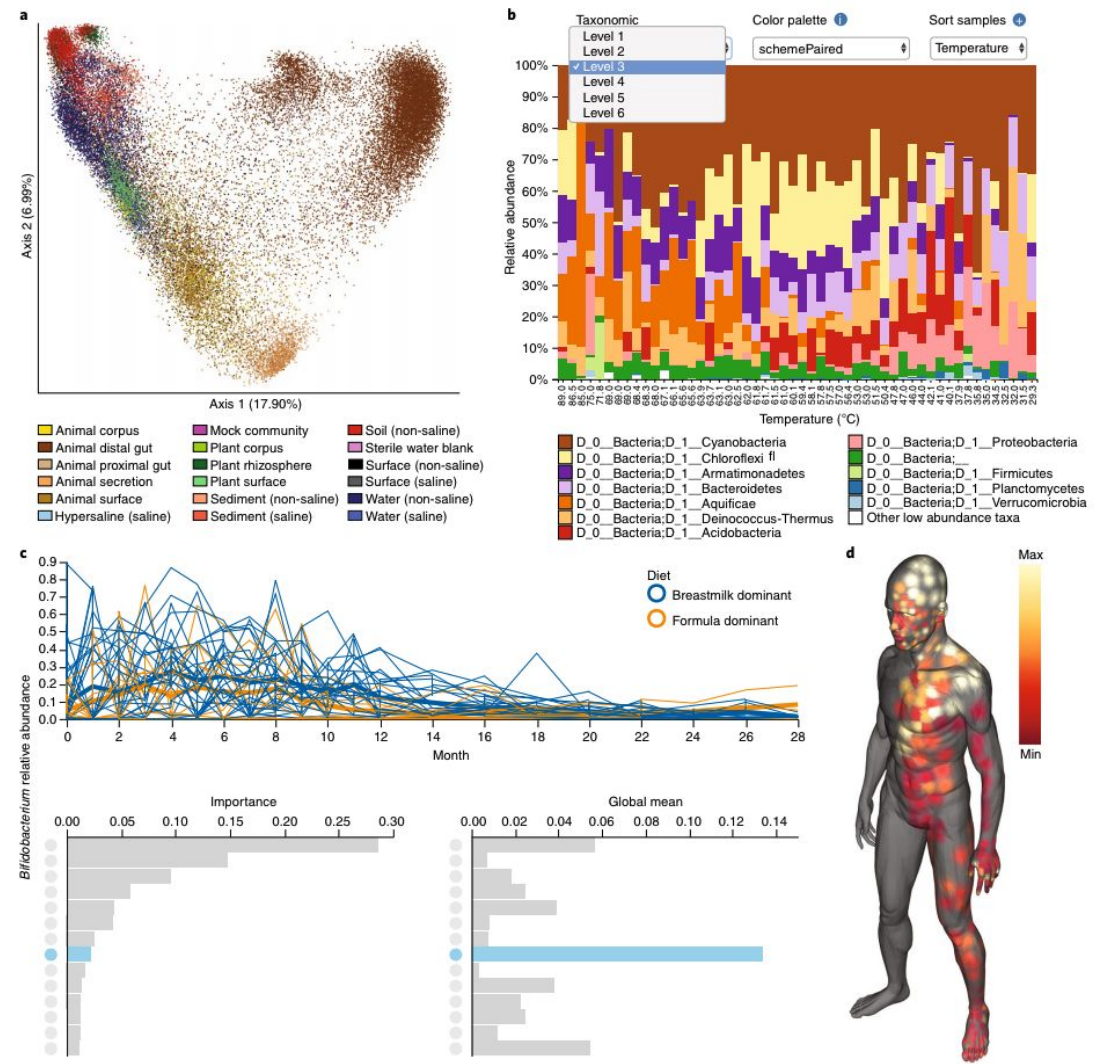


Advancing our understanding of cancer and the human microbiome with



Greg Caporaso, Ph.D.
 Center for Applied Microbiome Science
 Pathogen and Microbiome Institute
 Northern Arizona University
 Flagstaff, Arizona, USA
<http://caporasolab.us>

These slides: <https://bit.ly/q2-itcr-2022>

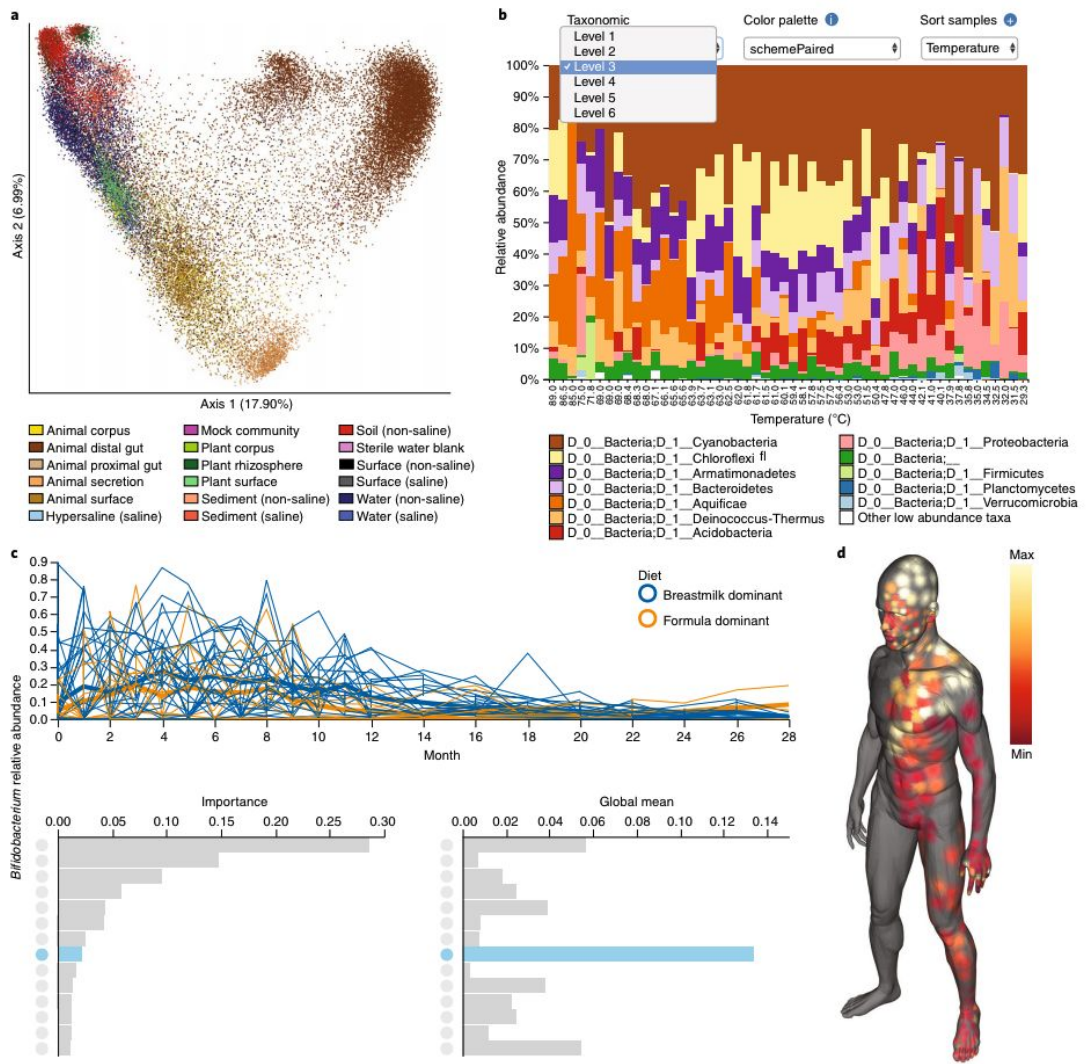


Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2

Evan Bolyen, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, Eric J. Alm, Manimozhiyan Arumugam, Francesco Asnicar, Yang Bai, Jordan E. Bisanz, Kyle Bittinger, Asker Brejnrod, Colin J. Brislaw, C. Titus Brown, Benjamin J. Callahan, Andrés Mauricio Caraballo-Rodríguez, John Chase, Emily K. Cope, Ricardo Da Silva, Christian Diener, Pieter C. Dorrestein, Gavin M. Douglas, Daniel M. Durall, Claire Duvallet, Christian F. Edwards, Madeleine Ernst, Mehrbod Estaki, Jennifer Fouquier, Julia M. Gauglitz, Sean M. Gibbons, Deanna L. Gibson, Antonio Gonzalez, Kestrel Gorlick, Jiarong Guo, Benjamin Hillmann, Susan Holmes, Hannes Holste, Curtis Huttenhower, Gavin A. Huttley, Stefan Janssen, Alan K. Jarmusch, Lingjing Jiang, Benjamin D. Kaehler, Kyo Bin Kang, Christopher R. Keefe, Paul Keim, Scott T. Kelley, Dan Knights, Irina Koester, Tomasz Kosciolko, Jorden Kreps, Morgan G. I. Langille, Joslynn Lee, Ruth Ley, Yong-Xin Liu, Erika Loftfield, Catherine Lozupone, Massoud Maher, Clarisse Marotz, Bryan D. Martin, Daniel McDonald, Lauren J. McIver, Alexey V. Melnik, Jessica L. Metcalf, Sydney C. Morgan, Jamie T. Morton, Ahmad Turan Naimey, Jose A. Navas-Molina, Louis Felix Nothias, Stephanie B. Orchanian, Talima Pearson, Samuel L. Peoples, Daniel Petras, Mary Lai Preuss, Elmar Pruesse, Lasse Buur Rasmussen, Adam Rivers, Michael S. Robeson II, Patrick Rosenthal, Nicola Segata, Michael Shaffer, Arron Shiffer, Rashmi Sinha, Se Jin Song, John R. Spear, Austin D. Swafford, Luke R. Thompson, Pedro J. Torres, Pauline Trinh, Anupriya Tripathi, Peter J. Turnbaugh, Sabah Ul-Hasan, Justin J. van der Hoof, Fernando Vargas, Yoshiki Vázquez-Baeza, Emily Vogtmann, Max von Hippel, William Walters, Yunhu Wan, Mingxun Wang, Jonathan Warren, Kyle C. Weber, Charles H. D. Williamson, Amy D. Willis, Zhenjiang Zech Xu, Jesse R. Zaneveld, Yilong Zhang, Qiyun Zhu, Rob Knight & J. Gregory Caporaso

Show fewer authors

<https://doi.org/10.1038/s41587-019-0252-6>



Our microbiomes impact the efficacy of medical treatments and suggest new treatment protocols.

How gut bacteria could boost cancer treatments

Faecal transplants have helped some people to overcome resistance to powerful immunotherapies. Now dozens of trials are taking aim at the cancer-microbiome connection.

Jeanne Erdmann




Cell

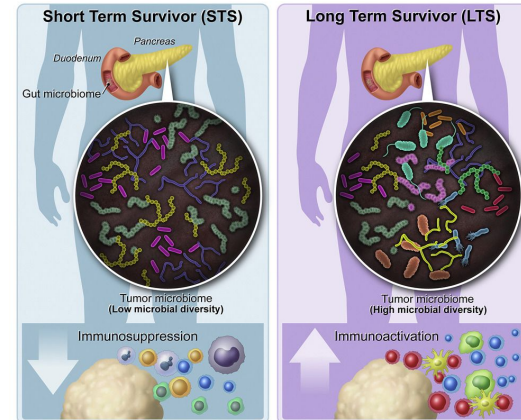
Volume 178, Issue 4, 8 August 2019, Pages 795-806.e12



Article

Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes

Erick Riquelme^{1, 2, 18}, Yu Zhang^{1, 18}, Liangliang Zhang^{3, 4}, Maria Montiel¹, Michelle Zoltan¹, Wenli Dong³, Pompeyo Quesada¹, Ismet Sahin⁵, Vidhi Chandra¹, Anthony San Lucas⁶, Paul Scheet⁶, Hanwen Xu¹, Samir M. Hanash^{1, 7}, Lei Feng³, Jared K. Burks⁸, Kim-Anh Do³, Christine B. Peterson⁹, Deborah Nezman⁹, Ching-Wei D. Tzeng¹⁰, Michael P. Kim¹⁰, Cynthia L. Sears¹¹, Nadim Ajami¹², Joseph Petrosino¹², Laura D. Wood¹³, Anirban Maitra¹⁴, Ravid Straussman⁹, Matthew Katz¹⁰, James Robert White¹⁵, Robert Jenq⁴, Jennifer Wargo^{4, 10}, Florencia McAllister^{1, 16, 17, 19} 



Visual Art: © 2019 The University of Texas MD Anderson Cancer Center

[q2-FMT](#): a new plugin, currently in alpha release, for use in fecal microbiota transplant studies ([interactive plot](#))

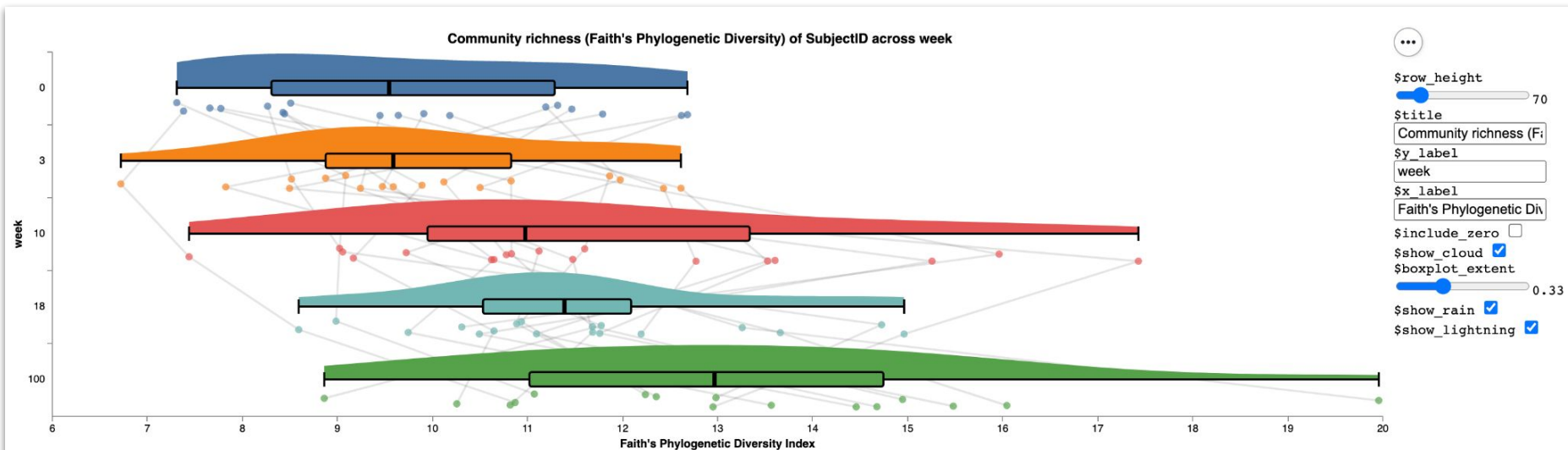
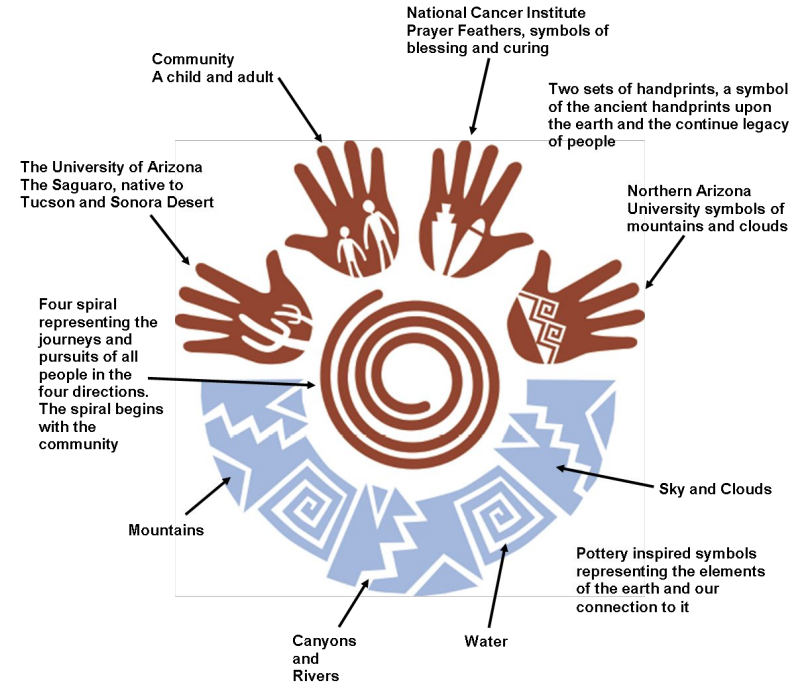


Figure 1. Raincloud plots showing the distribution of subjects' measure of faith_pd across week. Kernel density estimation performed using a bandwidth calculated by Scott's method. Boxplots show the min and max of the data (whiskers) as well as the first, second (median), and third quartiles (box). Points and connecting lines represent individual subjects with a consistent jitter added across groups such that slopes across adjacent groups are visually comparable between subjects.

Group A week	Group B week	A Median faith_pd	B Median faith_pd	n	test-statistic	p-value	q-value
0 (n=18)	3 (n=17)	9.549735	9.592980	17	70.0	0.758312	0.758312
0 (n=18)	10 (n=18)	9.549735	10.981772	18	20.0	0.004337	0.005783
0 (n=18)	18 (n=18)	9.549735	11.393924	18	4.0	0.000386	0.001545
0 (n=18)	100 (n=16)	9.549735	12.972867	16	5.0	0.001123	0.002247

Table 1. Wilcoxon Signed Rank tests between groups (week vs week), with two-sided, asymptotic p-value calculations and Benjamini–Hochberg correction for multiple comparisons (q-value).

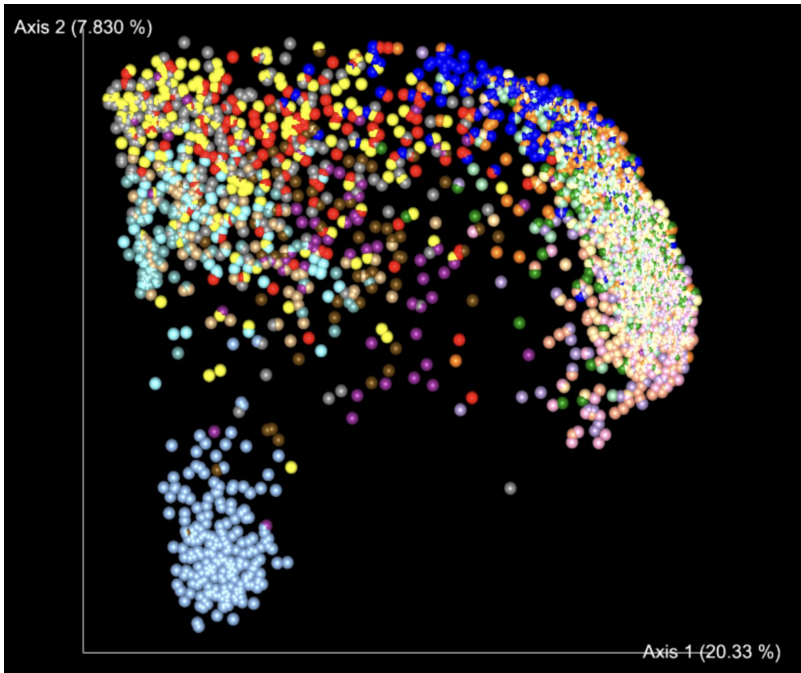
Partnership for Native American Cancer Prevention: Viewing Native American Cervical Cancer Disparities through the Lens of the Vaginal Microbiome



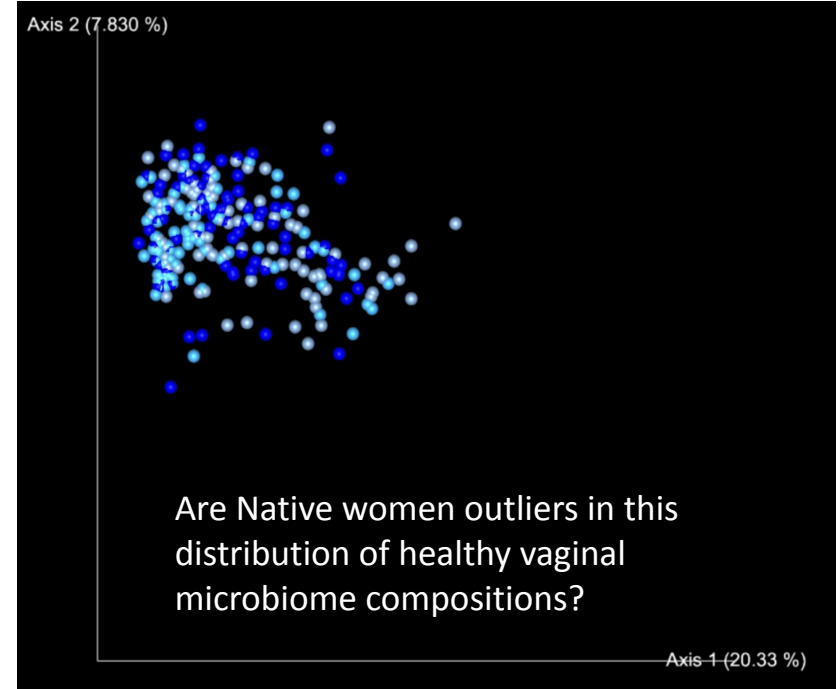
Reducing ethnic bias in public microbiome data. Are Native American women outliers in this distribution of healthy microbiome samples?

([Interactive visualization](#))

All eighteen Human Microbiome Project body sites.



Three vaginal sites only.



Four Corners Science and Computing Club

Leadership team

J Gregory Caporaso (PI)
Department of Biological Sciences
Northern Arizona University

Jani Ingram (PI)
Department of Chemistry
Northern Arizona University

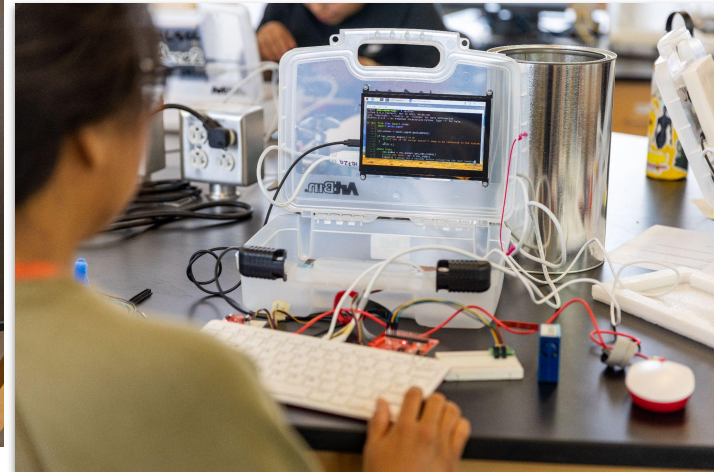
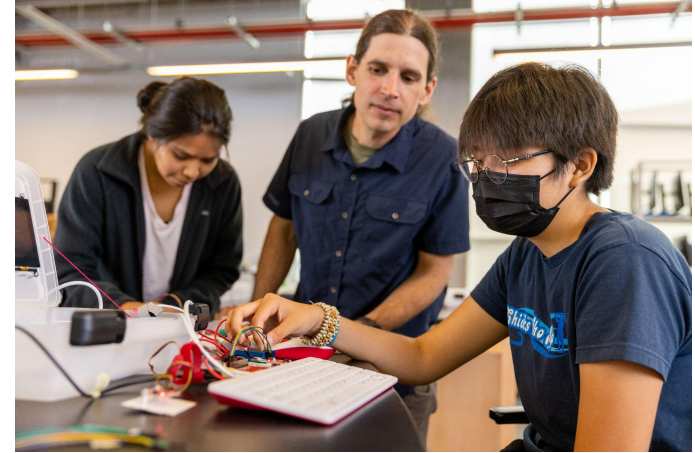
Joslynn Lee (PI)
Department of Chemistry
Fort Lewis College

Naomi Lee (PI)
Department of Chemistry
Northern Arizona University

Jeff Meilander
Department of Biological Sciences
Northern Arizona University

Mansel Nelson
Institute for Tribal Environmental
Professionals
Northern Arizona University

**Chan
Zuckerberg
Initiative** 



Some details at <https://bit.ly/3euJoAB>

cancer microbiome data portal

Will enable access to microbiome marker gene, shotgun, metabolome, and other data for use in meta-analysis.



redbiom

Qiimp

Help ▾

Software ▾

More Info ▾

Sign in

Reset Password

Sign up

Qiita Spots Patterns

Qiita (*canonically pronounced cheetah*) is an entirely **open-source** microbial study management platform. It allows users to keep track of multiple studies with multiple 'omics data. Additionally, Qiita is capable of supporting multiple analytical pipelines through a 3rd-party plugin system, allowing the user to have a single entry point for all of their analyses.

Qiita provides database and compute resources to the global community, alleviating the technical burdens that are typically limiting for researchers studying microbial ecology (e.g. familiarity with the command line or access to compute power).

Qiita's platform allows for quick reanalysis of the datasets that have been deposited using the latest analytical technologies. This means that Qiita's internal datasets are living data that is periodically re-annotated according to current best practices.

For more information about how to use Qiita, visit the [documentation](#).

Note that you should be logged into the system to access any studies and files available.



OPEN ACCESS



PEER-REVIEWED

RESEARCH ARTICLE

Multi-omics data integration reveals metabolome as the top predictor of the cervicovaginal microenvironment

Nicholas A. Bokulich , Paweł Łaniewski , Anja Adamov, Dana M. Chase, J. Gregory Caporaso,
Melissa M. Herbst-Kralovetz 

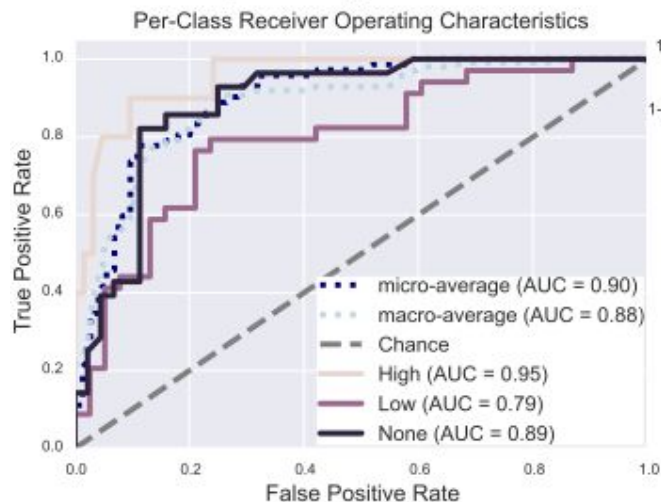
Version 2



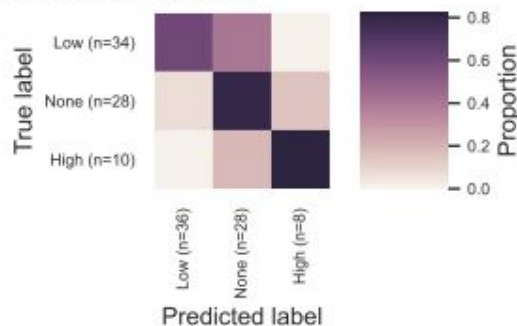
Published: February 23, 2022 • <https://doi.org/10.1371/journal.pcbi.1009876>

Genital inflammation

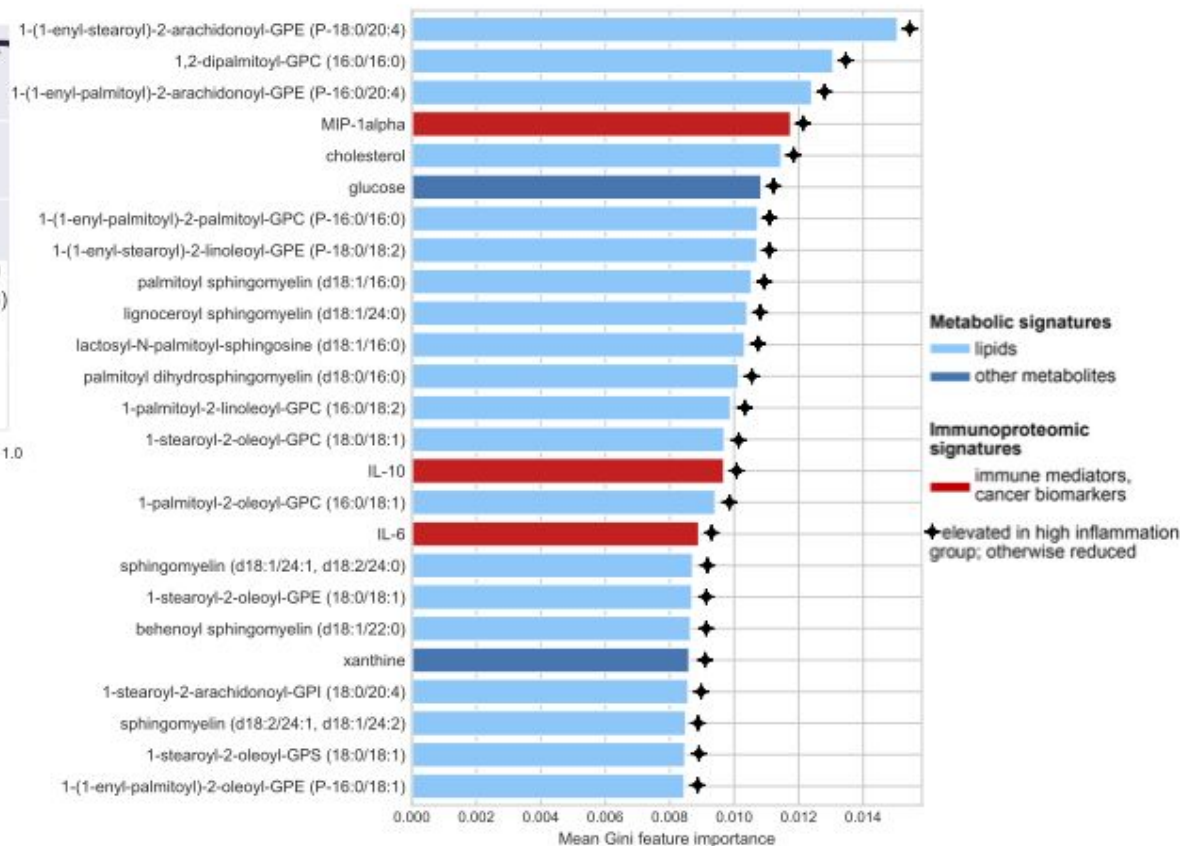
A. Receiver operating curve



B. Confusion matrix



C. Most predictive features



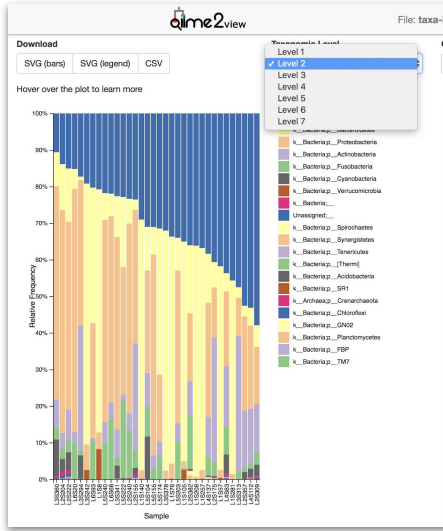


Low-level features

- **Decentralized provenance tracking** automates bioinformatics record keeping facilitating reproducibility.

QIIME 2 integrated data provenance ensures reproducibility

(try this out by clicking the *Provenance* tab [here](#))



The image shows the QIIME2 interface with the 'Provenance' tab selected. The main area displays a 'Provenance Graph' showing the flow of data from 'seqs' and 'sequences' through 'demultiplexed_seqs' and 'reference_reads' to a 'reads classifier', which then outputs 'tab' and 'taxonomy'. The 'reads classifier' node is highlighted in blue. To the right, the 'Action Details' panel provides metadata for the 'denoise_single' action, including execution UUID, runtime, start/end times, duration, and parameters like trunc_len, trim_left, max_ee, trim_q, chimera_method, min_fold_parent_over_abundance, and n_threads.

```
graph TD
  seqs --> demultiplexed_seqs
  sequences --> reference_reads
  demultiplexed_seqs --> reads_classifier
  reference_reads --> reads_classifier
  reads_classifier --> tab
  reads_classifier --> taxonomy
```

Action Details

- execution: uid: "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 microseconds"
- 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_ee: 2, 3: trim_q: 2, 4: chimera_method: "consensus", 5: min_fold_parent_over_abundance: 1, 6: n_threads: 1

Provenance “replay” enables generation of executable code from data provenance, making it easier to re-run analyses that you (or others) have run before. ([Instructional video](#) on [QIIME 2 YouTube Channel](#))

The screenshot shows the QIIME 2 interface with the file 'taxa-bar-plots.qzv' open. The 'Provenance' tab is active, displaying a 'Provenance Graph' on the left and 'Action Details' on the right. The graph shows a flow from 'seqs' and 'sequences' to 'reference_taxonomy', then to 'demultiplexed_seqs' and 'reference_reads', which both feed into a highlighted 'reads classifier' action. This action then feeds into 'tab', which finally feeds into 'taxonomy'. The 'Action Details' for the 'reads classifier' (method 'denoise_single') include:

- execution:** uuid: "3897fb5c-55ed-46b1-a48d-aa0651d2b597"
- runtime:** start: 2017-09-28T21:14:34.374Z, end: 2017-09-28T21:23:05.935Z, duration: "8 minutes, 31 seconds, and 561708 microseconds"
- 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_ee: 2; 3: trunc_q: 2; 4: chimera_method: "consensus"; 5: min_fold_parent_over_abundance: 1; 6: n_threads: 1



The screenshot shows a Python script named 'your_pipeline.py' generated from QIIME 2 provenance. The script is a function 'your_pipeline' that takes several arguments and performs a series of steps to generate a taxa bar plot. The code includes comments and uses the 'ctx' object to retrieve actions and their parameters. The script is as follows:

```
1 #
2 # This file was generated from QIIME 2 Provenance. Cool!
3 #
4 #
5
6 def your_pipeline(ctx, table, phylogeny, sampling_depth, metadata,
7                   n_jobs_or_threads=1):
8     faith_pd = ctx.get_action('diversity_lib', 'faith_pd')
9     unweighted_unifrac = ctx.get_action('diversity_lib', 'unweighted_unifrac')
10    weighted_unifrac = ctx.get_action(
11        'diversity_lib',
12        'weighted_unifrac')
13    pcoa = ctx.get_action('diversity', 'pcoa')
14    emperor_plot = ctx.get_action('emperor', 'plot')
15    core_metrics = ctx.get_action('diversity', 'core_metrics')
16
17    cr = core_metrics(table=table, sampling_depth=sampling_depth,
18                    metadata=metadata, n_jobs=n_jobs_or_threads)
19
20    faith_pd_vector = faith_pd(table=cr.rarefied_table,
21                              phylogeny=phylogeny)
22
23    dms = []
24    dms += unweighted_unifrac(table=cr.rarefied_table, phylogeny=phylogeny,
25                             threads=n_jobs_or_threads)
26    dms += weighted_unifrac(table=cr.rarefied_table,
27                            phylogeny=phylogeny,
28                            threads=n_jobs_or_threads)
29
30    pcoas = []
31    for dm in dms:
32        pcoas += pcoa(distance_matrix=dm)
33
34    plots = []
35    for pcoa in pcoas:
36        plots += emperor_plot(pcoa=pcoa, metadata=metadata)
37
38    return (
39        cr.rarefied_table, faith_pd_vector, cr.observed_features_vector,
40        cr.shannon_vector, cr.evenness_vector, *dms,
41        cr.jaccard_distance_matrix, cr.bray_curtis_distance_matrix,
42        *pcoas, cr.jaccard_pcoa_results, cr.bray_curtis_pcoa_results,
43        *plots, cr.jaccard_emperor, cr.bray_curtis_emperor)
44
```



Low-level features

- **Decentralized provenance tracking** automates bioinformatics record keeping facilitating reproducibility.
- **Multiple user interfaces.** The same functionality is accessible through graphical interface, command line interface, and API, which target different types of users.

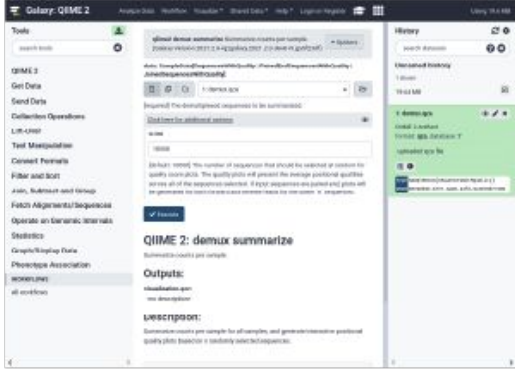
Choose the QIIME 2 interface that helps you work most efficiently.

a) QIIME 2 View



Related Software:
Galaxy

b) QIIME 2 on Galaxy



Related Software:
EBI Metagenomics Portal
QIITA
NIH Nephel

c) QIIME 2 CLI

```

2. ~ (zsh)
$ qiime info
System versions
Python version: 3.5.3
QIIME 2 release: 2017.6
QITMF 2 version: 2017.6.0
q2cli version: 2017.6.0

Installed plugins
alignment 2017.6.0
composition 2017.6.0
dada2 2017.6.0
    
```

Related Software:
QIIME 1

d) QIIME 2 Artifact API

```

Untitled - idle
Python 3 | idle
Not saved yet

(1) import pandas as pd
    from qiime2 import Artifact

(2) t = Artifact.load('table.qzv')
    t.view(pd.DataFrame)

    
```

	4bf5eeb300368260019e1fbc7a3c718fc
L1S105	2222.0
L1S140	0.0
L1S208	0.0
L1S257	0.0
L1S281	0.0

Related Software:
phyloseq



COMPUTATIONAL SOPHISTICATION

Data analyst (clinician, policy maker, research subject) Domain scientists Power users Data scientists

Galaxy Platforms News

The [Galaxy Platform Directory](#) lists resources for easily running your analysis on Galaxy, including publicly available servers, cloud services, and containers and VMs that run Galaxy. Here's the recent platform news we know about:



Protologger

Protologger is an all-in-one genome description tool, aimed at simplifying gathering the data required for writing protologues. This includes providing taxonomic, functional and ecological insights. See the [Protologger video](#) for an introduction.



q2Galaxy (Alpha)

Starting with the [2021.4 release](#) of QIIME 2 (see releases below), this [Docker image](#) is automatically updated as QIIME 2 is updated. It contains the full set of QIIME 2 actions as Galaxy tools, and is maintained by the [QIIME 2 community](#). QIIME 2 is a next-generation microbiome bioinformatics platform that is extensible, free, and open source. QIIME 2 has [extensive documentation](#), an [online \(and free\) workshop](#), and a [support forum](#). There is also a [video](#) describing how to launch the QIIME 2 Galaxy Docker image (on Windows!). Finally, this is an *Alpha* release. Please report any issues you notice on the [forum](#).



Center for Phage Technology (CPT)

Four videos about the Texas A&M Center for Phage Technology's Galaxy server came out recently. They include three how-to videos, and a webinar by Jason Gill. The Videos highlight doing annotation with Galaxy and Apollo.



New Platforms Section on YouTube

Protologger and CPT are not the only Galaxy platforms creating videos. We have gathered or created YouTube playlists for every Galaxy platform that we could find videos for. There are [52 videos about 15 different platforms](#), ranging from Galaxy Australia to Sciensano.

QIIME 2's Galaxy support was initially released in April 2021, and available on the Galaxy Tool Shed as of August 2022.

A new [written](#) and [28-video](#) course, the QIIME 2 Cancer Microbiome Intervention Tutorial, illustrates the use of QIIME 2 on Galaxy to analyze public cancer microbiome data.

Galaxy Tool Shed

Repositories Groups Help User

8681 valid tools on Apr 04, 2022

All Repositories

Browse by category

Repository Actions

Repository `suite_qiime2_core`

Name: suite_qiime2_core

Owner: q2d2

Synopsis: The QIIME 2 'core' distribution

Link to this repository: https://toolshed.g2.bx.psu.edu/view/q2d2/suite_qiime2_core/3f011fac899d

Clone this repository: `hg clone` https://toolshed.g2.bx.psu.edu/repos/q2d2/suite_qiime2_core

Type: repository_suite_definition

Revision: 0:3f011fac899d

New Plants flavour in UseGalaxy.eu

In preparation for the workshop [Plant Transcriptome Analysis in Galaxy](#) that took place in mid April, a new flavour of the Galaxy Server has been set up, filtering the tools by the plant research field.



QIIME 2 multi-interface JupyterBook tutorials allow you to learn QIIME 2 and microbiome analysis using the interface that you will be most effective with, including the new QIIME 2 Galaxy interface, the command line interface, or the Python 3 API. Access the same functionality regardless of your interface preference.

For example, see the QIIME 2 *Cancer Microbiome Intervention* tutorial at <https://bit.ly/3vZ0nle>:

Generate summaries of full table and sequence data

Next, it's useful to generate summaries of the feature table and sequence data. We did this after running DADA2 previously, but since we're now working with a new feature table and new sequence data, we should look at a summary of this table as well.

Using the `qiime2 feature-table summarize` tool:

1. Set `"table"` to #: `feature-table.qza`
2. Expand the `additional options` section
 - For `"sample_metadata"`:
 - Press the + `Insert sample_metadata` button to set up the next steps.
 1. Leave as `Metadata from TSV`
 2. Set `"Metadata Source"` to `sample-metadata.tsv`
3. Press the `Execute` button.

Once completed, for the new entry in your history, use the `Edit` button to set the name as follows:

(Renaming is optional, but it will make any subsequent steps easier to complete.)

History Name	"Name" to set (be sure to press <code>Save</code>)
--------------	-----------------------------------------------------

```
#: qiime2 feature-table summarize [...] : table.qzv  
visualization.qzv
```

Using the `qiime2 feature-table tabulate-seqs` tool:

1. Set `"data"` to #: `rep-seqs.qza`
2. Press the `Execute` button.

Once completed, for the new entry in your history, use the `Edit` button to set the name as follows:

(Renaming is optional, but it will make any subsequent steps easier to complete.)

Generate summaries of full table and sequence data

Next, it's useful to generate summaries of the feature table and sequence data. We did this after running DADA2 previously, but since we're now working with a new feature table and new sequence data, we should look at a summary of this table as well.

```
qiime feature-table summarize \  
  --i-table feature-table.qza \  
  --m-sample-metadata-file sample-metadata.tsv \  
  --o-visualization table.qzv  
qiime feature-table tabulate-seqs \  
  --i-data rep-seqs.qza \  
  --o-visualization rep-seqs.qzv
```

Generate summaries of full table and sequence data

Next, it's useful to generate summaries of the feature table and sequence data. We did this after running DADA2 previously, but since we're now working with a new feature table and new sequence data, we should look at a summary of this table as well.

```
table_viz, = feature_table_actions.summarize(  
    table=feature_table,  
    sample_metadata=sample_metadata_md,  
)  
rep_seqs_viz, = feature_table_actions.tabulate_seqs(  
    data=rep_seqs,  
)
```




Low-level features

- **Decentralized provenance tracking** automates bioinformatics record keeping facilitating reproducibility.
- **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.

Add new plugin

DEICODE

0.2.3

(pronounced /de.ko.de/) Robust Aitchison PCA for sparse omics datasets, linking specific features to beta-diversity ordination through the use of compositional biplots.



mmvec

v1.0.1

A software package for learning microbe-metabolite interactions.



q2-aldex2

0.1.1

Compositional differential abundance analysis. ALDEx2 provides a framework that encompasses essentially all high-throughput sequencing data types by modelling the data as a log-ratio transformed probability distribution rather than as counts.



q2-breakaway

1.0

“breakaway” is the premier package for statistical analysis of microbial diversity. “breakaway” implements the latest and greatest estimates of richness, as well as the most commonly used estimates. The “breakaway” philosophy is to estimate diversity, to put error bars on diversity estimates, and to perform hypothesis tests for diversity that use those error bars.



q2-clawback

0.0.3

Assembles taxonomic weights to increase classification accuracy with q2-feature-classifier. Can download data from Qiita or use your data.



q2-coordinates

2018.11

A qiime2 plugin supporting methods for geographic mapping of qiime2 artifact data or metadata.



q2-dbotu

2018.4.2

q2-feature-classifier

2019.1

q2-fragment-insertion

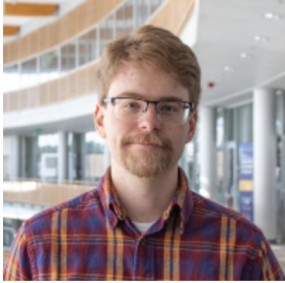
2019.1

Chan
Zuckerberg
Initiative 

[CZI project description](#)

(source: NAU News)

Current Caporaso Lab team members on the ITCR project



Evan Bolyen

*Senior Research
Software Engineer*



Keegan Evans

*Research Software
Engineer*



Liz Gehret

*Research Software
Engineer*



Chloe Herman

PhD Student



Anthony Simard

*Research Software
Engineer*

We're currently [hiring a research scientist](#).

Acknowledgements

Caporaso Lab

Evan Bolyen
Keegan Evans
Liz Gehret
Chloe Herman
Jeff Meilander
Anthony Simard
Franklin Stewart
and former lab members...

QIIME and QIIME 2 communities

QIIME 1 Development Team

QIIME 2 Development Team

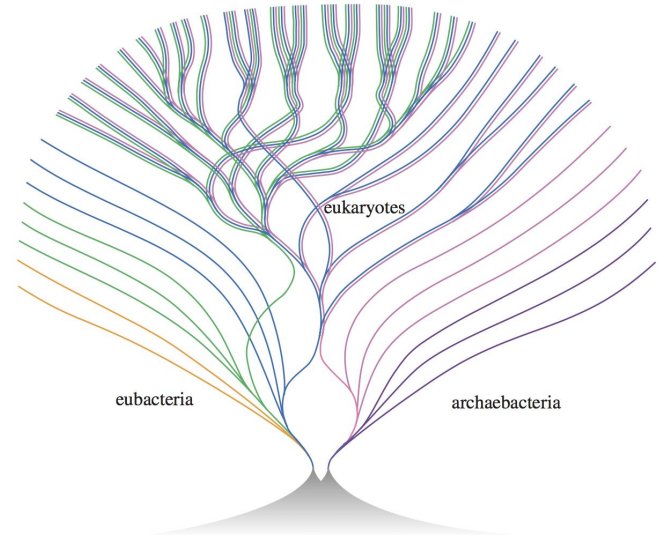


Image source: [Martin \(1999\)](#)

We're currently hiring a research scientist.

QIIME 2 project funding

National Cancer Institute

- ITCR ([1U24CA248454-01](#))
- [Partnership for Native American Cancer Prevention](#) (U54CA143925)

National Science Foundation ([1565100](#))

Chan-Zuckerberg Initiative

- Essential Open Source Software (2019-207342)
- [EOSS D&I supplement \(2021-237226\)](#)

Alfred P. Sloan Foundation

QIIME 2 resources

Docs: <https://docs.qiime2.org>

Developer resources: <https://dev.qiime2.org>

Tech support: <https://forum.qiime2.org>

Twitter: [@qiime2](#)

GitHub: <https://github.com/qiime2>

Workshops: <https://workshops.qiime2.org>

YouTube: <https://www.youtube.com/c/qiime2>

These slides: <https://bit.ly/q2-itcr-2022>