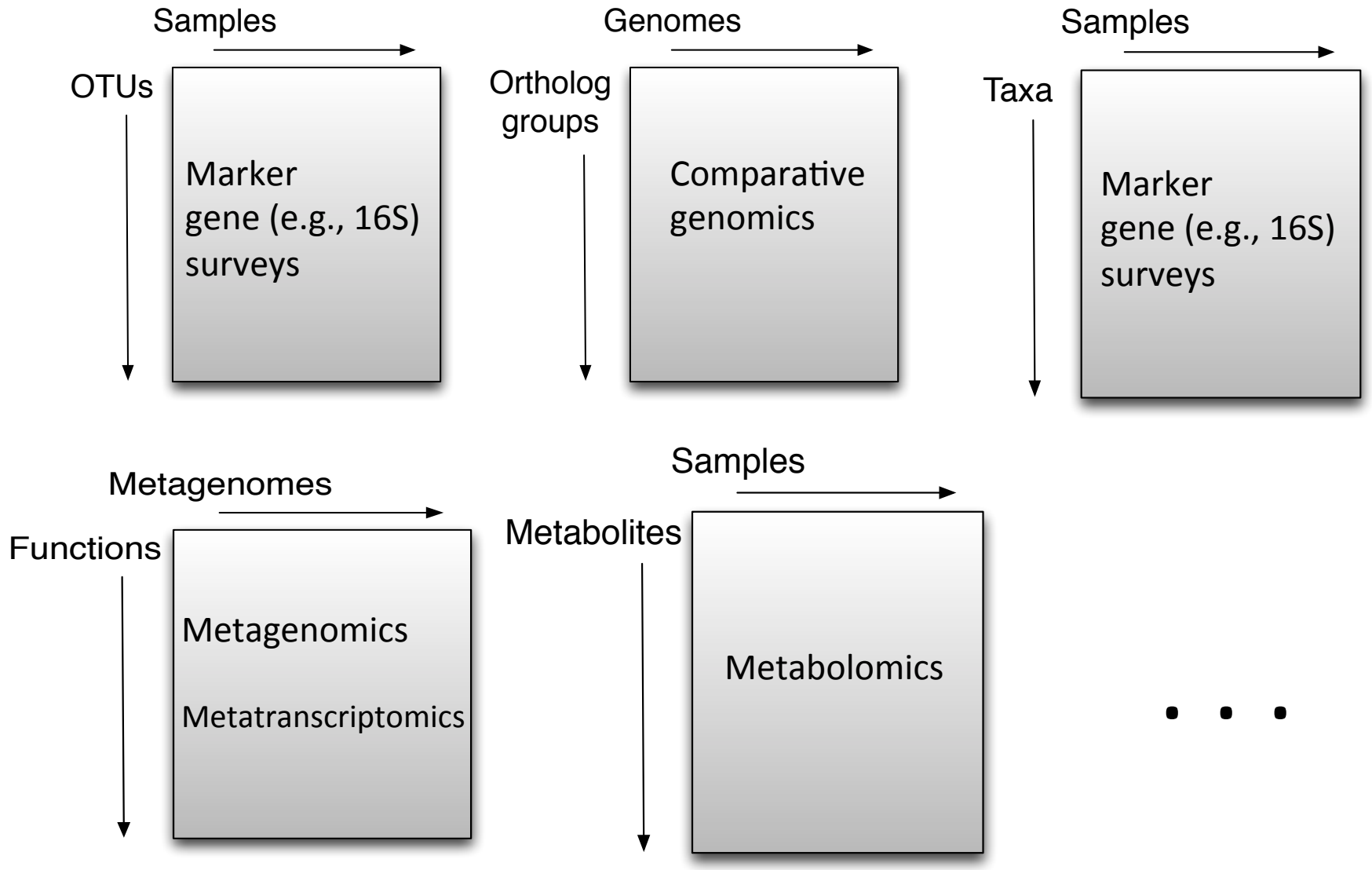


sample x observation contingency matrix



MG-RAST
metagenomics analysis server

sample x observation contingency matrix



The Biological Observation Matrix (BIOM) Format or: How I Learned To Stop Worrying and Love the Ome-ome

Format for representing arbitrary sample x observation contingency tables with optional metadata

MG-RAST
metagenomics analysis server

qiime
Quantitative Insights Into Microbial Ecology

VAMPS
The Visualization and Analysis of Microbial Population Structures

http://www.biom-format.org

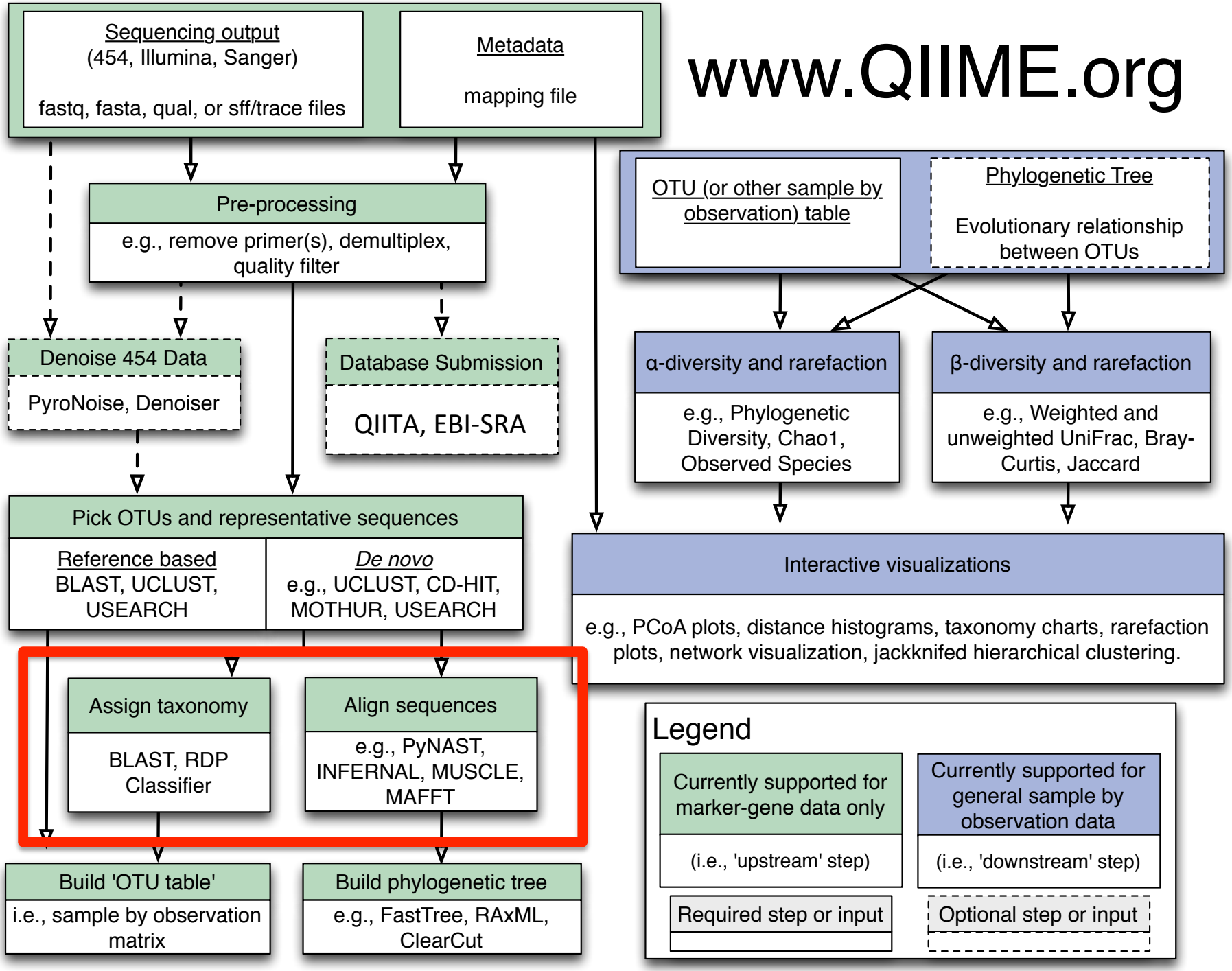
McDonald et al., *GigaScience* (2012).

Reference databases

- Greengenes (16S)
  - Default reference database included in QIIME 1.9.0 and later
- SILVA (16S/18S)
- UNITE (ITS)
- IMG (protein sequences)

Reference databases

- Latest QIIME-compatible databases:
  - http://qiime.org/home_static/dataFiles.html
- Tutorials
  - 16S
    - http://qiime.org/tutorials/illumina_overview_tutorial.html
    - http://qiime.org/tutorials/tutorial.html
  - ITS
    - http://qiime.org/tutorials/fungal_its_analysis.html
  - 18S
    - http://qiime.org/tutorials/processing_18S_data.html
  - Shotgun data (metagenome data)
    - http://qiime.org/tutorials/shotgun_analysis.html


Sequencing output
(454, Illumina, Sanger)
fastq, fasta, qual, or sff/trace files
Metadata
mapping file
www.QIIME.org
Pre-processing
e.g., remove primer(s), demultiplex, quality filter
OTU (or other sample by observation) table
Phylogenetic Tree
Evolutionary relationship between OTUs
Denoise 454 Data
PyroNoise, Denoiser
Database Submission
QIITA, EBI-SRA
α-diversity and rarefaction
e.g., Phylogenetic Diversity, Chao1, Observed Species
β-diversity and rarefaction
e.g., Weighted and unweighted UniFrac, Bray-Curtis, Jaccard
Pick OTUs and representative sequences
Reference based
BLAST, UCLUST, USEARCH
De novo
e.g., UCLUST, CD-HIT, MOTHUR, USEARCH
Interactive visualizations
e.g., PCoA plots, distance histograms, taxonomy charts, rarefaction plots, network visualization, jackknifed hierarchical clustering.
Assign taxonomy
BLAST, RDP Classifier
Align sequences
e.g., PyNAST, INFERNAL, MUSCLE, MAFFT
Build 'OTU table'
i.e., sample by observation matrix
Build phylogenetic tree
e.g., FastTree, RAxML, ClearCut
Legend
Currently supported for marker-gene data only
(i.e., 'upstream' step)
Currently supported for general sample by observation data
(i.e., 'downstream' step)
Required step or input
Optional step or input

Representative Sequences

If de novo or open-reference OTU picking is used, a representative sequence will be selected for each OTU. By default, this will be the most abundant read for that OTU. These reads are used for taxonomic assignment and tree building.

Closed-reference OTU picking does not generated a tree or taxonomic assignment. The tree and taxonomies in the reference files are used instead.

Taxonomic Assignment

With workflow scripts, you do not have to manually assign taxonomy, but the underlying script being used is assign_taxonomy.py.

Each representative sequence is assigned-some methods can have ambiguous results, while others return the best match.

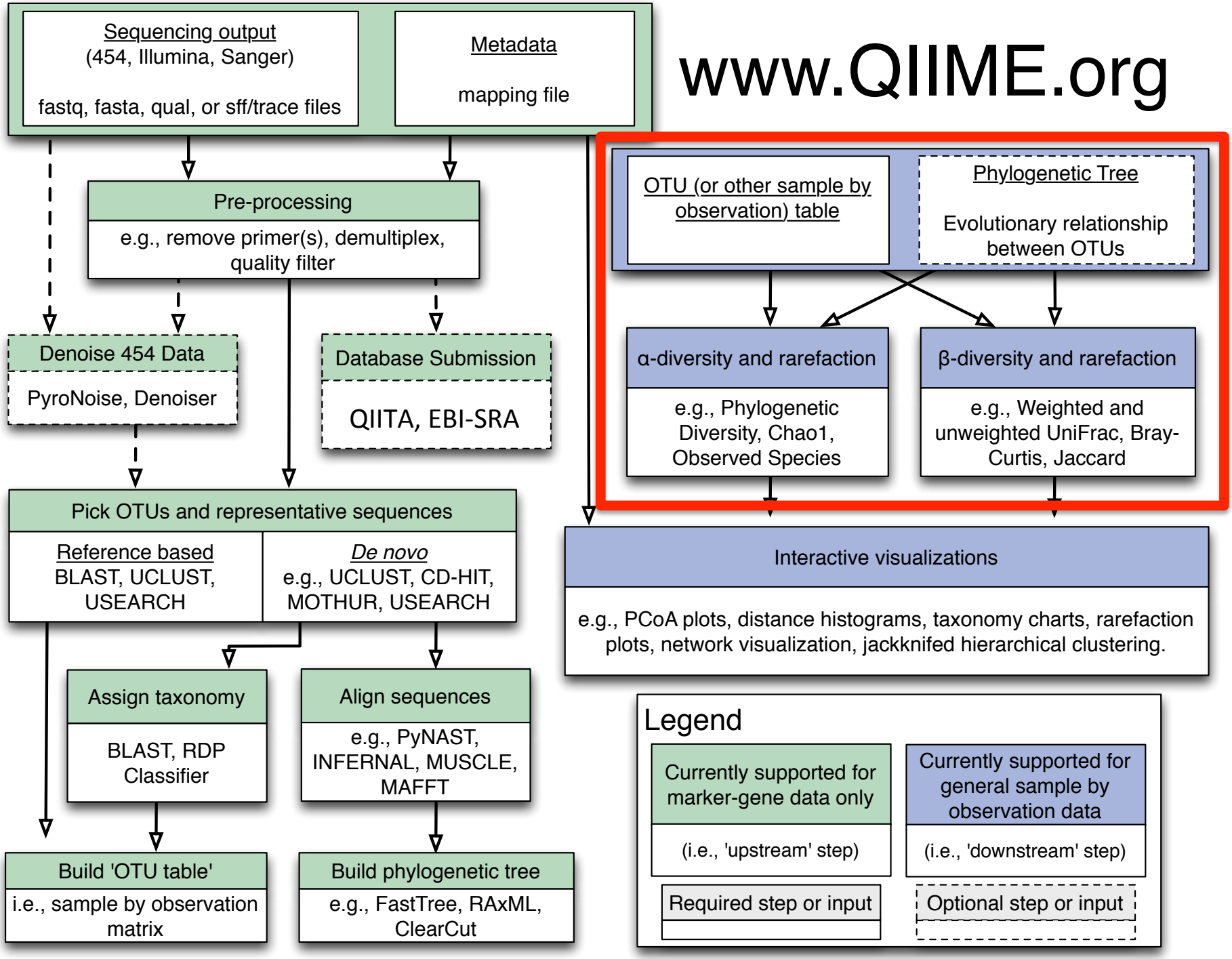
Phylogenetic Trees

The representative sequences are aligned, then filtered for highly gapped/variable bases, and finally a tree is built.

The sequences that fail to align can be filtered out of the final OTU table.

Scripts used:
align_seqs.py
filter_alignment.py
make_phylogeny.py

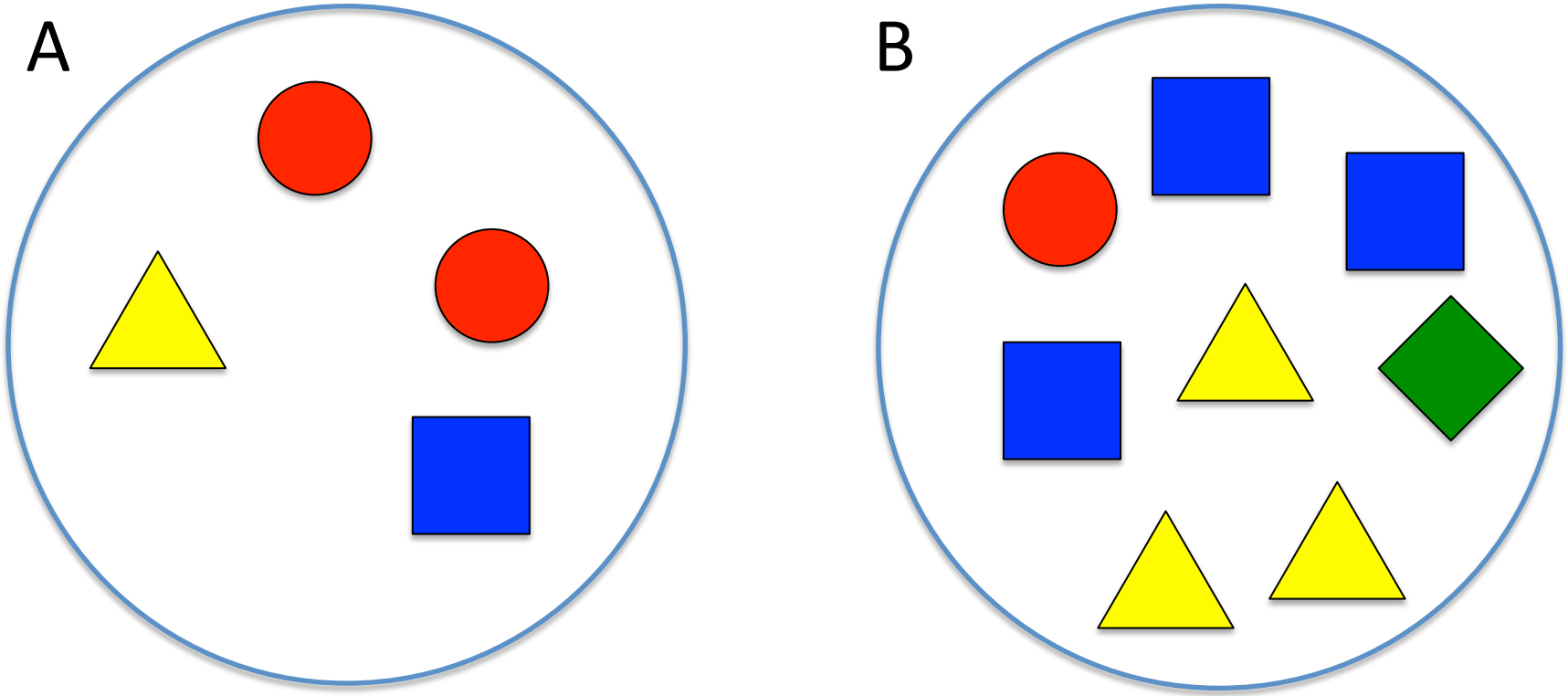
Diversity and statistical analysis
Part I


Sequencing output
(454, Illumina, Sanger)
fastq, fasta, qual, or sff/trace files
Metadata
mapping file
www.QIIME.org
Pre-processing
e.g., remove primer(s), demultiplex, quality filter
Denoise 454 Data
PyroNoise, Denoiser
Database Submission
QIITA, EBI-SRA
OTU (or other sample by observation) table
Phylogenetic Tree
Evolutionary relationship between OTUs
α-diversity and rarefaction
e.g., Phylogenetic Diversity, Chao1, Observed Species
β-diversity and rarefaction
e.g., Weighted and unweighted UniFrac, Bray-Curtis, Jaccard
Pick OTUs and representative sequences
Reference based
BLAST, UCLUST, USEARCH
De novo
e.g., UCLUST, CD-HIT, MOTHUR, USEARCH
Interactive visualizations
e.g., PCoA plots, distance histograms, taxonomy charts, rarefaction plots, network visualization, jackknifed hierarchical clustering.
Assign taxonomy
BLAST, RDP Classifier
Align sequences
e.g., PyNAST, INFERNAL, MUSCLE, MAFFT
Build 'OTU table'
i.e., sample by observation matrix
Build phylogenetic tree
e.g., FastTree, RAxML, ClearCut
Legend
Currently supported for marker-gene data only
(i.e., 'upstream' step)
Currently supported for general sample by observation data
(i.e., 'downstream' step)
Required step or input
Optional step or input

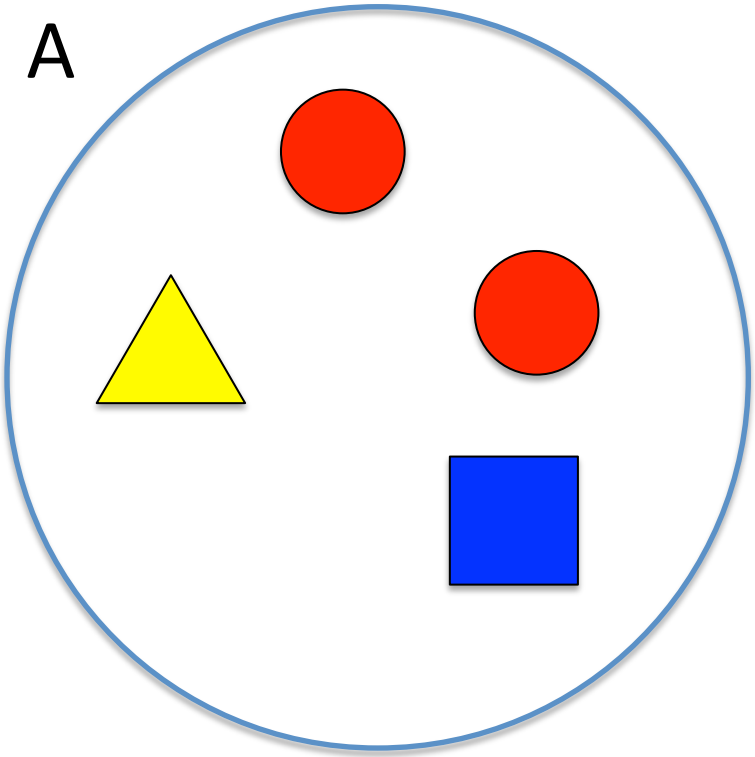
Alpha and beta diversity

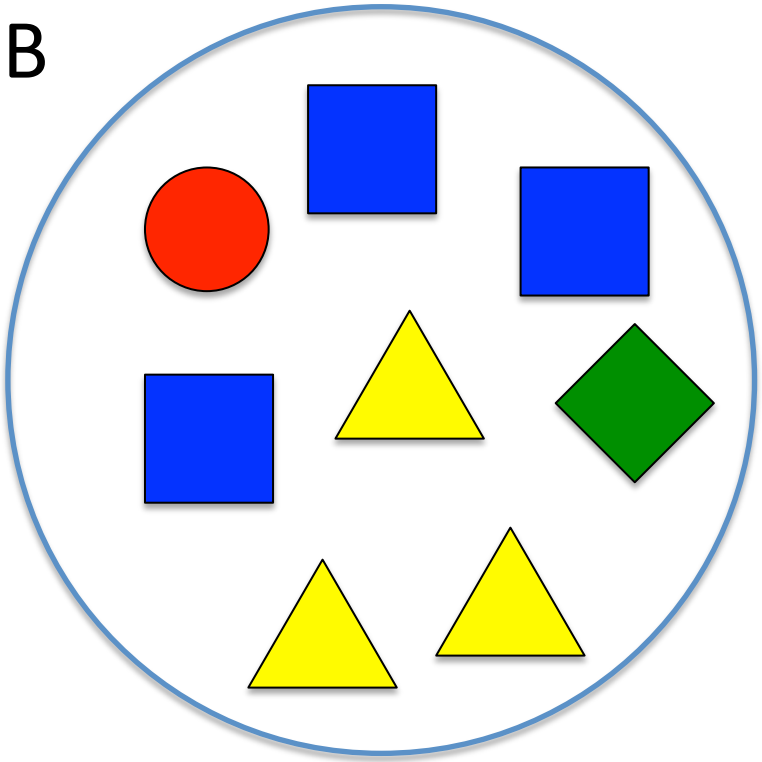
- Alpha is within a sample
  - E.g., how many species are in a sample
- Beta is between samples
  - E.g., how similar are two samples
- Lots of ways to calculate these

Alpha diversity



Rarefaction



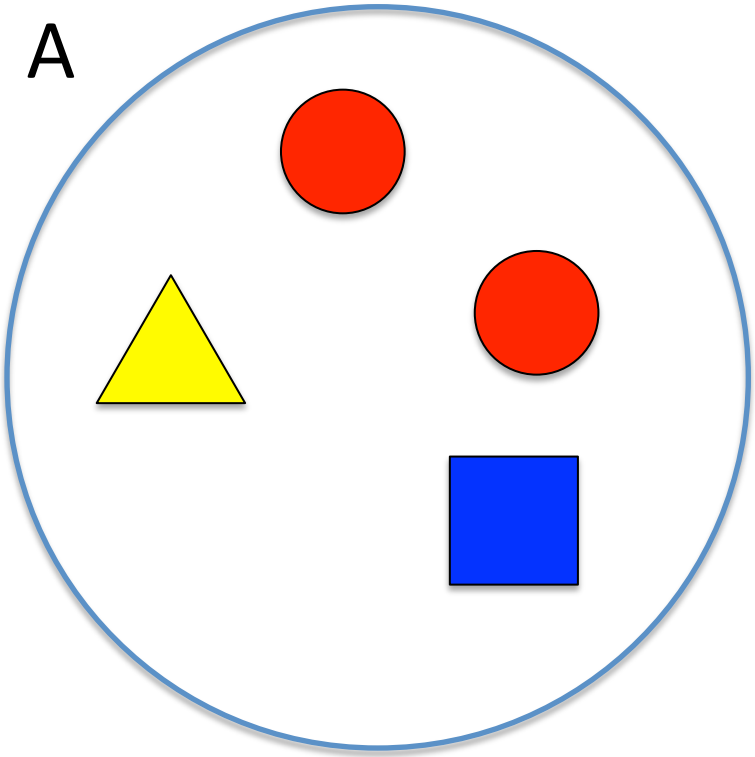


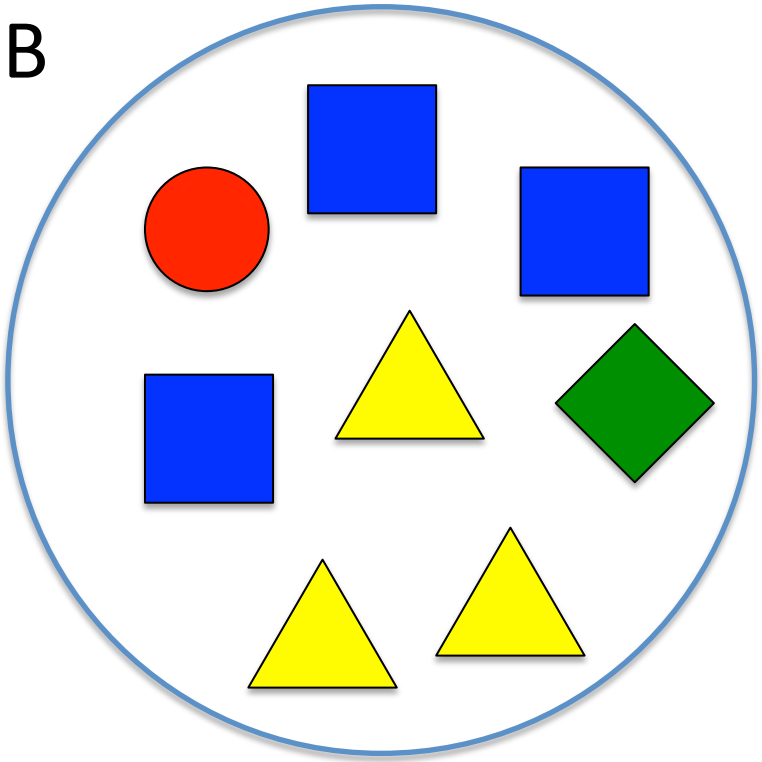
Alpha diversity = 3
Total sequences = 4

Alpha diversity = 4
Total sequences = 8

Randomly select 4 sequences from B

Rarefaction



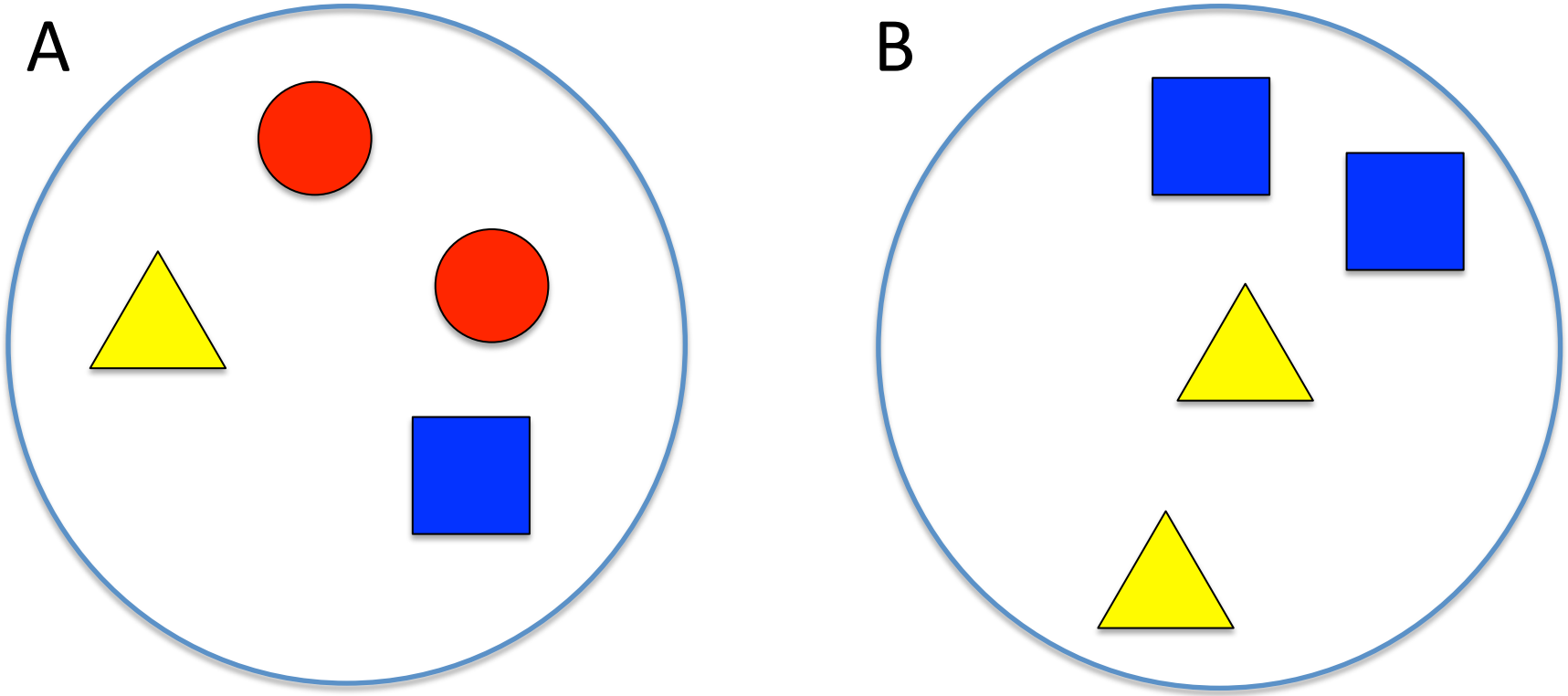


Alpha diversity = 3
Total sequences = 4

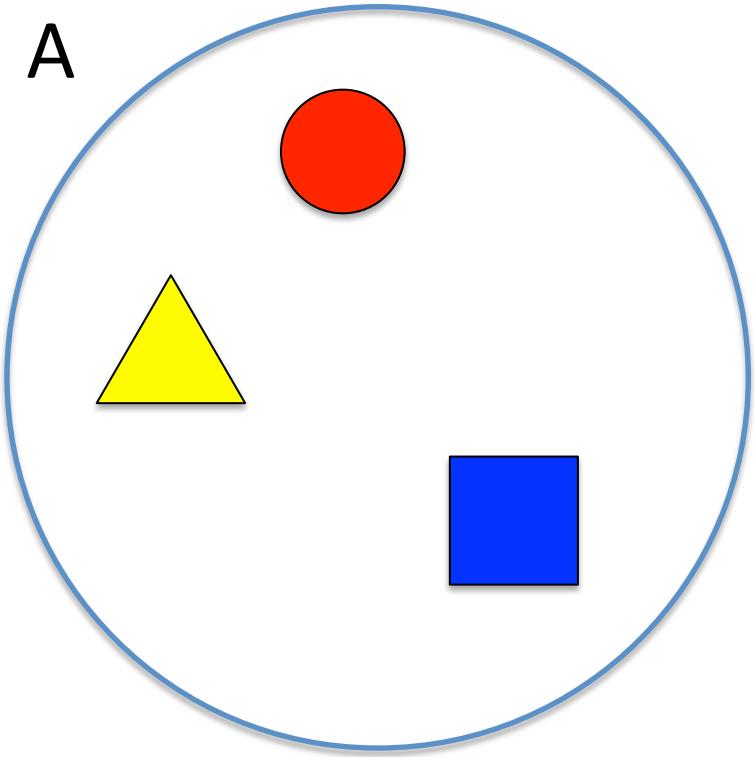
Alpha diversity = 4
Total sequences = 8

Rarefy to 4 sequences

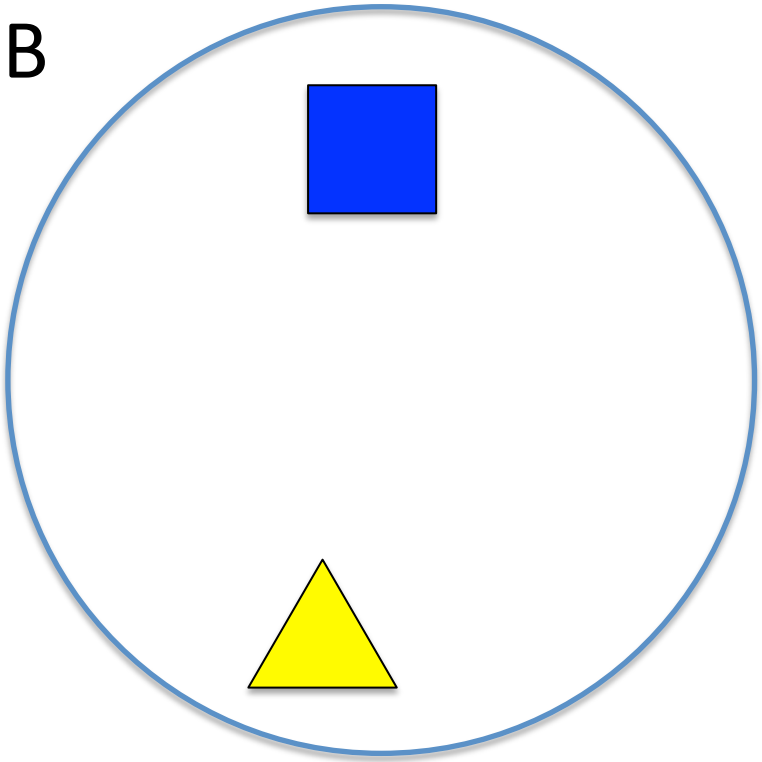
Rarefaction



Rarefaction

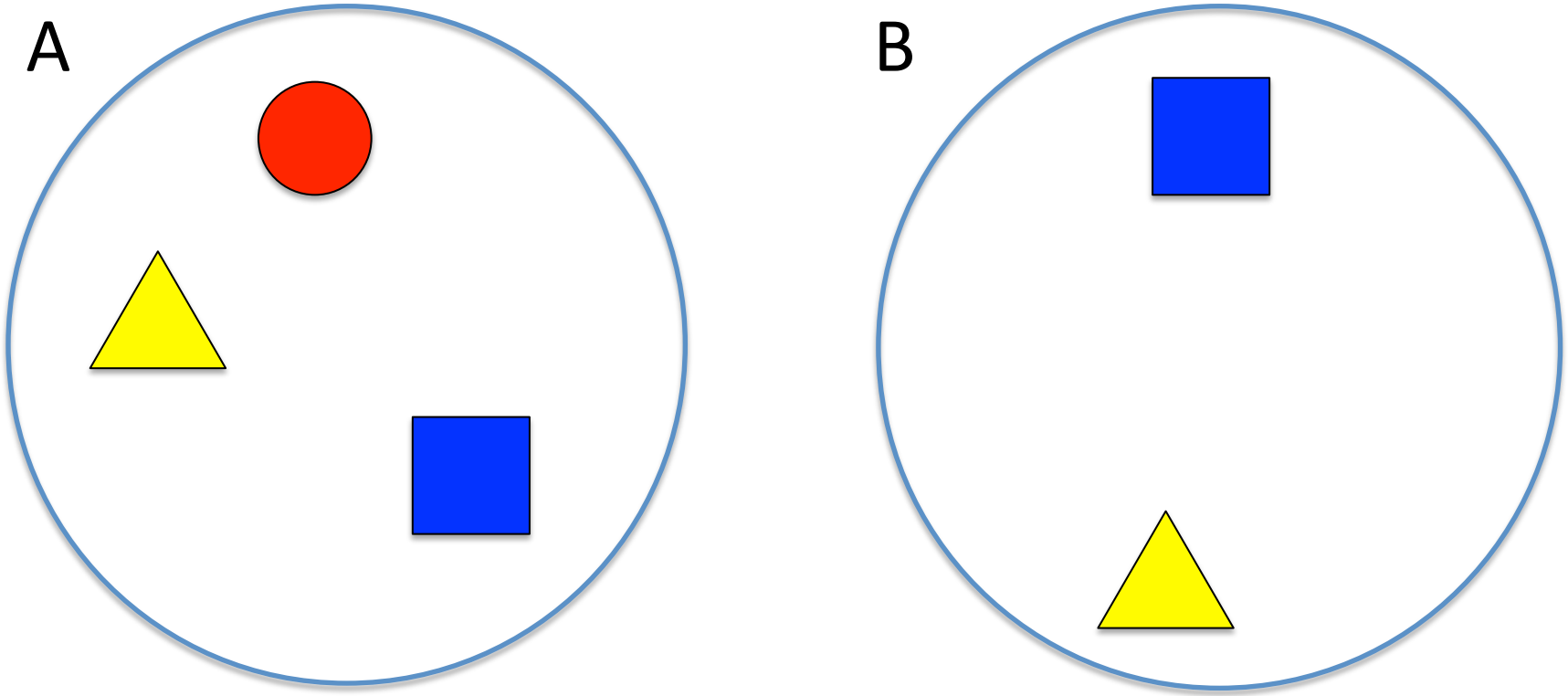


Alpha diversity = 3



Alpha diversity = 2

Rarefaction

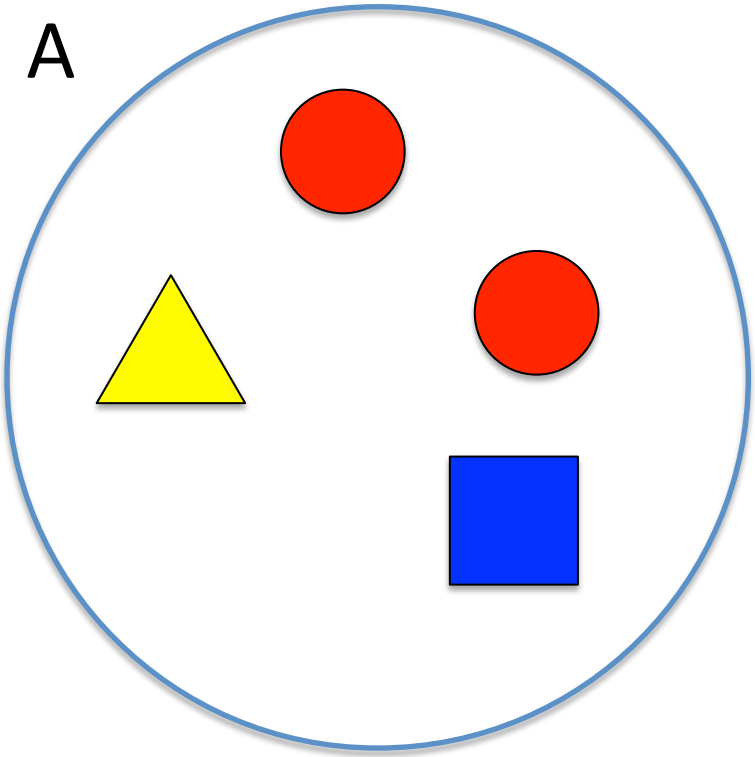


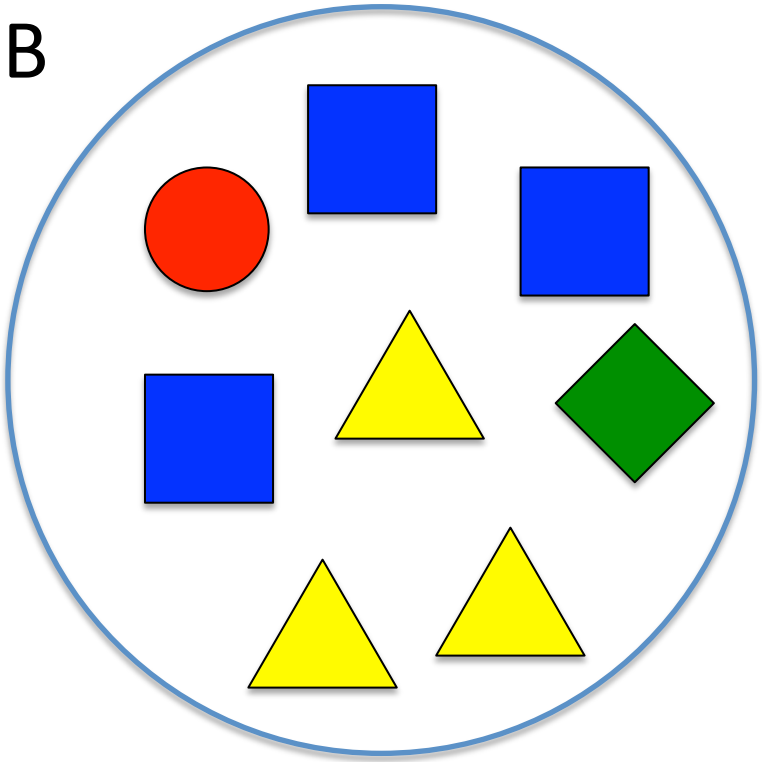
Alpha diversity = 3

Alpha diversity = 2

Sample A is more diverse than sample B

Rarefaction



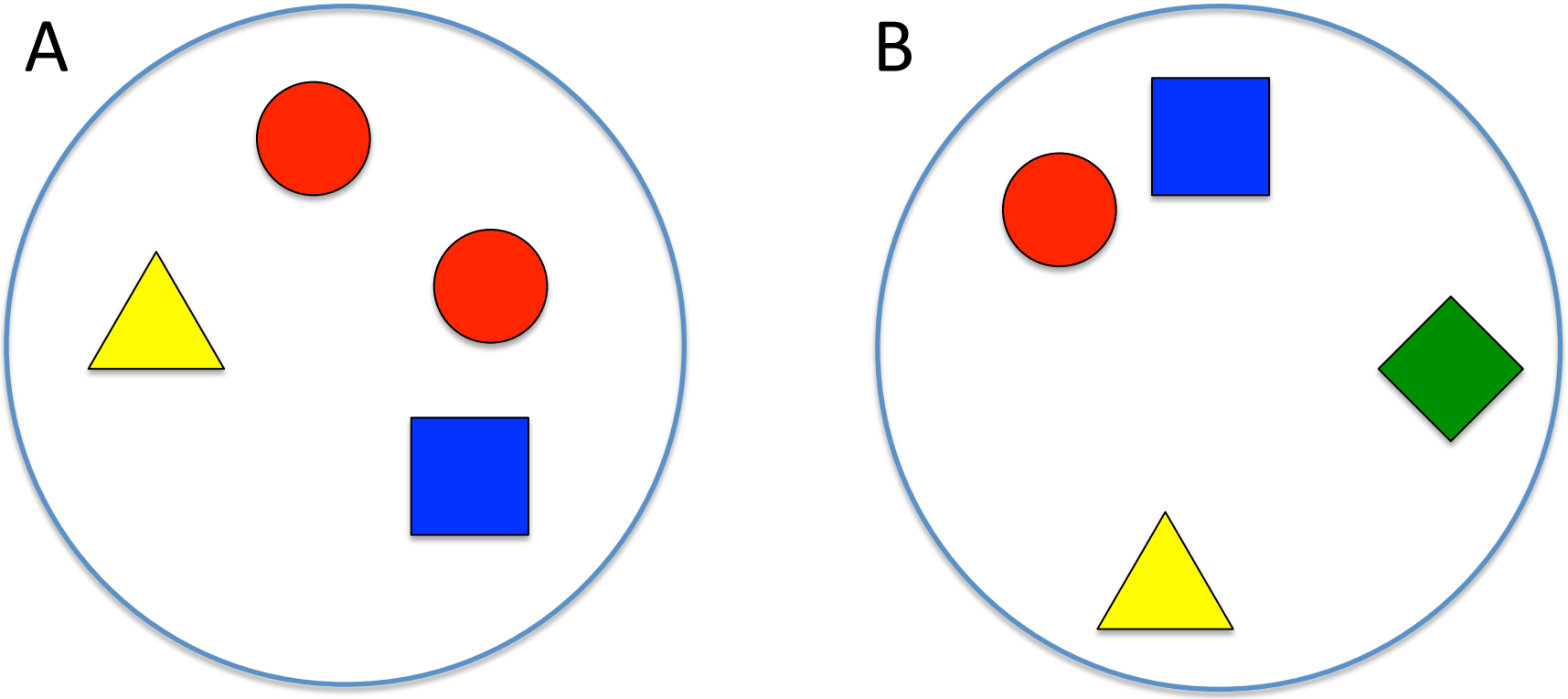


Alpha diversity = 3
Total sequences = 4

Alpha diversity = 4
Total sequences = 8

Rarefy to 4 sequences

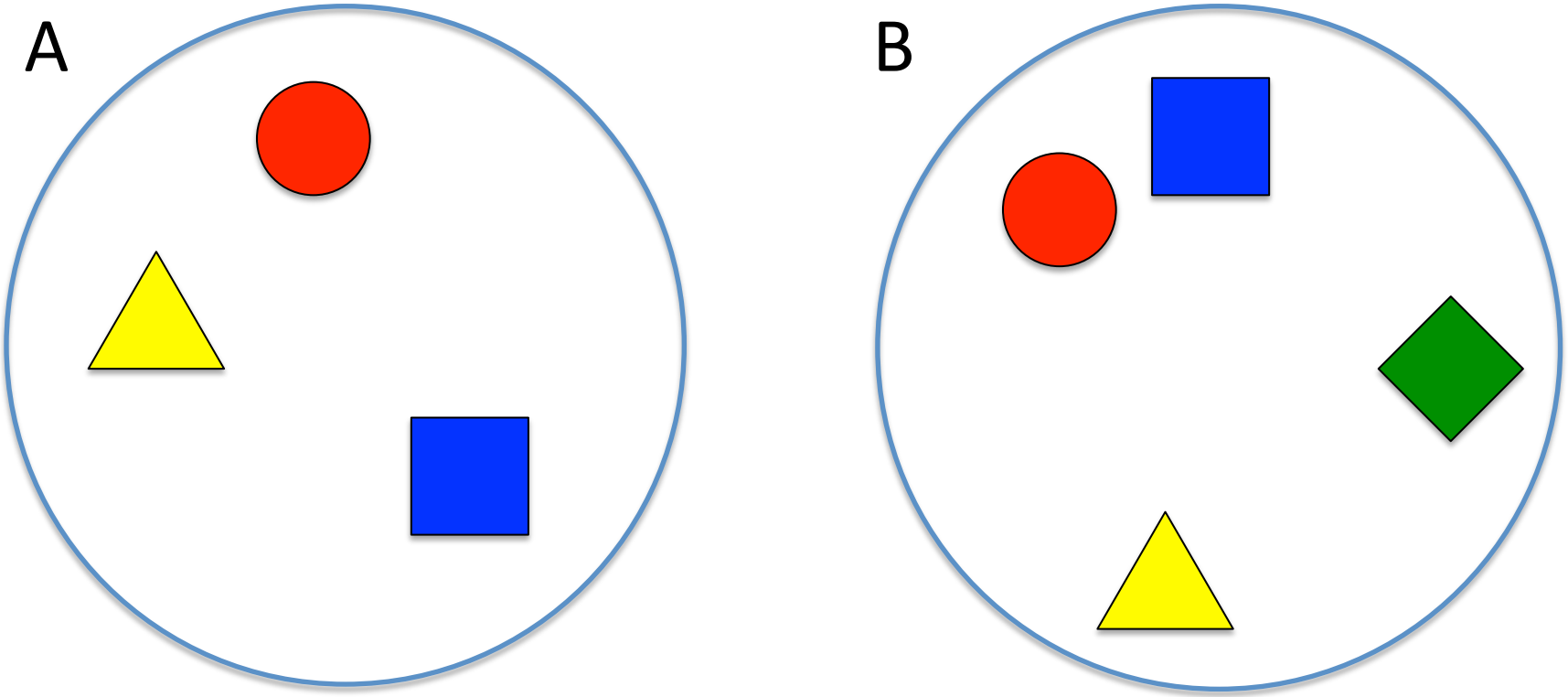
Alpha diversity



Alpha diversity = 3

Alpha diversity = 4

Alpha diversity

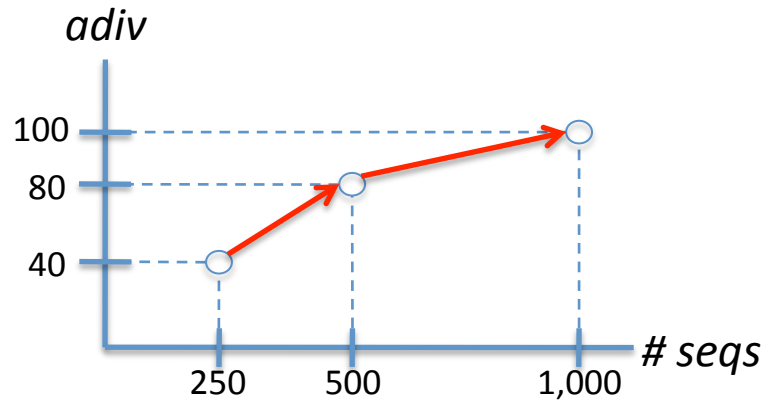


Alpha diversity = 3

Alpha diversity = 4

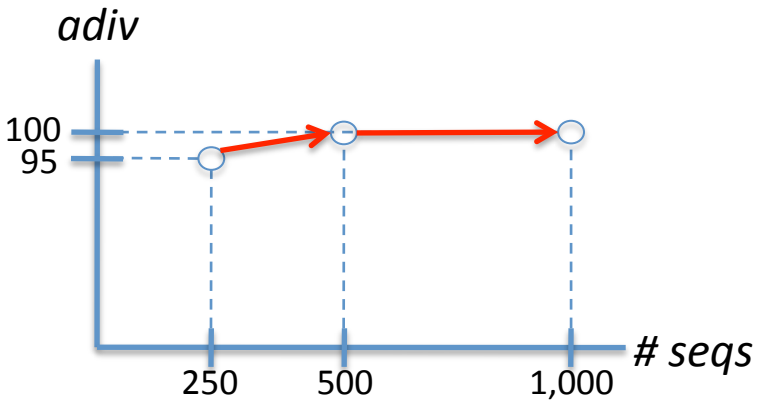
Sample B is more diverse than sample A

Multiple Rarefaction



Higher sequencing effort will probably result in higher observed diversity

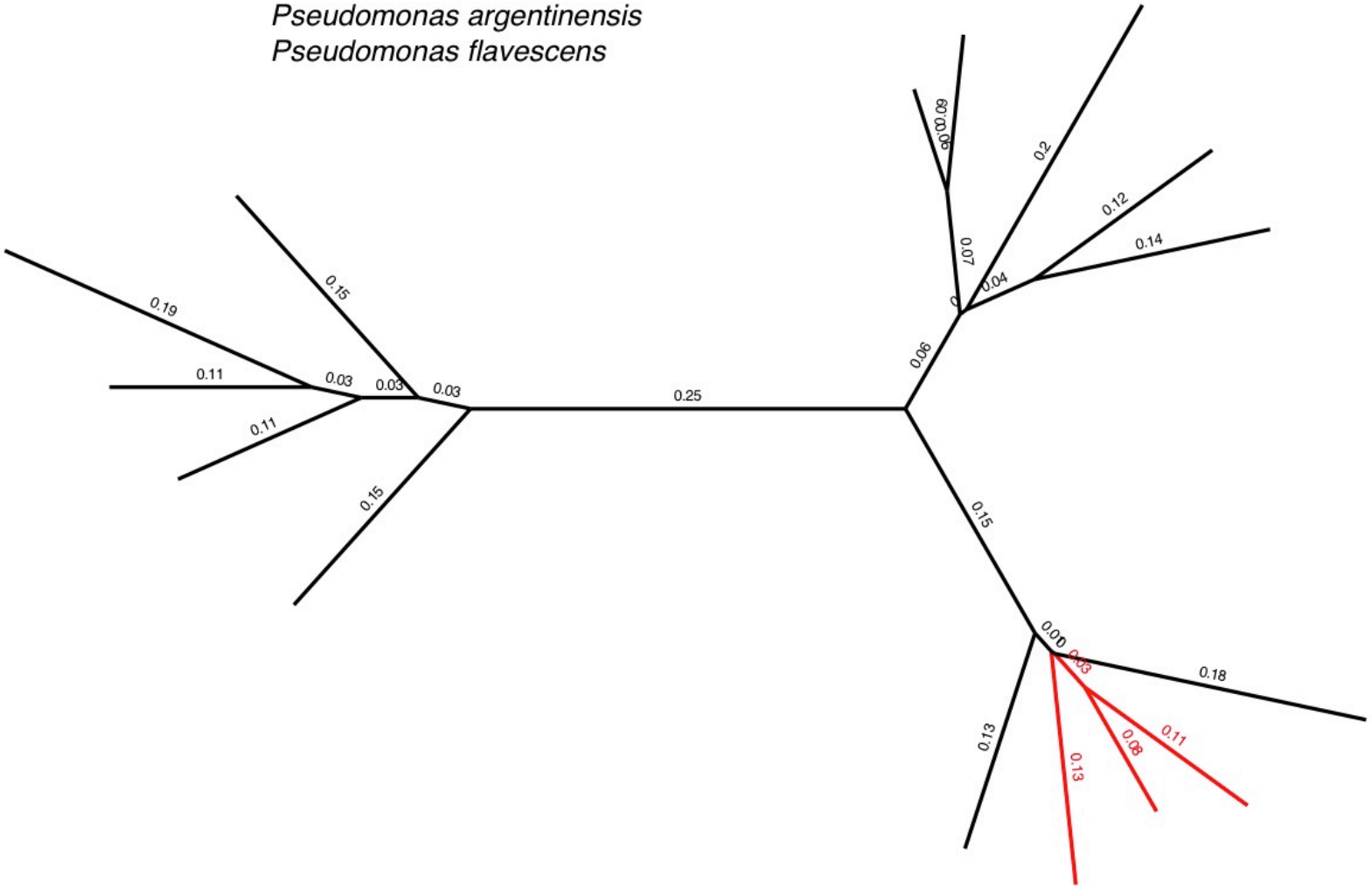
Multiple Rarefaction



Higher sequencing effort will probably not add to observed diversity

Sample A

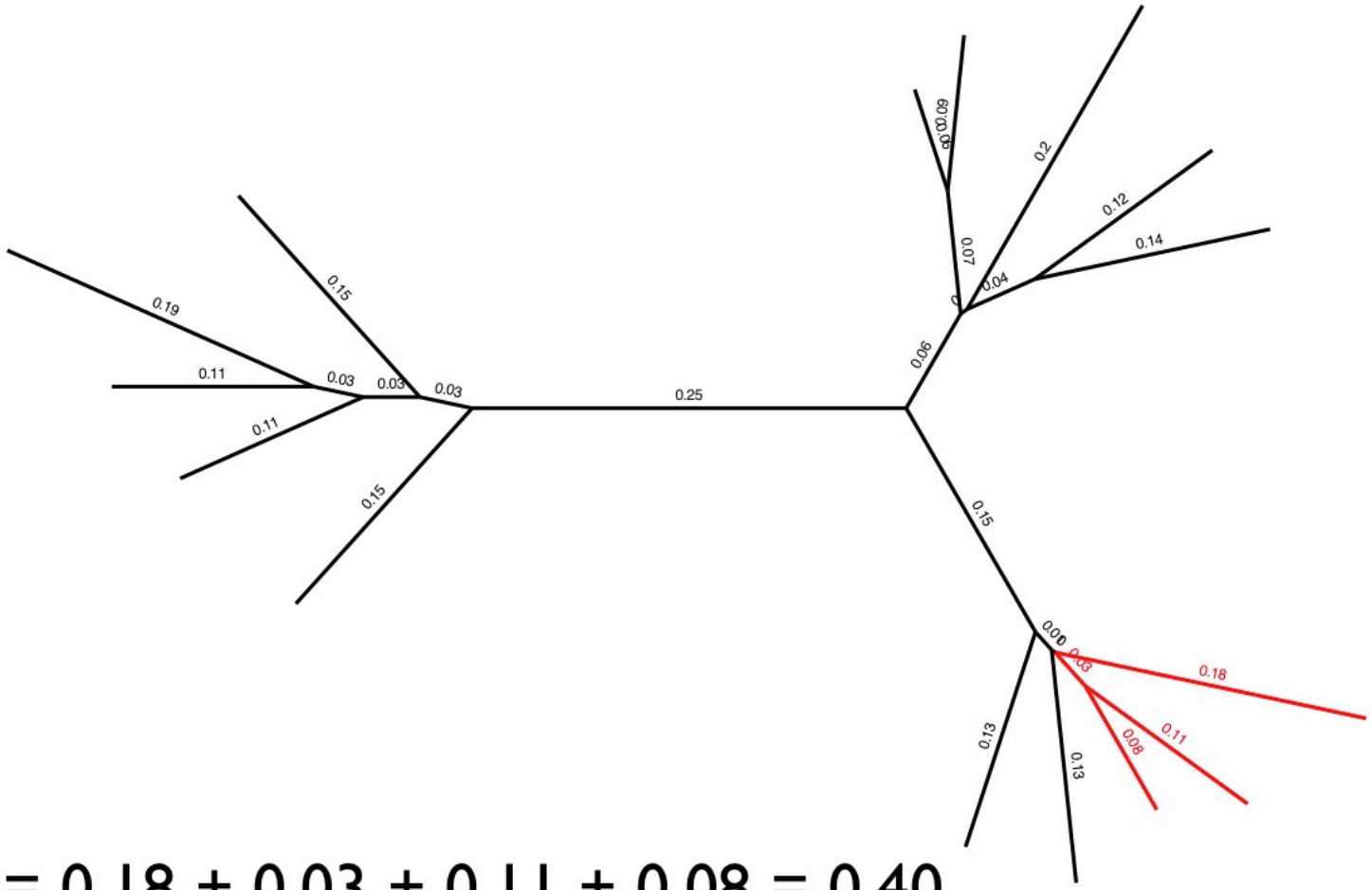
Pseudomonas aeruginosa
Pseudomonas argentinensis
Pseudomonas flavescens



PD = 0.13 + 0.03 + 0.11 + 0.08 = 0.35

Sample B

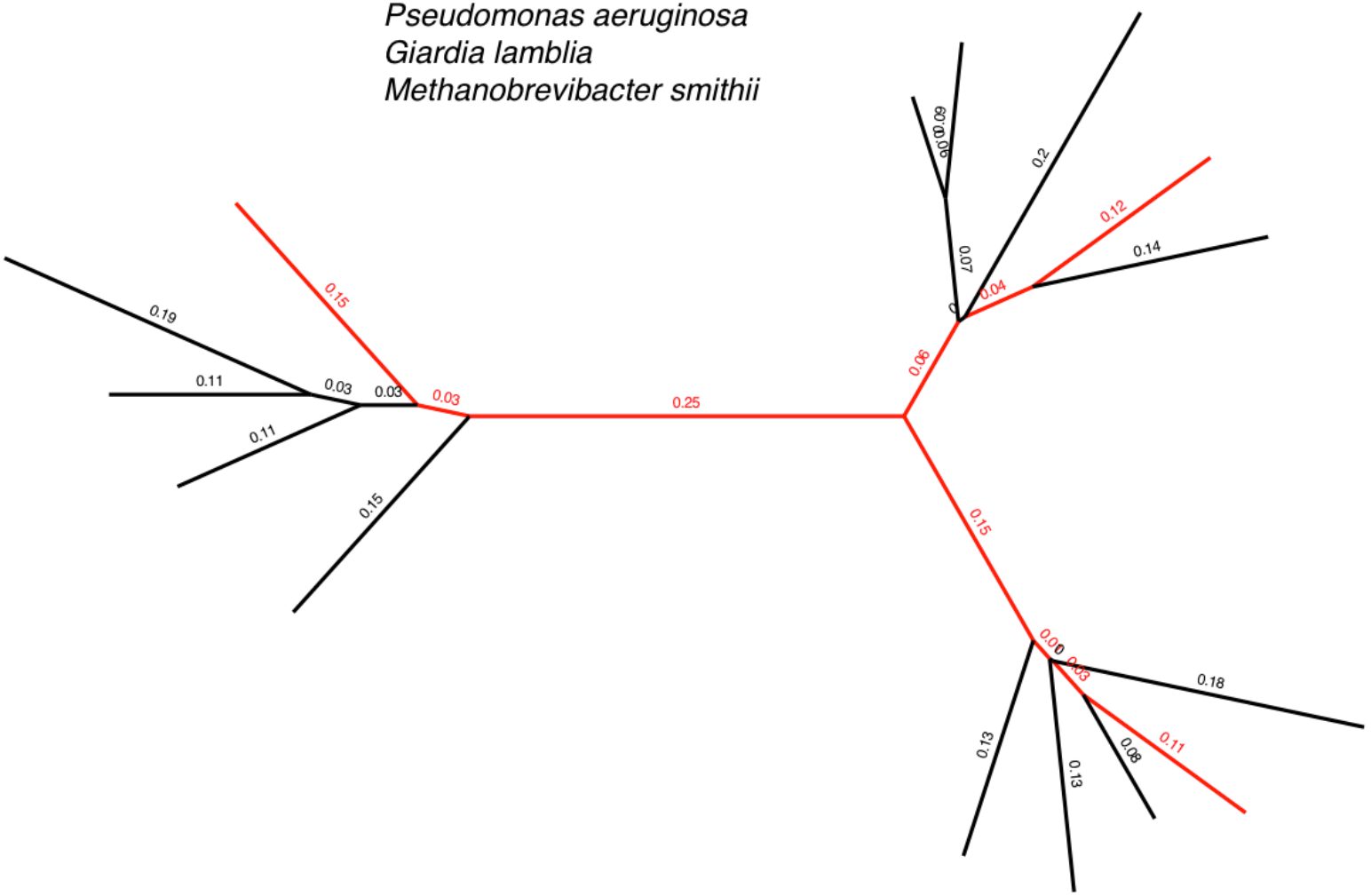
Pseudomonas aeruginosa
Pseudomonas argentinensis
Escherichia coli



PD = 0.18 + 0.03 + 0.11 + 0.08 = 0.40

Sample C

Pseudomonas aeruginosa
Giardia lamblia
Methanobrevibacter smithii



PD = 0.15+0.03+0.25+0.06+0.04+0.12+0.15+0.01+0.03+0.11= 0.95

Phylogenetic Diversity (PD)

Sample A

Pseudomonas aeruginosa
Pseudomonas argentinensis
Pseudomonas flavescens

PD = 0.35

Sample B

Pseudomonas aeruginosa
Pseudomonas argentinensis
Escherichia coli

PD = 0.40

Sample C

Pseudomonas aeruginosa
Giardia lamblia
Methanobrevibacter smithii

PD = 0.95

Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biological Conservation. 61:1-10.

Phylogenetic Diversity (PD)

Sample A
Pseudomonas aeruginosa
Pseudomonas argentinensis
Pseudomonas flavescens

Sample B
Pseudomonas aeruginosa
Pseudomonas argentinensis
Escherichia coli

Sample C
Pseudomonas aeruginosa
Giardia lamblia
Methanobrevibacter smithii

PD = 0.35 < PD = 0.40 < PD = 0.95

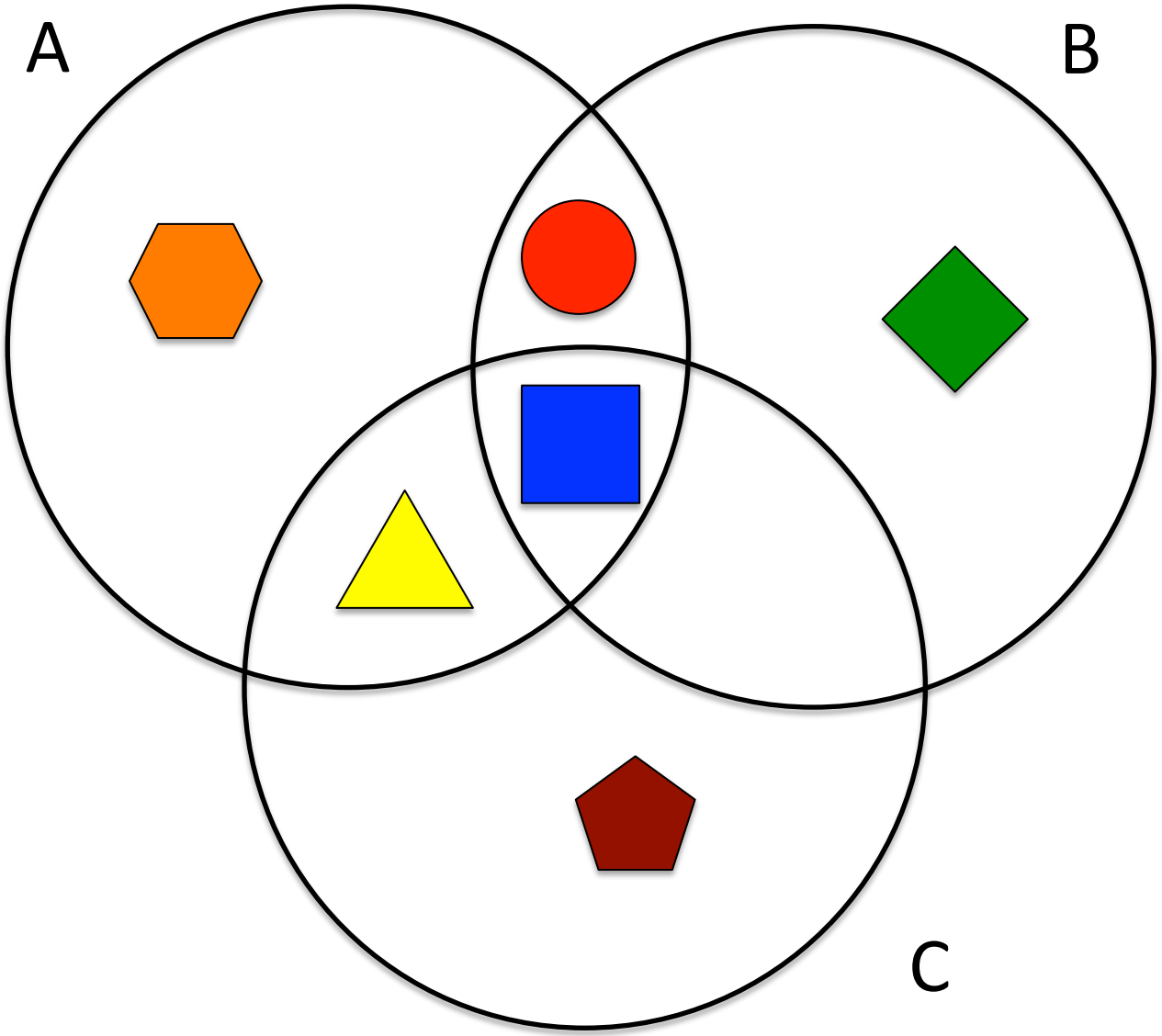
Conclusion:

Sample C is more diverse than sample B, which is more diverse than sample A.

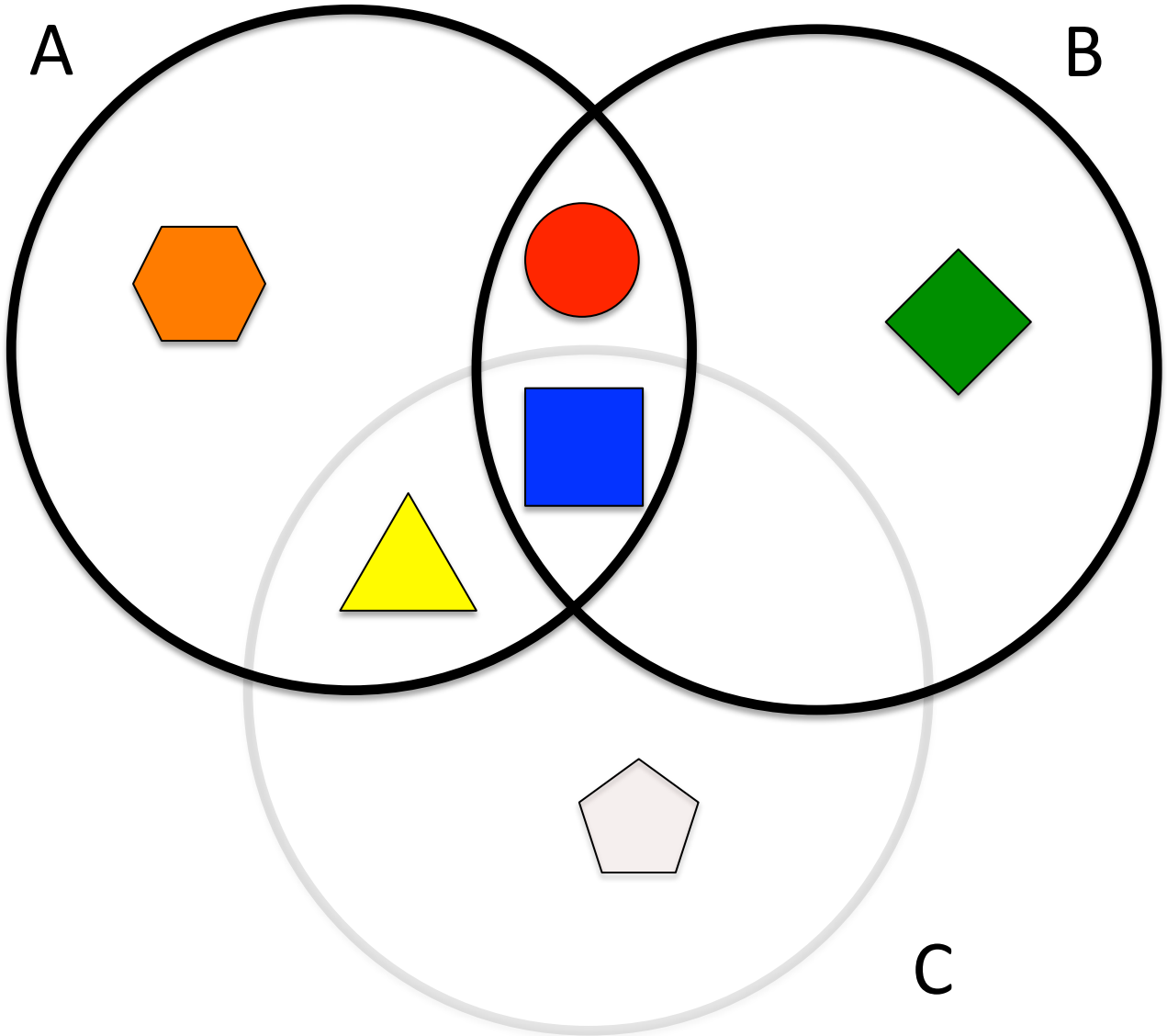
Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biological Conservation. 61:1-10.

Diversity and statistical analysis Part II

Beta diversity

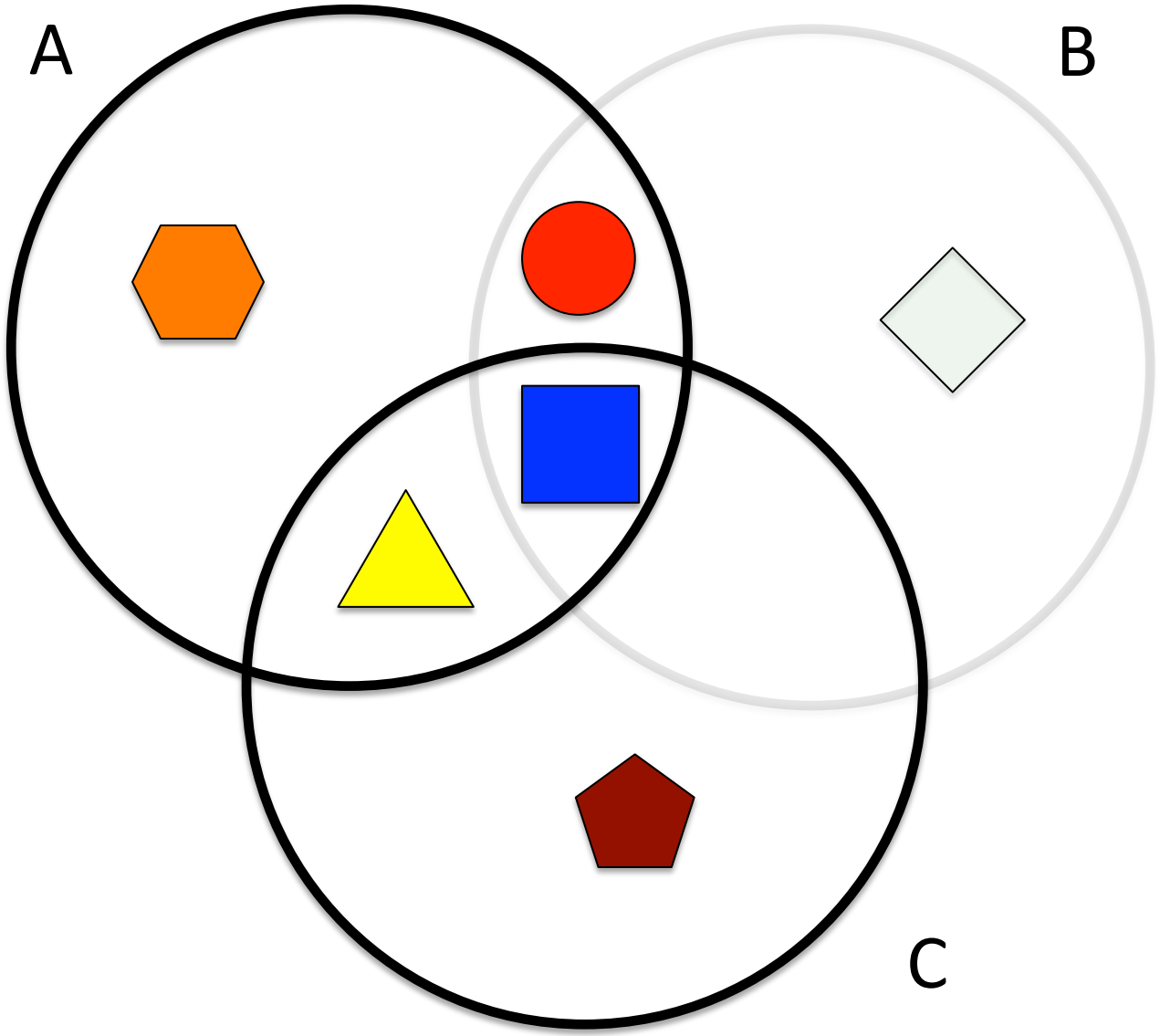


Beta diversity



A x B = 2

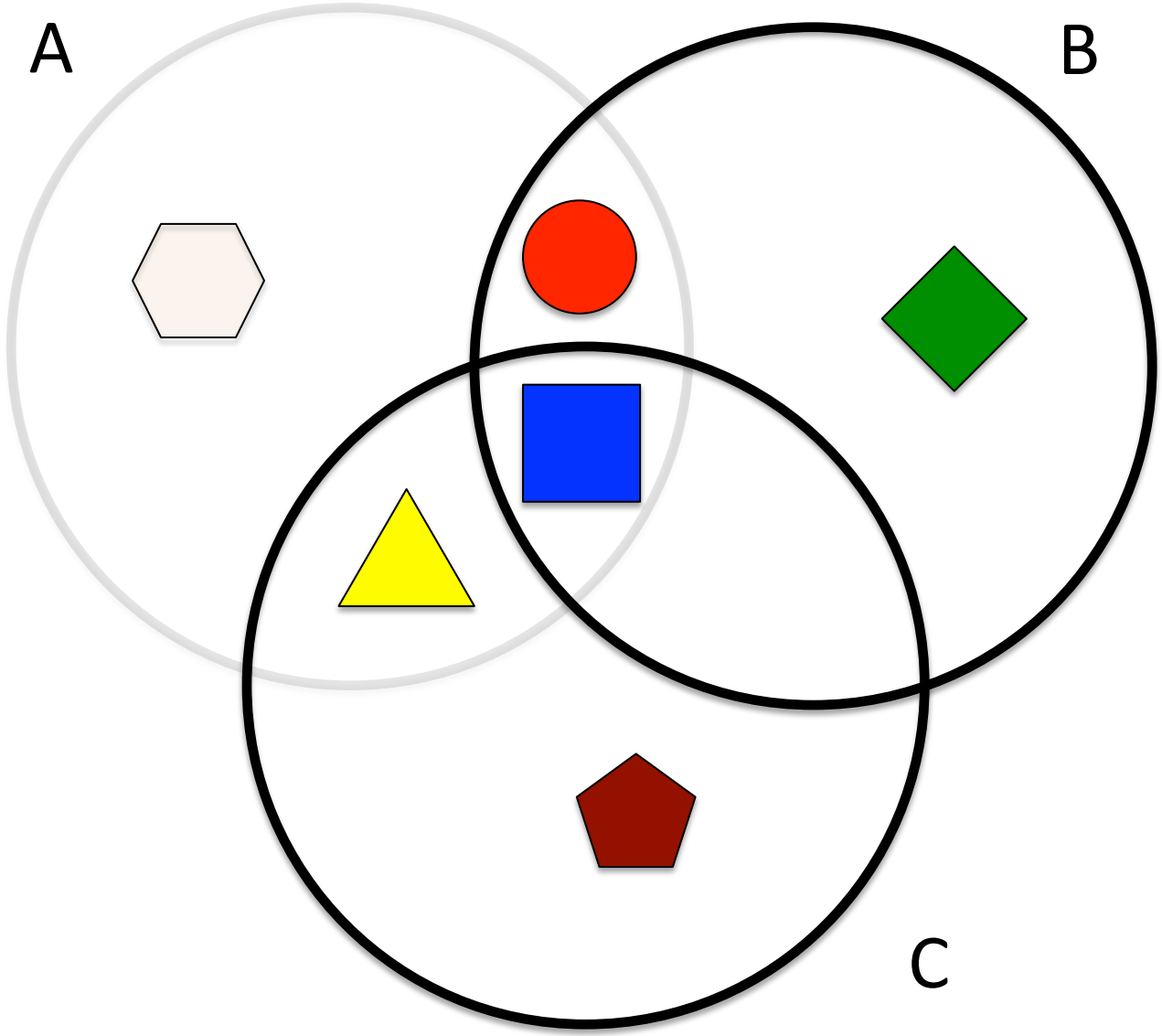
Beta diversity



A x B = 2

A x C = 2

Beta diversity

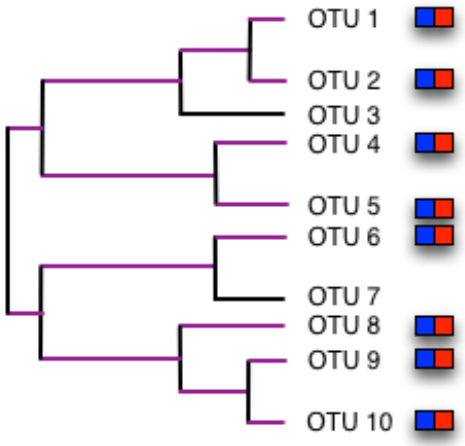


A x B = 2

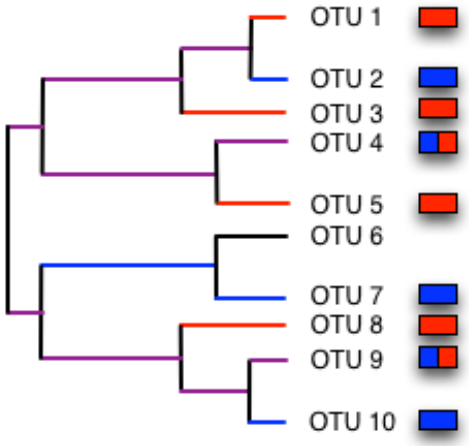
A x C = 2

B x C = 1

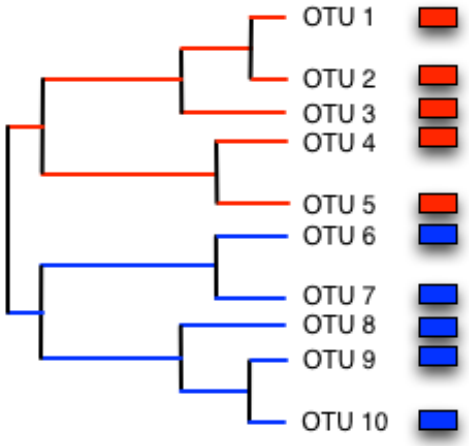
Unweighted UniFrac: a phylogenetic measure of the dissimilarity of microbial communities



U = 0.0
Identical communities



U ≈ 0.5
Related communities

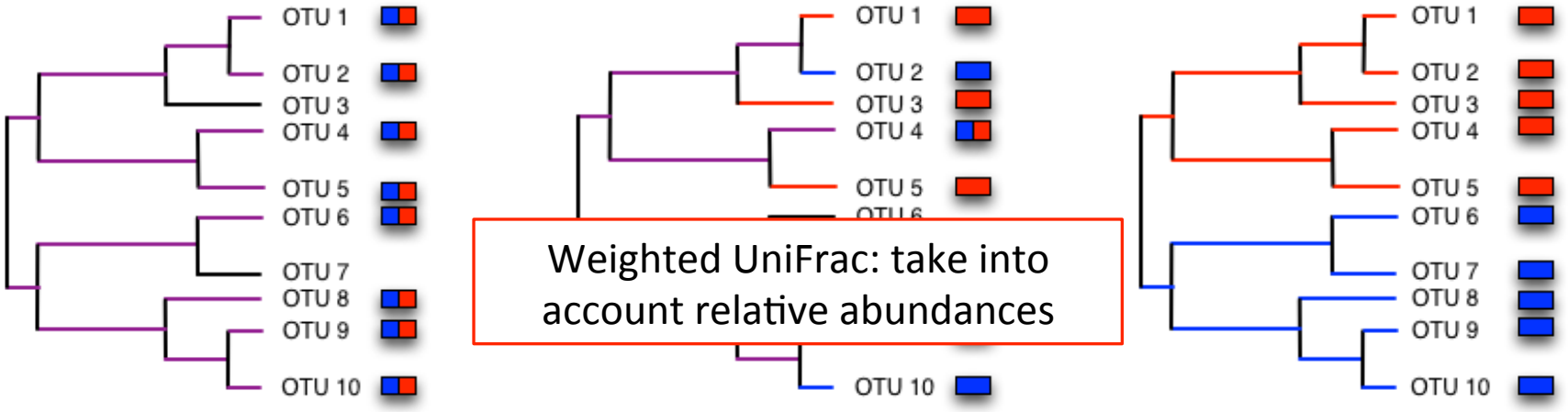


U = 1.0
Unrelated communities

Percent of observed branch length that is unique to either sample

Lozupone and Knight, 2005, Appl Environ Microbiol 71:8228

Unweighted UniFrac: a phylogenetic measure of the dissimilarity of microbial communities



U = 0.0
Identical communities

U ≈ 0.5
Related communities

U = 1.0
Unrelated communities

Percent of observed branch length that is unique to either sample

Lozupone and Knight, 2005, Appl Environ Microbiol 71:8228

Beta diversity comparison:

visually with ordination plots (e.g., PCoA, NMDS)

statistically (e.g., PERMANOVA, ANOSIM*)

Sample ID	Sample Type
A	Plant (yellow)
B	Plant (yellow)
C	Turtle (red)
D	Turtle (red)
E	Human (green)
F	Dog (blue)

Unweighted UniFrac distance matrix:

	A	B	C	D	E	F
A	0.00	0.35	0.83	0.83	0.90	0.90
B	0.35	0.00	0.86	0.85	0.92	0.91
C	0.83	0.86	0.00	0.25	0.88	0.87
D	0.83	0.85	0.25	0.00	0.88	0.88
E	0.90	0.92	0.88	0.88	0.00	0.50
F	0.90	0.91	0.87	0.88	0.50	0.00

* In QIIME 1.9.1, use `compare_categories.py`

(Not real data!)

Beta diversity comparison:

visually with ordination plots (e.g., PCoA, NMDS)

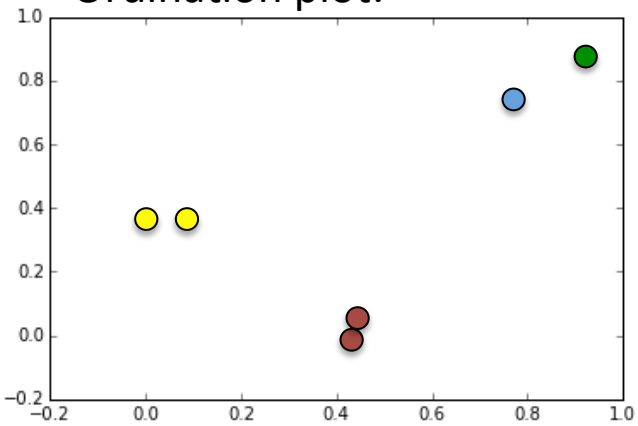
statistically (e.g., PERMANOVA, ANOSIM*)

Sample ID	Sample Type
A	Plant (yellow)
B	Plant (yellow)
C	Turtle (red)
D	Turtle (red)
E	Human (green)
F	Dog (blue)

Unweighted UniFrac distance matrix:

	A	B	C	D	E	F
A	0.00	0.35	0.83	0.83	0.90	0.90
B	0.35	0.00	0.86	0.85	0.92	0.91
C	0.83	0.86	0.00	0.25	0.88	0.87
D	0.83	0.85	0.25	0.00	0.88	0.88
E	0.90	0.92	0.88	0.88	0.00	0.50
F	0.90	0.91	0.87	0.88	0.50	0.00

Ordination plot:



(Not real data!)

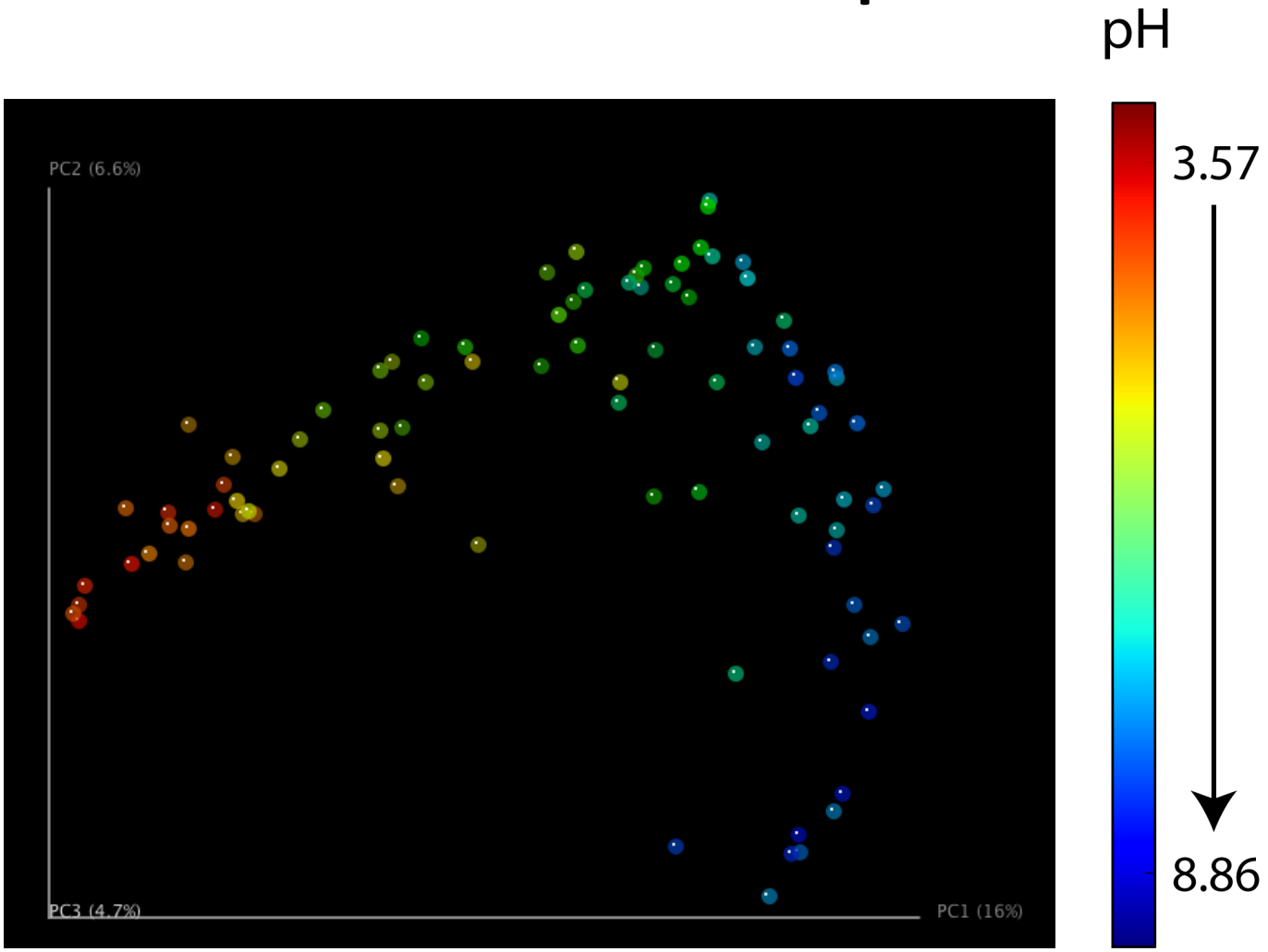
* In QIIME 1.9.1, use `compare_categories.py`

So we've got a distance matrix: now what?

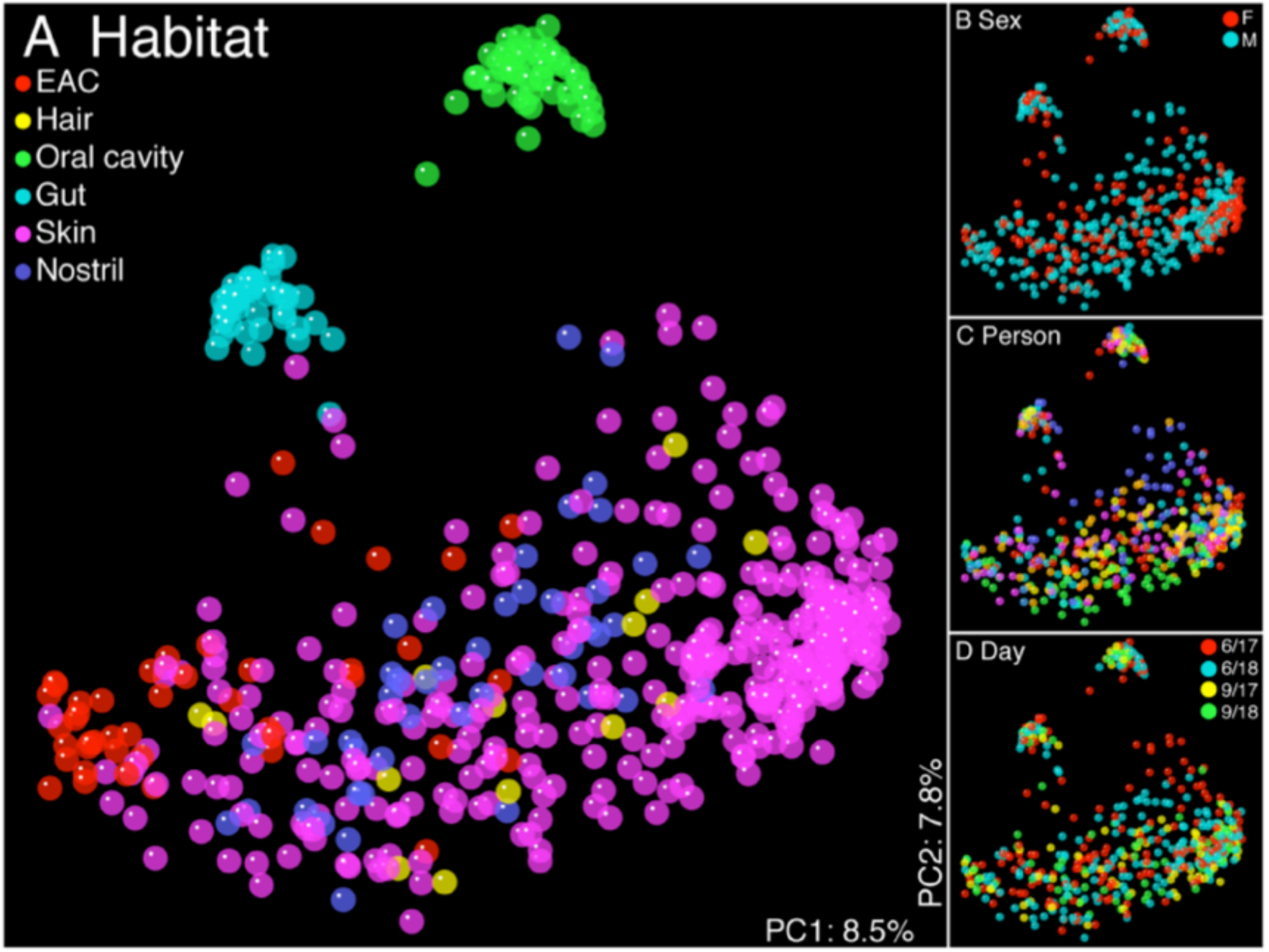
unweighted_unifrac_dm.txt

	Stillton4R2	Stillton4R3	Stillton4R1	HCanyon3R3	HCanyon3R2	HCanyon3R1	Halls9R2	HCanyon2R2	HCanyon2R3	HCanyon2R1	Halls9R1	Stillton10R3	HCanyon0R1	HCanyon0R2	HCanyon0R3	HCa
Stillton4R2	0.0	0.382273294624	0.391416675288	0.560309484808	0.553938232028	0.566136031815	0.557134987546	0.531719852875	0.53655901824	0.567041909667	0.574831935502	0.474244845718				
Stillton4R3	0.382273294624	0.0	0.394399899497	0.586303332083	0.583969320358	0.589363298118	0.588066262733	0.560923487354	0.57767776062	0.601572335069	0.609758664376	0.479910046195				
Stillton4R1	0.391416675288	0.394399899497	0.0	0.580628929798	0.583767102601	0.584463513906	0.573828216898	0.550892933103	0.566056393448	0.584611276938	0.607144746402	0.469209424282				
HCanyon3R3	0.560309484808	0.586303332083	0.580628929798	0.0	0.354600316745	0.345731612479	0.458450145834	0.412488628393	0.385127601086	0.384525100123	0.46823652766	0.585402372425				
HCanyon3R2	0.553938232028	0.583969320358	0.583767102601	0.354600316745	0.0	0.36256949665	0.462707750398	0.414528760004	0.384537766442	0.380087766632	0.46109426257	0.574488221279				
HCanyon3R1	0.566136031815	0.589363298118	0.584463513906	0.345731612479	0.36256949665	0.0	0.452351806146	0.415483559719	0.398796424875	0.381981707704	0.485596583752	0.580926122772				
Halls9R2	0.557134987546	0.588066262733	0.573828216898	0.458450145834	0.462707750398	0.452351806146	0.0	0.447883445295	0.429943464459	0.409064513124	0.344264504725	0.561110815583				
HCanyon2R2	0.531719852875	0.560923487354	0.550892933103	0.412488628393	0.414528760004	0.415483559719	0.447883445295	0.0	0.404179520995	0.388727659604	0.468345488157	0.549328940024				
HCanyon2R3	0.53655901824	0.57767776062	0.566056393448	0.385127601086	0.385453766442	0.398796424875	0.429943464459	0.404179520995	0.0	0.361444528983	0.439367245599	0.566330894525				
HCanyon2R1	0.567041909667	0.601572335069	0.584611276938	0.384525100123	0.380087766632	0.381981707704	0.409064513124	0.388727659604	0.361444528983	0.0	0.454633280595	0.582635621853				
Halls9R1	0.574831935502	0.609758664376	0.607144746402	0.46823652766	0.46109426257	0.485596583752	0.344264504725	0.468345488157	0.439367245599	0.454633280595	0.0	0.578254150137				
Stillton10R3	0.474244845718	0.479910046195	0.469209424282	0.585402372425	0.574488221279	0.580926122772	0.561110815583	0.549328940024	0.566330894525	0.582635621853	0.578254150137					
HCanyon0R1	0.776189708931	0.779315452075	0.781074100354	0.69071344042	0.692644612288	0.683086513009	0.727520276229	0.709578414217	0.711736423222	0.707912554104	0.745189206089	0.7				
HCanyon0R2	0.797241309582	0.795032626325	0.797809039119	0.713804444578	0.714717508733	0.697490417772	0.75335887058	0.723508997775	0.728837315629	0.722257776622	0.761081079305	0.7				
HCanyon0R3	0.784124036272	0.791711071574	0.795527048923	0.693476495058	0.699412891734	0.688360016409	0.732489014053	0.713925200799	0.708354096579	0.708462144382	0.747778043238	0.7				
HCanyon7R3	0.539194289149	0.563745794923	0.557216919336	0.457709184042	0.467073058252	0.473888308307	0.42997132225	0.455576263999	0.448029116277	0.441566044862	0.444788573194	0.5				
HCanyon7R2	0.665877547192	0.683833494738	0.674471865383	0.55308292175	0.563316168524	0.562611936812	0.546938312206	0.558741865535	0.562533096516	0.534453165883	0.558141564634	0.6				
HCanyon7R1	0.554857668962	0.578652540309	0.570402839551	0.475303182122	0.474309745536	0.462587318718	0.43076590681	0.463222699451	0.452021582021	0.42664191335	0.44552755863	0.5				
HCanyon1R1	0.746066596617	0.754645795641	0.759778994683	0.654674045733	0.651121732227	0.640930093945	0.703134531793	0.67682938985	0.675957137688	0.676955919301	0.714539249748	0.7				
HCanyon1R3	0.751522854919	0.764074726397	0.767226018039	0.641176135477	0.65043468158	0.629924888988	0.702785457767	0.675191090098	0.668258446703	0.66343689429	0.713256418398	0.7				
HCanyon1R2	0.748585453295	0.753698061339	0.759152313976	0.666322115311	0.666043304584	0.654075859985	0.711448710816	0.680563093126	0.679964026606	0.678341781539	0.728364004942	0.7				
HCanyon10R2	0.56835786215	0.584149104488	0.584183423897	0.466055758676	0.464168845592	0.470680372129	0.458866733827	0.4814190318	0.460896823727	0.442543616102	0.4619867362	0.5				
HCanyon10R3	0.570798042641	0.588843077275	0.591942476159	0.50320590177	0.498220269993	0.503094520399	0.460532212318	0.496962587514	0.490816424808	0.497919293873	0.463277414447	0.5				
HCanyon10R1	0.505926071682	0.528347698837	0.529350507225	0.472413250039	0.479553556468	0.478832889376	0.455918418773	0.458027517992	0.428733419477	0.456719821741	0.457712834343	0.5				
HCanyon11R3	0.582831143875	0.594413158276	0.60681715155	0.463621293703	0.45842280462	0.486656215192	0.430493752745	0.47439497384	0.445815433526	0.432013882772	0.437061115715	0.5				
HCanyon11R2	0.52855255543	0.55774407831	0.555177188694	0.432763917311	0.429877059678	0.440567014981	0.41332040453	0.426696806769	0.398447088329	0.396059530935	0.433778641456	0.5				
HCanyon11R1	0.555250897098	0.579991048797	0.575126874355	0.453486254954	0.44588780311	0.456712334902	0.419232496891	0.446896791255	0.430325765526	0.428966840449	0.431167398777	0.5				
HCanyon6R2	0.568527068157	0.58513246508	0.583868790939	0.430174126838	0.419369519662	0.432708499988	0.416828113391	0.445815626173	0.426928785146	0.410839210322	0.422526407071	0.5				
HCanyon6R3	0.566346578975	0.587780110737	0.586940661236	0.438634245182	0.425883025257	0.449519180577	0.424862337935	0.447100684715	0.425434475646	0.426218627557	0.424693270252	0.5				
HCanyon6R1	0.583972679946	0.586778207344	0.581534515565	0.434912987895	0.424307243199	0.441765941119	0.415236907762	0.428454635151	0.428113957425	0.406954252133	0.422536839493	0.5				
HCanyon5R1	0.558910457622	0.57470972671	0.562821643814	0.440681773256	0.451596106791	0.444216261284	0.424754580111	0.435763316753	0.417737787339	0.416631599644	0.455877139948	0.5				
HCanyon5R3	0.559987492698	0.5820196982	0.584250873735	0.450336948189	0.426574240559	0.448524323812	0.439745925494	0.440643510478	0.417613041188	0.413563495656	0.464820696561	0.5				
HCanyon5R2	0.577951660858	0.593568671618	0.596389572645	0.422545765978	0.409515246412	0.434403209926	0.429247944328	0.453870629905	0.431590973836	0.41366014112	0.436175614058	0.5				
HCanyon4R1	0.516189922282	0.555158871988	0.544024708263	0.416449154884	0.420465768689	0.437380878857	0.46015301609	0.459537309642	0.42519226313 2	0.439771880651	0.474121489459	0.5				
HCanyon4R2	0.601827528628	0.632734373589	0.628005145571	0.414666304996	0.417716489314	0.443390188238	0.497188855337	0.473766452031	0.447416361784	0.454050692681	0.502442274303	0.6				
HCanyon4R3	0.624113459554	0.653406187405	0.649892943176	0.438848993112	0.452700670776	0.4566604112	0.506388612937	0.493725468949	0.479954328128	0.462327913958	0.510888415471	0.6				
HCanyon12R1	0.513376891954	0.547194076841	0.519066521533	0.471737648119	0.464638931219	0.461841527684	0.430362902794	0.439277983407	0.419105789781	0.41992903626	0.466361101825	0.5				
HCanyon12R2	0.552892298156	0.560516044793	0.549257756132	0.504271040435	0.504749255647	0.517681811493	0.453014756694	0.480827841616	0.489487879426	0.485996535935	0.465436661211	0.5				
HCanyon12R3	0.553185666616	0.585465060164	0.585665024134	0.439712101711	0.431336995109	0.463612059381	0.409326408365	0.435088785406	0.414920465304	0.407840817122	0.422739541765	0.5				
HCanyon8R1	0.552014316498	0.559504417239	0.562836483618	0.483602818412	0.470838841272	0.485053828711	0.446654297414	0.465192835869	0.464892486774	0.441733317446	0.466837938413	0.5				
HCanyon8R2	0.547616166832	0.576958110793	0.578527645361	0.458907582912	0.452778768072	0.456800424019	0.420176245854	0.449481849055	0.430019204227	0.413022684081	0.432050808518	0.5				
HCanyon8R3	0.549015567252	0.573444281355	0.580500957664	0.446205572204	0.446999609932	0.453171155433	0.4327591563	0.439870648001	0.425623148959	0.408822227859	0.446705079363	0.5				
Halls8R2	0.535707795107	0.569114464296	0.574510728774	0.46528819939	0.452053418068	0.460087502347	0.379587194176	0.458489263366	0.421181481291	0.440297005744	0.366881097176	0.5				
Stillton10R1	0.462296292442	0.468444771913	0.470923059335	0.575228561838	0.585038958921	0.5846812234	0.56445075535	0.557710461145	0.559334675708	0.5881972363	0.584771138758					
Stillton10R2	0.538798221769	0.543866219051	0.557970880727	0.543811509839	0.541537645813	0.56301479339	0.486181550953	0.53366160907	0.52362238711	0.535021530589	0.480837812931					
Halls8R1	0.614648837651	0.6401217906	0.644104393777	0.493196498338	0.492370158337	0.486233469981	0.445575362279	0.508540989086	0.484457264025	0.473964462468	0.458044156786	0.6				
Halls6R1	0.484233643835	0.518598404668	0.519082685954	0.476689255656	0.489545391911	0.479649954411	0.484643817074	0.479596366267	0.472472195267	0.47020063594	0.505956019295	0.5				
Halls6R2	0.496757718878	0.528784811375	0.521274731439	0.45457903264	0.434119427081	0.447597701249	0.457381539248	0.442946410151	0.453605919538	0.450370943322	0.476499991041	0.5				
Halls4R2	0.509683393195	0.495879096417	0.488951752223	0.609282443564	0.612813488861	0.603872514737	0.603436206122	0.590101531162	0.603855889606	0.609643606143	0.639977381353	0.5				
Stillton5R2	0.552954619991	0.583608736216	0.583498177592	0.471532456639	0.464725650565	0.479581685887	0.521034266478	0.507044011114	0.481356798038	0.487411548649	0.523238604314	0.5				
Halls4R1	0.549133039964	0.534016529239	0.54216719333	0.597412160187	0.596104351076	0.583900421014	0.611452811208	0.594261248789	0.610285284262	0.62138814384	0.619094392421	0.5				
Halls3R1	0.473322285448	0.512621151343	0.492765444818	0.486663734375	0.479618250755	0.49005497706	0.495762448414	0.481055233469	0.485571399694	0.495623098076	0.512078468475	0.5				

Ordination Techniques



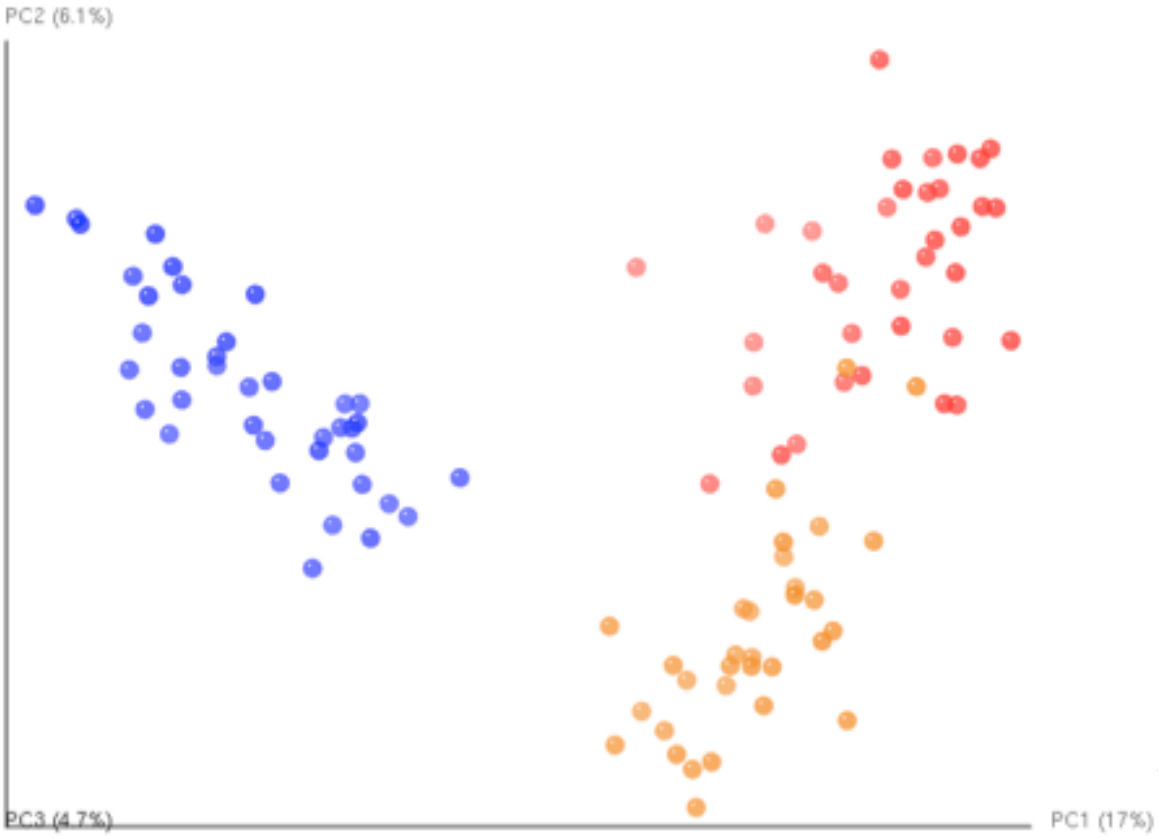
Lauber CL et al. Appl. Environ. Microbiol. 2009;75:5111-5120



Elizabeth K. Costello, et al. Science 2009.
Bacterial Community Variation in Human Body Habitats Across Space and Time.

Variation in sampling depth also needs to be controlled for beta diversity!

Variation in sampling depth is an important consideration

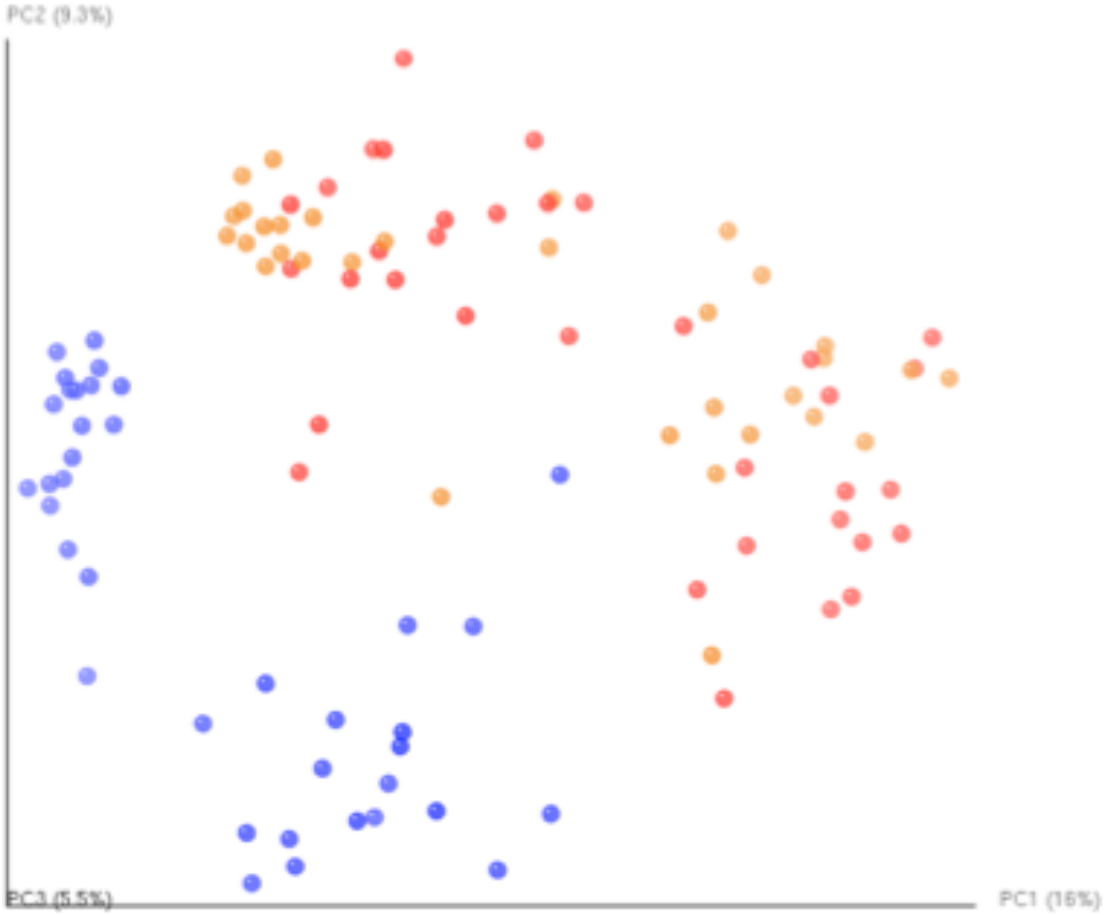


Human skin, colored by **individual**, at 500 sequences/sample

Image/analysis credit: Justin Kuczynski

Data reference:
Forensic identification using skin bacterial communities. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Proc Natl Acad Sci U S A. 2010 Apr 6;107(14):6477-81.

Variation in sampling depth is an important consideration



Human skin, colored by **individual**, at either 50 or 500 sequences/ sample

Image/analysis credit: Justin Kuczynski

Data reference:
Forensic identification using skin bacterial communities. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Proc Natl Acad Sci U S A. 2010 Apr 6;107(14):6477-81.

Variation in sampling depth is an important consideration

Human skin, colored by **sampling depth**, at either 50 (blue) or 500 (red) sequences/sample

Image/analysis credit: Justin Kuczynski

Data reference:
Forensic identification using skin bacterial communities. Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Proc Natl Acad Sci U S A. 2010 Apr 6;107(14):6477-81.

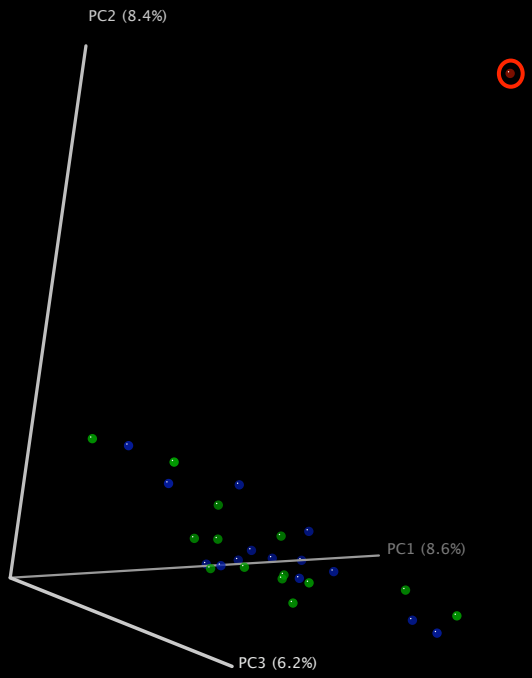
How deep is deep enough?

It depends on the question…

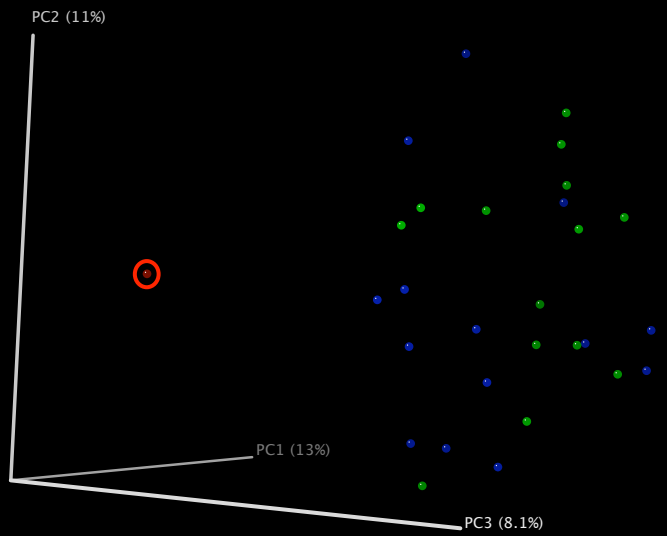
- Differences between community types: not many sequences.
- Rare biosphere: more (but be careful about sequencing noise!)

How deep is deep enough?

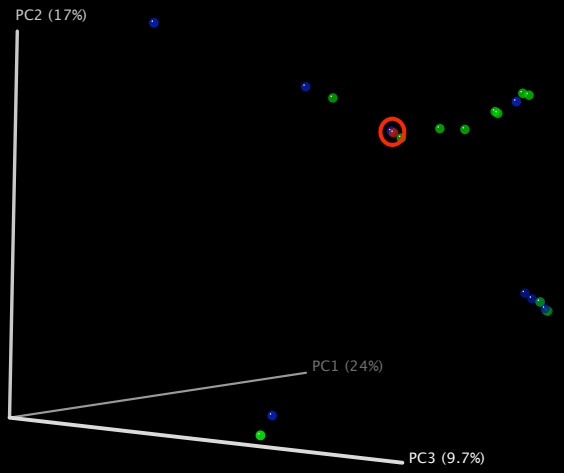
100 sequences/sample



10 sequences/sample



1 sequence/sample



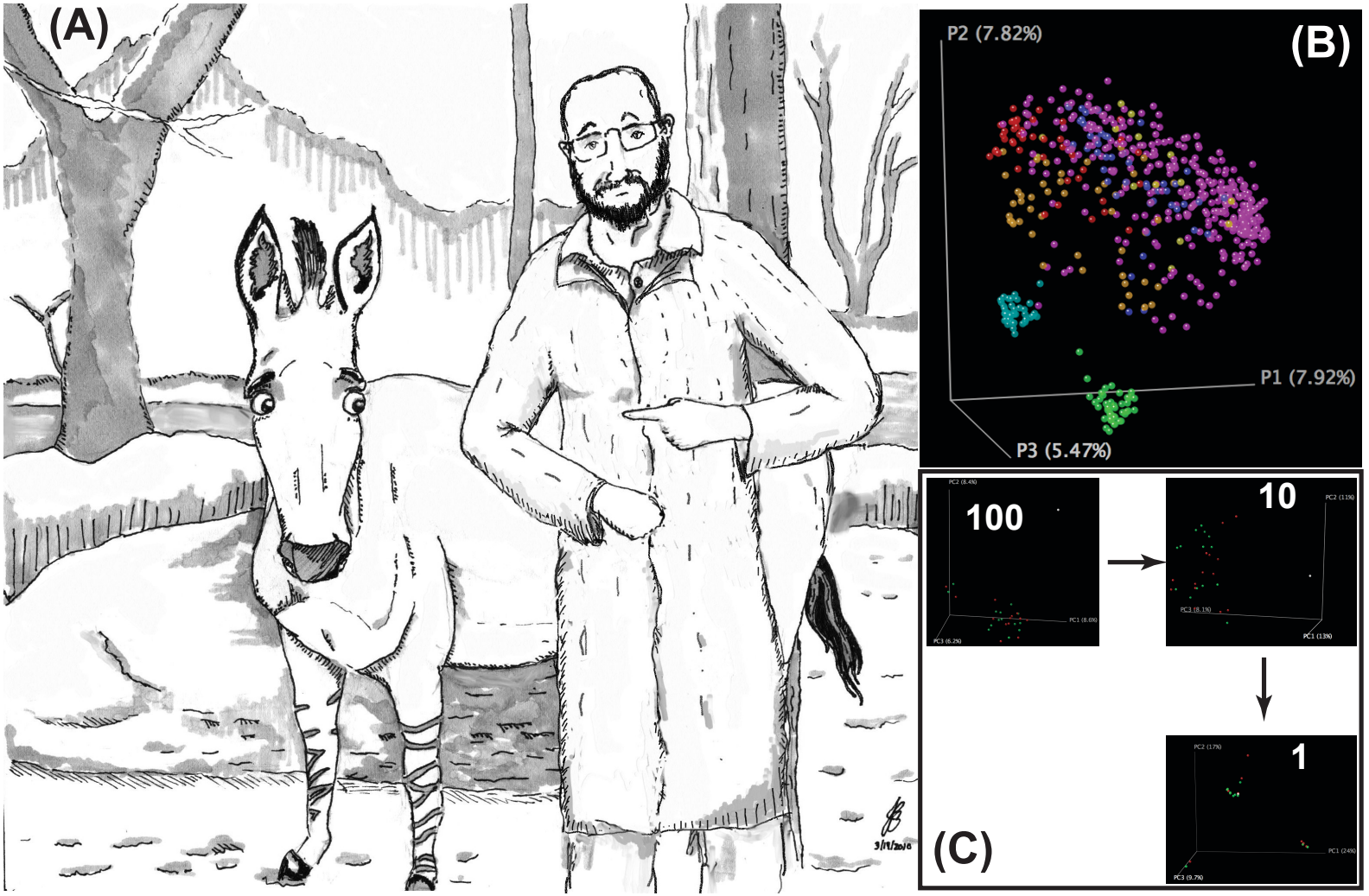
Direct sequencing of the human microbiome readily reveals community differences.
J Kuczynski et al. Genome Biology (2011).

Genome Biol. 2010;11(5):210. Epub 2010 May 5.

Direct sequencing of the human microbiome readily reveals community differences.

Kuczynski J, Costello EK, Nemergut DR, Zaneveld J, Lauber CL, Knights D, Koren O, Fierer N, Kelley ST, Ley RE, Gordon JI, Knight R.

Figure 1



Differentially abundant OTUs

Which OTUs have statistically different abundances between groups of samples?

Example: which OTUs are differentially abundant across human body sites?

Differentially abundant OTUs

group_significance.py

- Kruskal-Wallis (non-parametric, QIIME default)
  - Groups must contain > 4 samples
  - Commonly used for marker gene surveys
- ANOVA (parametric, previous QIIME default)
  - Equal variance
  - Normality of residuals
  - Independence
  - Usually violated by marker gene surveys
- *Many others*

Differentially abundant OTUs

Currently an active area of research, recent evidence suggests other approaches are better:

Waste not, want not: why rarefying microbiome data is inadmissible. McMurdie and Holmes, 2014.
http://www.ncbi.nlm.nih.gov/pubmed/24699258
(Available in QIIME: `differential_abundance.py`)

Analysis of composition of microbiomes… Mandal et al., 2015.
http://www.ncbi.nlm.nih.gov/pubmed/26028277
(Not yet available in QIIME, but coming soon.)

core_diversity_analyses.py

Site index

- Home
- Install
- Documentation
- Tutorials
- Blog
- Developer

Quick search

Go

Enter search terms or a module, class or function name.

core_diversity_analyses.py – A workflow for running a core set of QIIME diversity analyses.

Description:

This script plugs several QIIME diversity analyses together to form a basic workflow beginning with a BIOM table, mapping file, and optional phylogenetic tree.

The included scripts are those run by the workflow scripts alpha_rarefaction.py, beta_diversity_through_plots.py, summarize_taxa_through_plots.py, plus the (non-workflow) scripts make_distance_boxplots.py, compare_alpha_diversity.py, and otu_category_significance.py. To update parameters to the workflow scripts, you should pass the same parameters file that you would pass if calling the workflow script directly.

Usage: `core_diversity_analyses.py [options]`

Input Arguments:

```
[REQUIRED]

-i, --input_biom_fp
    The input biom file [REQUIRED]

-o, --output_dir
    The output directory [REQUIRED]

-m, --mapping_fp
    The mapping filepath [REQUIRED]

-e, --sampling_depth
    Sequencing depth to use for even sub-sampling and maximum rarefaction depth. You should review the output of print_biom_table_summary.py to decide on this value.

[OPTIONAL]

-p, --parameter_fp
    Path to the parameter file, which specifies changes to the default behavior. For more information, see www.qiime.org/documentation/qiime_parameters_files.html [if
    omitted, default values will be used]

-a, --parallel
    Run in parallel where available. Specify number of jobs to start with -O or in the parameters file. [default: False]

--nonphylogenetic_diversity
    Apply non-phylogenetic alpha (chao1 and observed_species) and beta (bray_curtis) diversity calculations. This is useful if, for example, you are working with non-
    amplicon BIOM tables, or if a reliable tree is not available (e.g., if you're working with ITS amplicons) [default: False]

--suppress_taxa_summary
    Suppress generation of taxa summary plots. [default: False]
```

core_diversity_analyses.py

```
$ core_diversity_analyses.py –i otu_table.biom
                              –o core_output
                              –m map.txt
                              –c SampleType
                              –t rep_set.tre
                              –e 20
```

core_diversity_analyses.py

```
$ core_diversity_analyses.py –i otu_table.biom
                              –o core_output
                              –m map.txt
                              –c SampleType
                              –t rep_set.tre
                              –e 20
```

core_diversity_analyses.py

```
$ core_diversity_analyses.py –i otu_table.biom
                             –o core_output
                             –m map.txt
                             –c SampleType
                             –t rep_set.tre
                             –e 20
```

core_diversity_analyses.py

```
$ core_diversity_analyses.py –i otu_table.biom
                              –o core_output
                              –m map.txt
                              –c SampleType
                              –t rep_set.tre
                              –e 20
```

core_diversity_analyses.py

```
$ core_diversity_analyses.py –i otu_table.biom
                              –o core_output
                              –m map.txt
                              –c SampleType
                              –t rep_set.tre
                              –e 20
```

core_diversity_analyses.py

```
$ core_diversity_analyses.py –i otu_table.biom
                              –o core_output
                              –m map.txt
                              –c SampleType
                              –t rep_set.tre
                              –e 20
```

core_diversity_analyses.py

```
$ core_diversity_analyses.py –i otu_table.biom
                              –o core_output
                              –m map.txt
                              –c SampleType
                              –t rep_set.tre
                              –e 20
```

Hands-on with `core_diversity_analyses.py` output

Research

Highly accessed | Open Access

Moving pictures of the human microbiome

J Gregory Caporaso[1], Christian L Lauber[2], Elizabeth K Costello[3], Donna Berg-Lyons[2], Antonio Gonzalez[4], Jesse Stombaugh[1], Dan Knights[4], Pawel Gajer[5], Jacques Ravel[5], Noah Fierer[2,6], Jeffrey I Gordon[7] and Rob Knight[1,8]*

* Corresponding author: Rob Knight rob.knight@colorado.edu

▸ Author Affiliations

For all author emails, please log on.

Genome Biology 2011, **12**:R50 doi:10.1186/gb-2011-12-5-r50

The electronic version of this article is the complete one and can be found online at:
http://genomebiology.com/2011/12/5/R50

Full tutorial available at:
http://qiime.org/tutorials/illumina_overview_tutorial.html

Moving Pictures of the Human Microbiome

- Two human subjects provided daily microbiome samples, one for 6 months and the other for 18 months.
- Sampled tongue, left and right palms, and gut (via feces).
- Tutorial data is a subset of this full data set.
- *Illumina HiSeq 2000* 16S sequencing with Earth Microbiome Project protocols (earthmicrobiome.org)

core_diversity_analyses.py precomputed results demonstration

http://bit.ly/1IL5WpX

Advanced topics

Chimera checking

- Usage in QIIME described here: http://qiime.org/tutorials/chimera_checking.html
- Should not be used if one is using closed-reference OTU picking.
- Demonstration of chimera checking on the QIIME Illumina tutorial data with usearch61.

Supervised Learning

This approach is useful if one needs to categorize samples based upon microbial data. QIIME implementation is a random forest approach.

The script used is supervised_learning.py.

Demonstration of supervised learning with QIIME tutorial data.

http://qiime.org/tutorials/running_supervised_learning.html

SourceTracker

QIIME tutorial page: http://qiime.org/tutorials/source_tracking.html

Describes an older command for converting one's OTU table to tab-delimited format. A command that works with the new version of the biom softare is:

```
biom convert --to-tsv -i INPUT -o OUTPUT
```

Plates from two fields of my maize study were mixed up for a particular week (Aurora and Lansing fields, week 9). Can see predictions from SourceTracker which indicate the source of the microbes from the correct field.

PICRUSt

- Predicts metagenomes from 16S data
- Useful for exploring metagenomic data without shotgun sequences
  - Save money and time
- PICRUSt workflow
- Website: http://picrust.github.io

Microbial Study Design Considerations

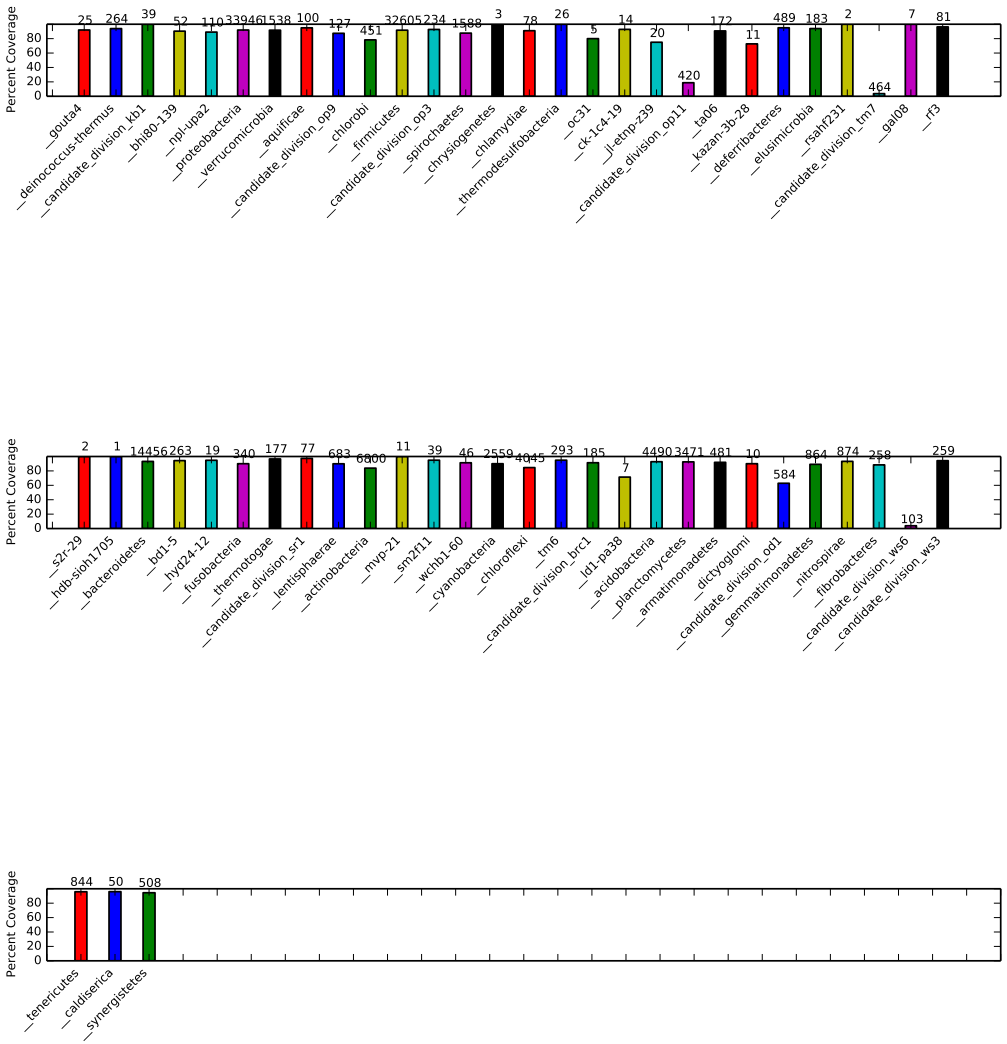
You probably already have sequencing data, ready to analyze, but…

PCR primer choice

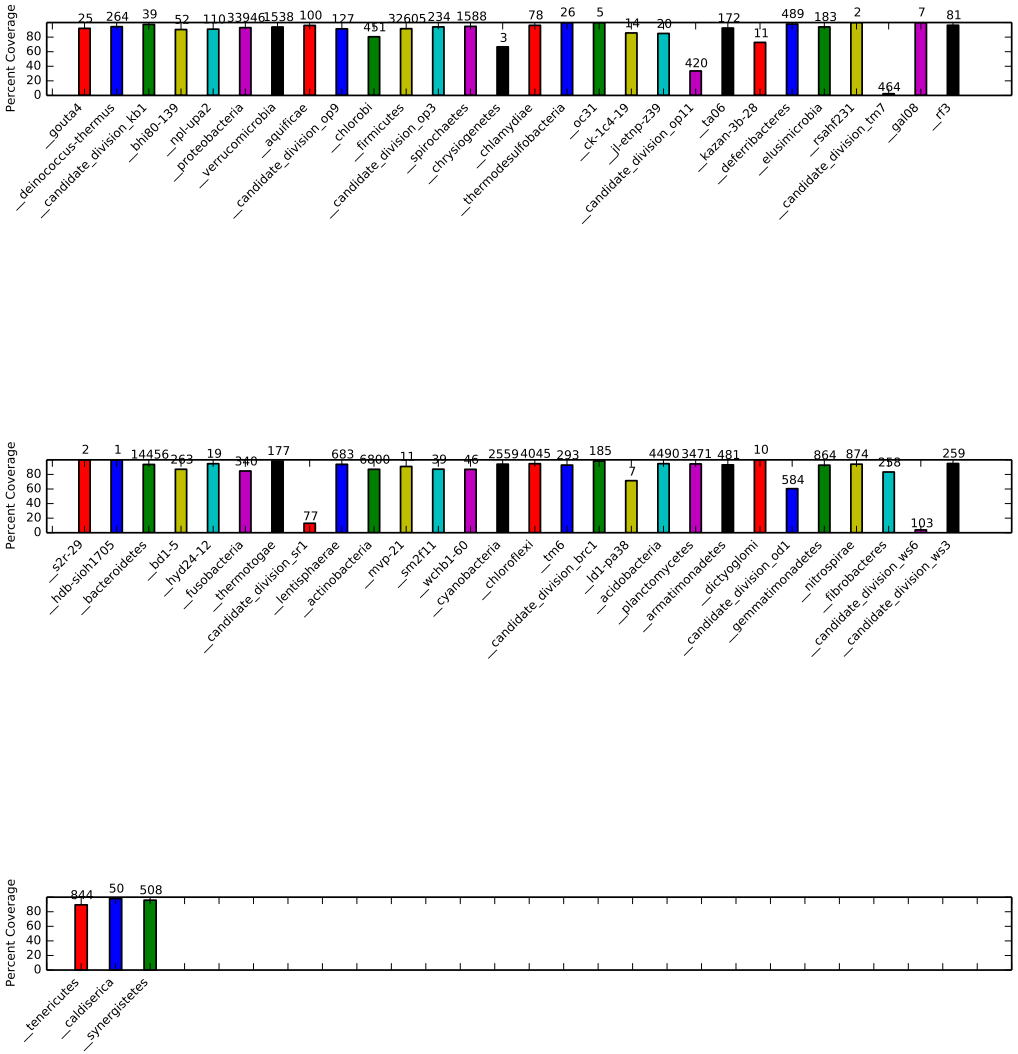
- Taxa
- Amplicon Length
- Target gene

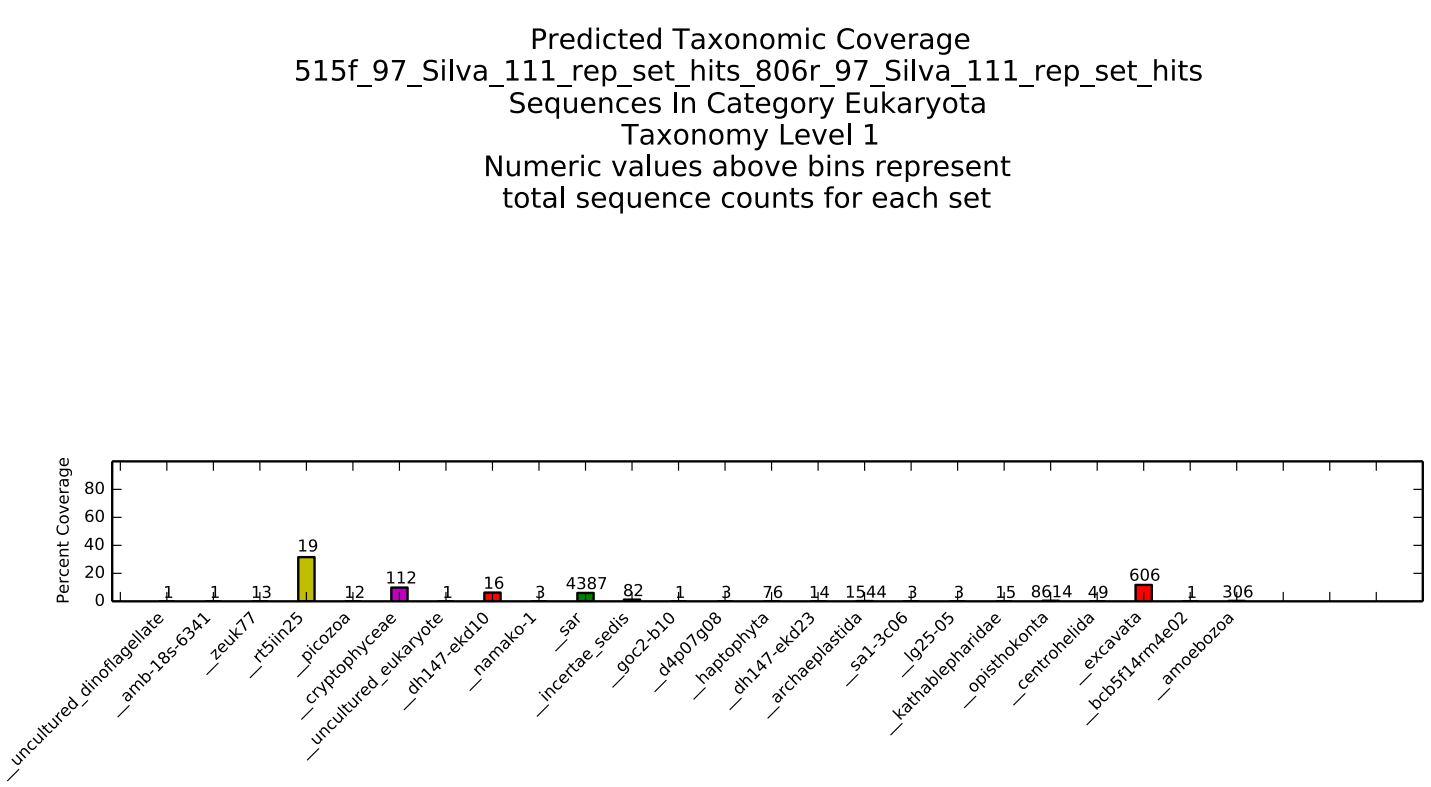
We can address these with Primer Prospector: http://pprospector.sourceforge.net/

V4 Bacteria

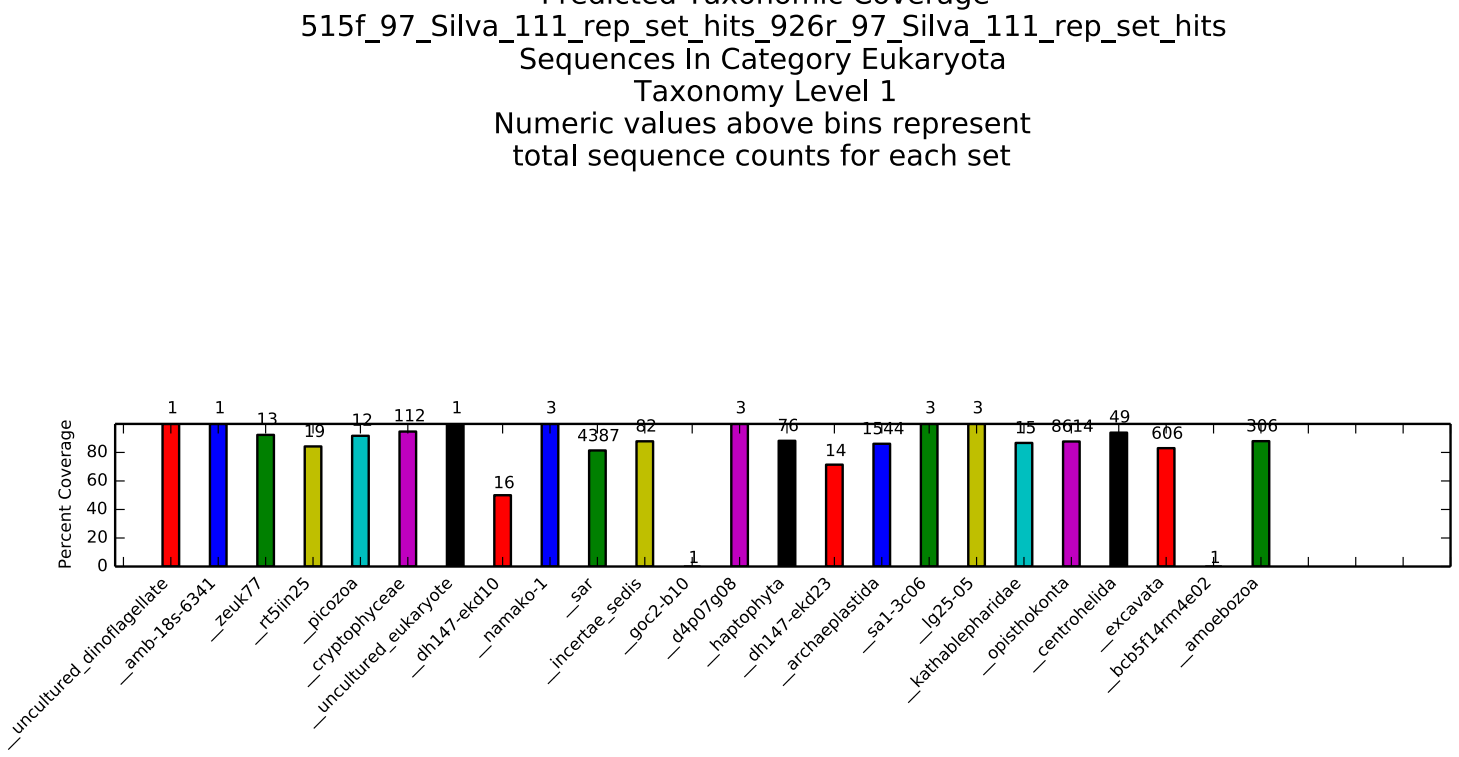

Predicted Taxonomic Coverage
515f_97_Silva_111_rep_set_hits_806r_97_Silva_111_rep_set_hits
Sequences In Category Bacteria
Taxonomy Level 1
Numeric values above bins represent
total sequence counts for each set
Percent Coverage
0
20
40
60
80
25
264
39
52
110
339
461
538
100
127
451
3260
5234
1588
3
78
26
5
14
20
420
172
11
489
183
2
464
7
81
__gouta4
__deinococcus-thermus
__candidate_division_kb1
__bhi80-139
__npl-upa2
__proteobacteria
__verrucomicrobia
__aquificae
__candidate_division_op9
__chlorobi
__firmicutes
__candidate_division_op3
__spirochaetes
__chrysiogenetes
__chlamydiae
__thermodesulfobacteria
__oc31
__ck-1c4-19
__ji-etnp-239
__candidate_division_op11
__ta06
__kazan-3b-28
__deferribacteres
__elusimicrobia
__rsahf231
__candidate_division_tm7
__gal08
__rf3
2
1
1445
6263
19
340
177
77
683
6800
11
39
46
2559
4045
293
185
7
4490
3471
481
10
584
864
874
259
103
259
__s2r-29
__hdb-sioh1705
__bacteroidetes
__bd1-5
__hyd24-12
__fusobacteria
__thermotogae
__candidate_division_sr1
__lentisphaerae
__actinobacteria
__nvp-21
__smzf11
__wchb1-60
__cyanobacteria
__chloroflexi
__tm6
__candidate_division_brc1
__id1-pa38
__acidobacteria
__planctomycetes
__armatimonadetes
__dictyoglomi
__candidate_division_od1
__gemmatimonadetes
__nitrospirae
__fibrobacteres
__candidate_division_ws6
__candidate_division_ws3
844
50
508
__tenericutes
__caldiserica
__synergistetes

V4-5 Bacteria





V4 (515f/806r) predicted Eukaryotic amplification



V4-5 (515f/926r) predicted Eukaryotic amplification

Validating metadata

- validate_mapping_file.py demonstration
- Keemei demonstration
  - http://keemei.qiime.org

Reproducible bioinformatics

Why is it important for a scientific experiment to be reproducible?

- Differentiate real results from experimental artifacts

Reproducible versus replicable

- Replicable: exact same conditions lead to concordant results
- Reproducible: some experimental variation is allowed, but results are concordant

What does it mean for a bioinformatics experiment to be replicable?

- Our experimental methods are not as 'noisy' as most.
- Same commands on the same system should give you the same results (if the algorithm is *deterministic*).

Deterministic versus non-deterministic

- Deterministic algorithm: a given input produces the same series of internal states and results in the same output.
- Non-deterministic algorithm: a given input may produce different internal states and/or result in a different output.
  - Commonly probabilistic algorithms in bioinformatics

Deterministic

- Smith-Waterman alignment: if properly implemented, aligning two sequences will always give the same result

Non-deterministic/Probabilistic

- Sub-sampling a data set (e.g., rarefaction of an OTU table)
- Jackknifed analyses
- Permutation-based p-values (Monte Carlo)

How can we develop software that supports reproducibility?

Open source software

- Consider making your software open source
  - Avoid “black box”
  - Benefits other researchers/community
  - Encourages collaboration
  - Extendibility
- Place it under public hosted revision control
  - GitHub
  - Bitbucket
  - SourceForge

Version control systems

- Git
- Subversion (svn)
- Mercurial (hg)
- CVS (ancient history)

- Allow for viewing history of changes, obtaining previous versions.
- Example: https://github.com/biocore/qiime

Virtual Machines

- Publish virtual machine images
  - Gives access to exact software, configuration, and data used in analyses

What's in a good log file?

- Ideally will supplement your lab notebook (for successful runs)
  - Version information
  - Exact commands that were run
  - Details on input files (path, md5)
  - System configuration details
- Publish these as supplementary material

MD5

- A cryptographic hash function: deterministic function which takes some input and returns a fixed-size string – changing the input should change the return value

From Wikipedia:

- it is easy (but not necessarily quick) to compute the hash value for any given message
- it is infeasible to generate a message that has a given hash
- it is infeasible to modify a message without changing the hash
- it is infeasible to find two different messages with the same hash

IPython Notebook

- Interactive, executable documents: code, text, images, etc.
- Easy to share, publish, convert
- Great for keeping track of analysis commands, code, descriptions/comments, etc.
- Make your methods section *executable*!

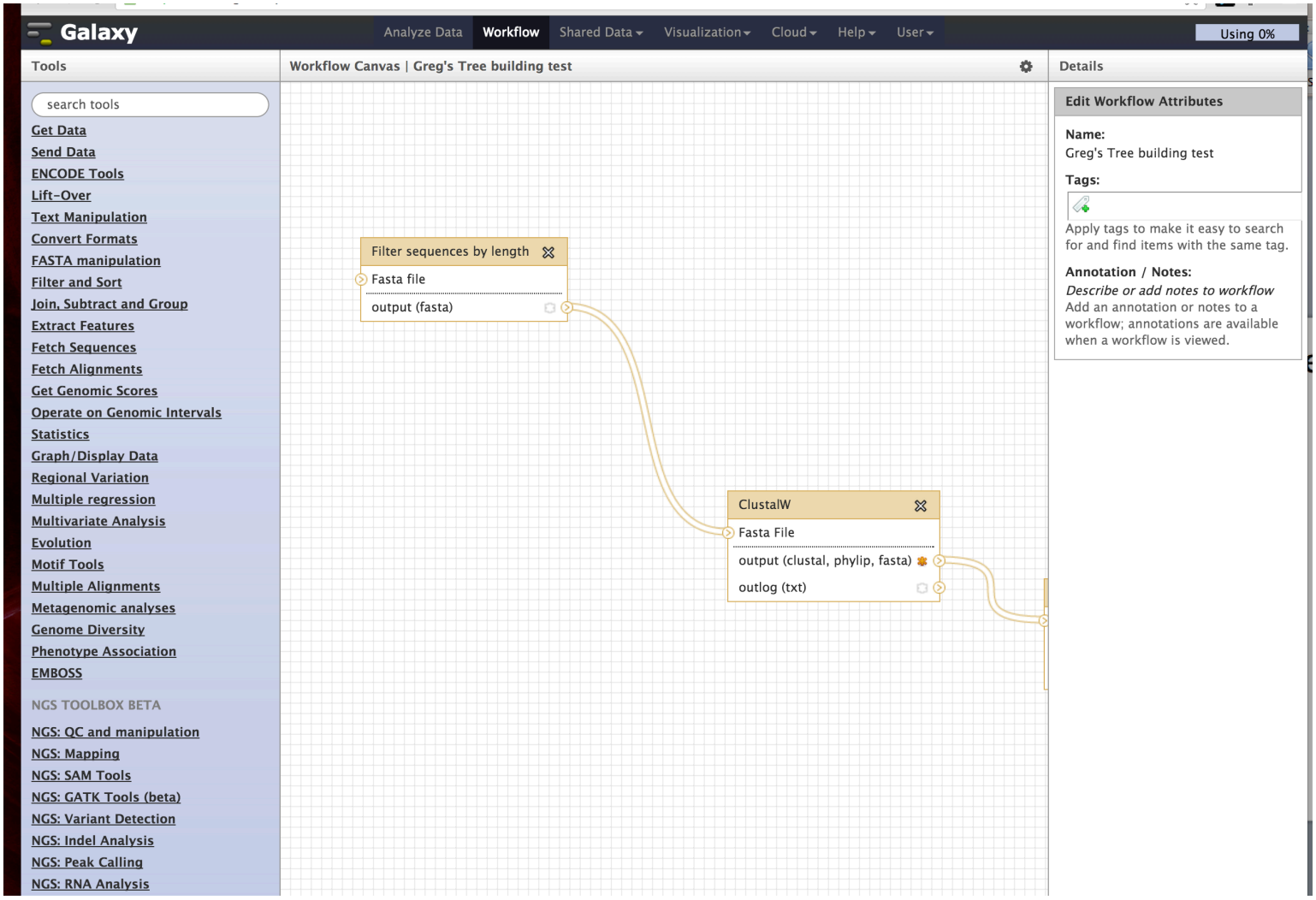
http://ipython.org/

IPython Notebook

- Workshop exercises are written in IPython Notebooks
- An Introduction to Applied Bioinformatics
  - Online interactive bioinformatics book
  - http://readIAB.org

http://ipython.org/

Reproducible computing through workflow engines, e.g. Galaxy



https://main.g2.bx.psu.edu

Future directions: QiiTA and QIIME 2

QiiTA pre-history

- Previous iteration known as the “QIIME Database” (QIIME-DB)
- Suffered database crash
  - Efforts focused on rewrite instead of recover

QiiTA: QIIME-DB Reboot

- System for:
  - Depositing/archiving microbiome data
  - Performing meta-analysis
    - Combine data from a variety of sources (marker gene, metagenomic, metabolomic, etc.)
- Goals
  - Easy-to-use web interface
  - User-deployable in a variety of environments (e.g., laptops to clusters)
  - Powerful meta-analysis capabilities

QiiTA

- Currently supports:
  - Querying studies
  - Closed-reference OTU picking
  - Submission of data to EBI-SRA (needed for publishing studies)
- Demonstration of querying and downloading study data

Moving toward QIIME 2

- QIIME is currently:
  - Command-line only (very limited Galaxy support)
  - Simplistic execution of workflows
  - Hard to extend and maintain (for both users and devs)
  - Can be difficult to install

Moving toward QIIME 2

- Most requested feature: graphical interface
- Most support efforts: command-line issues

Users spend too much time grappling with the command line and less time performing awesome microbiome research.

Devs spend too much time helping users with installation and command-line issues, and less time answering users' research questions.

QIIME 2 Overview

- Complete redesign/rebuild of QIIME
- Powered by scikit-bio (http://scikit-bio.org)
- Graphical web-based interface
- Command-line interface and Python API
- Interactive visualizations
- Deployable on laptops -> clusters
- Extendable by users/devs via plugin system

QIIME 2 Overview

- Currently being developed
  - Initial alpha release scheduled for 2016
- Follow updates at http://blog.qiime.org and on Twitter by following @qiime_

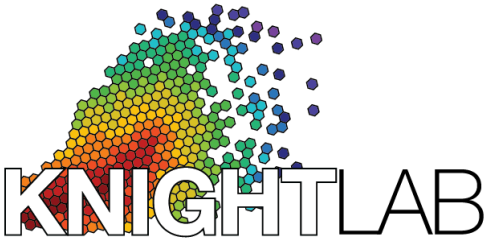
Preparation for day 2

- Download and install Cyberduck
  - http://cyberduck.io
- Download and install Google Chrome
  - https://www.google.com/chrome/

Note: We'll upload the latest version of today's slides this evening.

Acknowledgements

Evan Bolyen
Nick Bokulich
Katy Califf
Greg Caporaso
Jose Clemente
John Chase
Kevin Cohen
Antonio Gonzalez
Crystal Hepp
Rob Knight
Bruce Hungate
Larry Hunter
Paul Keim
Scott Kelley
Justin Kuczynski
Cathy Lozupone
Daniel McDonald
David Mills
Norm Pace
Fernando Perez
Jai Ram Rideout
Egbert Schwartz
Karen Schwarzberg
Jeffrey Siegel
Jesse Stombaugh
Yoshiki Vazquez
Tony Walters



ALFRED P. SLOAN FOUNDATION

william.a.walters@gmail.com
jai.rideout@gmail.com

Slides compiled by:

Greg Caporaso
John Chase
Jose Clemente
Antonio Gonzalez Peña
Rob Knight
Cathy Lozupone
Daniel McDonald
Jai Ram Rideout
Yoshiki Vázquez Baeza
Tony Walters

www.qiime.org