

Slides: goo.gl/QXVa4h
Slides (pdf): goo.gl/B1Qnk7
Schedule: goo.gl/XDrhf3



qiime2

Workshop

STAMPS

August 3, 2018

Rob Knight
[Knight Lab](#)

UC San Diego

Evan Bolyen
[Caporaso Lab](#)

NAU NORTHERN ARIZONA
UNIVERSITY
The Pathogen and Microbiome Institute

Project funding:

Slides: goo.gl/QXVa4h
Slides (pdf): goo.gl/B1Qnk7
Schedule: goo.gl/XDrhf3



Award number: [1565100](#)



What do you **see**
when you look in the
mirror?



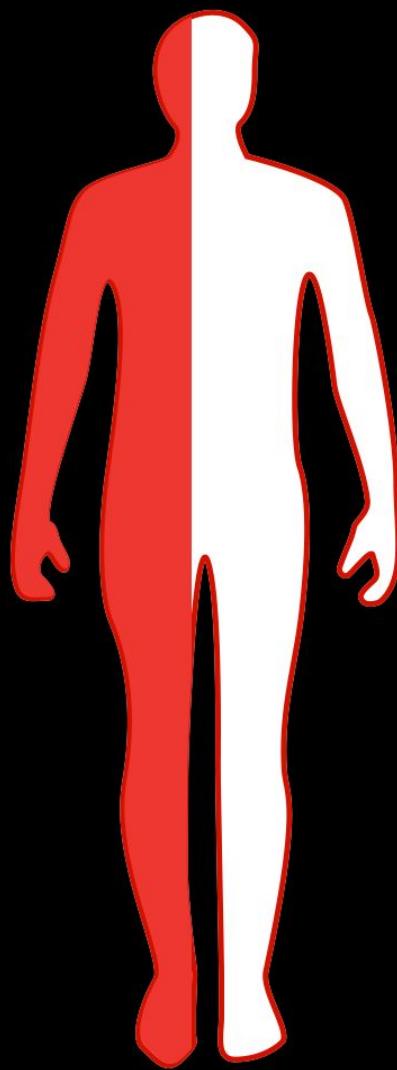
A composite image showing a woman in a white robe applying cream to her face. The image is split vertically down the center. The left side shows her from the chest up, facing right, with her hands near her chin. The right side shows her from the shoulders up, facing left, also with her hands near her chin. She has blonde hair tied back in a ponytail with a yellow hair tie.

We see an organism that is
43% human

30 trillion
human cells

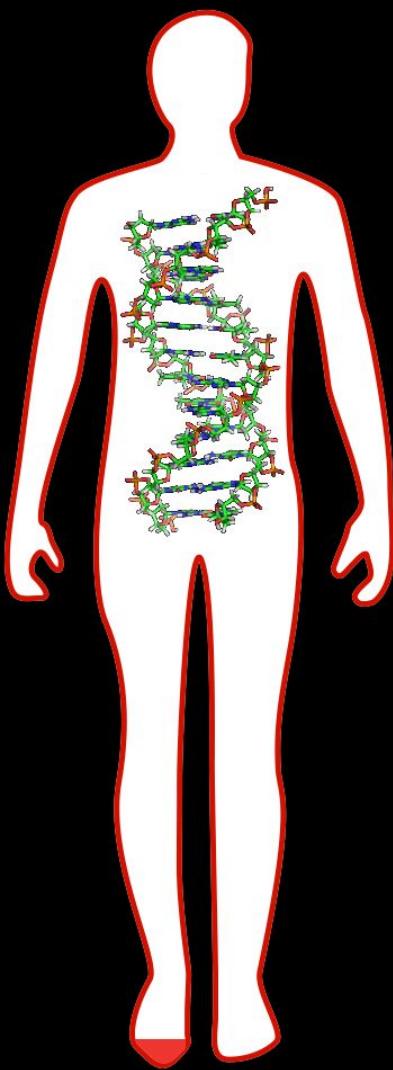
39 trillion
microbial cells

43%



20,000
human genes

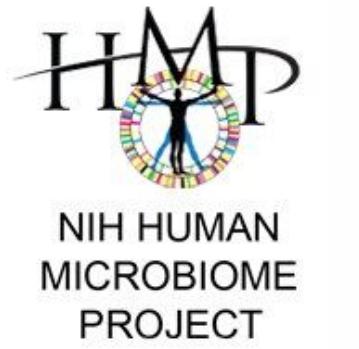
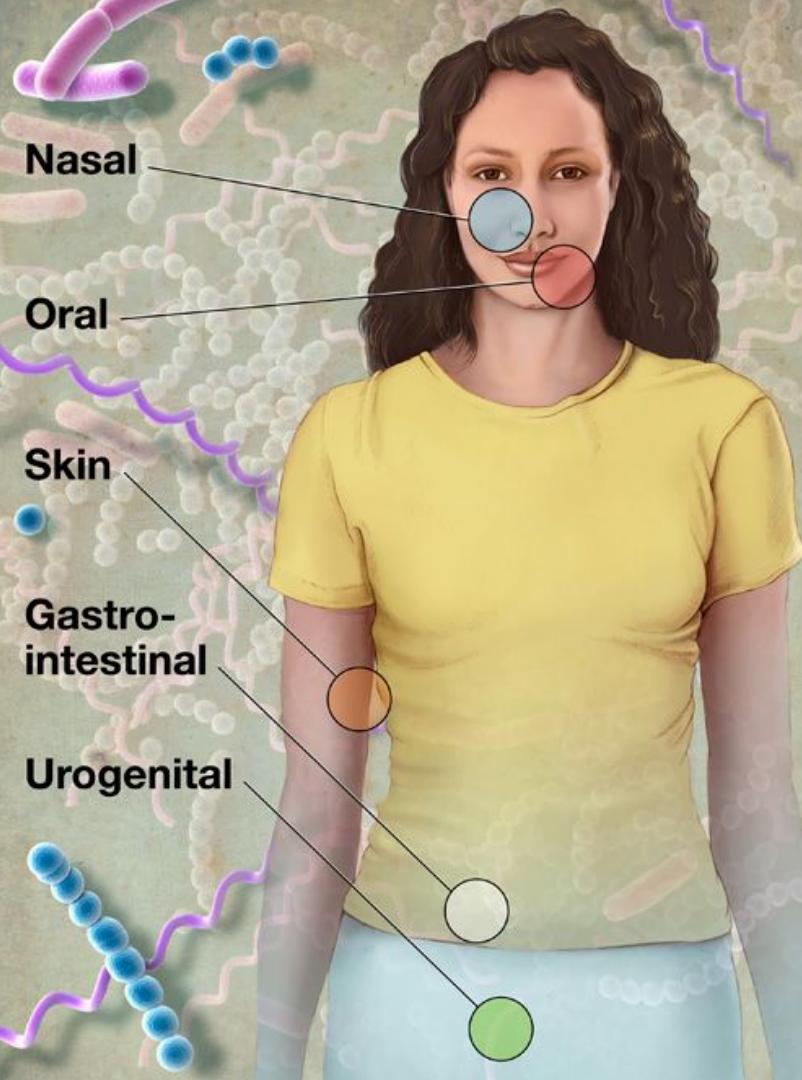
1 %



2-20 million
microbial genes

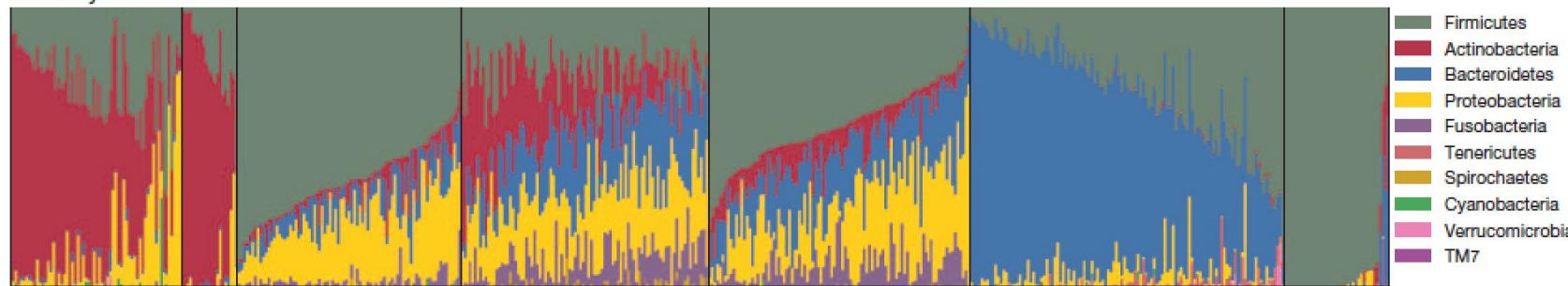
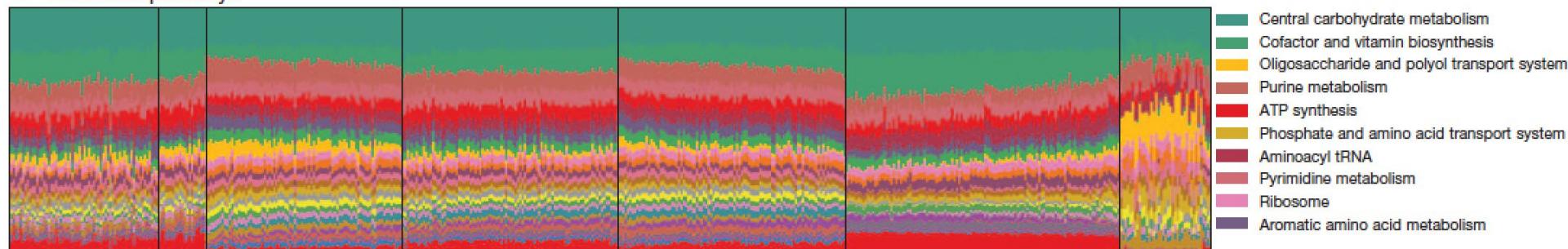
We neglect 99%
of the system!

How can we map our microbes?



\$173M
up to 18 sites
250 people
up to 3 timepoints
4.5 trillion bases
Shotgun + 16S

The image features a large, bold letter 'A' in black, which is partially filled with a circular emblem. The emblem contains a golden-yellow crown at the top and a golden-yellow sword with a decorative hilt below it. The background of the image is white.

a Phyla**b Metabolic pathways**

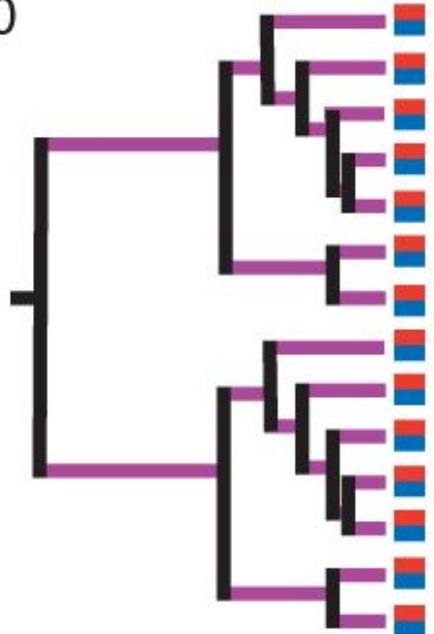
Anterior nares RC Buccal mucosa Supragingival plaque Tongue dorsum Stool Posterior fornix



UniFrac: Exploiting Evolution to Compare Communities

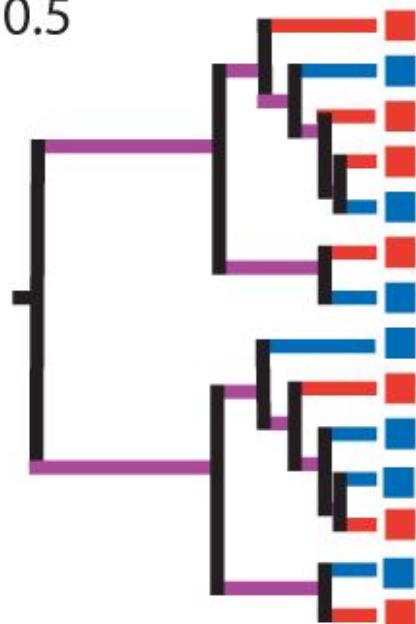
Identical communities:

$$D = 0$$



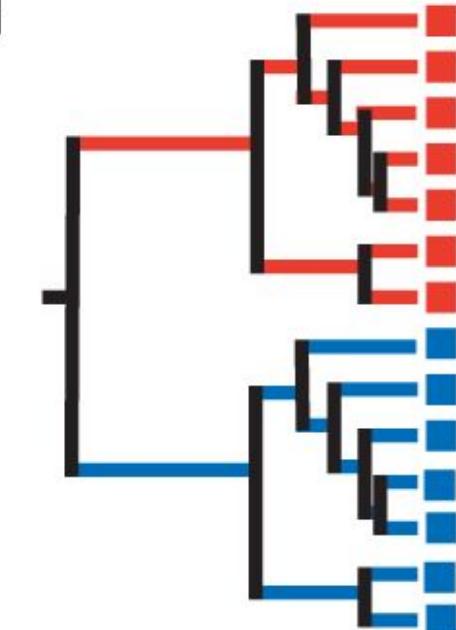
Related communities:

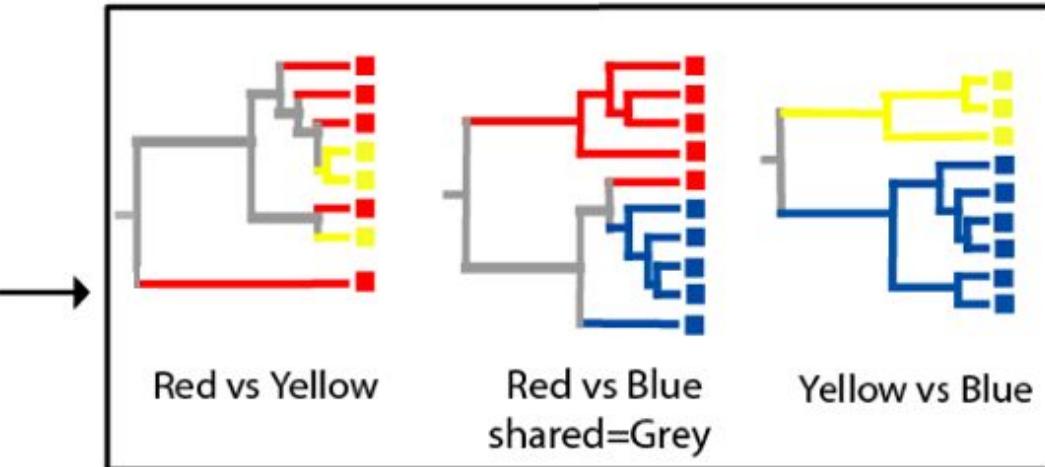
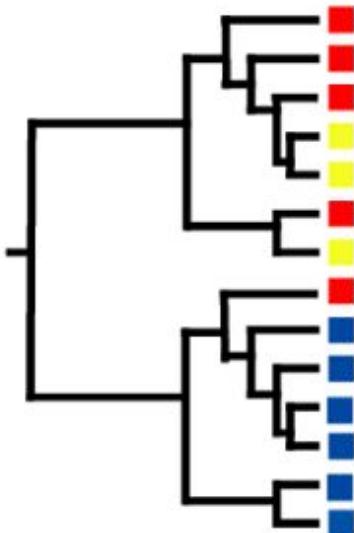
$$D \sim 0.5$$



Unrelated communities:

$$D = 1$$

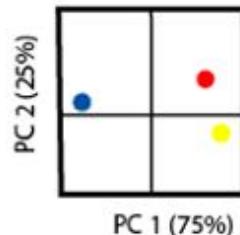




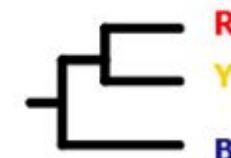
	R	Y	B
R	0	.	.7
Y	.	0	1
B	.7	1	0

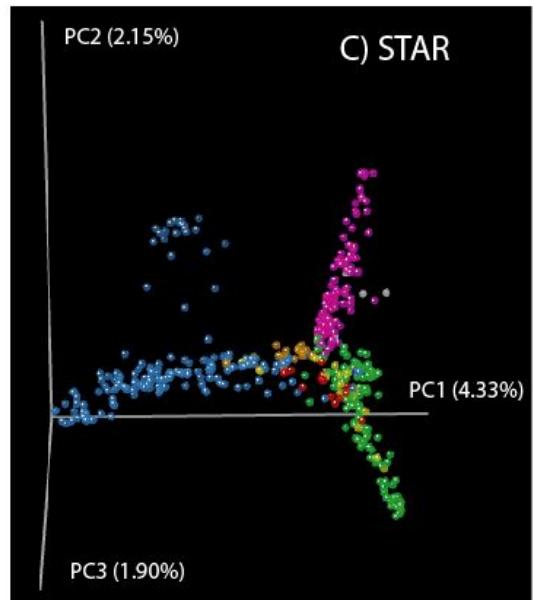
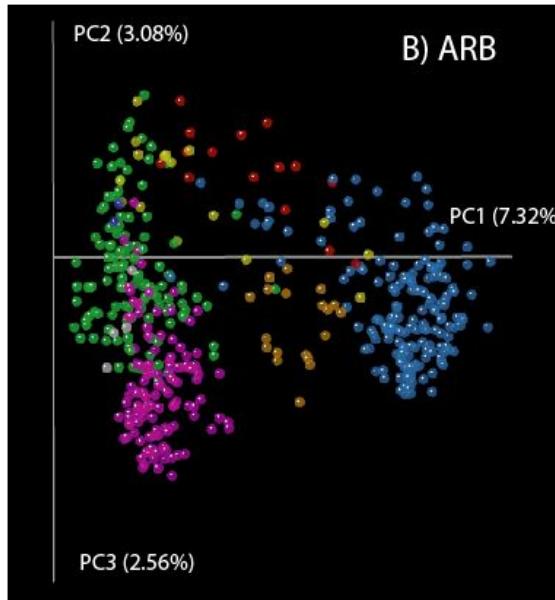
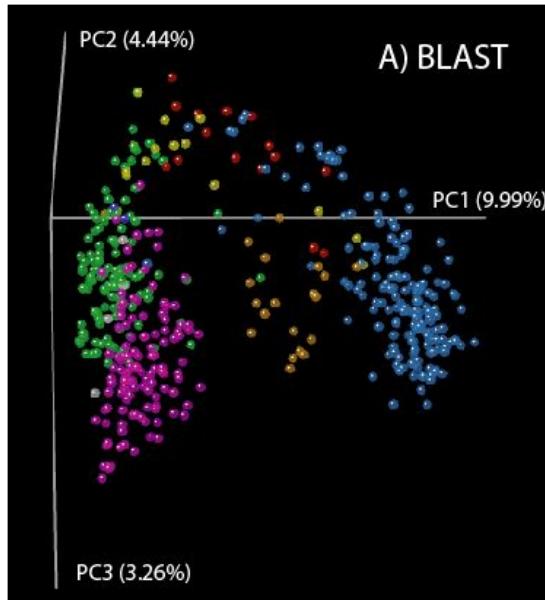
Distance Matrix

PCoA

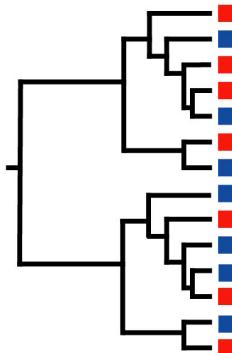


Hierarchical Cluster

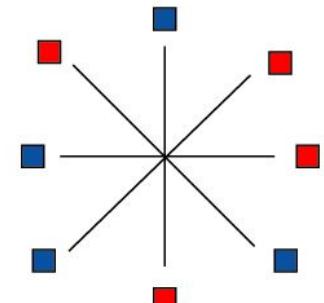


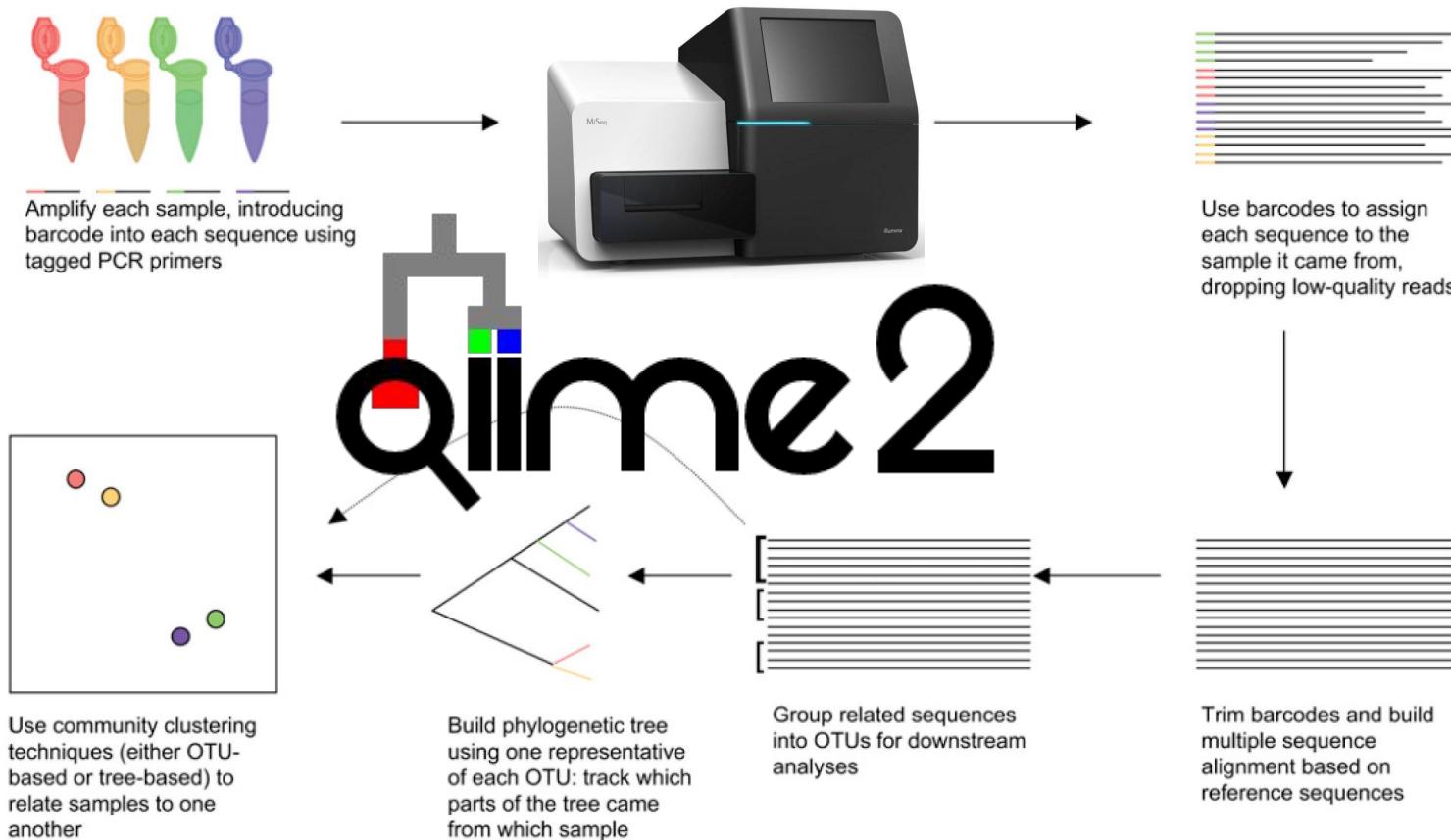


- Vertebrate Gut ● Termite Gut ● Human Skin/Mouth ● Invertebrate Associated
- Saline Freeliving ● NonSaline Freeliving ● Mixed Freeliving

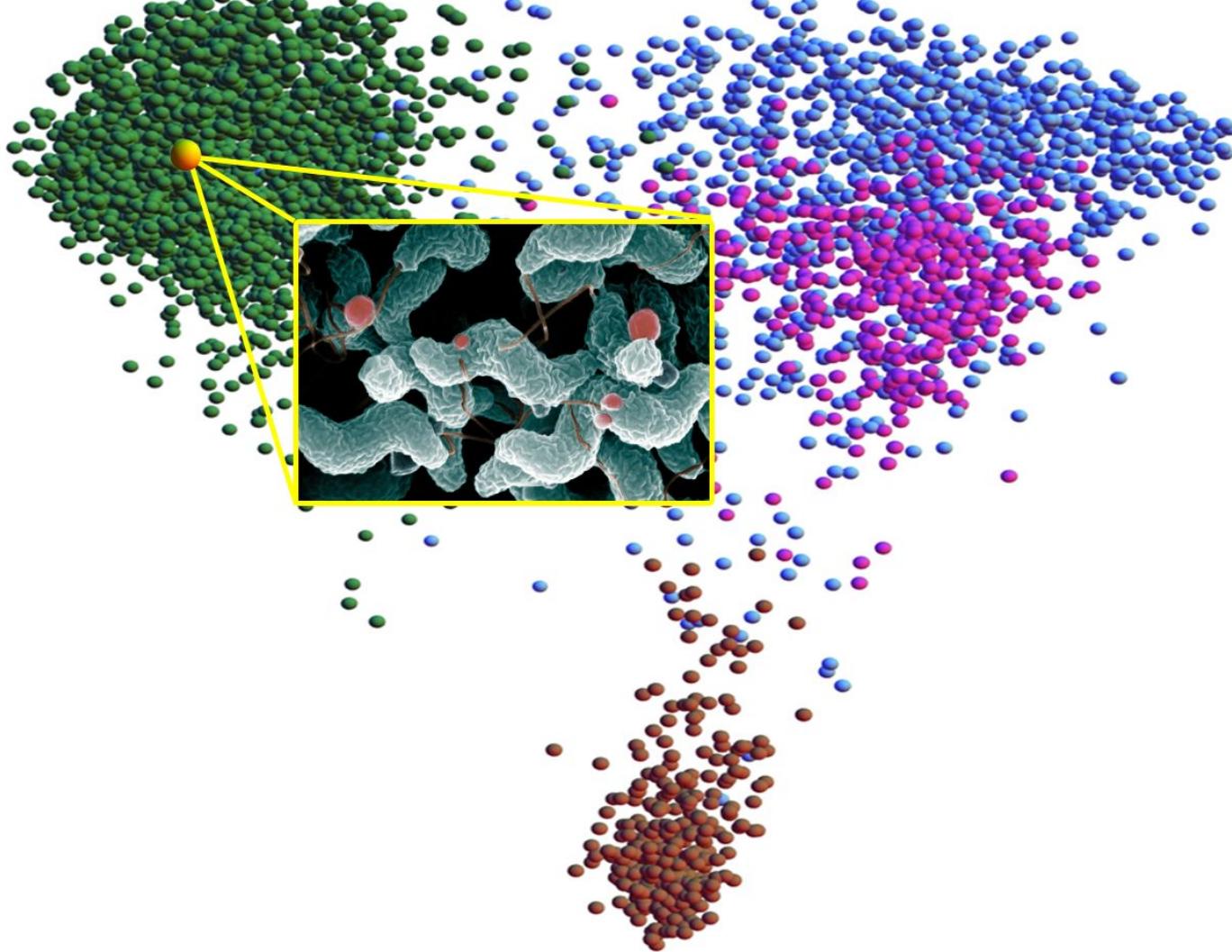


BLAST to greengenes core set tree replicates Arb parsimony insertion results when phylogenetic relationships matter







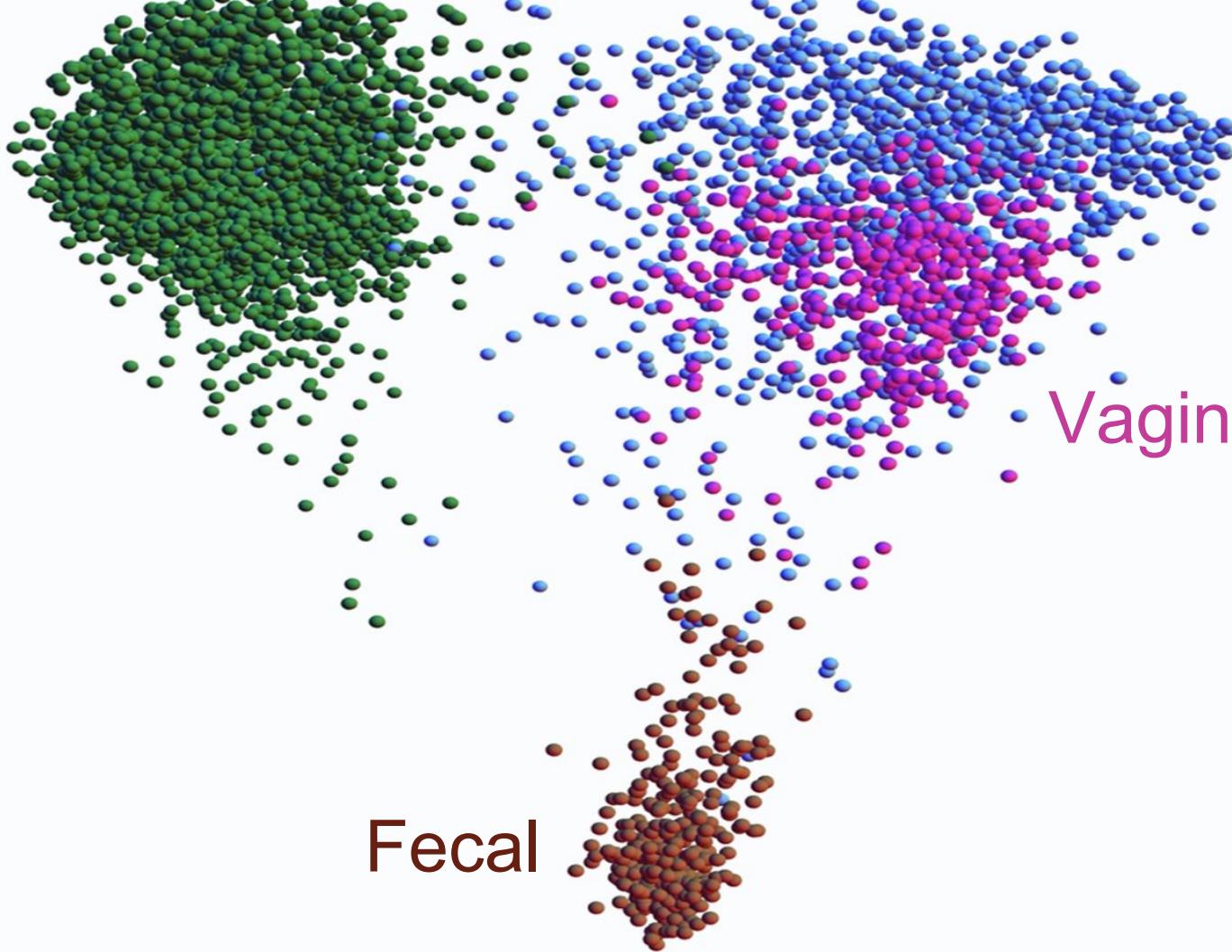


Oral

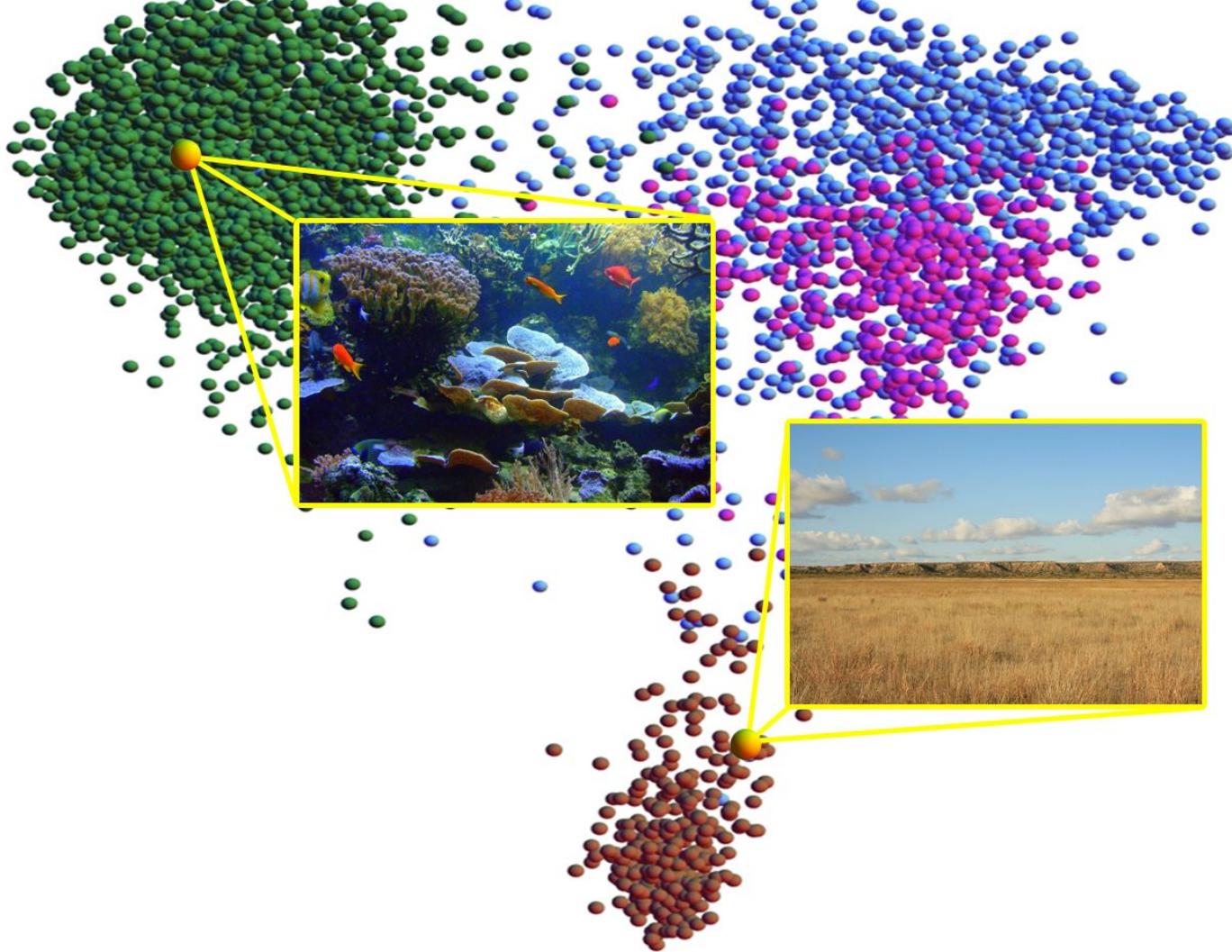
Skin

Vaginal

Fecal

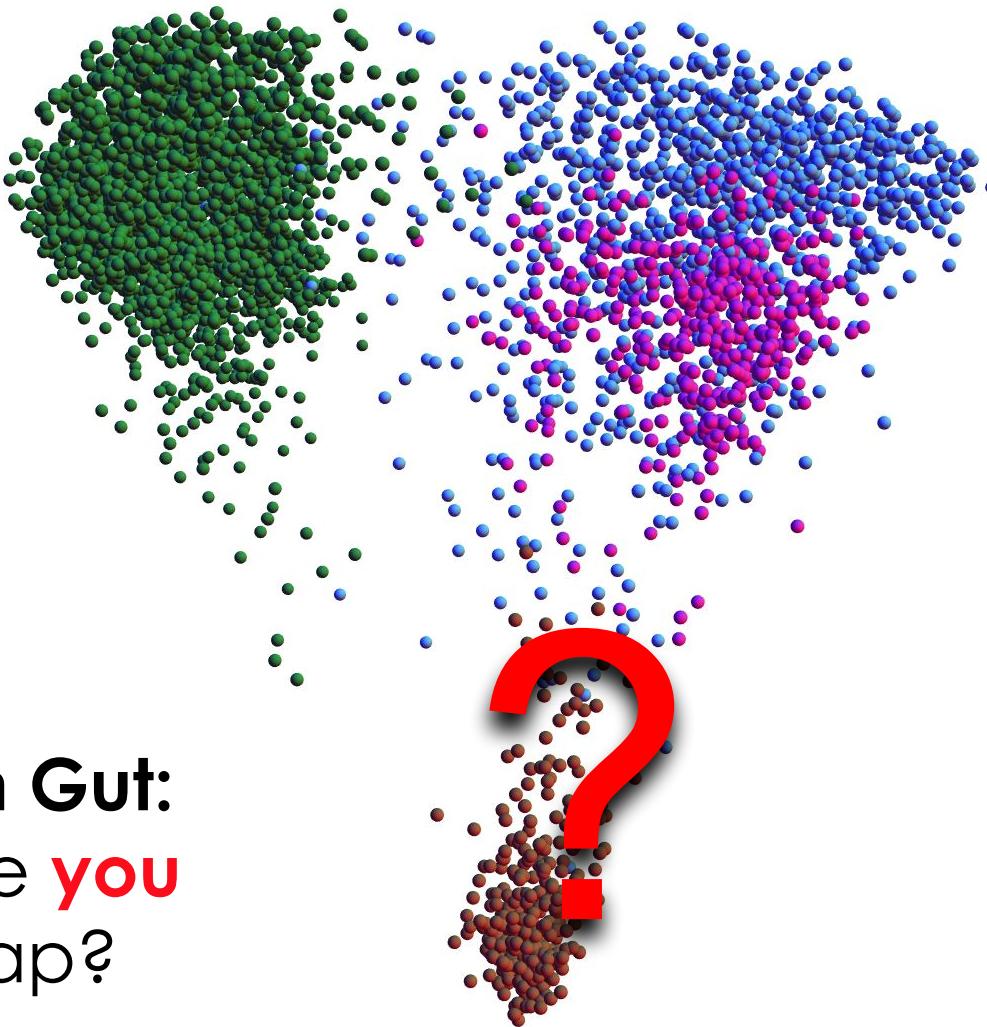






How can I place myself on the
microbial map?

American Gut: Where are **you** on the map?



>\$1.5M raised

>12000 kits

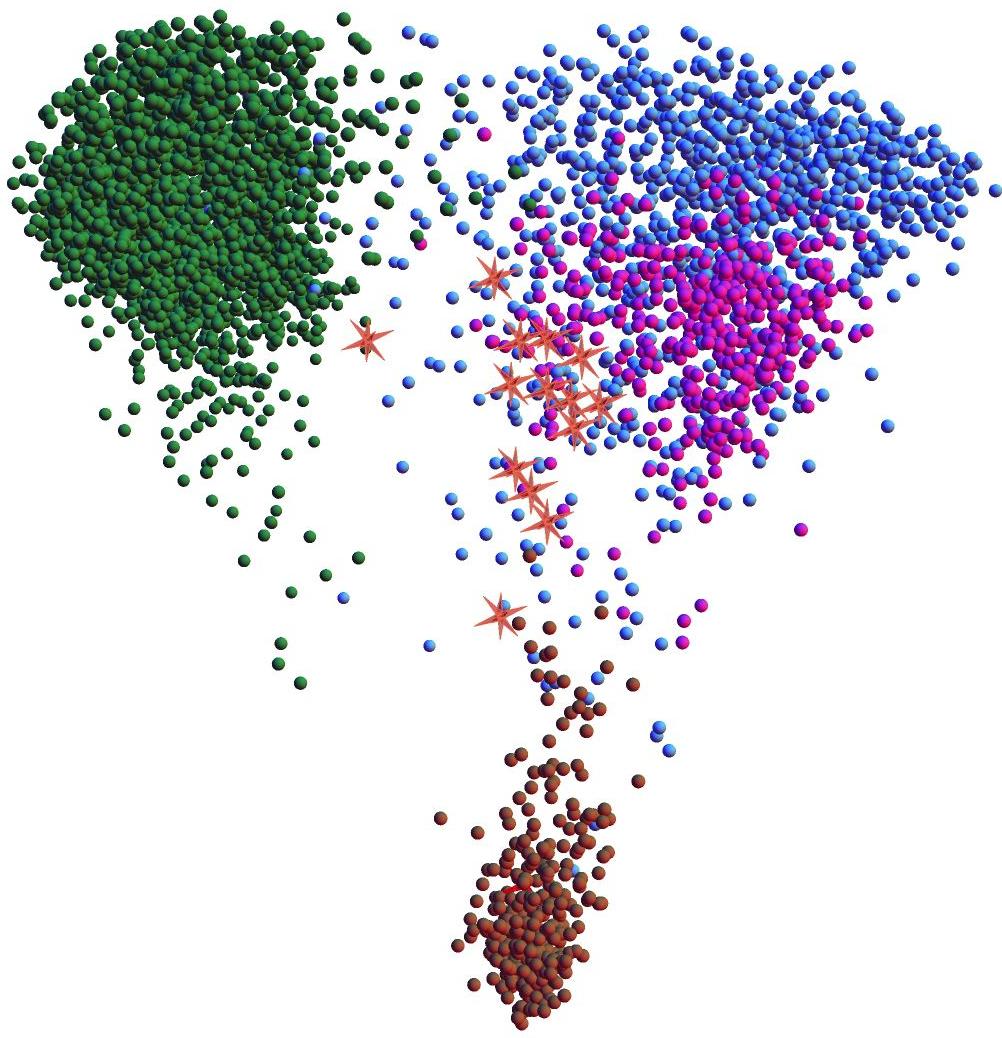
>10000 samples
sequenced and data
in EBI

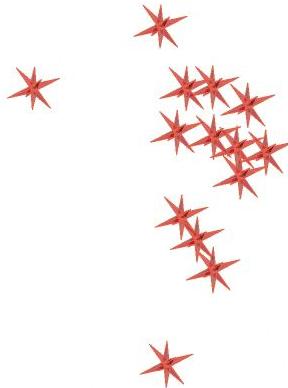
...but we still need
more!



microbiome

MO BIO
research

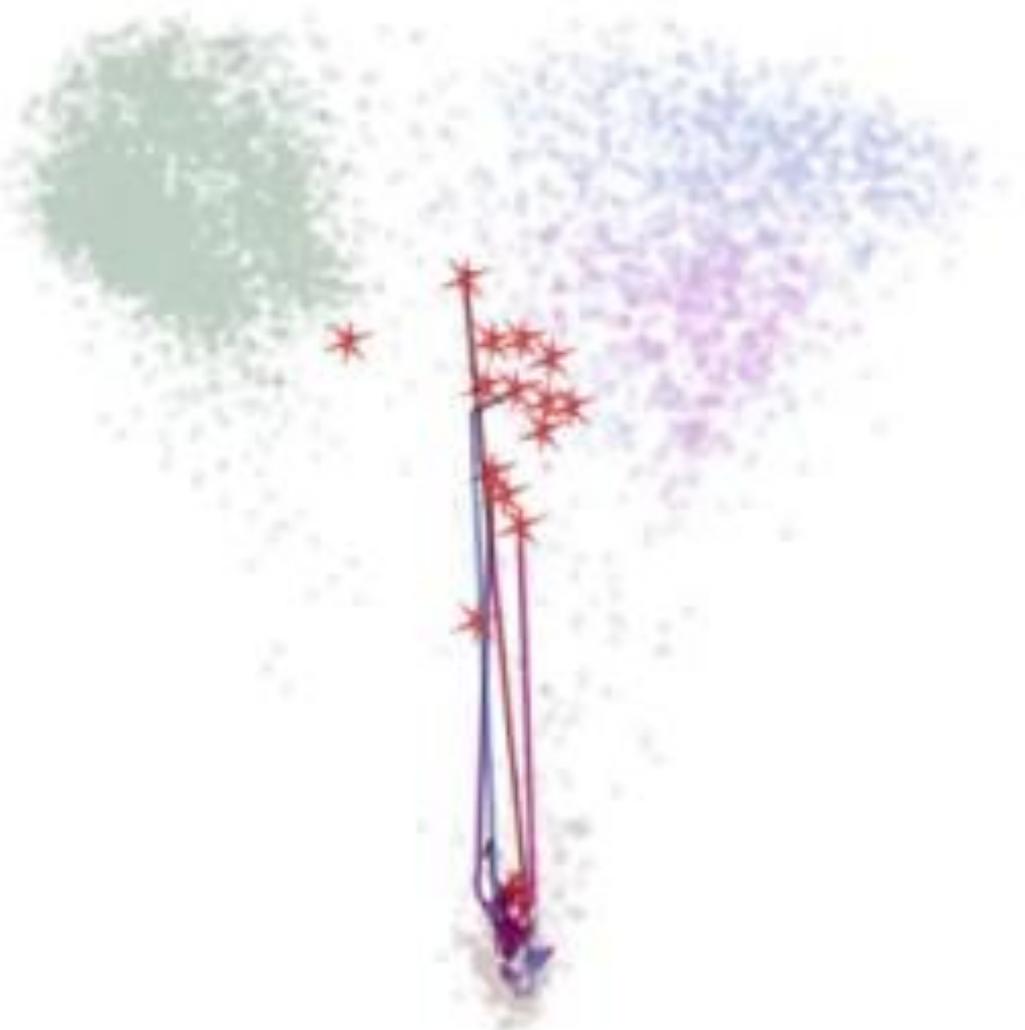


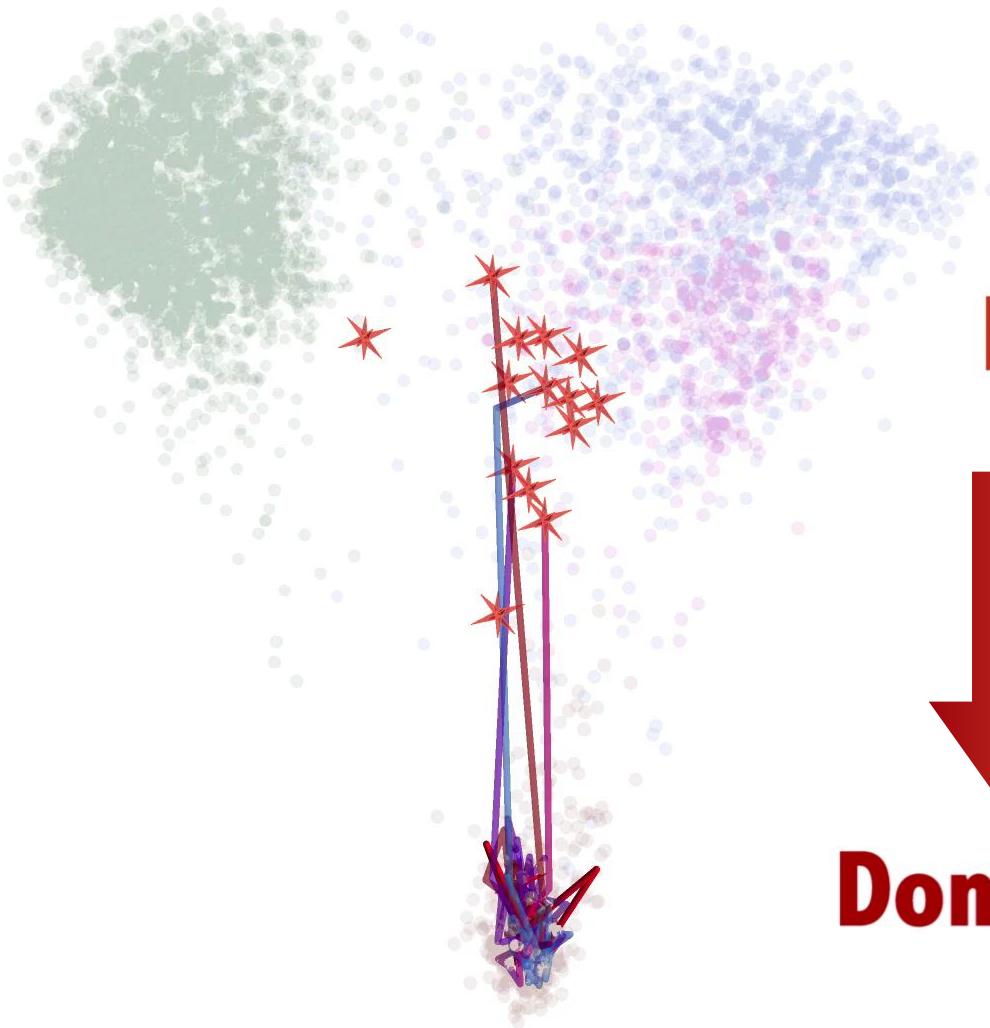


Patients

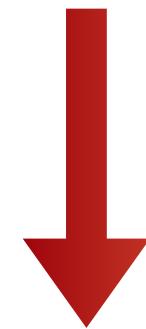


Donor





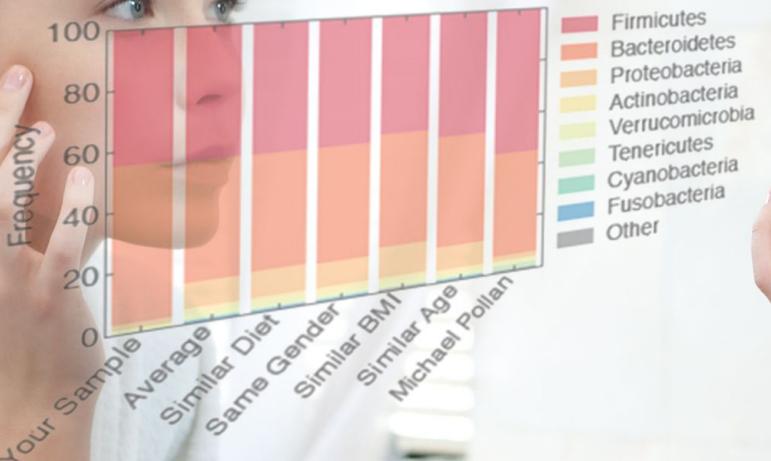
Patients



Donor



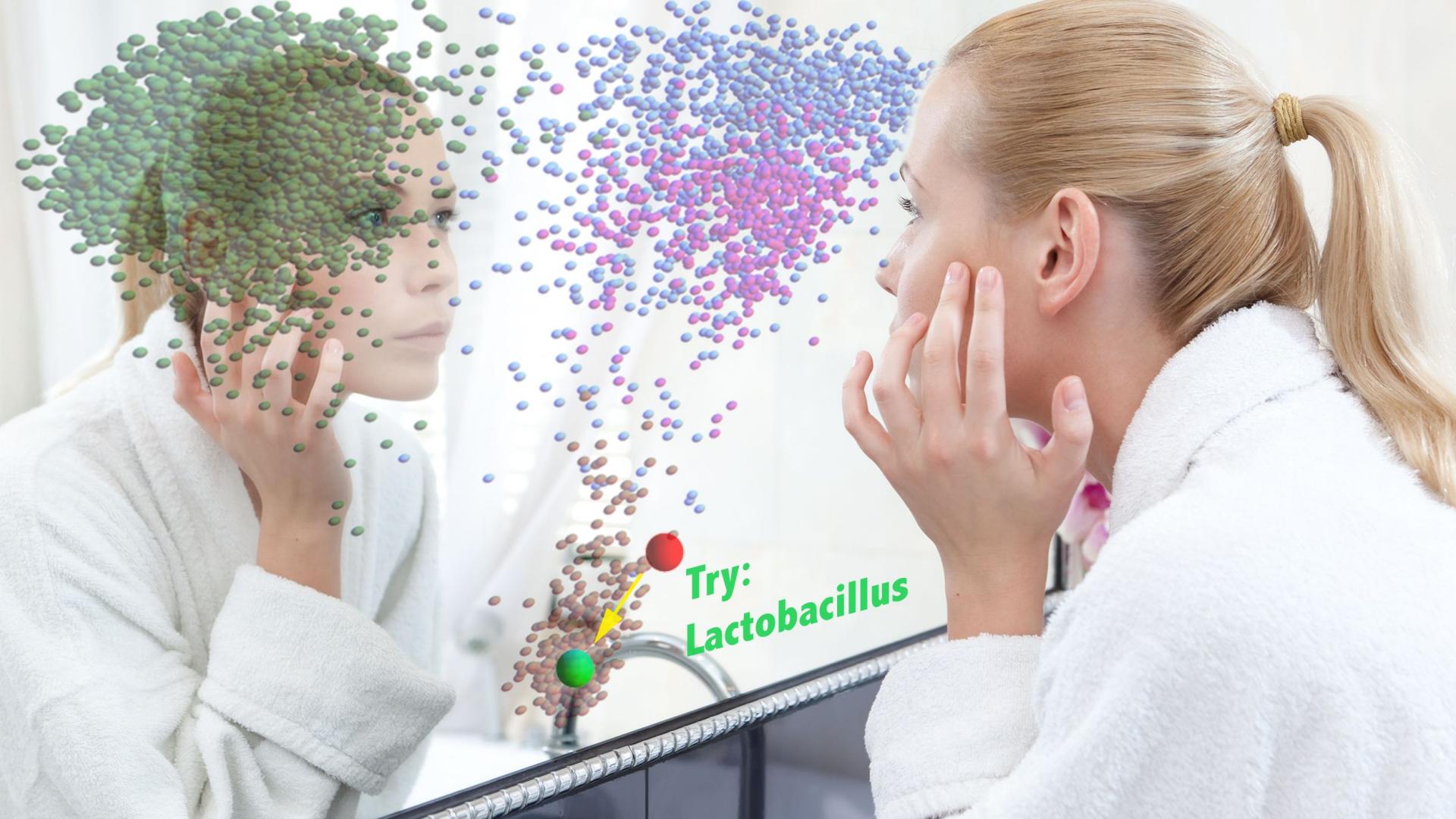








DANGER:
IBD RISK!



Try:
Lactobacillus

Where there is data,
there is hope

QIIME 1 citations

Google Scholar: >11,000

Structure, function and diversity of the healthy human microbiome

The Human Microbiome Project Consortium

Affiliations | Contributions | Corresponding author

Nature 486, 207–214 (14 June 2012) | doi:10.1038/nature11234

Received 02 November 2011 | Accepted 16 May 2012 | Published online 13 June 2012

Intestinal microbiome is related to lifetime antibiotic use in Finnish pre-school children

Katri Korppela, Anne Salonen, Lauri J. Virta, Riina A. Kekkonen, Kristoffer Forslund, Peer Bork & Willem M. de Vos

Affiliations | Contributions | Corresponding author

Nature Communications 7, Article number: 10410 | doi:10.1038/ncomms10410

Mx1 reveals innate pathways to antiviral resistance and lethal influenza disease

Padmini S. Pillai¹, Ryan D. Molony¹, Kimberly Martinod², Huiping Dong³, Iris K. Pang¹, Michal C. Tai^{1,*}, Angel G. Solis¹, Piotr Bielecki¹, Subhasis Mohanty³, Mark Trentalange⁴, Robert J. Homer⁵, Richard A. Flavell^{1,8}, Denisa D. Wagner², Ruth R. Montgomery⁶, Albert C. Shaw³, Peter Staeheli⁷, Akiko Iwasaki^{1,8,†}

+ Author Affiliations

↔ Corresponding author. Email: akiko.iwasaki@yale.edu

✉ Present address: Institute of Stem Cell Biology and Regenerative Medicine, School of Medicine, Stanford, CA 94305, USA.

Science 22 Apr 2016:

Vol. 352, Issue 6284, pp. 463–466

DOI: 10.1126/science.aaf3926

Cell Host & Microbe

Volume 19, Issue 5, 11 May 2016, Pages 731–743

Resource

Genetic Determinants of the Gut Microbiome in UK Twins

Julia K. Goodrich¹, Emily R. Davenport¹, Michelle Beaumont², Matthew A. Jackson², Rob K. Oberg³, Tim D. Spector², Jordana T. Bell², Andrew G. Clark¹, Ruth E. Ley^{1,5}, ▲, ■

Marine mammals harbor unique microbiota by and yet distinct from the sea

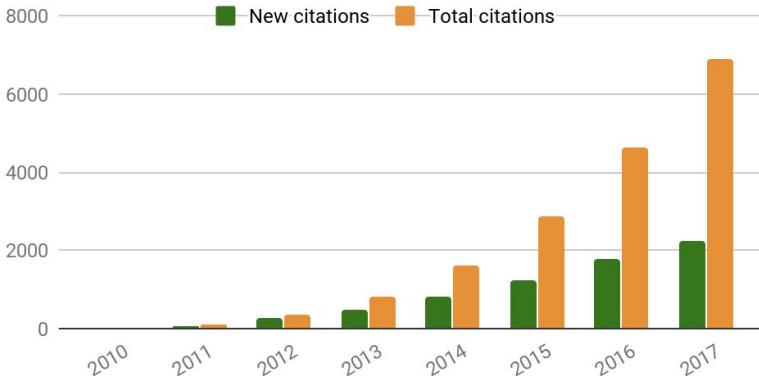
Elisabeth M. Bik, Elizabeth K. Costello, Alexandra D. Switzer, Benjamin J. Callahan, Susan P. Holmes, Randall S. Wells, Kevin P. Carlin, Eric D. Jensen, Stephanie Venn-Watson & David A. Relman

Affiliations | Contributions | Corresponding author

Nature Communications 7, Article number: 10516 | doi:10.1038/ncomms10516

Received 04 August 2015 | Accepted 18 December 2015 | Published 03 February 2016

New citations and total citations by year (source: Web of Science)



The ISME Journal (2016) 10, 1308–1322; doi:10.1038/ismej.2015.221

Wind and sunlight shape microbial diversity in surface waters of the North Pacific Subtropical Gyre
OPEN

Jessica A Bryant^{1,2}, Frank O Aylward^{2,3}, John M Eppley^{2,3}, David M Karl^{2,3}, Matthew J Church^{2,3} and Edward F DeLong^{1,2,3}

Microbial community assembly and metabolic function during mammalian corpse decomposition

Jessica L. Metcalf^{1,2,*}, Zhenjiang Zech Xu², Sophie Weiss³, Simon Lax^{4,5}, Will Van Treuren⁶, Embrette R. Hyde⁶, Se Jin Song^{1,2}, Amnon Amir², Peter Larsen^{4,7}, Naseer Sangwan^{4,7,8}, Daniel Haarmann⁹, Greg C. Humphrey², Gail Ackermann², Luke R. Thompson², Christian Lauber¹⁰, Alexander Bibat¹¹, Catherine Nicholas¹¹, Matthew J. Geber¹¹, Joseph F. Petrosino¹², Sasha C. Reed¹³, Jack A. Gilbert^{4,5,7,8,14}, Aaron M. Lynne⁹, Sibyl R. Bucheli⁹, David O. Carter¹⁵, Rob Knight^{2,16,*}

+ Author Affiliations

↔ Corresponding author. E-mail: robknight@ucsd.edu (R.K.); jessica.metcalf@colorado.edu (J.L.M.)

Science 08 Jan 2016;
Vol. 351, Issue 6269, pp. 158–162
DOI: 10.1126/science.aad2546

Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients

V. Gopalakrishnan^{1,2,*}, C. N. Spencer^{2,3,*}, L. Nezi^{3,7}, A. Reuben¹, M. C. Andrews¹, T. V. Karpinets³, P. A. Prieto^{1,1}, D. Vicente^{1...}

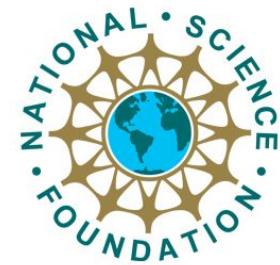
* See all authors and affiliations

Science 20 Nov 2017;
eaan4236
DOI: 10.1126/science.aan4236



Peer Reviewed
← see details





[NSF Award: 1565100](#)

High-level features (these attract users to system)

- **Latest and greatest microbiome bioinformatics methods and visualizations.**
- **Accessibility** through accurate, detailed, and interesting documentation and well-designed interfaces.
- **A community** of microbiome scientists, developers, and bioinformaticians.



<https://forum.qiime2.org>

Please read our [Code of Conduct](#) when joining.

all categories ► all tags ► **Categories** Latest New (193) Unread (37) Top Bookmarks My Posts

Category

User support

This category is for QIIME 2 user support. Use this category if you have questions about QIIME 2 or think you found a bug.

Developer discussion

This category is for software developer technical support. If you have questions, comments or announcements about developing QIIME 2 plugins, interfaces or the QIIME 2 framework, this is the place to post them.

Ideas and Suggestions

This category is for ideas and suggestions. If you have thoughts about how QIIME 2 can be improved, this is the place to post them. Well received ideas will be moved to the appropriate issue tracker.

Announcements

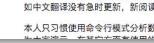
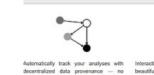
Announcements related to QIIME 2 workshops, etc.

Community Tutorials

The Community Tutorials category contains tutorials that are not part of the official QIIME 2 documentation.

QIIME2中文帮助文档 (Chinese Manual)

Yong_Xin_Liu
QIIME 2 2017.7



声明：本文档QIIME2官方帮助文档的中文字版，由中科院植物所刘多鑫博士翻译并亲测有效。文档翻译已获QIIME2项目组授权，由QIIME2更新贡献组，如使用或遇到问题请访问QIIME2官方论坛阅读最新版中文帮助。如中文翻译没有及时更新，烦请读英文原版。新的读英文原版：<https://docs.qiime2.org>

本人只习惯使用命令行模式数据分析，图形界面和ipython模式下使用暂不介绍。本系列的教程主要以命令行方式

Topics

Latest

DADA2 multiple runs with different numbers of samples

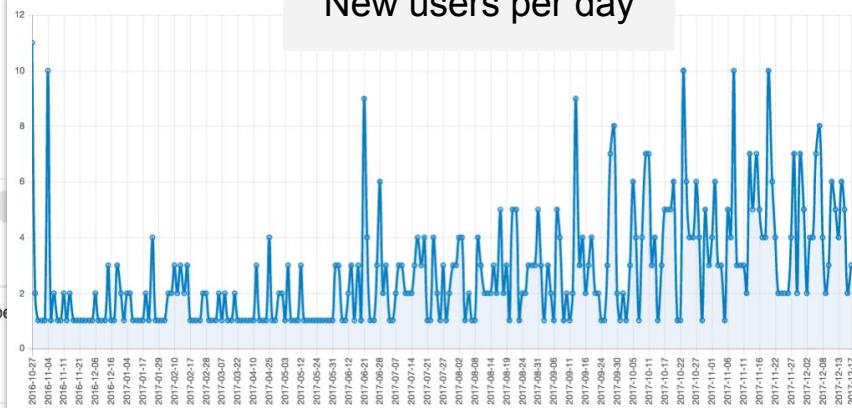
new
User support
dada2 queued

Boxplot of base quality scores after quality filtering

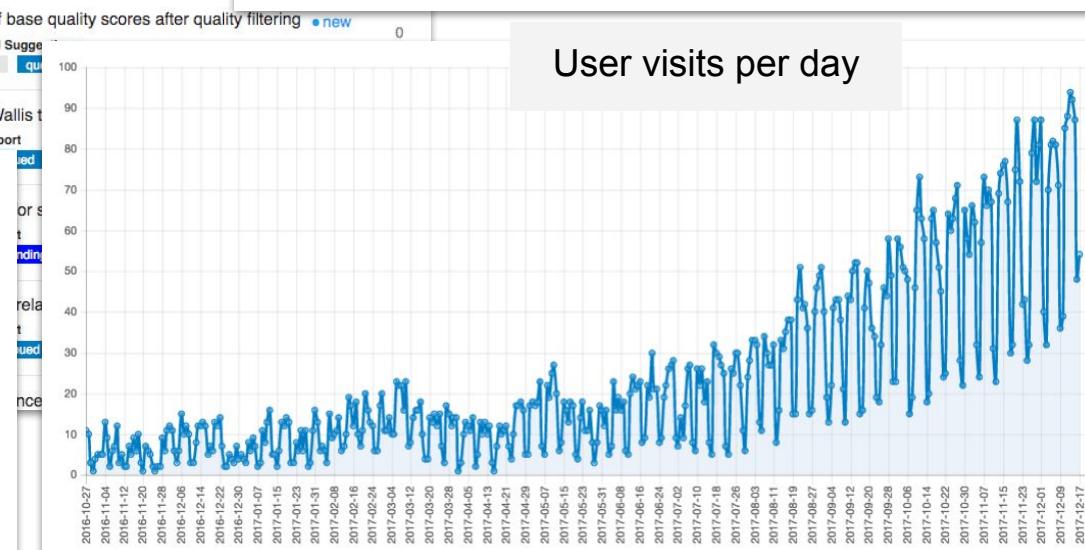
Ideas and Sugg...
quality-filter que...

Kruskal-Wallis t...
User support

New users per day



User visits per day



QIIME 2 workshops

<https://workshops.qiime2.org>

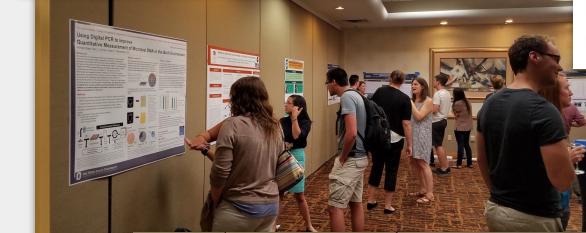
Upcoming Workshops

Title	Location	Workshop Dates
Teaching and Developing QIIME 2 (FULL)	San Diego, CA, USA	May 9, 2018 - May 11, 2018
Microbial Communities Profiling via QIIME 2 and QIita	Columbia University Medical Center	June 4, 2018 - June 5, 2018

Past Workshops

Title	Location	Workshop Dates
Microbiome Bioinformatics with QIIME 2 (SOLD OUT!)	Bethesda, Maryland	Feb. 22, 2018 - Feb. 23, 2018
Workshop on Genomics	Český Krumlov, Czech Republic	Jan. 7, 2018 - Jan. 20, 2018
DZIF bioinformatics workshop: 16S Community Profiling with QIIME 2	BRICS, Braunschweig, Germany	Dec. 19, 2017 - Dec. 22, 2017
Advanced Bioinformatics workshop for early career researchers.	Australian Centre for Ancient DNA, University of Adelaide, Adelaide, Australia	Nov. 13, 2017 - Nov. 17, 2017
Strategies and Techniques for Analyzing Microbial Population Structures (STAMPS); (includes a QIIME 2 session)	Woods Hole, MA, USA	July 30, 2017 - Aug. 9, 2017
Microbial Communities Profiling via QIIME 2 and QIita	Columbia Medical School	June 28, 2017 - June 29, 2017
Microbiome Bioinformatics with QIIME 2	Embassy Suites by Hilton Convention Center, Las Vegas, Nevada	June 21, 2017 - June 23, 2017
Microbiome Analysis in the Cloud (includes a QIIME 2 session)	Institute for Genome Sciences, University of Maryland School of Medicine	June 15, 2017 - June 16, 2017
QIIME 2 Workshop	University of New Hampshire	Feb. 13, 2017 - Feb. 14, 2017
Microbiome Analysis with QIIME 2	Northern Arizona University	Jan. 18, 2017 - May 10, 2017
Workshop on Genomics (includes a QIIME 2 session)	Český Krumlov, Czech Republic	Jan. 17, 2017 - Jan. 17, 2017

Las Vegas Workshop, June 2017



Register for the June 2017 *Microbiome Bioinformatics with QIIME 2* workshop in Las Vegas by 11:59pm Pacific Time on May 5th for a chance to win one of three limited-edition QIIME 2 cross-stitches. Be the envy of your lab by owning this beautiful piece of microbiome bioinformatics art!

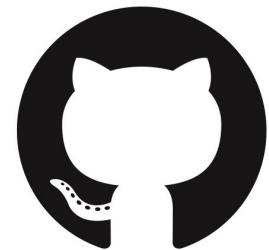
<https://workshops.qiime.org/qiime-2-workshop-2017-06-21>



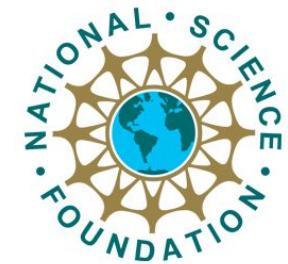
Fine art by Matthew Dillon (@thermokarst).

Completely Free and Open Source

- You have a *licensed right* to use, redistribute, and even modify QIIME 2
 - (Consider contributing to the QIIME 2 project via GitHub Pull Requests!)
- You can audit every line of code
- Tested on every code change
 - Unit tests, code coverage, linting
 - Tutorial commands automatically tested
- World-wide [network of collaborators](#)



Find all of our code at: <https://github.com/qiime2/>



[NSF Award: 1565100](#)

Low-level features (these get our users hooked)

- **Multiple user interfaces.** The same functionality is accessible through graphical interface, command line interface, and API, which target different types of users.
- **Plugin architecture** allows the software to keep pace with the field. Any developer can create and distribute a QIIME 2 plugin.
- **Decentralized provenance tracking** automates bioinformatics record keeping facilitating reproducibility.

"What did you do 5 months ago?"

notes.txt

```
echo "core_diversity_analyses.py -i  
/home/caporaso/analysis/atacama-7may2013/slout_r1/or_otus/otu_table_mc2_w_tax_no_pynast_failures.  
biom -o /home/caporaso/analysis/atacama-7may2013/slout_r1/or_otus/cd_16662/ -e 16662 -m  
/home/caporaso/analysis/atacama-7may2013/map.txt -ao 26 -t  
/home/caporaso/analysis/atacama-7may2013/slout_r1/or_otus/rep_set.tre -c  
SiteName,Depth,ExtractGroupNo, TransectID, Vegetation" | qsub -keo -N ata-cd
```

The above failed during OTU category significance (see the log file for the error - maybe one of these categories has a value that is observed only once?), so re-running without that step for now...

```
echo "core_diversity_analyses.py -i  
/home/caporaso/analysis/atacama-7may2013/slout_r1/or_otus/otu_table_mc2_w_tax_no_pynast_failures.  
biom -o /home/caporaso/analysis/atacama-7may2013/slout_r1/or_otus/cd_16662/ -e 16662 -m  
/home/caporaso/analysis/atacama-7may2013/map.txt -ao 26 -t  
/home/caporaso/analysis/atacama-7may2013/slout_r1/or_otus/rep_set.tre -c  
SiteName,Depth,ExtractGroupNo, TransectID, Vegetation --suppress_otu_category_significance  
--recover_from_failure" | qsub -keo -N ata-cd
```

```
echo "pick_open_reference_otus.py -i /home/caporaso/analysis/atacama-7may2013/slout_r2/seqs.fna  
-r /data/gg_13_5_otus/rep_set/97_otus.fasta -o  
/home/caporaso/analysis/atacama-7may2013/slout_r2/or_otus/ -ao 28 -p  
/home/caporaso/analysis/atacama-7may2013/uc_fast_params.txt" | qsub -keo -N or-3
```

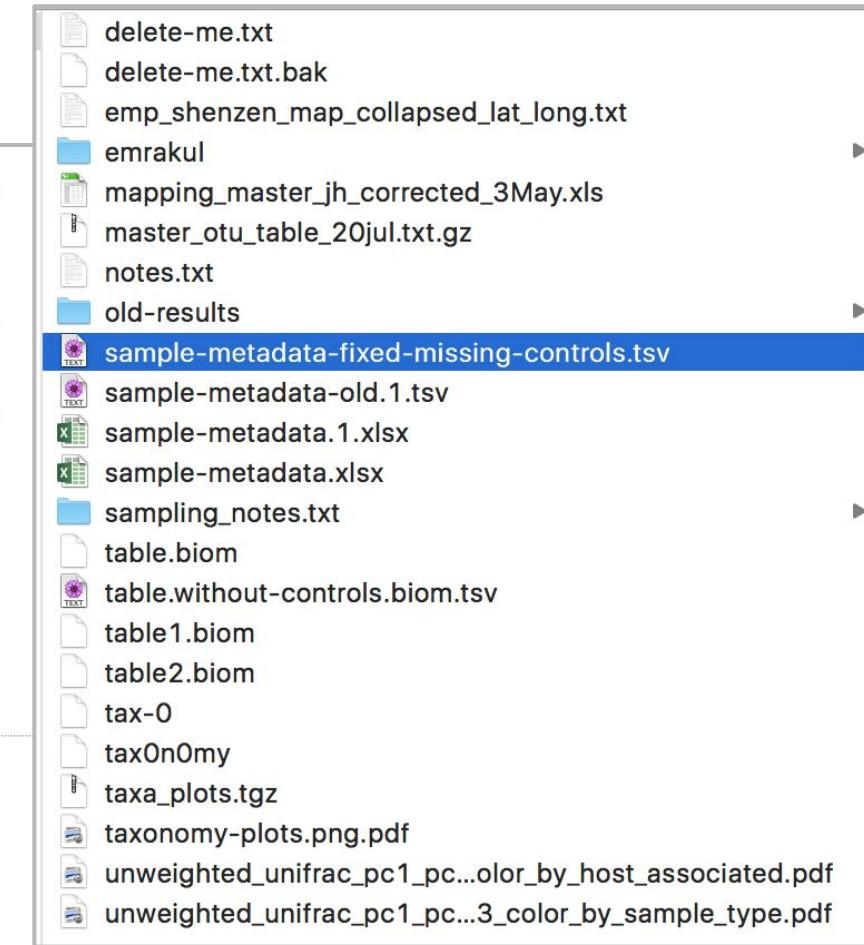
Katy's analysis on first and second sequencing runs

You have read access to all files in /home/caporaso/analysis/atacama-~~7may2013~~.

The split_libraries_fastq.py input files you need are:

sequences:

```
/home/caporaso/analysis/atacama-7may2013/2014.04.30/Undetermined_S0_L0  
01_R1_001.fastq.gz
```



QIIME 2: Dynamic, Interactive Visualizations with Provenance! (Live demo)

Taxonomic Composition Bar Plots

*Make sure you are using
Google Chrome or Firefox!*

QIIME 2 integrated data provenance ensures reproducibility (backup slide)

Two screenshots of the QIIME 2 interface illustrating data provenance.

The left screenshot shows a taxonomic bar plot. A context menu is open over the plot, showing a dropdown menu with "Level 1" through "Level 7" options. "Level 2" is selected. A color palette and sorting controls are also visible at the top.

The right screenshot shows the "Provenance Graph" and "Action Details" panels. The Provenance Graph illustrates the data processing workflow:

```
graph TD; seqs[seqs] --> demultiplexed_seqs[demultiplexed_seqs]; sequences[sequences] --> reference_taxonomy[reference_taxonomy]; demultiplexed_seqs --> reference_reads[reference_reads]; reference_reads --> reads_classifier[reads classifier]; reads_classifier --> tab[tab]; taxonomy[taxonomy] --> tab
```

The Action Details panel provides detailed information about the execution of the "denoise_single" method:

- execution:** uuid: "3897fb5c-55ed-46b1-a48d-ae0651d2b597"
- runtime:** start: 2017-09-28T21:14:34.374Z, end: 2017-09-28T21:23:05.935Z, duration: "8 minutes, 31 seconds, and 561708 microsecond s"
- action:** type: "method", plugin: "environment:plugins:dada2", action: "denoise_single"
- inputs:**
 - 0: demultiplexed_seqs: "ce7e102e-4b8c-455c-b2af-a7fb342b7fa1"
- parameters:**
 - 0: trunc_len: 120
 - 1: trim_left: 0
 - 2: max_eee: 2
 - 3: trunc_q: 2
 - 4: chimera_method: "consensus"
 - 5: min_fold_parent_over_abundance: 1
 - 6: n_threads: 1

How did QIIME 2 know what happened?

.qza and .qzv files are just structured zip files

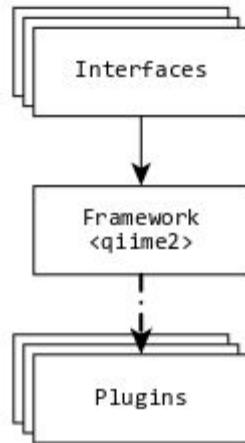
- Data, in the usual formats (e.g., newick, fasta, tsv, ...), is in the data/ directory.
- The rest is human-readable QIIME-specific metadata.

```
13:15:05 ~/Downloads evan@evan-laptop:~/Downloads
$ unzip rep-seqs.qza
Archive: rep-seqs.qza
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/metadata.yaml
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/VERSION
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/data/dna-sequences.fasta
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/metadata.yaml
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/VERSION
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/action/action.yaml
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/artifacts/150dbad9-9772-4aab-be03-c4c9c89d6573/metadata.yaml
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/artifacts/150dbad9-9772-4aab-be03-c4c9c89d6573/VERSION
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/artifacts/150dbad9-9772-4aab-be03-c4c9c89d6573/action/action.yaml
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/artifacts/5abbd33-13f7-4d46-87ac-064556741c49/metadata.yaml
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/artifacts/5abbd33-13f7-4d46-87ac-064556741c49/VERSION
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/artifacts/5abbd33-13f7-4d46-87ac-064556741c49/action/action.yaml
  inflating: f19ba9ad-ead3-4991-b4cf-e6d7ece0e1ae/provenance/artifacts/5abbd33-13f7-4d46-87ac-064556741c49/action/barcodes.tsv
```

Future plans for provenance

- Automatic data integrity checks:
“The reference database you used has been retracted by the author...”
- Replay analysis actions on a new dataset

QIIME 2 Doesn't Know Very Much:



- Interfaces are responsible for translating user intent into action.
- The framework provides the interface SDK and Plugin API.
- The plugins, which define *all* domain-specific functionality.

Choose the interface that fits your needs

q2cli the command line interface

```
2. ~ (zsh)  
$ qiime info  
System versions  
Python version: 3.5.3  
QIIME 2 release: 2017.6  
QIIME 2 version: 2017.6.0  
q2cli version: 2017.6.0  
  
Installed plugins  
alignment 2017.6.0  
composition 2017.6.0  
dada2 2017.6.0
```

q2studio the graphical user interface (PROTOTYPE)

q2studio is a functional prototype of a graphical user interface for QIIME 2, and is not necessarily feature-complete with respect to q2cli and the Artifact API.

QIIME 2 Studio

Active Jobs 1 Finished Jobs Failed Jobs

Action	Started	Elapsed
Denoise and dereplicate paired-end sequences	17-07-07 01:57:27	00:00:05

Artifacts 2 Visualizations Metadata 1

Name	UUID	Type	Actions
demux	043bdcdf-9f32-48ce-8c6d-4403bf550a59	SampleData[PairedEndSequencesWithQuality]	<button>Delete</button>
emp-paired-end-sequences	9c70333e-82d6-4f4a-9cf7-baebae8b642f	EMPPairedEndSequences	<button>Delete</button>

Untitled - idle

```
[1] import pandas as pd  
from qiime2 import Artifact  
  
[2] t = Artifact.load('table.qza')  
t.view(pd.DataFrame)  
  
4b5eeb300368260019c1fbc7a3c718fc  
  
L1S105 2222.0  
L1S140 0.0  
L1S208 0.0  
L1S257 0.0  
L1S281 0.0
```

Python 3 | idle Not saved yet



This interface can view .qza and .qzv files directly in your browser without uploading to a server. [Click here](#) to learn more.

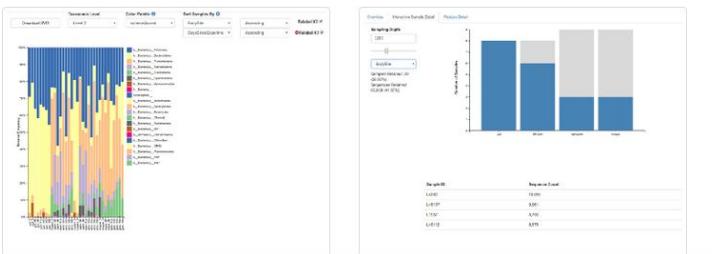
Drag and drop or click here

to view a QIIME 2 Artifact or Visualization (.qza/.qzv) from your computer.

You can also provide a link to a [file on Dropbox](#) or a [file from the web](#).

Gallery

Don't have a QIIME 2 result of your own to view? Try one of these!



q2view: a read-only web interface for viewing results without having QIIME 2 installed

- Reads QIIME 2 outputs (QZAs and QZVs)
- No installation required
- Easily share links using Dropbox with collaborators
- No uploading (your data stays on your computer)!

<https://view.qiime2.org> (live demo)

Interfaces Coming Soon!



COMMON
WORKFLOW
LANGUAGE

Anyone can create and distribute a plugin.

- They define *all* bioinformatics analysis functionality.
- They're Python 3 "method annotations" that QIIME 2 interprets.
- They can wrap methods not written in Python 3 (e.g., DADA2 is written in R, and mafft is a binary).

```
8
9 import qiime2.plugin
10 from q2_types.per_sample_sequences import (
11     SequencesWithQuality, PairedEndSequencesWithQuality)
12 from q2_types.sample_data import SampleData
13 from q2_types.feature_data import FeatureData, Sequence
14 from q2_types.feature_table import FeatureTable, Frequency
15
16 import q2_dada2
17
18 plugin = qiime2.plugin.Plugin(
19     name='dada2',
20     version=q2_dada2.__version__,
21     website='http://benjneb.github.io/dada2/',
22     package='q2_dada2',
23     description='This QIIME 2 plugin wraps DADA2 and supports '
24         'sequence quality control for single-end and paired-end '
25         'reads using the DADA2 R library.'),
26     short_description='Plugin for sequence quality control with DADA2.',
27     citations=qiime2.plugin.Citations.load('citations.bib', package='q2_dada2')
28 )
29
30
31 plugin.methods.register_function(
32     function=q2_dada2.denoise_single,
33     inputs={'demultiplexed_seqs': SampleData[SequencesWithQuality | 
34                                                 PairedEndSequencesWithQuality]},
35     parameters={'trunc_len': qiime2.plugin.Int,
36                 'trim_left': qiime2.plugin.Int,
37                 'max_ee': qiime2.plugin.Float,
38                 'trunc_q': qiime2.plugin.Int,
39                 'chimera_method': qiime2.plugin.Str %
40                     qiime2.plugin.Choices(_CHIM_OPT),
41                 'min_fold_parent_over_abundance': qiime2.plugin.Float,
42                 'n_threads': qiime2.plugin.Int,
43                 'n_reads_learn': qiime2.plugin.Int,
44                 'hashed_feature_ids': qiime2.plugin.Bool},
45     outputs=[('table', FeatureTable[Frequency]),
46              ('representative_sequences', FeatureData[Sequence])],
47     input_descriptions={
48         'demultiplexed_seqs': ('The single-end demultiplexed sequences to be '
49                               'denoised.')
50     },
51     parameter_descriptions={
52         'trunc_len': ('Position at which sequences should be truncated due to '
53                       'decrease in quality. This truncates the 3\' end of the '
54                       'sequences.')
55     }
56 )
```



Slides: goo.gl/QXVa4h
Slides (pdf): goo.gl/B1Qnk7
Schedule: goo.gl/XDrhf3

Coffee Break!

Moving Pictures Intro and Sample Metadata*

* a.k.a. *sample data, variables, mapping file*

For more on this topic, see [Conducting a Microbiome Study](#), by Goodrich et al.

Moving Pictures of the Human Microbiome

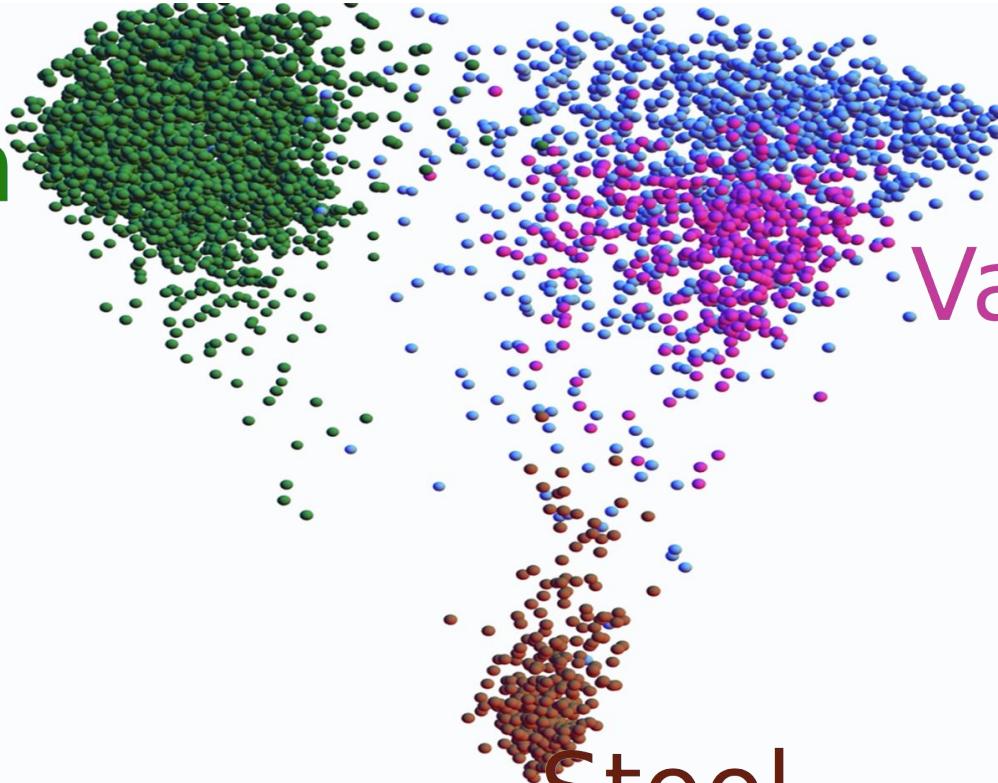
- Two human subjects provided daily microbiome samples, one for 6 months and the other for 18 months.
- Sampled tongue, left and right palms, and gut (via feces).
- Tutorial data is a subset of this full data set.
- Illumina GAIx sequencing with a prototype of what later became the Earth Microbiome Project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

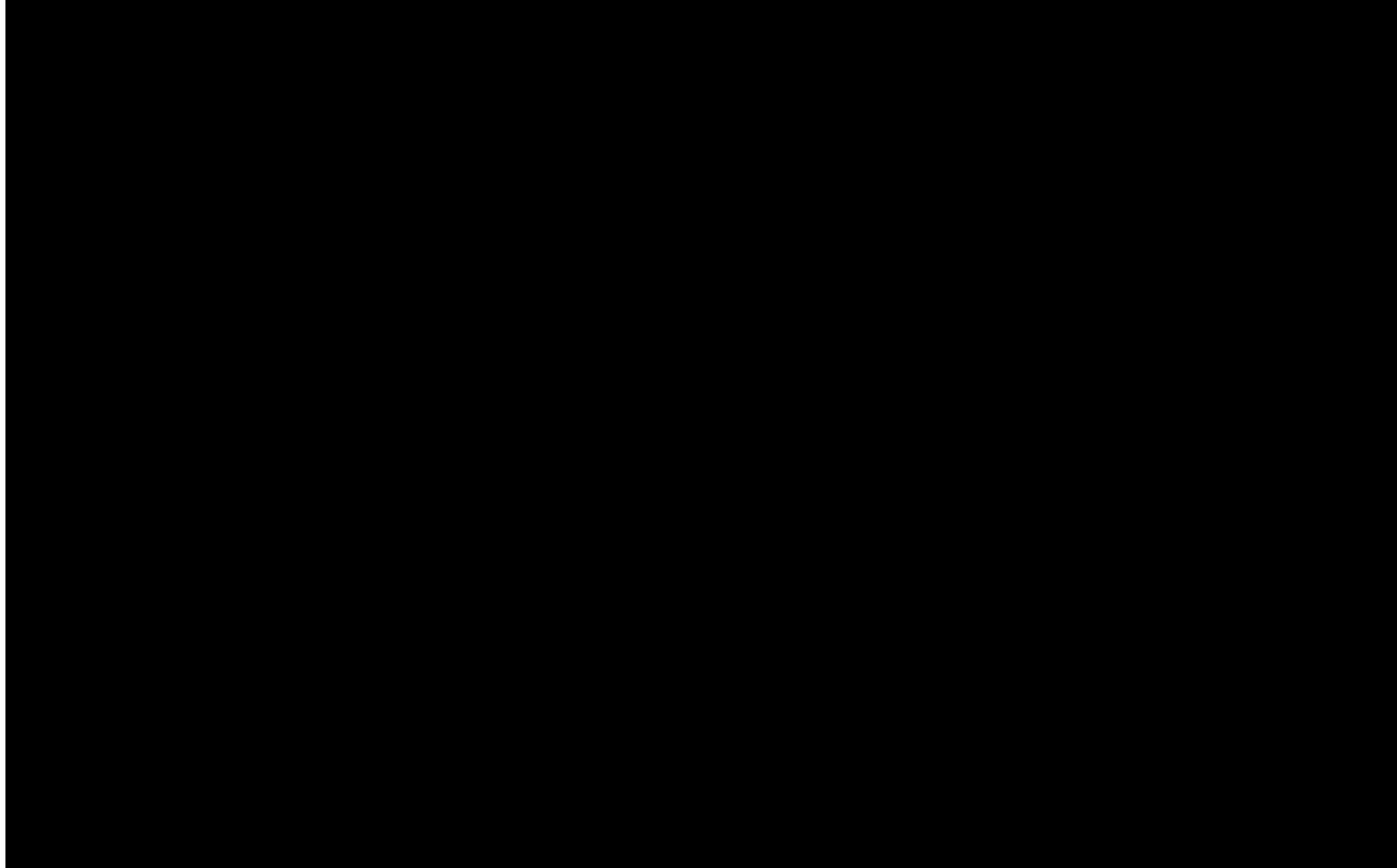
Mouth

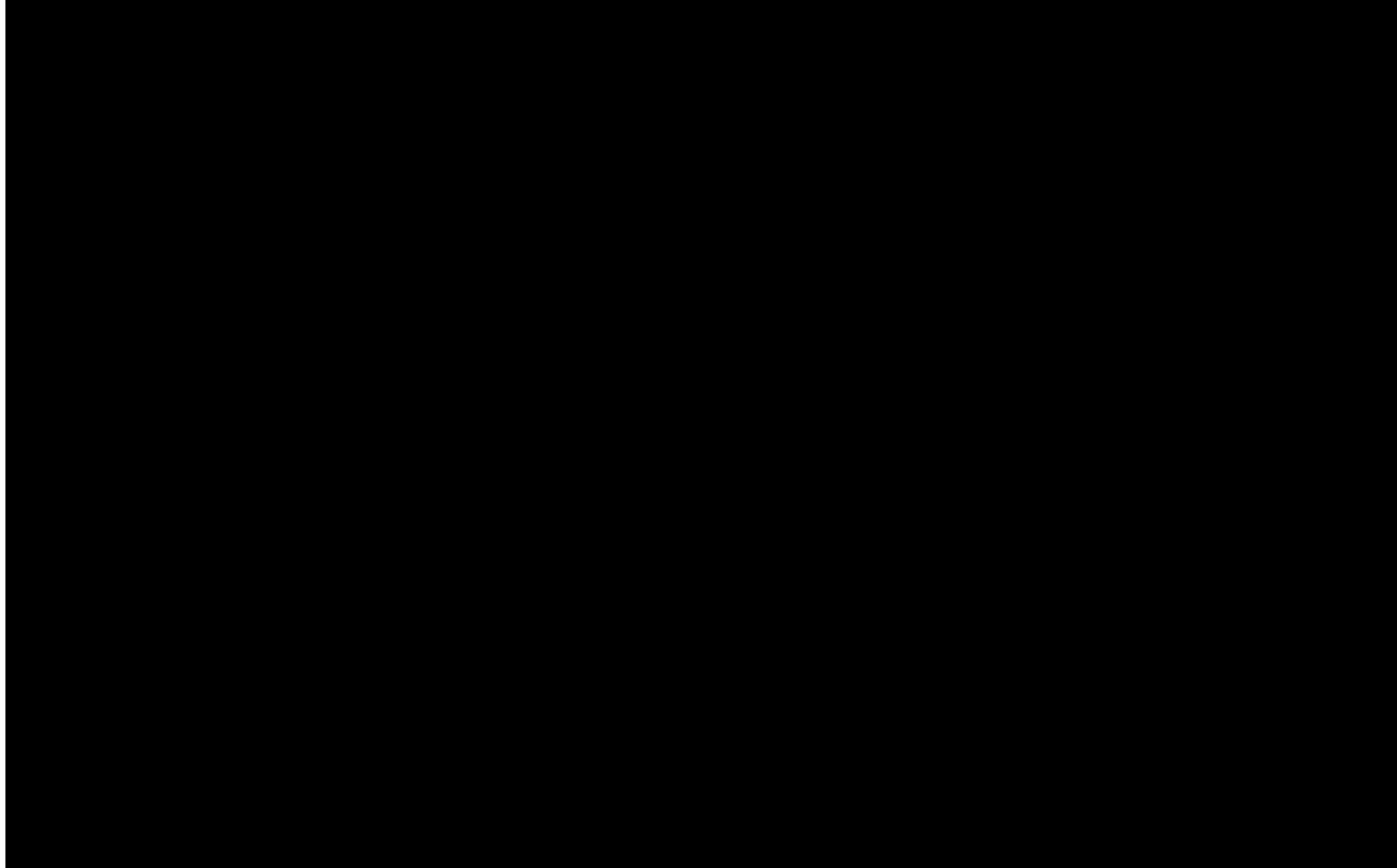
Skin

Vagina

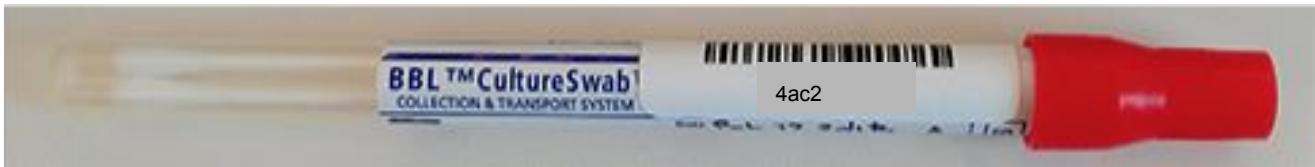
Stool







Collect samples and metadata



sample-metadata.tsv		
SampleID	Date of collection	Body site
4ac2	1 Nov 2016	Gut
e375	2 Nov 2016	Left palm
4gd8	2 Nov 2016	Right palm
9872	2 Nov 2016	Tongue

Choose your sample ids carefully! The examples here would not be good in practice, see [cual-id](#) (Chase et al., 2015) for help with this.

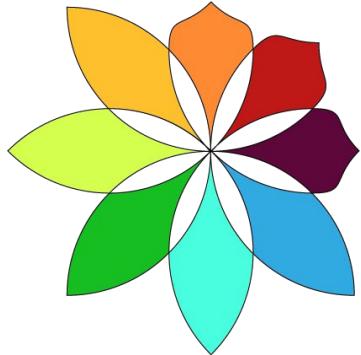
Sample metadata: what to include

Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIS) specifications.

Yilmaz et al. Nature Biotechnology 29, 415–420 (2011).

Specification projects	MIGS					MIMS	MIMARKS		New checklists
Checklists	EU	BA	PL	VI	ORG	metagenomes	survey	specimen	e.g., pan-genomes
Shared descriptors	collection date, environmental package, environment (biome), environment (feature), environment (material), geographic location (country and/or sea, region), geographic location (latitude and longitude), investigation type, project name, sequencing method, submitted to INSDC								
Checklist-specific descriptors	assembly, estimated size, finishing strategy, isolation and growth condition, number of replicons, ploidy, propagation, reference for biomaterial			target gene					
Applicable environmental packages (measurements and observations)	Air Host-associated Human-associated Human-oral Human-gut Human-skin Human-vaginal			Microbial mat/biofilm Miscellaneous natural or artificial environment Plant-associated Sediment Soil Wastewater/sludge Water					

Sample metadata: formatting and validating



[Keemei: cloud-based validation of tabular bioinformatics file formats in Google Sheets.](#)

Rideout JR, Chase JH, Bolyen E, Ackermann G, González A, Knight R, Caporaso JG. GigaScience. 2016;5:27.

<http://keemei.qiime2.org>

Moving Pictures sample-metadata (QIIME 2.0.6)

File Edit View Insert Format Data Tools Add-ons Help Last edit was yesterday at 12:02 PM

#	#SampleID	BarcodeSequence	LinkerPrimerSeq	BodySite	Year	Month?	Day	Subject	ReportedAntibioticUsage	DaysSinceExperimentStart
2	L1S8	ERRORS:		ut	2008	10	28	1	Yes	0
3	L1S140			ut	2008	10	28	2	Yes	0
4	L1S57	Duplicate sample ID. Duplicates in A2, A21		ut	2009	1	20	1	No	84
5	L1S208			ut	2009	1	20	2	No	84
6	L1S76	ACTACGTGTGC GTGCCAGCMG	gut		2009	2	17	1	No	112
7	L1S105	AGTGCATGCA	GTGCCAGCMG	gut	2009	3	17	1	No	140
8	L1S257	CCGACTGAGAT	GTGCCAGCMG	gut	2009	3	17	2	No	140
9	L1S281	CCTCTCGTGT	GTGCCAGCMG	gut	2009	4	14	2	No	168
10	L2S240	CATATCGCAGT	GTGCCAGCMG	left palm	2008	10	28	2	Yes	0
11	L2S155	ACGATGGACG	GTGCCAGCMG	left palm	2009	1	20	1	No	84
12	L2S309	CGTGCATTATC	GTGCCAGCMG	left palm	2009	1	20	2	No	84
13	L2S175	AGCTATCCACG	GTGCCAGCMG	left palm	2009	2	17	1	No	112
14	L2S204	ATGCAGCTCAC	GTGCCAGCMG	left palm	2009	3	17	1	No	140

Logging into the servers

Logging into the servers

We'll use QIIME 2 on Amazon EC2 instances for this workshop.

We're happy to help you install QIIME 2 on your computer, just come find us during the course or try out the [installation instructions](#) yourself (it's much easier than installing QIIME 1).

Logging into the servers

```
ssh <username>@workshop-server.qiime2.org
```

<username>: your STAMPS username (back of nametag)

Password: stamps2018

Our First Plugin

Interactive example of q2-metadata

To download your first visualization, go to:

<https://workshop-server.qiime2.org/<username>>

Moving Pictures tutorial: A Very Simple Plugin

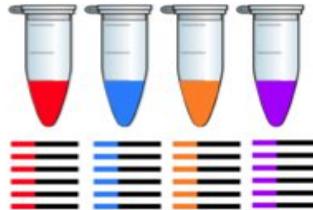
Open this link in a new tab/window and keep it open,
we'll be following along with the online tutorial.

<https://docs.qiime2.org/2018.6/tutorials/moving-pictures/>

Demultiplexing, Denoising, and Phylogenetic Reconstruction

Extract DNA, isolate and amplify the rRNA from all samples using barcoded PCR, and sequence.

Barcoded per-sample
rRNA

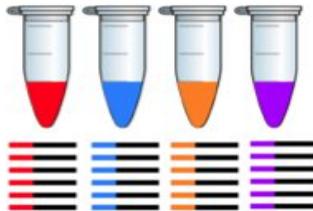


Track per-sample
barcodes (e.g., in
spreadsheet)

sample-metadata.tsv	
SampleID	BarcodeSequence
4ac2	AACGCAC
e375	AAGAGAT
4gd8	ACAGCAG
9872	ACAGCTA

Extract DNA, isolate and amplify the rRNA from all samples using barcoded PCR, and sequence.

Barcoded per-sample rRNA



Track per-sample
barcodes (e.g., in
spreadsheet)

Pool and sequence samples



sample-metadata.tsv	
SampleID	BarcodeSequence
4ac2	AACGCAC
e375	AAGAGAT
4gd8	ACAGCAG
9872	ACAGCTA

sequences.fastq(.gz

```
T barcodes.fastq(.gz)
```



Demultiplexed sequence data

4ac2.fastq(.gz)
e375.fastq(.gz)
4gd8.fastq(.gz)
9872.fastq(.gz)
<pre> @HWI-6X 9267:1:1:25:1109 TACGGAGGGTGCAGCGTTAACGGAATTACTGGCGTAA AGCGTACGTAGCGGTTAGGTAAAGTCAGATGTGAAAGCCC CGGGCTCCACCTGGGAATGG + aaaba^`a^N `\\ ``a a]Zaa^^\Z`[M]a`[VY a^ X ^ Z]NZ`^]TY\] ^RVH PHOWZM[PTRPTRYUBBBBBB BBBBBBBBBBBBBBBBBBBBBB </pre>

or

Multiplexed sequence data

sample-metadata.tsv			
SampleID	BarcodeSequence		
4ac2	AACGCAC		
e375	<table border="1"> <tr> <td>sequences.fastq(.gz)</td> </tr> <tr> <td> <pre> @HWI-6X_9267:1:1:25:1051 GACGAAGGTGACGACCGTTGCTCGGAATCACTGGGCATAAAGCGCGTAGGTG GCTTGGTAAGTCCA + abaaaaaa^`a_]`SVYGYVDXOZVT\T @HWI-6X_9267:1:TACGTATGGGGCAA GTGGCTTAAGCGCA + aa^^[___^__^_ _ _XUWWURZUYYY]XXR @HWI-6X_9267:1:TACGTAGGGGCAA GATGGACAAGTCTG </pre> </td></tr> </table>	sequences.fastq(.gz)	<pre> @HWI-6X_9267:1:1:25:1051 GACGAAGGTGACGACCGTTGCTCGGAATCACTGGGCATAAAGCGCGTAGGTG GCTTGGTAAGTCCA + abaaaaaa^`a_]`SVYGYVDXOZVT\T @HWI-6X_9267:1:TACGTATGGGGCAA GTGGCTTAAGCGCA + aa^^[___^__^_ _ _XUWWURZUYYY]XXR @HWI-6X_9267:1:TACGTAGGGGCAA GATGGACAAGTCTG </pre>
sequences.fastq(.gz)			
<pre> @HWI-6X_9267:1:1:25:1051 GACGAAGGTGACGACCGTTGCTCGGAATCACTGGGCATAAAGCGCGTAGGTG GCTTGGTAAGTCCA + abaaaaaa^`a_]`SVYGYVDXOZVT\T @HWI-6X_9267:1:TACGTATGGGGCAA GTGGCTTAAGCGCA + aa^^[___^__^_ _ _XUWWURZUYYY]XXR @HWI-6X_9267:1:TACGTAGGGGCAA GATGGACAAGTCTG </pre>			
4gd8	<table border="1"> <tr> <td>barcodes.fastq(.gz)</td> </tr> <tr> <td> <pre> @HWI-6X_9267:1:1:25:1051 AACGCAC + bbbbbbb @HWI-6X_9267:1:1:25:267 AAGAGAT + bbbbbbb @HWI-6X_9267:1:1:25:609 AACGCAC + </pre> </td></tr> </table>	barcodes.fastq(.gz)	<pre> @HWI-6X_9267:1:1:25:1051 AACGCAC + bbbbbbb @HWI-6X_9267:1:1:25:267 AAGAGAT + bbbbbbb @HWI-6X_9267:1:1:25:609 AACGCAC + </pre>
barcodes.fastq(.gz)			
<pre> @HWI-6X_9267:1:1:25:1051 AACGCAC + bbbbbbb @HWI-6X_9267:1:1:25:267 AAGAGAT + bbbbbbb @HWI-6X_9267:1:1:25:609 AACGCAC + </pre>			
9872			

If data is still multiplexed, assign sequence reads to samples (i.e., *demultiplex* or *demux*).

sample-metadata.tsv

SampleID	BarcodeSequence
4ac2	AACGCAC
e375	AAGAGAT
4gd8	ACAGCAG
9872	ACAGCTA

barcodes.fastq(.gz)

@HWI-6X 9267:1:1:25:105

AACGCAC
+
Bbbbbbb

@HWI-6X_9267:1:1:25:267

AAGAGAT
+
bbbbbbb

@HWI-6X_9267:1:1:25:609

AACGCAC
+
bbbbbbb

©HWI-6X 9267:1:1:25:519

ACAGCAG
+
hhhhhhh

@HWI-6X 9267:1:1:25:110

ACAGCTA
+
bbbbb

sequences.fastq(.gz)

@HWI-6X 9267:1:1:25:1051

GACGAAGGTGACGACCGTTGCTCGGAATCACTGGCATAAAGCGCGCTAGGTG
GCTTGGTAAGTCCATGGTGAATCCTCGGCTAACCGAGGAACTG

+ 1 2 3 4 5 6 7 8 9 10

TACGTATGGGCAAGCGTTATCCGAATTATTGGCGTAAAGAGTCGTAGGTG
GTGGCTTAAGCGCAGGGTTAACGCATGGCTTAACATTGTTCTC
+

©HWI-6X 9267:1:1:25:609

TACGTAGGGGCAAGCGTTATCCGGATTACTGGGTAAAGGGACGGTAGACG
GATGGACAAGTCTGTAGTGAAAGGCTGGGCTCAACCCCCGGGACGG
+
+

aaaab`aaa`aaaaaaaaaaaaaaaaaaaa^aaaaaaaaYQ^` ^]a]\a`a]\`Z_[] [I^`azz^WW^ `^ZZ T]XY^`^ZX\ZJS[W[V^`HOVYTET@HWI-6X 9267:1:1:25:519

GACGGAGGATGCAAGTGTATCCGAATCACTGGCGTAAAGCGTCTGTAGGTG
GTTTACTAAGTCAACTGTTAACATCTGAGGCTAACCTCGAAATCG
+

aaaaaaaa`aaaaaa\aaaaaaaa` `` aa aaaa^Z [ZY^aa\U[^\[]YZ]
WY]] Z XX\[\[]]^`(\XTVX)`T_VZ[\]ZXVXYFX_VYJWWZL

@HWI-6X 9267:1:1:25:1109

TACGGAGGGTGCAGCGTTAACGAACTACTGGCGTAAAGCGTAGTAGCGC
GTTAGGTAAGTCAGATGTGAAAGCCCCGGGCTCACCTGGGAATGG
+

QIIME 2 supports multiplexed or demultiplexed (per sample) FASTQ files as input.

Sequence quality control, and optionally grouping sequences by similarity (i.e., OTU picking).

SampleData [SequencesWithQuality]

4ac2.fastq(.gz)

```
@HWI-6X 9267:1:1:25:1109
GAC
AGC
TAC
TCG
AGA
+
aba
ATG
GAC
AGC
GAG
+
ABA
XYU
U^V
@HWI-6X 9267:1:1:25:1109
TACGGAGGGTGCAGCGTTAACGGAAATTACTGGCGTAA
AGCGTACGTAGCGGTAGGTAAGTCAGATGTGAAAGCCC
CGGGCTCCACCTGGGATGG
+
aaaba^`a^N ` ``a a]Zaa^^\Z`[M]a`[VY a^ X
^ Z]NZ``]TY\] ^RVH PHOWZM[PTRPTRYUBBBBBB
BBBBBBBBBBBBBBBBBBBBBB
```

FeatureTable [Frequency]

	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

FeatureData [Sequence]

```
>feature5
GACGAAGGTGACGACC GTT GCT CGGA ATCA CTGGG CATA AGCG CGCTAGGTG
GCTTGGTAAGTCCATGGT GAA ATCCCTCGGCTCAACCGAGGA ACTG

>feature4
TACGTAGGGGCAAGCGTTATCCGGATTACTGGGTG TAAAGGGAGCGTAGACG
GATGGACAAGTCTGATGTGAAAGGCTGGGCTCAACCCGGGACGG

>feature2
TACGTATGGGCAAGCGTTATCCGGATTATGGCGTAAAGAGTGCCTAGGTG
GTGGCTTAAGCGCAGGGTTAAGGCAATGGCTTA ACT ATT GTT CTC

>feature1
GACGGAGGATGCAAGTGTATCCGGAAATCACTGGGCTAAAGCGTCTGTAGGTG
GTTTACTAAGTCAACTGTTAAATCTTGAGGCTCAACCTCGAAATCG
```

In a perfect world...

GGCGAGCGTT
GGCGAGCGGT
GGACGGCGTT
GGACGGCGTT
GGACGGCGTT
GGACGGCTTT
GGACGGCTTT
GGACGGCTTT
GGACGGCTTT
GGACGGCTTT
GGACGGCTGT
GGACGGCTGT

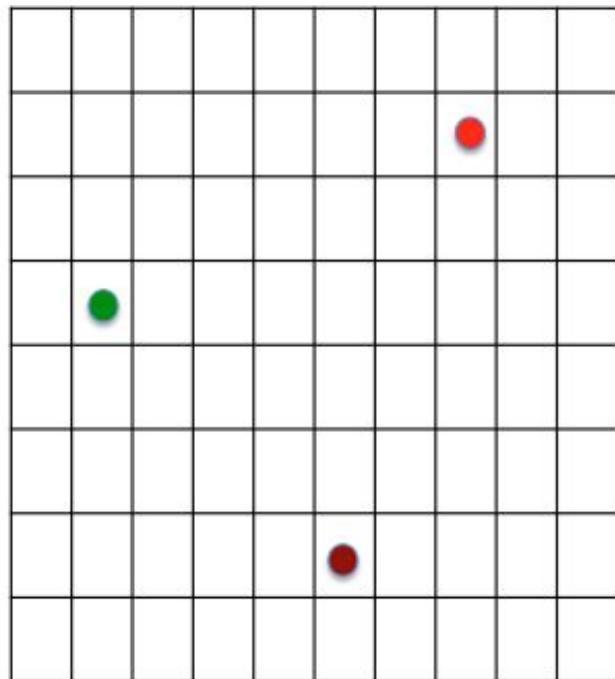
90% OTU Clustering

FeatureTable[Frequency]		
	OTU1	OTU2
4ac2	100	79
e375	88	35
4gd8	86	51
9872	12	87

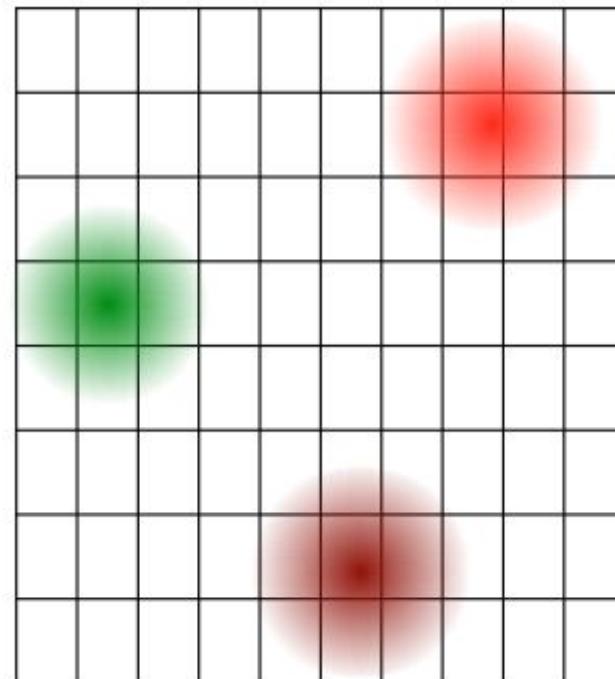
Amplicon Sequence Variants
(100% OTU Clustering)

FeatureTable[Frequency]					
	SV1	SV2	SV3	SV4	SV5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

What normally happens during sequencing?



True sequences

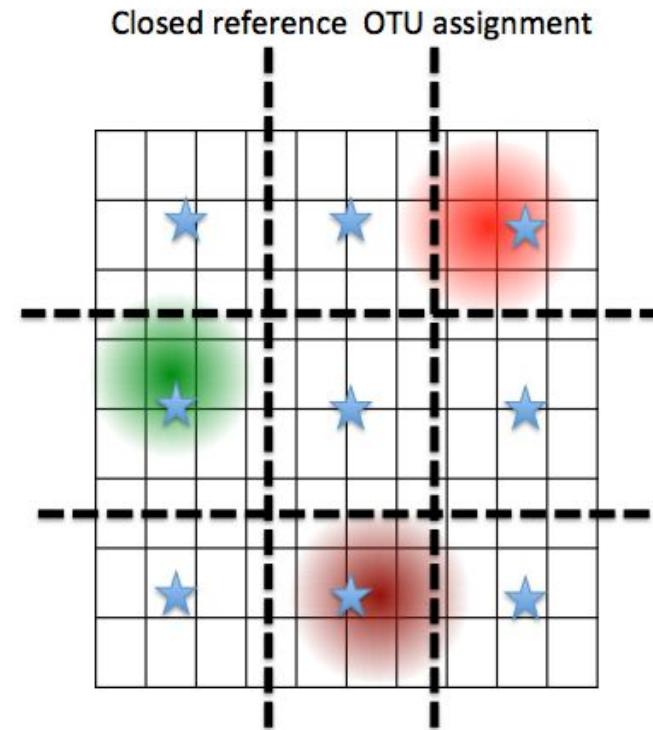
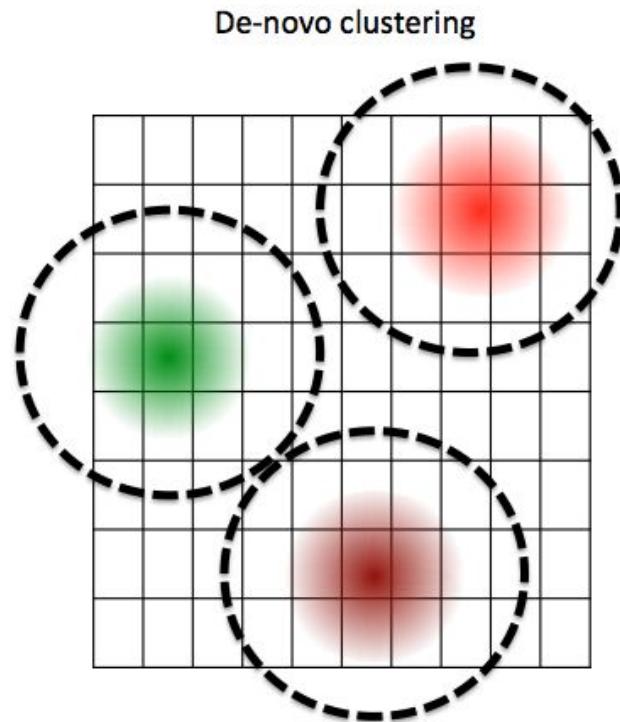


After Sequencing

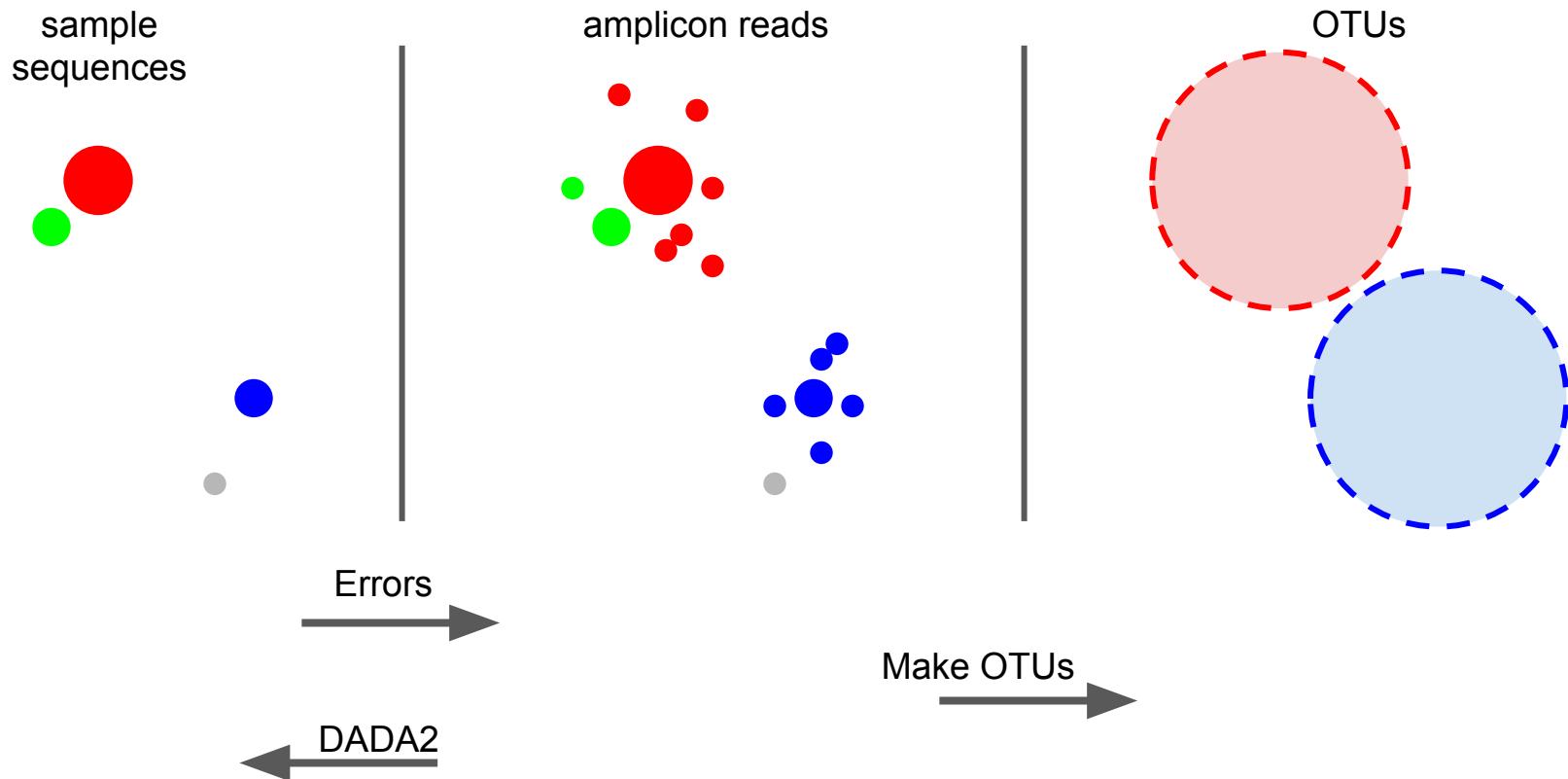
Cleaning and manipulating raw sequences

- Clustering
 - Remove noisy sequences and reduce the amount of sequences to process
 - Works based on a given threshold, i.e. 97% similarity but other exist like Oligotyping
 - There are different methods (closed or open reference) and algorithms (sortmerna, vclust)
- Remove noise
 - Find the cleanest sequence
 - Correct and/or discard super noisy sequences
 - Examples are: DADA2 and Deblur

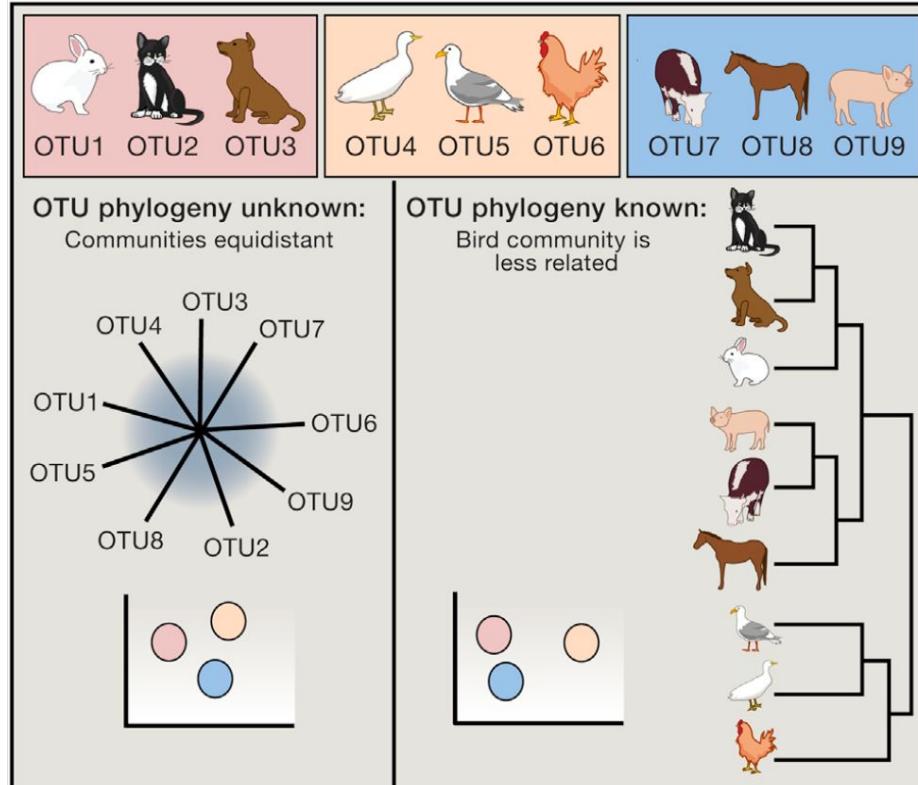
Clustering methods **ideal** situation



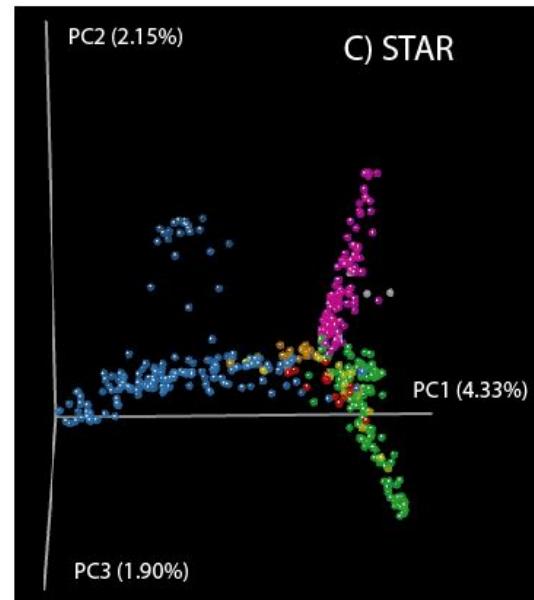
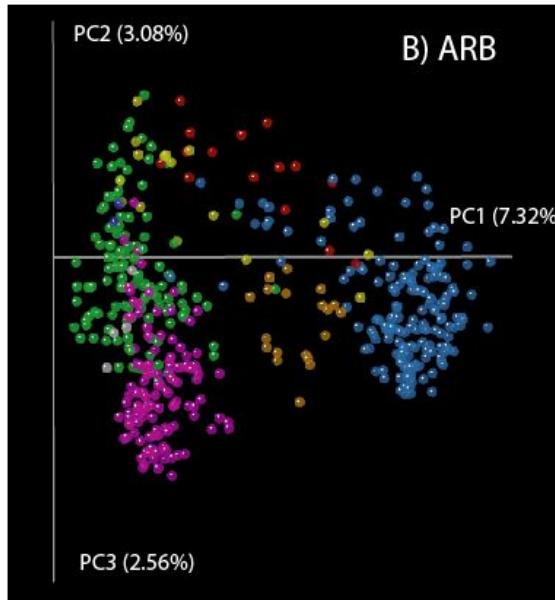
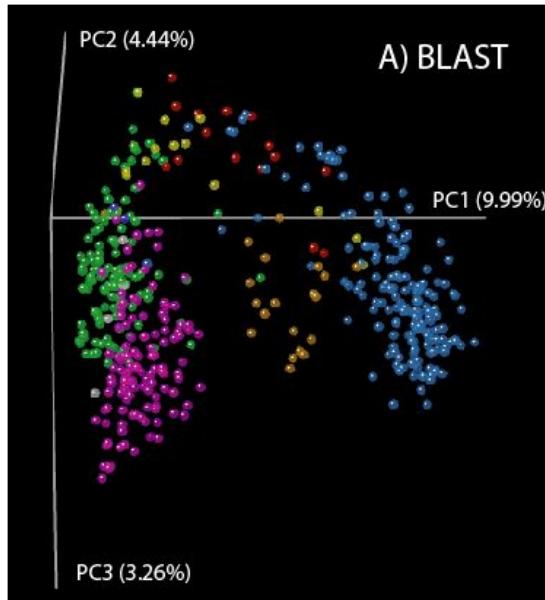
DADA2



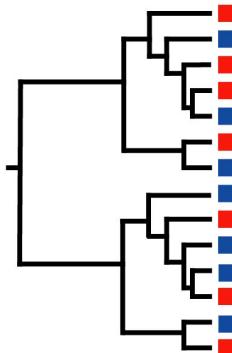
Operational Taxonomic Units (OTUs)



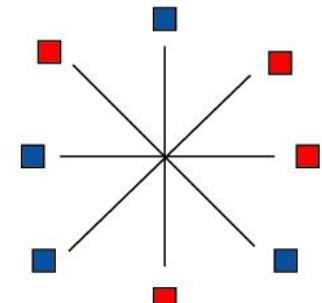
Operational taxonomic units are more generally referred to as features.



- Vertebrate Gut ● Termite Gut ● Human Skin/Mouth ● Invertebrate Associated
- Saline Freeliving ● NonSaline Freeliving ● Mixed Freeliving



BLAST to greengenes core set tree replicates Arb parsimony insertion results when phylogenetic relationships matter



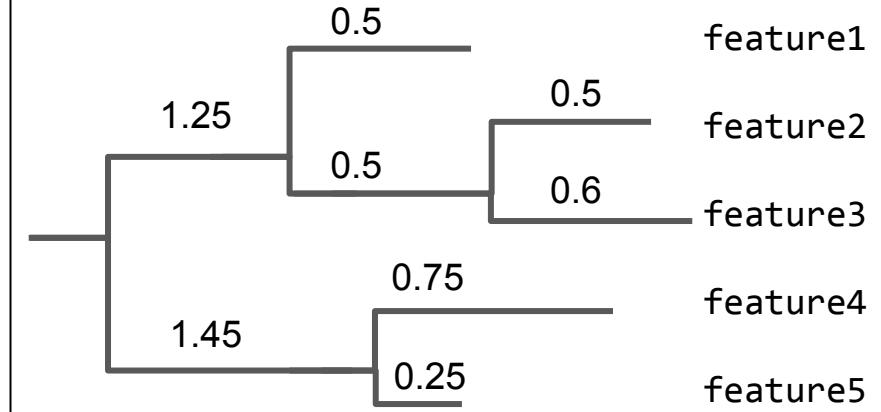
Phylogenetic reconstruction of observed sequences (optional).

FeatureData [Sequence]

```
>taxon5
GACGAAGGTGACGCCGTTGCTCGGAATCACTGGGCATAAAGCGCGTAGGTGGCTTGGTAAGTCCATGGTGAA
ATCCCTGGCTCAACCGAGGAACGT
>taxon4
TACGTAGGGGCAAGCGTTATCCGGATTACTGGGTGAAAGGGAGCGTAGACGGATGGACAAGTCTGATGTGAA
AGCGTGGGCTCAACCCGGGACGG
>taxon2
TACGTATGGGCAAGCGTTATCCGGATTATTGGCGTAAAGAGTGCCTAGGTGGCTTAAGCGCAGGGTTA
AGGCAATGGCTTAACATTGTTCTC
>taxon1
GACCGAGGATCCAAGCTTATCCGAATCACTGGCGTAAAGCGCTGTAGGTGGTTACTAAGTCAACTGTTAA
ATCTGAGGCTAACCTCGAAATCG
>taxon3
TACGGAGGGTGCAGCGTTAATCGGAAATTACTGGCGTAAAGCGTACGTAGGCAGTTAGTAAGTCAGATGTGAA
AGCCCCGGCTCCACCTGGGAATGG
```

Align sequences,
filter highly variable
(i.e., randomly
evolving) positions,
and build
phylogenetic tree.

Phylogeny [Rooted]



Moving Pictures tutorial: Demultiplexing, Denoising, and Building a Phylogeny

[https://docs.qiime2.org/2018.6/tutorials/moving-pictures/
#demultiplexing-sequences](https://docs.qiime2.org/2018.6/tutorials/moving-pictures/#demultiplexing-sequences)



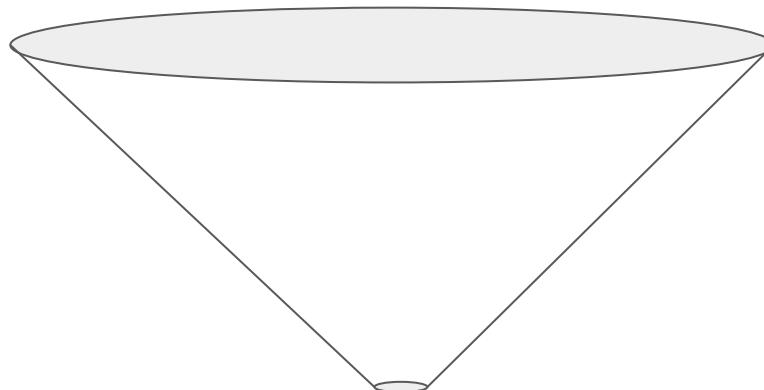
Slides: goo.gl/QXVa4h
Slides (pdf): goo.gl/B1Qnk7
Schedule: goo.gl/XDrhf3

Lunch Time

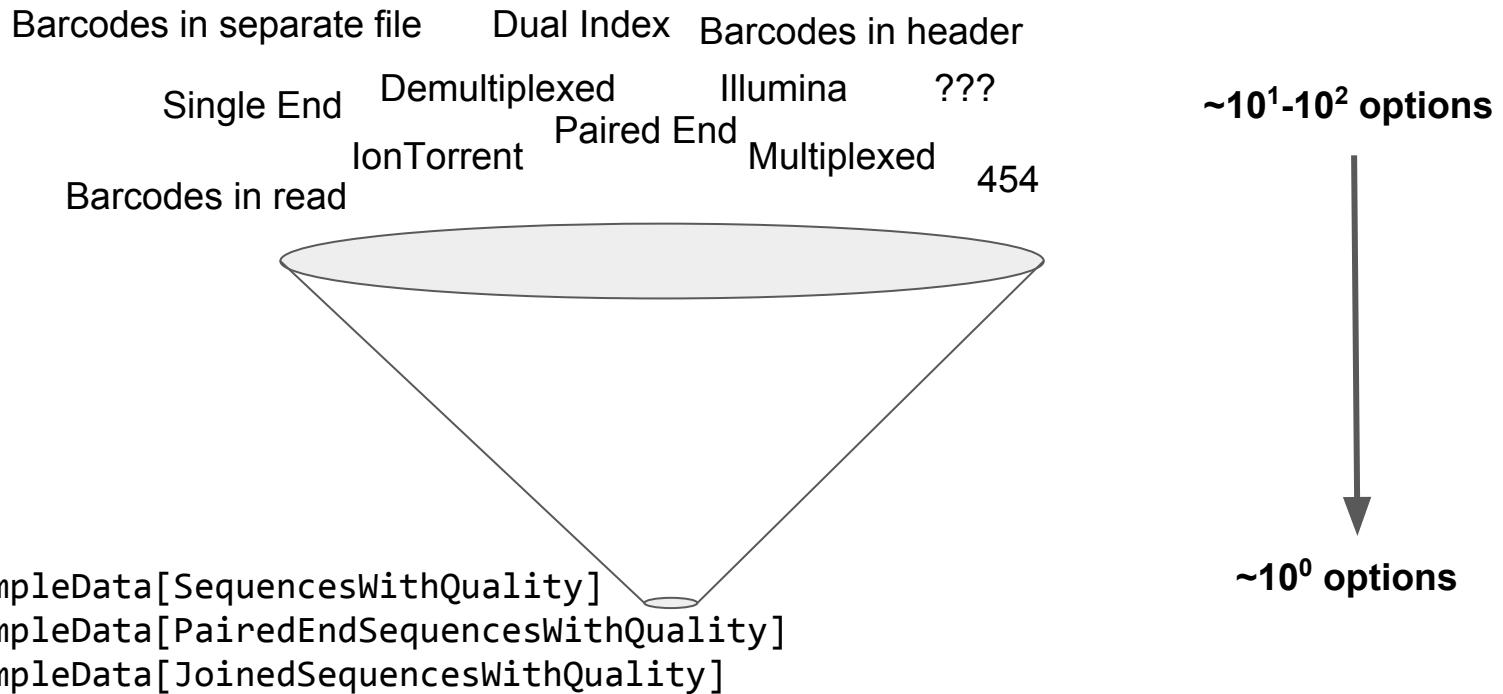
Getting Data into (and out of) QIIME 2

Opinion: Getting sequence data into QIIME 2 is the hardest part of QIIME 2 (or any software package)

Barcodes in separate file	Dual Index	Barcodes in header	
Single End	Demultiplexed	Illumina	???
		Paired End	
	IonTorrent	Multiplexed	
Barcodes in read			454



Opinion: Getting sequence data into QIIME 2 is the hardest part of QIIME 2 (or any software package)



Opinion: Getting sequence data into QIIME 2 is the hardest part of QIIME 2 (or any software package)

Once your data is in one of these formats, your analysis in QIIME 2 can begin!

`SampleData[SequencesWithQuality]`
`SampleData[PairedEndSequencesWithQuality]`
`SampleData[JoinedSequencesWithQuality]`

Types and Formats

- Formats are a *way to lay-out data on-disk*
 - A particular format or objects is a *view* of the data.
 - Many ways to view the data, different tools have different rules for input
 - Formats can be negotiated via transformers
- Types are *what the data represents*
 - The types define the *valid composition* of actions.

Getting Data into QIIME 2

[https://docs.qiime2.org/2018.6/tutorials/moving-pictures/
#obtaining-and-importing-data](https://docs.qiime2.org/2018.6/tutorials/moving-pictures/#obtaining-and-importing-data)

Getting Data Out of QIIME 2

Going to lose provenance :(

```
qiime tools export table.qza --output-dir  
my_table
```

QIIME 2 - 2018.8

- `qiime tools import/export` are being revamped
- Export to multiple formats!
 - (This was always possible via the API)

QIIME 2 uses Artifacts and Visualizations

- Where did my FASTA file go?
 - It's inside a ZIP file

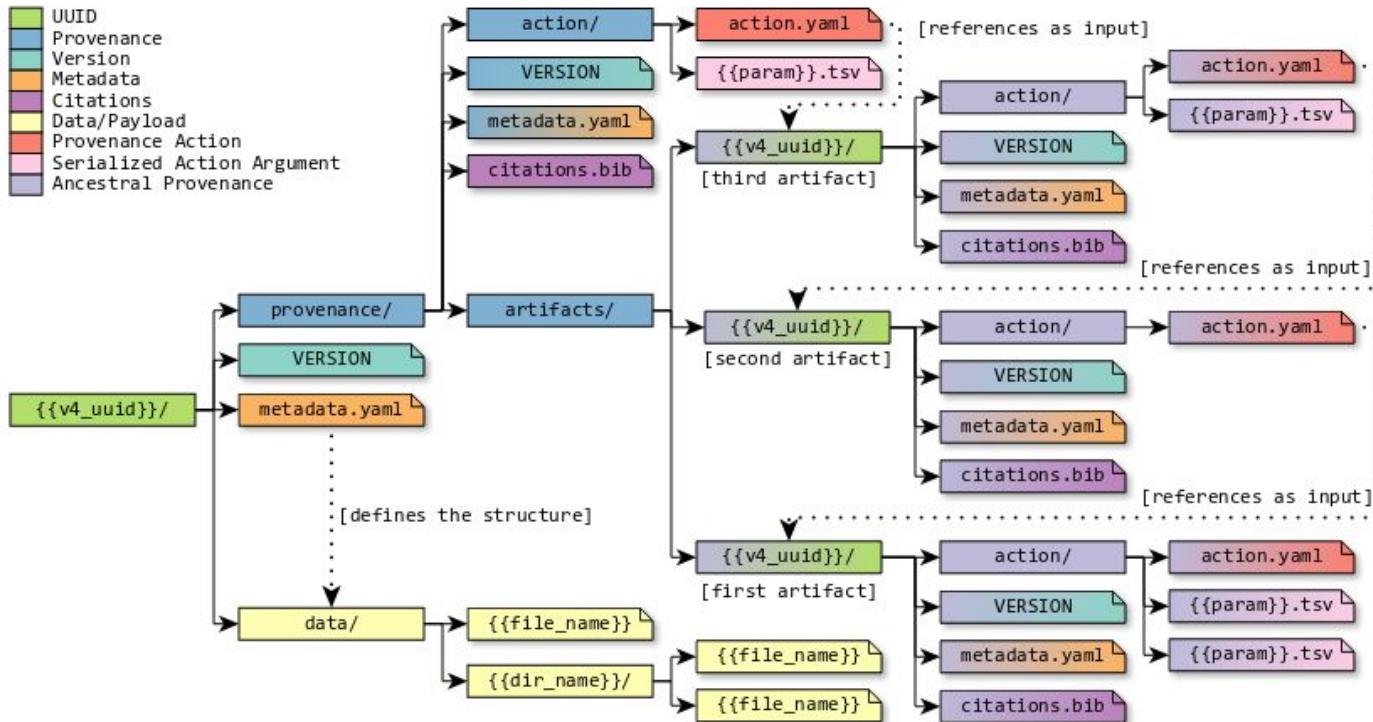
.qza and .qzv files are just zip files

- We aren't inventing new formats
 - A QIIME 2 result can be "used" on almost any computer
- ZIP64 (OS X by default only knows ZIP32, but other OSes work well)
 - Use `qiime tools extract` if unzipping isn't working.

Learn More:

<https://dev.qiime2.org/latest/storing-data/>

Archive Structure



QIIME 2 Results Are Deeply Annotated

- Provenance is captured by the framework
 - You can't escape it! No matter what interface you use, the SDK will always track this information
 - Even visualizations and statistics retain full provenance
- Citations are tracked by *usage*
 - If the framework used the code, it will attempt to cite it to the best of its ability (requires plugin developers to provide annotation)
- Transformations and other runtime information captured as well
 - No random seeds yet (this is hard to do in a robust way)
- The data directory is described by the format
 - If you know the format, there's no guesswork
- Intention is described by the type

q2view Helps Keep the Data in the ZIP File

- Visualizations should be simple to use
- ZIP files aren't a convenient way to send a PDF
- Extracting ZIP files separates your data from your provenance
- q2view bridges the gap so that provenance sticks around with the data

Sample Normalization

Does anything concern you about this table?

FeatureTable[Frequency]					
	feature1	feature2	feature3	feature4	feature5
4ac2	84	1	73	198	2
e375	24	2	44	176	1
4gd8	11	0	10	30	0
9872	0	0	25	2	0

Diversity metrics in ordinations are often impacted by the total frequency observed in samples, such that in this example 4gd8 (right palm) might look more similar to 9872 (tongue) than to e375 (left palm).

FeatureTable[Frequency]					
	feature1	feature2	feature3	feature4	feature5
4ac2	84	1	73	198	2
e375	24	2	44	176	1
4gd8	11	0	10	30	0
9872	0	0	25	2	0

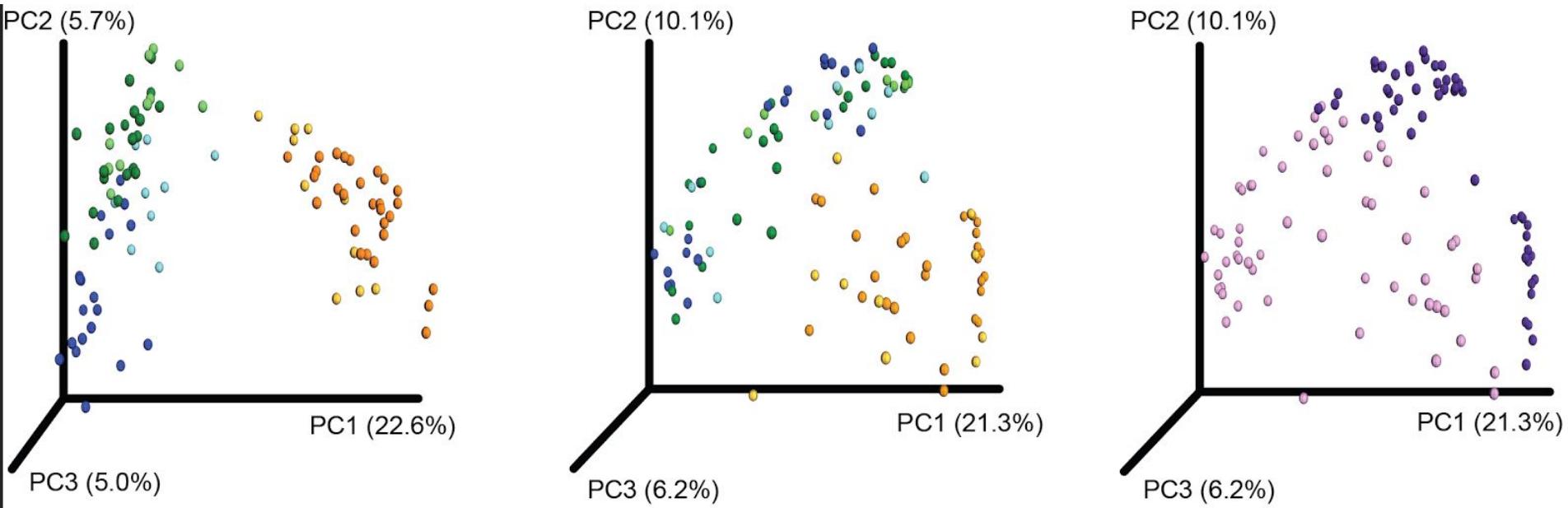
	Total frequency
4ac2	358
e375	247
4gd8	51
9872	27

This is most commonly handled by *rarefaction*, which is currently* a necessary evil. Frequencies are subsampled without replacement until all samples have the same total. Samples with fewer sequences than your *even sampling depth* will be filtered out of the feature table.

FeatureTable[Frequency]					
	feature1	feature2	feature3	feature4	feature5
g345	11	1	10	29	0
c5d7	4	0	7	40	0
f6ee	11	0	10	30	0
efd3	0	0	0	0	0

	Total frequency
g345	51
c5d7	51
f6ee	51
efd3	0

* A good project would be developing diversity metrics that are not sensitive to total frequency.



■ M2_finger

■ M2_key

■ M3_finger

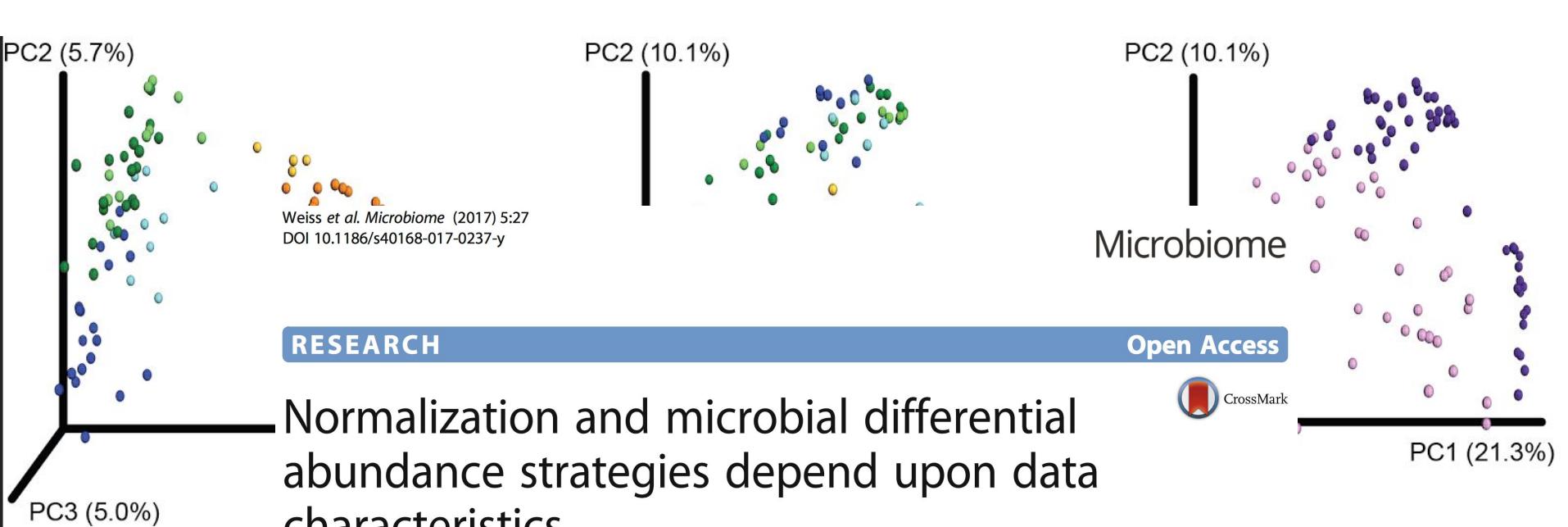
■ M3_key

■ M9_finger

■ M9_key

■ 50 sequences per sample

■ 500 sequences per sample



Normalization and microbial differential abundance strategies depend upon data characteristics

Sophie Weiss¹, Zhenjiang Zech Xu², Shyamal Peddada³, Amnon Amir², Kyle Bittinger⁴, Antonio Gonzalez², Catherine Lozupone⁵, Jesse R. Zaneveld⁶, Yoshiki Vázquez-Baeza⁷, Amanda Birmingham⁸, Embriette R. Hyde² and Rob Knight^{2,7,9*}

M2_finger

M2_key

M3_finger

M9_finger

M9_key

50 sequences per sample

500 sequences per sample

Moving Pictures tutorial: Sample Normalization and Diversity Analyses

View your feature-table summary visualization to explore different sampling depths for rarefying.

α -diversity Metrics, Statistics, and Visualizations

Comparing microbial communities

How many “species” are there? *Alpha diversity (richness, evenness, or both).*

How similar are pairs of samples? *Beta diversity*

Who is there? *Taxonomic profiling, differential abundance testing.*

Alpha diversity metrics operate on a single sample (i.e., within sample diversity).

Beta diversity metrics operate on a pair of samples (i.e., between sample diversity).

Qualitative diversity metrics only consider the presence/absence of features.

Quantitative diversity metrics consider abundance of features.

Phylogenetic diversity metrics incorporate evolutionary relationships between taxa, but assume that we know what those relationships are.

Non-phylogenetic diversity metrics assume that all taxa are equally related, so don't make assumptions about evolutionary relationships.

Observed OTUs (or Observed Species): a qualitative, non-phylogenetic, alpha diversity metric

FeatureTable [Frequency]		feature1	feature2	feature3	feature4	feature5
4ac2		25	30	15	0	0
e375		0	17	33	25	0



SampleData [AlphaDiversity]	
	Observed OTUs
4ac2	
e375	

Count the number of different features in a sample.

Observed OTUs (or Observed Species): a qualitative, non-phylogenetic, alpha diversity metric

FeatureTable [Frequency]		feature1	feature2	feature3	feature4	feature5
4ac2		25	30	15	0	0
e375		0	17	33	25	0



SampleData [AlphaDiversity]	
	Observed OTUs
4ac2	3
e375	3

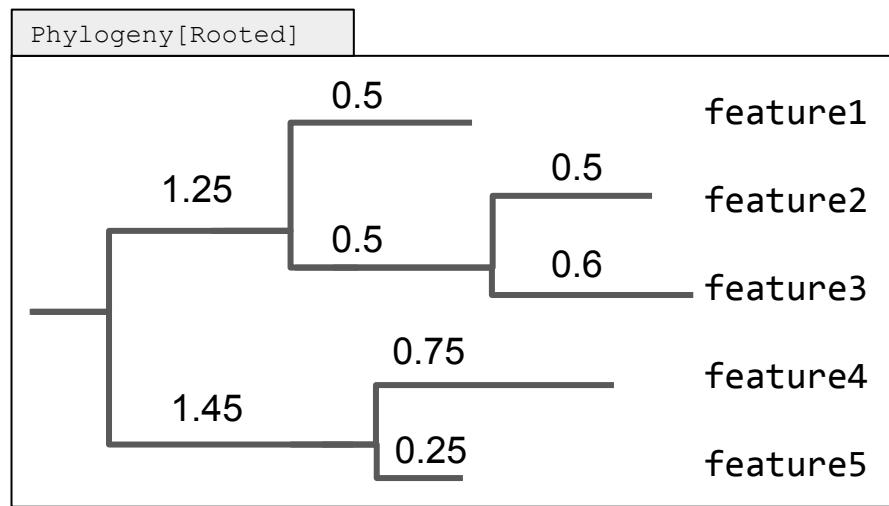
Count the number of different features in a sample.

Why incorporate phylogeny in a diversity metric?

FeatureTable [Frequency]					
	feature1	feature2	feature3	feature4	feature5
4ac2	25	30	15	0	0
e375	0	17	33	25	0

→

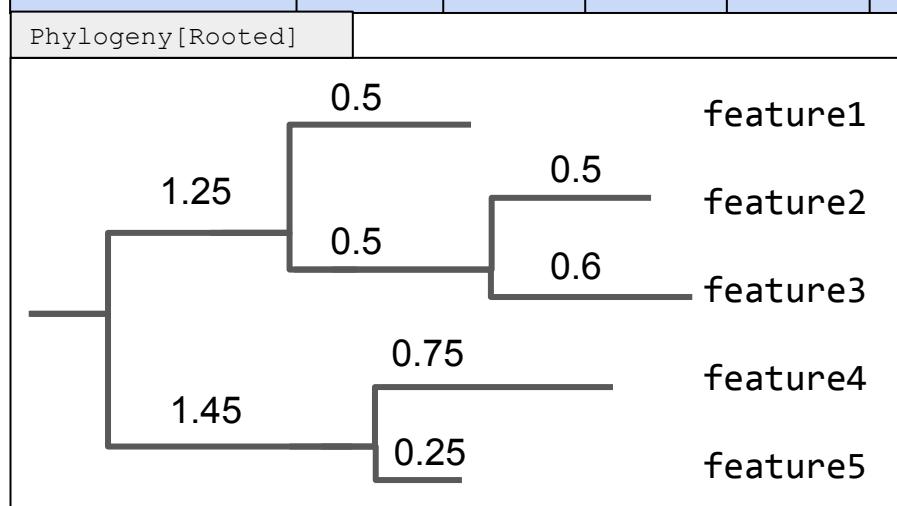
SampleData [AlphaDiversity]	
	Observed OTUs
4ac2	3
e375	3



FeatureData [Taxonomy]	
	Domain
feature1	Bacteria
feature2	Bacteria
feature3	Bacteria
feature4	Archaea
feature5	Archaea

Faith's Phylogenetic Diversity (PD): a qualitative, phylogenetic, alpha diversity metric

FeatureTable [Frequency]					
	feature1	feature2	feature3	feature4	feature5
4ac2	25	30	15	0	0
e375	0	17	33	25	0



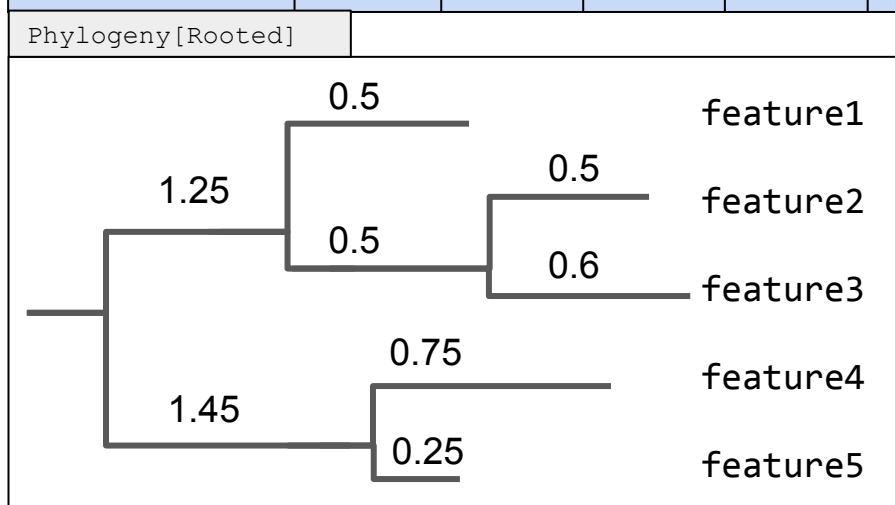
SampleData [AlphaDiversity]	
	Faith's PD
4ac2	
e375	

Sum of branch length covered by a sample.

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

Faith's Phylogenetic Diversity (PD): a qualitative, phylogenetic, alpha diversity metric

FeatureTable [Frequency]					
	feature1	feature2	feature3	feature4	feature5
4ac2	25	30	15	0	0
e375	0	17	33	25	0



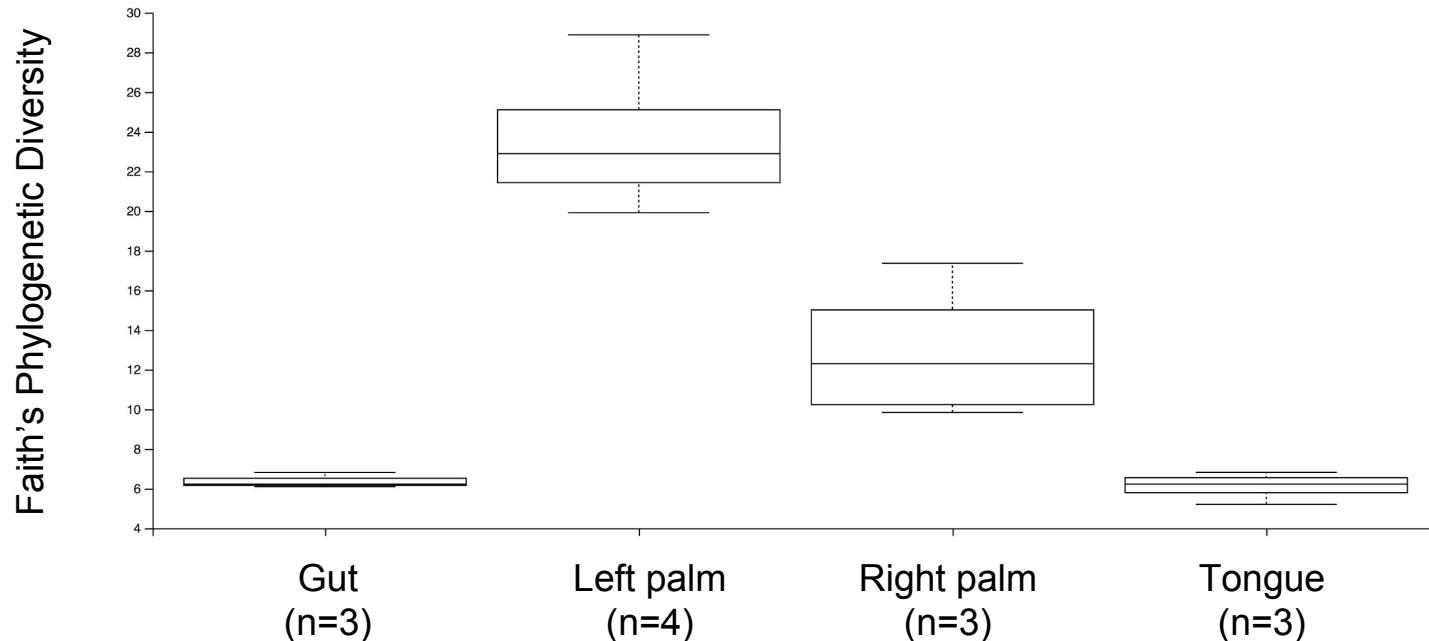
SampleData [AlphaDiversity]	
	Faith's PD
4ac2	3.35
e375	5.05

Sum of branch length covered by a sample.

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

Alpha diversity comparison

- visually with distribution comparison plots (discrete data) or scatter plots (continuous data)
- statistically with Kruskal-Wallis (discrete data) or Spearman correlation (continuous data)



Moving Pictures tutorial: alpha diversity statistics and visualizations

[https://docs.qiime2.org/2018.6/tutorials/moving-pictures/
#alpha-and-beta-diversity-analysis](https://docs.qiime2.org/2018.6/tutorials/moving-pictures/#alpha-and-beta-diversity-analysis)



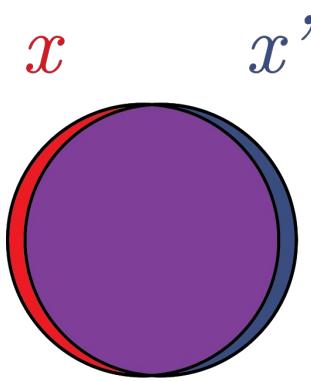
Slides: goo.gl/QXVa4h
Slides (pdf): goo.gl/B1Qnk7
Schedule: goo.gl/XDrhf3

Coffee Break!

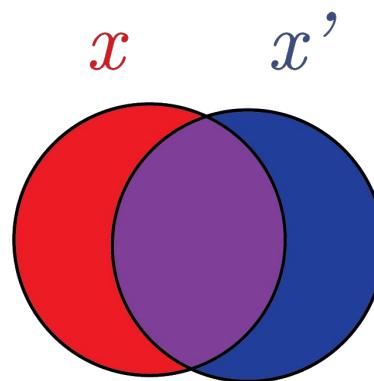
β -diversity Metrics, Statistics, and Visualizations

Jaccard distance (1908)

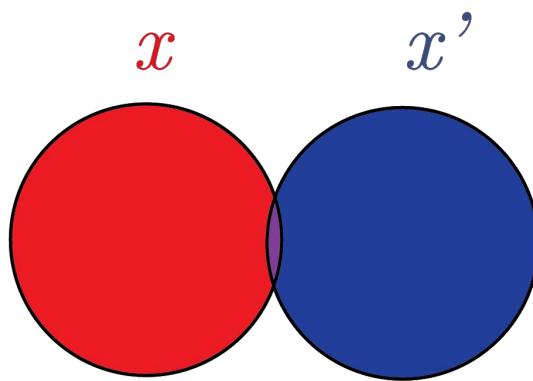
Fraction of unique features, regardless of abundance.



$$d_j \approx 0$$



$$d_j \approx 0.5$$



$$d_j \approx 1$$

Jaccard distance:

a qualitative, non-phylogenetic beta diversity metric

FeatureTable [Frequency]

	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$



DistanceMatrix

	4ac2	e375	4gd8	9872
4ac2				
e375				
4gd8				
9872				

Jaccard distance:

a qualitative, non-phylogenetic beta diversity metric

FeatureTable [Frequency]

	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$



	4ac2	e375	4gd8	9872
4ac2	0.0			
e375		0.0		
4gd8			0.0	
9872				0.0

Jaccard distance:

a qualitative, non-phylogenetic beta diversity metric

FeatureTable [Frequency]

	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$



DistanceMatrix

	4ac2	e375	4gd8	9872
4ac2	0.0			
e375	0.4	0.0		
4gd8	0.4	0.0	0.0	
9872	0.5	0.5	0.5	0.0

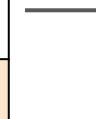
Jaccard distance:

a qualitative, non-phylogenetic beta diversity metric

FeatureTable [Frequency]

	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

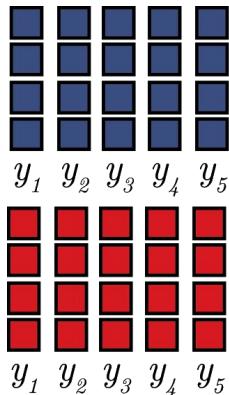
$$J(A, B) = 1 - \frac{A \cap B}{A \cup B}$$



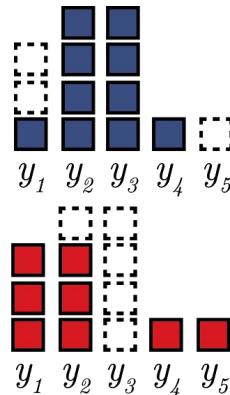
	4ac2	e375	4gd8	9872
4ac2	0.0	0.4	0.4	0.5
e375	0.4	0.0	0.0	0.5
4gd8	0.4	0.0	0.0	0.5
9872	0.5	0.5	0.5	0.0

Bray-Curtis distance (1948)

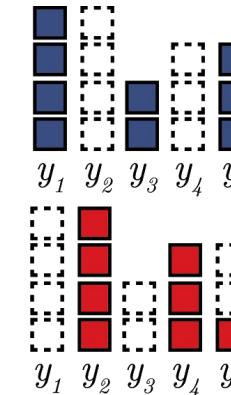
Fraction of overabundant counts.



$$d_{bc} = 0$$



$$d_{bc} \approx 0.5$$



$$d_{bc} \approx 1$$

$$\frac{0 \times \boxed{}}{20 \times \boxed{\textcolor{red}{\square}} + 20 \times \boxed{\textcolor{blue}{\square}}}$$

$$\frac{8 \times \boxed{}}{8 \times \boxed{\textcolor{red}{\square}} + 10 \times \boxed{\textcolor{blue}{\square}}}$$

$$\frac{16 \times \boxed{}}{8 \times \boxed{\textcolor{red}{\square}} + 9 \times \boxed{\textcolor{blue}{\square}}}$$

Sorenson, Th. "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content." Kongelige Danske Videnskabernes Selskab 5.1-34 (1948): 4-7.

Bray-Curtis distance:

a quantitative, non-phylogenetic beta diversity metric

FeatureTable[Frequency]

	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

$$BC(A, B) = \frac{\sum_i |X_{iA} - X_{iB}|}{\sum_i (X_{iA} + X_{iB})}$$

X_{iA} : frequency of feature i in sample A

	4ac2	e375	4gd8	9872
4ac2	0.0			
e375		0.0		
4gd8			0.0	
9872				0.0

Bray-Curtis distance:

a quantitative, non-phylogenetic beta diversity metric

FeatureTable[Frequency]

	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

$$BC(A, B) = \frac{\sum_i |X_{iA} - X_{iB}|}{\sum_i (X_{iA} + X_{iB})}$$

X_{iA} : frequency of feature i in sample A

	4ac2	e375	4gd8	9872
4ac2	0.0			
e375	0.19	0.0		
4gd8	0.15	0.07	0.0	
9872	0.65	0.69	0.70	0.0

Bray-Curtis distance:

a quantitative, non-phylogenetic beta diversity metric

FeatureTable[Frequency]

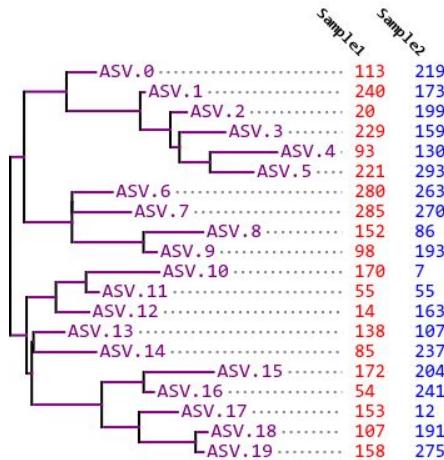
	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

$$BC(A, B) = \frac{\sum_i |X_{iA} - X_{iB}|}{\sum_i (X_{iA} + X_{iB})}$$

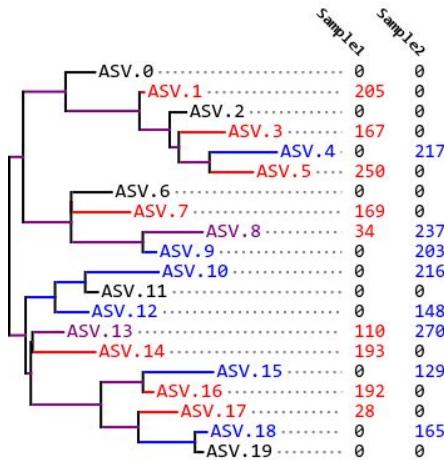
X_{iA} : frequency of feature i in sample A

	4ac2	e375	4gd8	9872
4ac2	0.0	0.19	0.15	0.65
e375	0.19	0.0	0.07	0.69
4gd8	0.15	0.07	0.0	0.70
9872	0.65	0.69	0.70	0.0

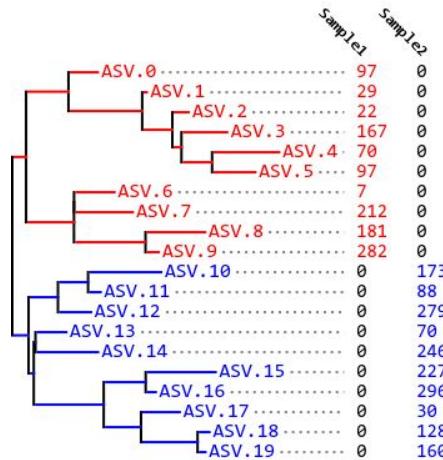
Unweighted UniFrac distance: a qualitative, phylogenetic beta diversity metric



Unweighted UniFrac = 0.000000
Weighted UniFrac = 0.118186



Unweighted UniFrac = 0.526659
Weighted UniFrac = 0.560364



Unweighted UniFrac = 1.000000
Weighted UniFrac = 1.000000

$$UU(A, B) = \frac{\text{sum of unique branch length}}{\text{sum of observed branch length}}$$

Unweighted UniFrac distance: a qualitative, phylogenetic beta diversity metric

FeatureTable [Frequency]

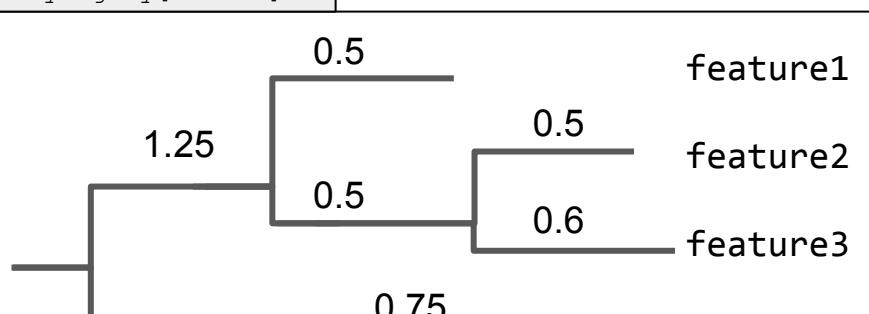
	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

$$UU(A, B) = \frac{\text{sum of unique branch length}}{\text{sum of observed branch length}}$$

DistanceMatrix

	4ac2	e375	4gd8	9872
4ac2	0.0			
e375		0.0		
4gd8			0.0	
9872				0.0

Phylogeny [Rooted]

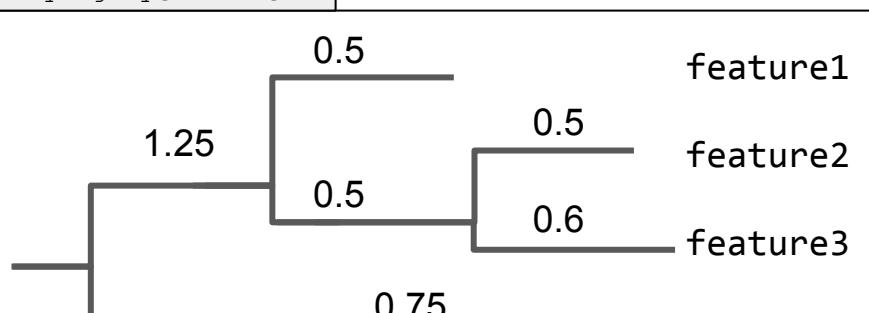


Unweighted UniFrac distance: a qualitative, phylogenetic beta diversity metric

FeatureTable [Frequency]

	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

Phylogeny [Rooted]



$$UU(A, B) = \frac{\text{sum of unique branch length}}{\text{sum of observed branch length}}$$

DistanceMatrix

	4ac2	e375	4gd8	9872
4ac2	0.0			
e375	0.13	0.0		
4gd8	0.13	0.0	0.0	
9872	0.14	0.18	0.18	0.0

Unweighted UniFrac distance: a qualitative, phylogenetic beta diversity metric

FeatureTable [Frequency]

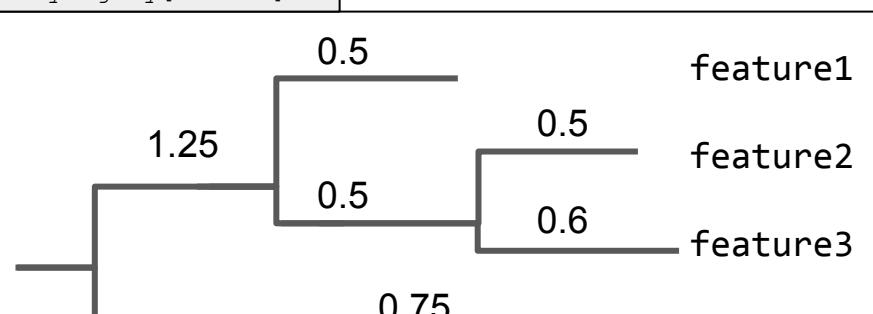
	feature1	feature2	feature3	feature4	feature5
4ac2	42	0	37	99	1
e375	12	1	22	88	0
4gd8	25	3	23	86	0
9872	0	0	87	12	0

$$UU(A, B) = \frac{\text{sum of unique branch length}}{\text{sum of observed branch length}}$$

DistanceMatrix

	4ac2	e375	4gd8	9872
4ac2	0.0	0.13	0.13	0.14
e375	0.13	0.0	0.0	0.18
4gd8	0.13	0.0	0.0	0.18
9872	0.14	0.18	0.18	0.0

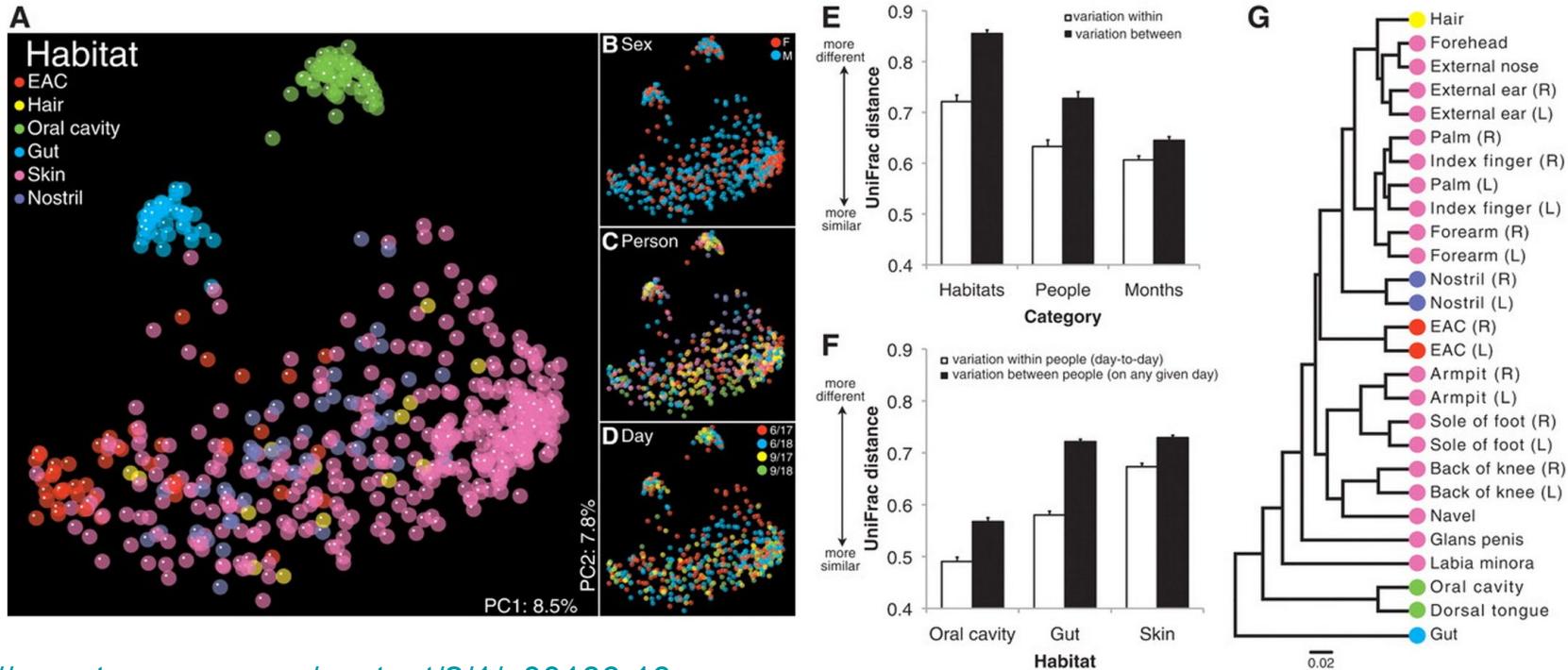
Phylogeny [Rooted]



The first few lines of a relatively small distance matrix...

1	> Stillton4R2	Stillton4R3	Stillton4R1	H Canyon3R3	H Canyon3R2	H Canyon3R1	Halls9R2	H Canyon2R2	H Canyon2R3	H Canyon2R1	Halls9R1	Stillton10R3	H Canyon0R1	H Canyon0R2	H Canyon0R3	H Canyon7R2	H Canyon7R1	H Can
2	Stillton4R2	0.0	0.382273294624	0.391416675288	0.56036938484088	0.553938232028	0.566136031815	0.55171934987546	0.53655901824	0.56704190567	0.57481395502	0.474244845718	0.776189708931	0.797241309582	0.71			
3	Stillton4R3	0.382273294624	0.0	0.394399899497	0.58630332083	0.58630332083	0.58630332083	0.58630332083	0.58630332083	0.5860224873354	0.577776776062	0.601572353059	0.609758664376	0.779315452075	0.795032626253	0.75		
4	Stillton4R1	0.391416675288	0.394399899497	0.0	0.586028929798	0.583767102601	0.584463513906	0.573822816898	0.550892933103	0.566056393448	0.584611276938	0.607144746402	0.469209424282	0.7810714010354	0.797809309119	0.75		
5	H Canyon3R3	0.56036938484088	0.58630332083	0.58630332083	0.586028929798	0.583767102601	0.354600316745	0.345731612479	0.458450145834	0.412488628393	0.385127601086	0.48625352766	0.58540237425	0.69071344042	0.71380444578	0.65		
6	H Canyon3R2	0.55393823208	0.58630332083	0.58630332083	0.586028929798	0.583767102601	0.354600316745	0.362656949665	0.462707750398	0.414528760004	0.385453766442	0.380087766632	0.46109426257	0.57448821279	0.692644612288	0.714717508733	0.65	
7	H Canyon3R1	0.566136031815	0.58630332083	0.58630332083	0.586028929798	0.583767102601	0.354731612479	0.362565949665	0.0	0.452351806146	0.415483559719	0.38976424875	0.381981707704	0.485596583752	0.580926122772	0.683086513009	0.697490417772	0.65
8	Halls9R2	0.557134987546	0.586022627323	0.573822816898	0.458450145834	0.462707750398	0.452531806146	0.0	0.447883445295	0.429943464459	0.409604513124	0.344264504725	0.561110815583	0.727520276229	0.75335887058	0.75		
9	H Canyon2R2	0.531719852875	0.560923487354	0.550892933103	0.412488628393	0.414528760004	0.415483559719	0.447883445295	0.0	0.404179520995	0.388727659604	0.468345488157	0.549328940024	0.709578414217	0.723508997775	0.73		
10	H Canyon2R3	0.53655901824	0.5776776062	0.566056393448	0.385127601086	0.385453766442	0.398796424875	0.429943464459	0.404179520995	0.0	0.36144452893	0.439367245599	0.566330894525	0.711736423222	0.72873315629	0.76		
11	H Canyon2R1	0.567041909667	0.601572330569	0.584611276938	0.384525100123	0.380087766632	0.381871707704	0.40964513124	0.388727659604	0.361444528983	0.570912554101	0.722577576622	0.722577576622	0.722577576622	0.722577576622	0.722577576622	0.722577576622	
12	Halls9R1	0.574831935502	0.607958664376	0.607144746402	0.46826352676	0.46109426257	0.485596583752	0.344264504725	0.463845488157	0.439367245599	0.5045633208595	0.0	0.578254150137	0.745189206089	0.761081073905	0.74		
13	Stillton10R3	0.474244845718	0.479910461959	0.469294024282	0.585402372425	0.574848822129	0.580926122772	0.561110815583	0.549328940024	0.566330894525	0.582635621853	0.578254150137	0.0	0.77519377947	0.79775580641	0.79775580641	0.79775580641	0.79775580641
14	H Canyon0R1	0.776189708931	0.779315452075	0.781074100354	0.690713440402	0.69264612288	0.683086513009	0.727520276229	0.709578414217	0.711736423222	0.707912554104	0.745189206089	0.779519377947	0.0	0.387183864476	0.31		
15	H Canyon0R2	0.797241309582	0.795032626232	0.797809391919	0.713804445478	0.7147175087833	0.697490417772	0.75353887058	0.723508997775	0.728837315629	0.761081079305	0.79775580641	0.387183864476	0.0	0.37			
16	H Canyon0R3	0.784124036272	0.791711071574	0.795527049823	0.693476418734	0.688360016409	0.71324895014053	0.713225020799	0.708340569579	0.708462144382	0.747778043238	0.78904482767	0.38461076947	0.3715375100	0.73			
17	H Canyon0R3	0.539192489149	0.563745794923	0.557216919336	0.457709184402	0.467073058252	0.473888383037	0.42997132225	0.455576263999	0.448029116277	0.441566044862	0.444788573194	0.561681574886	0.740008927446	0.7662626	0.75		
18	H Canyon0R7	0.665877541792	0.68383349738	0.674471865383	0.53083292175	0.563316168524	0.562611936812	0.546938321206	0.558741865355	0.562533006516	0.534451654634	0.657732384662	0.723340784923	0.743399175	0.73			
19	H Canyon0R1	0.554857668962	0.578652503409	0.570402839551	0.475303182122	0.474309745536	0.462587318718	0.43076590681	0.463222699451	0.452021582021	0.42664191335	0.44552755863	0.564686335148	0.727341798211	0.7505363	0.75		
20	H Canyon1R1	0.746065696617	0.75464579561	0.759778994683	0.654674045733	0.65112173227	0.649030039345	0.703134513793	0.67682938985	0.6759575137688	0.676955919301	0.714539249748	0.752511987622	0.451032388712	0.48191137688	0.71		
21	H Canyon1R3	0.7515225854919	0.764074726397	0.767226018039	0.641176135477	0.65043461858	0.62924888988	0.702785457767	0.675191090098	0.668258446703	0.663436849294	0.713256418398	0.763047504609	0.440489328365	0.45701904579	0.75		
22	H Canyon1R2	0.748585453295	0.753698061339	0.759152313976	0.66632211531	0.666043304584	0.654057859985	0.711448710816	0.680563093126	0.6799460426606	0.678341781539	0.7283640404942	0.752981637948	0.445883860778	0.4733377	0.75		
23	H Canyon1R2	0.56835786215	0.584149104488	0.581483432897	0.46605578676	0.464168854592	0.470680372129	0.458666733287	0.48141903126	0.46098682372	0.4425453616102	0.4619687362	0.573815947327	0.71020961042	0.734399175	0.75		
24	H Canyon10R3	0.570798402641	0.588433072775	0.591942467159	0.503205901977	0.49822069939	0.490354052309	0.4650352212318	0.4969625857514	0.490816424808	0.497919293873	0.463273144447	0.585101517174	0.732074178077	0.76453933	0.75		
25	H Canyon10R1	0.505926071682	0.528347698573	0.529350507225	0.472413250399	0.479553556468	0.478832883976	0.455918418773	0.452087517992	0.4527334419477	0.456718219721	0.4577312834343	0.528604962934	0.733759437162	0.75381372	0.75		
26	H Canyon11R3	0.582831143875	0.594413158276	0.60668171515	0.463621293703	0.458422804642	0.486656215192	0.430493752745	0.47439493784	0.445815433526	0.432018827272	0.43706115175	0.591646247274	0.737206924727	0.7571514	0.75		
27	H Canyon11R2	0.528552555543	0.55774407831	0.555177188694	0.432763917311	0.429877059678	0.440567014981	0.41332040453	0.426696806769	0.398447088329	0.396059530935	0.433778641456	0.546155178322	0.730957367352	0.7533215	0.75		
28	H Canyon11R1	0.555250987098	0.579991048797	0.575126874355	0.4534362549549	0.44588780311	0.456712334902	0.419232496981	0.4468697971255	0.430235265526	0.428966404849	0.431167398777	0.579943394066	0.730267428808	0.76081576981	0.75		
29	H Canyon6R2	0.568527068157	0.585152346508	0.583867897093	0.430174126838	0.419365916562	0.432708499988	0.448168213391	0.445815626173	0.426928785154	0.410839210322	0.42252640701	0.5747408501553	0.724938493714	0.7494427	0.75		
30	H Canyon6R3	0.566346578975	0.587780110737	0.586940661236	0.438634245182	0.425883025257	0.449519180577	0.424863237935	0.447100684715	0.425434475646	0.426218627557	0.4246233207252	0.570052188601	0.7227254724469	0.7484881	0.75		
31	H Canyon6R1	0.583972679946	0.586787203744	0.5815345315565	0.434912987895	0.4243037243199	0.441765941119	0.415236907762	0.428454635151	0.426954235119	0.4226329389493	0.469549521533	0.5732303234932	0.71559691738	0.7418845	0.75		
32	H Canyon5R1	0.558914057622	0.574709726761	0.562821643814	0.440681773256	0.415966156213	0.435765316753	0.4173561610791	0.424754580111	0.435765316753	0.417737738339	0.455877139946	0.563966325588	0.708246555272	0.7363125	0.73		
33	H Canyon5R3	0.559874926968	0.5820196982	0.582450873735	0.4503363948198	0.4265742405959	0.448524328318	0.4397459524594	0.440643510478	0.417613041188	0.4356153495565	0.464820696561	0.585182947948	0.710049190032	0.7373235	0.73		
34	H Canyon5R2	0.577951660858	0.5935368761716	0.59633895726545	0.422545756978	0.409515246412	0.434403209962	0.429247944328	0.453873029965	0.431590733736	0.43166014112	0.4361756140548	0.583732320655	0.717174395032	0.7323225	0.73		
35	H Canyon4R1	0.516189922282	0.555158871798	0.550442708263	0.4164491545884	0.420465768689	0.437380878857	0.46015301609	0.459537309642	0.4251922653132	0.439771880651	0.4747121489459	0.562799867876	0.715927187809	0.7375584	0.75		
36	H Canyon4R2	0.601827582628	0.632734373598	0.628005145571	0.414666304994	0.417716489314	0.443390182838	0.49718855337	0.4737664520531	0.447416361784	0.454051962981	0.502442273403	0.570052188601	0.67365762825	0.6942020	0.75		
37	H Canyon4R3	0.624113459545	0.653406174875	0.649892943176	0.4378488993112	0.457007607767	0.45666604112	0.4593726548949	0.47953237291395	0.460846248971	0.4623273913958	0.51088415471	0.6406842648971	0.694942617997	0.768424727297	0.77		
38	H Canyon12R1	0.513376919549	0.547194076841	0.519066521533	0.471737648119	0.464181526784	0.430362929794	0.439779783407	0.449105789781	0.41992093622	0.466361108125	0.520926634082	0.723250659142	0.7418837	0.75			
39	H Canyon12R2	0.552892291856	0.560516044793	0.549527557613	0.50472104035	0.517681811493	0.453017456949	0.488498787942	0.4895953593593	0.465436661211	0.495253187949	0.713659241861	0.75952187592	0.745795227592	0.754556	0.75		
40	H Canyon12R3	0.553185666616	0.585465606164	0.5856656024134	0.439712101711	0.431336995109	0.463612059381	0.409326408365	0.435088785406	0.414920465304	0.40784017122	0.422739541765	0.580361846776	0.7269663836109	0.7516425	0.75		
41	H Canyon8R1	0.552013416498	0.559504417239	0.526386483618	0.483602818412	0.470838441272	0.458053827811	0.446645294714	0.465192835689	0.464892445620	0.447733174476							

Various techniques are applied to interpret distance matrices



<http://msystems.asm.org/content/2/1/e00166-16>

Bacterial Community Variation in Human Body Habitats Across Space and Time.

Costello et al. Science (2009)

Beta diversity comparison

- visually with ordination plots (e.g., NMDS or PCoA; discrete or continuous data) or distribution box plots (discrete data)
- statistically with Kruskal-Wallis (discrete data), Mantel correlation (continuous data) or BEST (continuous data)

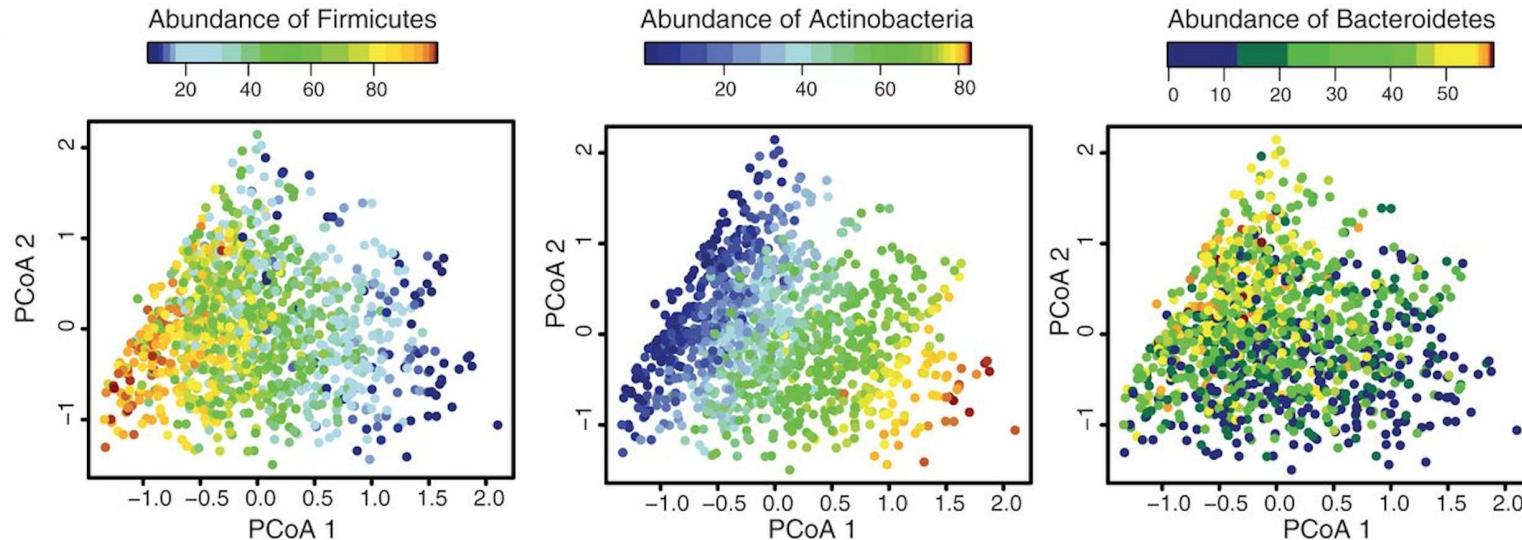


Image source: [Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity](#), Zhernakova et al., Science (2016).

Moving Pictures tutorial: beta diversity

[https://docs.qiime2.org/2018.6/tutorials/moving-pictures/
#alpha-and-beta-diversity-analysis](https://docs.qiime2.org/2018.6/tutorials/moving-pictures/#alpha-and-beta-diversity-analysis)

Taxonomic Assignment and Visualization

Taxonomic assignment of observed sequences.

FeatureData [Sequence]

```
>feature5  
GACGAAGGTGACGCCGTTGCTCGGAATCACTGGGCATAAAGCGCCGTAGGTGGCTTGGTAAGTCATGGTCAA  
ATCCCTCGGCTCAACCGAGGAACCTG  
>feature4  
TACGTAGGGGCAAGCGTTATCCGATTACTGGGTGAAAGGGAGCGTAGACGGATGGACAAGTCTGATGTGAA  
AGGCTGGGCTCAACCCGGGACGG  
>feature2  
TACGTATGGGCAAGCGTTATCCGAAATTGGGCGTAAAGAGTCGTAGGTGGCTTAAGCGCAGGGTTA  
AGGAATGGCTTAACCTATTGTTCTC  
>feature1  
GACGGAGGATCCAAGTGTATCCGGAATCACTGGGCCTAAAGCGCTGTAGGTGGTTACTAACGTAACTGTTAA  
ATCTTGAGGCTCAACCTCGAAATCG  
>feature3  
TACGGAGGGTGCAGCGTTATCGGAAATTACTGGGCGTAAAGCGTACGTAGGCCTTAGGTAAGTCAGATGTGAA  
AGCCCCGGGCTCACCTGGGATGG
```

FeatureData [Sequence]

```
>reference-sequence-1  
TTGAAGGTGGACGCCGTTGCTCGGAATCACTGGGCATAAAGCGCCGTAGGTGGCTTGGTAAGTCACATGGT  
GACTCAACCGAGGAACCTGAAAGTGGACGCCGTTGCTCGGAATCACTGGGCATAAAGCGCCGTAGGTG  
GCTTGGTAAGTCACATGGTACTCAACCGAGGAACCTGAA  
>reference-sequence-2  
AACGTAGGCAAGCGTTATCCGATTACTGGGTGAAAGGGAGCGTAGACGGATGGACAAGTCTGATGTGAAAGG  
CTGGGCTCAACCCGGGACCCCTGAAACCTGGGACCCCTTGCTCGGAATCACTGGGCATAAAGCGCCGTAGGT  
TGTG  
>reference-sequence-3  
Bacteria; Proteobacteria; Gammaproteobact  
>reference-sequence-4  
Bacteria; Bacteroidetes; Flavobacteria; Fl  
Bacteria; Proteobacteria; Deltaproteobact  
Archaea; Euryarchaeota; DSEG; 104A5
```

FeatureData [Taxonomy]

reference-sequence-1

Bacteria; Proteobacteria; Gammaproteobact

reference-sequence-2

Bacteria; Bacteroidetes; Flavobacteria; Fl

reference-sequence-3

Bacteria; Proteobacteria; Deltaproteobact

reference-sequence-4

Archaea; Euryarchaeota; DSEG; 104A5

Compare observed sequences to annotated reference sequences to make taxonomic assignments.

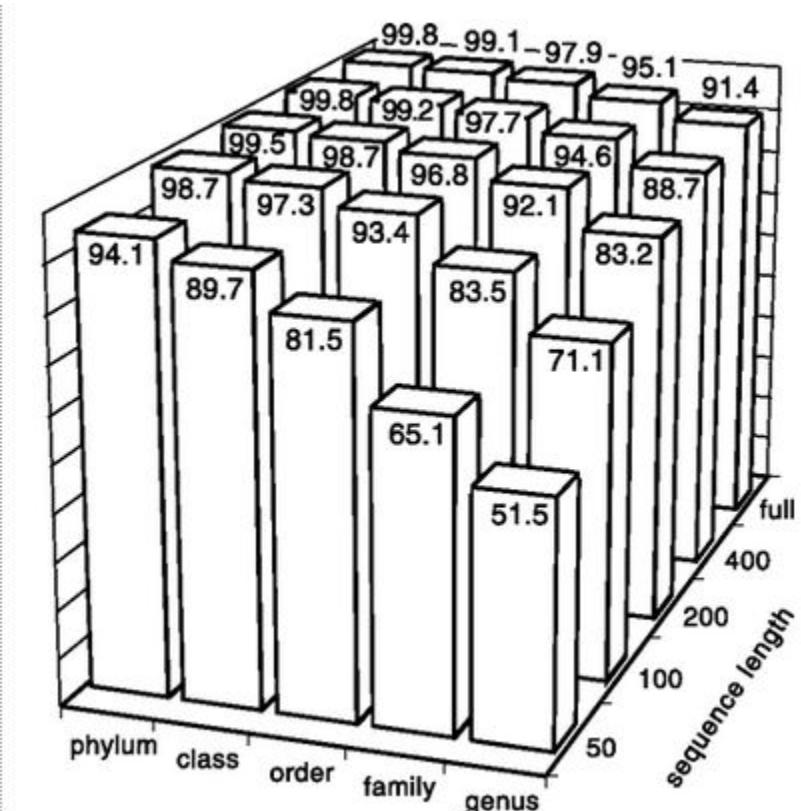
FeatureData [Taxonomy]

feature5	Bacteria; Proteobacteria
feature4	Bacteria; Proteobacteria
feature2	Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales
feature1	Bacteria; Proteobacteria
feature3	Bacteria; Proteobacteria; Deltaproteobacteria

Taxonomic assignment methods

- Compare directly to a reference database
 - UCLUST, VSEARCH, SortMERNA, BLAST.
- Predict using a machine learning classifier
 - RDP, q2-feature-classifier.
- Insert the sequences in a reference tree.
 - SEPP.

RDP Classifier



8 nucleotide words

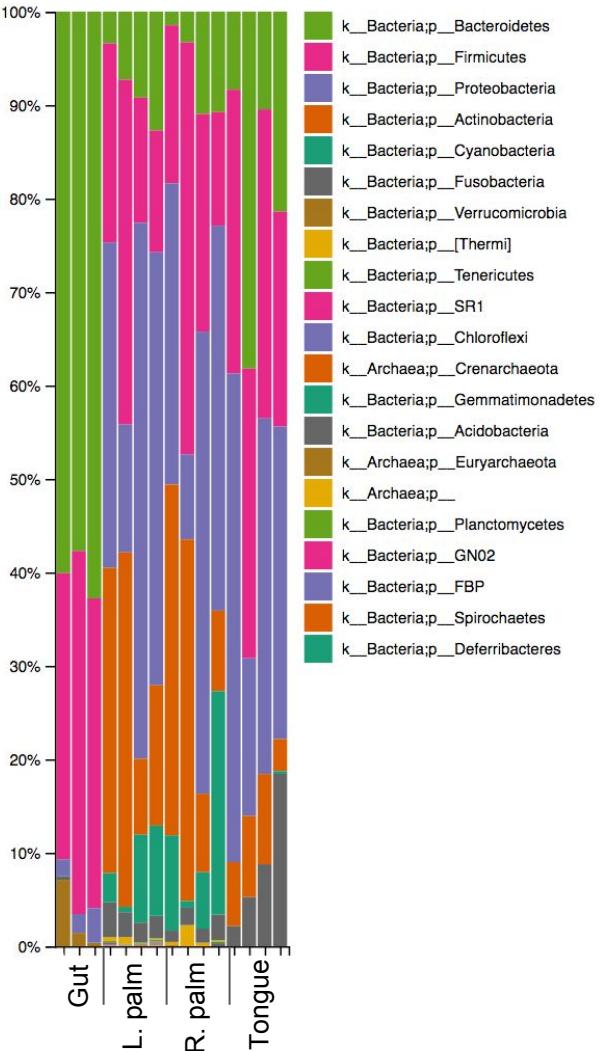
Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy^{o†}
Qiong Wang¹, George M. Garrity^{1,2}, James M. Tiedje^{1,2} and James R. Cole^{1,*}

Visualizing taxonomic profiles

Interactive barplots support:

- Taxonomic level selection
- Multi-level sorting
- Filtering
- Coloring
- Exporting plots (SVG) and raw data

Relative frequency



Moving Pictures tutorial: taxonomic assignment and visualization

[https://docs.qiime2.org/2018.6/tutorials/moving-pictures/
#taxonomic-analysis](https://docs.qiime2.org/2018.6/tutorials/moving-pictures/#taxonomic-analysis)

Slides: goo.gl/QXVa4h
Slides (pdf): goo.gl/B1Qnk7
Schedule: goo.gl/XDrhf3

Additional Resources

- <https://docs.qiime2.org/>
 - User Documentation
- <https://dev.qiime2.org/>
 - Developer Documentation
- <https://forum.qiime2.org/>
 - Come talk with us or ask questions!

mockrobiota: a Public Resource for Microbiome Bioinformatics Benchmarking

Nicholas A. Bokulich, Jai Ram Rideout, William G. Mercurio, Arron Shiffer, Benjamin Wolfe, Corinne F. Maurice, Rachel J. Dutton, Peter J. Turnbaugh, Rob Knight, J. Gregory Caporaso
Josh D. Neufeld, *Editor*

DOI: 10.1128/mSystems.00062-16

- Public repository of mock community datasets
 - Submit a Pull Request to add yours!
- Components:
 - Raw data
 - Sample metadata
 - Expected composition
- Current inventory:
 - 20 16S rRNA studies
 - 5 fungal ITS studies
 - 1 mock metagenome
 - Illumina HiSeq and MiSeq data
 - Even and staggered composition
 - Communities consist of 11-67 strains

An Introduction to Applied Bioinformatics

An Introduction to Applied Bioinformatics (or IAB) is a free, open source interactive text that introduces readers to core concepts of bioinformatics in the context of their implementation and application.



IAB is primarily being developed by Greg Caporaso (GitHub/Twitter: [@gregcaporaso](#)) in the Caporaso Lab at Northern Arizona University. You can find information on the courses I teach on my teaching website and information on my research and lab on [my lab website](#).

Download Local Copy

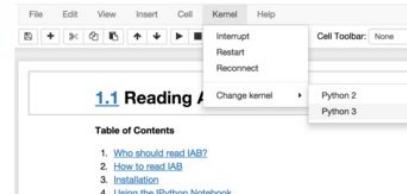
You can also download the latest copy of IAB as Jupyter Notebooks for your own PC.

Download

Read Interactively (recommended)

We're experimenting with Binder to simplify interactively reading IAB.

Note: On opening a notebook in Binder, you'll need to select the *Kernel* menu item, then *Change kernel*, and then *Python 3*. If you don't do this, you will get errors when you try to execute the notebooks. The screenshot to the right illustrates how to change to the Python 3 kernel.



[Read on Binder](#)

Read Statically (easiest)

Here you can find statically published copies of IAB with full output. There are several version to choose from:

- the latest version (you may encounter some instability, but will have the most recent content)
- version 0.1.1 (the most recent stable release)
- version 0.1.0 (the first version)

[Read Latest Online](#)



[readIAB.org](#)

Gut Check: Exploring Your Microbiome

by University of Colorado Boulder & University of Colorado System

i 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

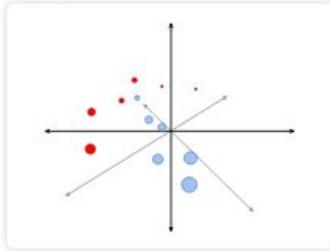
<https://www.coursera.org/learn/microbiome>

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

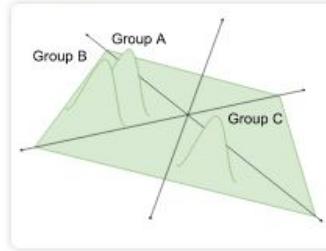
Where would you like to start?

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...

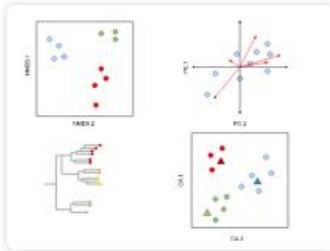
Explore data...



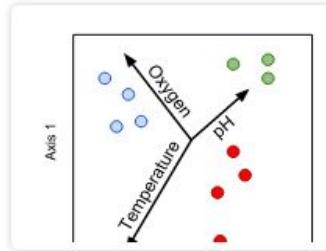
Test a hypothesis...



Browse visualisations...



Explore environmental influence...



QIIME is a community effort

Thanks to our [QIIME Forum masters](#), [QIIME 1 developers](#), [QIIME 2 developers](#), and plugin developers!

Thanks to our users for using and supporting QIIME!

License and contact information

(read this if you're interested in re-using these slides)

These slides were created and arranged by Greg Caporaso, Antonio Gonzalez, Ben Callahan, and other members of the QIIME development group.

For more bioinformatics educational content, see [*An Introduction to Applied Bioinformatics*](#) (IAB) and Dr. Caporaso's [teaching](#) and [lab](#) websites. For updates on IAB, scikit-bio, QIIME, and related projects, follow [@qiime2](#) and [@KnightLabNews](#) on Twitter.

This work is licensed under the Creative Commons Attribution 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Feel free to use or modify these slides, but please credit the QIIME developers by placing the following attribution information where you feel that it makes sense: QIIME 2, <https://qiime2.org>.