

# Metaanalysis in Qiita: quick-start

## Introduction

---

### What is Qiita?

[Qiita](#) is an open-source online repository and analysis tool for microbiome data, created by Antonio Gonzales, Jose Navas-Molina, *et al.* in the [Knight Lab](#) at UCSD. It's a terrifically useful tool for organizing the first steps of a microbiome analysis, particularly when you want to share data with collaborators or compare data to previous datasets. (You can find the publication describing this tool [here](#)).

### What is this document?

This document is more or less a shortcut to certain pieces of existing Qiita documentation, plus a little bit of extra detail for bits that should be particularly useful to you in the Hackathon.

### Qiita Documentation

Qiita is meant to be an accessible platform for microbiome analysis and exploration, even if you don't have prior experience with bioinformatics or command-line tools. Consequently the developers have put quite a lot of effort into writing documentation.

After you've walked through this quick-start, if you're interested in working more with Qiita, I'd recommend you more thoroughly explore the existing Qiita docs.

My recommendation for getting started is to follow the complete [Qiita Workshop Tutorial](#) from the Center for Microbiom Innovation at UCSD. This tutorial provides a small set of example data you can download, and walks you through step-by-step the creation and analysis of a new study.

There is additional, more-granular documentation available at the [central Qiita Documentation site](#). This includes information on some of Qiita's more advanced features, such as the ability to [easily submit studies](#) from Qiita to public nucleotide repositories.

## Create an account

---

To do analyses in Qiita, you'll need to get an account.

Follow [these instructions](#) from the CMI Workshop Tutorial to sign up.

# Check out a study

'Studies' form the basis of the Qiita database organization: researchers create a new study, and then upload their data to the study. Initially, studies are private, only viewable by their owner (and any other accounts the owner chooses to share it with). Eventually, users can choose to make these studies public.

If you click `Study / View Studies` in the nav bar at the top of the screen, you'll see a table of your private studies and a table of available public studies:

The screenshot shows the Qiita web interface with the following details:

**Your Studies**

Expand for analysis (artifact count)	Title	Study ID	Samples	Shared With These Users	Principal Investigator	Publications	Status	Qiita EBI submission
10	Gut microbiota utilize immunoglobulin A for mucosal colonization	11513	439	Modify Owner: gdonalds@caltech.edu Jon Sanders	Sarkis Mazmanian	29724905, 10.1126/science.aaq0926	public	ERP107727 (submitted)
No BIOMs	Metabolomics and Metagenomic Analysis of FINNRisk Subjects Who Develop Metabolic Disorders	11535	469	Modify Owner: Austin Swafford Jake, Jon Sanders, Keng Cher Soh	Rob Knight		sandbox	not submitted

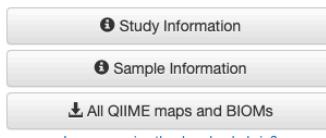
Showing 16 to 17 of 17 entries

**Public Studies**

Expand for analysis (artifact count)	Title	Study ID	Samples	Principal Investigator	Publications	Qiita EBI submission
6	A core gut microbiome in obese and lean twins.	77	281	Jeff Gordon	19043404, 10.1038/nature07540	not submitted
4	Soil bacterial and fungal communities across a pH gradient in an arable soil	94	27	Noah Fierer	20445636, 10.1038/ismej.2010.58	not submitted
8	Succession of microbial consortia in the developing infant gut microbiome <b>GOLD</b>	101	63	Ruth Ley	20668239, 25974302, 10.1073/pnas.1000081107, 10.1016/j.chom.2015.04.009	not submitted
4	Pyrosequencing-Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale	103	89	Noah Fierer	19502440, 10.1128/AEM.00335-09	not submitted
4	Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes	104	52	Noah Fierer	20561020, 10.1111/j.1462-2920.2010.02277.x	not submitted

Showing 1 to 5 of 459 entries

You can click on one of the example study titles to take a look inside. Here, I've clicked on study #101, "Succession of microbial consortia in the developing infant gut microbiome." This study was performed right here at Cornell!

**Data Types** (click on the tabs)**16S****Succession of microbial consortia in the developing infant gut microbiome - ID 101****koenig\_infant\_time\_series**Do you want to submit to EBI-ENA? Review the [submission checklist](#)**Abstract**

The colonization process of the infant gut microbiome has been called chaotic, but this view could reflect insufficient documentation of the factors affecting the microbiome. We performed a 2.5-year case study of the assembly of the human infant gut microbiome to relate life events to microbiome composition and function. Sixty fecal samples were collected from a healthy infant along with a diary of diet and health status. Analysis of >300,000 16S rRNA genes indicated that the phylogenetic diversity of the microbiome increased gradually over time and that changes in community composition conformed to a smooth temporal gradient. In contrast, major taxonomic groups showed abrupt shifts in abundance corresponding to changes in diet or health. Community assembly was nonrandom: we observed discrete steps of bacterial succession punctuated by life events. Furthermore, analysis of ~500,000 DNA metagenomic reads from 12 fecal samples revealed that the earliest microbiome was enriched in genes facilitating lactate utilization, and that functional genes involved in plant polysaccharide metabolism were present prior to the introduction of solid food, priming the infant gut for an adult diet. However, ingestion of table foods caused a sustained increase in the abundance of Bacteroidetes, elevated fecal short chain fatty acid levels, enrichment of genes associated with carbohydrate utilization, vitamin biosynthesis and xenobiotic degradation, and a more stable community composition, all of which are characteristic of the adult microbiome. This study revealed that seemingly chaotic shifts in the microbiome could be attributed to life events.

**Study ID:** 101**Owner:** gail.ackerman@colorado.edu**Publications:** 10.1073/pnas.1000081107, 10.1016/j.chom.2015.04.009, 20668239, 25974302**PI:** Ruth Ley (Cornell University)**Lab Contact:** Jesse Stombaugh (CU Boulder)**Shared With:****Samples:** 63**EBI:** not submitted**Study Tags**

Previously admin, Previously assigned, New

**GOLD** × Add more tags

New tags are linked to the user that created them. Report abuse.

Save tags

This study also has a 'Gold' tag, indicating that it has been curated by the Qiita development team as a great example.

If you'd like to try creating your own test study, try following [the tutorial](#) from the CMI Workshop docs.

## Find data for a meta-analysis

One of the great things about Qiita is that gives you access to data from hundreds of thousands of samples *that have all been processed the same way*. This makes it really simple to get up and going, without having to worry about downloading all the raw data and running it through your own computational pipeline. (You can also download the raw data if you want!)

There are two ways you can find data for analysis: searching for **studies**, and searching for **samples**.

### Finding studies from the study splash page

The table we just saw listed can be filtered by entering text into the **Filter by column data** search bar at the top of the page. This only searches the data you can read on the public splash page for the study, including the title, abstract, etc.

For example, I can find a study I did on the gut microbes of ants from the rainforests of Peru by typing 'rainforest ants' into the search bar:

Filter by column data (Title, abstract, PI, etc):

Rainforest Ants

Filter studies by tags: ([Admin](#), [User](#))

Select tags for filtering

## Your Studies

Show 5 ▾ entries

Expand for analysis (artifact count)	Title	Study ID	Samples	Shared With These Users	Principal Investigator	Publications	Status	Qiita EBI submission
No studies found								

Showing 0 to 0 of 0 entries (filtered from 1,721 total entries)

Previous Next

## Public Studies

Expand for analysis (artifact count)	Title	Study ID	Samples	Principal Investigator	Publications	Qiita EBI submission
10	   Gut bacteria of Peruvian rainforest ants	10343	487	Jon Sanders		ERP014516 (submitted)

Showing 1 to 1 of 1 entries (filtered from 466 total entries)

Previous 1 Next

## Finding samples using `redbiom`

In addition to finding complete studies, you can find *individual samples* from among the hundreds of thousands publically available in Qiita. This is a very new addition, and uses a clever search mechanism behind the scenes called `redbiom`.

The advantage to searching for particular samples is that you can start to explore more targeted questions built directly from data. For example, what if you wanted to compare all the samples that contained a particular microbial taxon, or even a particular 16S sequence? You can find those samples using the `redbiom` search functionality.

To search with `redbiom`, click the corresponding

## Finding samples by taxon

Let's say you want to search for all the samples that contain bacteria with a particular taxonomic label. Select 'Taxon' from the dropdown, enter the search term in the box, and click the search button on the right.

This search method uses the GreenGenes taxonomy, so you'll need to prepend the taxon label with the letter corresponding to the taxonomic level followed by two underscores. For example, if I want to search for samples containing bacteria belonging to the family Christensenellaceae, I would search for

`f__Christensenellaceae`:

Help and examples

f\_Christensenellaceae

Taxon



Show 10 entries

Filter results by column data:

Expand for analysis (artifact count)	Title	Study ID
Artifacts: 1   Samples: 1	Characterization of Airborne Microbial Communities at a High-Elevation Site and Their Potential To Act as Atmospheric Ice Nuclei	314
Artifacts: 1   Samples: 1	Widespread colonization of the lung by <i>Tropheryma whipplei</i> in HIV infection UPenn	1020
Artifacts: 1   Samples: 1	Myrold Alder Fir	1031
Artifacts: 1   Samples: 1	UPenn LHMP E3D01BAL	1171
Artifacts: 1   Samples: 1	Ileocecal resection dramatically affects the microbiome of the murine jejunum and colon	1564
Artifacts: 1   Samples: 1	Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing - DS6	1687
Artifacts: 1   Samples: 1	Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing - DS6	1688
Artifacts: 1   Samples: 1	Dominguez sleep deprived flies	1799
Artifacts: 1   Samples: 1	Intestinal Adaptation in Proximal and Distal Segments: Two Epithelial Responses Diverge After Intestinal Separation	10585
Artifacts: 1   Samples: 124	The effect of treatment with the helminth product Tuftsin-phosphorylcholine (TPC) on the microbiome of mice with active lupus	11625

Found 1707 artifacts with 63326 samples.

Showing 1 to 10 of 365 entries

Previous 1 2 3 4 5 ... 37 Next

This returns quite a lot of samples!

## Finding samples by ASV sequence

You can also search for a particular sequence. This allows you to very finely target microbial taxa, even if they don't correspond to particular taxonomic labels. For example, let's say I want to find samples that have close relatives of *Methylocystis parvus*, a methane-oxidizing (methanotrophic) bacterium.

I can go to NCBI and find the [16S ribosomal RNA gene](#) corresponding to this bacterium. Then I can find the portion of the gene that matches most of the samples in Qiita. This happens to be the V4 region of the 16S rRNA gene. The NCBI 16S rRNA sequence I found is:

```
>NR_044946.1 Methylocystis parvus OBBP 16S ribosomal RNA, partial sequence
AACGAACGCTGGCGGCAGGCCAACACATGCAAGTCGAACGCTGTAGCAATAACAGAGTGGCAGACGGGTG
AGTAACCGCTGGAACGTGCCCTTCGGAAATAACTCAGGGAAACTTGAGCTAATACGGATACGCC
CTTGGGGAAAGATTATTGCCGAAAGATCGGCCCGTCCGATTAGCTAGTTGGTGTGGTAATGGCGC
ACCAAGGGCAGCATCGGTAGCTGGTAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGAATTGGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGAGT
GATGAAGGCCCTAGGGTTGTAAAGCTTTGCCAGGGACGATAATGACGGTACCTGGATAAGAAGCCCC
GGCTAACCTCGTGCAGCAGCCGGTAATACGAAGGGGCTAGCGTTGGAAACTGGCTAA
GCGCACGTAGGCCAGTCTTAAGTCAGGGTGAATCCGAGGCTAACCTCGGAACTGCCTTGATACT
GGAGGTCTCGAGTCGGAGAGGTGAGTGGAACTGCGAGTGTAGAGGTGAAATTCTGAGATATTGCAAG
AACACCAGTGGCGAAGGCCGCTACTGCCCGTACTGACGCTGAGGTGCAAAGCGTGGGAGCAAACA
GGATTAGATAACCTGGTAGTCCACGCCAACGATGGATGCTAGCCGGTGGGAGCATGCTCTCAGTG
GCGCAGCTAACGCTTAAGCATCCGCCCTGGGAGTACGGTCGCAAGATTAAAACCAAAGGAATTGACG
GGGGCCCGACAAGCGGTGGAGCATGTGGTTAACATCGAACGCGCAGAACCTTACAGCTTTGAC
ATGCCCGGTATGATGCCAGAGATGGCTTCTCCGCAAGGGGCGGAGCACAGGTGCTGCATGGCTGT
CGTCAGCTCGTGTGAGATGGTTAACATCCGCAACGAGCGAACCCCTGCCCTAGTTGCCATC
ATTCAAGTGGGACTCTAGGGGACTGCCGTGATAAGCCCGAGGAAGGTGGGATGACGTCAAGTCCT
CATGCCCTACAGGCTGGCTACACACGTGCTACAATGGCGGTGACAATGGGAAGCGAAAGGGCGACCT
GGAGCAAATCTCAAAAGCCGCTCAGTCGGATTGCACTCTGCAACTCGAGTGCATGAAGGTGGAATCG
CTAGTAATCGCAGATCAGCACGCTCGGGTAATACGTTCCCGGCTTGTACACACCGCCGTCACACCA
TGGGAGTTGGTTTACCGAAGCGTTGCCAACCGCAAGGAGGCAGGCACCACGGTAGGGTCAGCGA
CTGGGTG
```

For `redbiom`, it's important the portion of our search sequence match *exactly* what's in the database: same start, same end. For most samples, that means selecting the 100 or 150 nucleotides immediately following the Earth Microbiome Project 515f primer, which is: `GTGYCAGCMGCCGCGGTAA`. I can replace the `Y` and `M` degenerate characters with a wildcard (`GTG*CAGC*GCCGCGGTAA`) and find this primer location using a text editor:

```
>NR_044946.1 Methylocystis parvus OBBP 16S ribosomal RNA, partial sequence
AACGAACGCTGGCGGCAGGCCAACACATGCAAGTCGAACGCTGTAGCAATAACAGAGTGGCAGACGGGTG
AGTAACCGCTGGAACGTGCCCTTCGGAAATAACTCAGGGAAACTTGAGCTAATACGGATACGCC
CTTGGGGAAAGATTATTGCCGAAAGATCGGCCCGTCCGATTAGCTAGTTGGTGTGGTAATGGCGC
ACCAAGGGCAGCATCGGTAGCTGGTAGAGGATGATCAGCCACACTGGGACTGAGACACGGCCCAGAC
TCCTACGGGAGGCAGCAGTGGGAATTGGACAATGGCGCAAGCCTGATCCAGCCATGCCCGTGAGT
GATGAAGGCCCTAGGGTTGTAAAGCTCTTCGCCAGGGACGATAATGACGGTACCTGGATAAGAACCCCC
GGCTAACCTTC**GTGCCAGCAGCCCGGTAA**TACGAAGGGGCTAGCGTTCTCGGAATCACTGGCGTAAA
GCGCACGTAGGCCGATCTTAAGTCAGGGTGAATCCGAGGCTCAACCTCGGAACTGCCTTGATACT
GGAGGTCTCGAGTCGGGAGAGGTGAGTGGAACTGCGAGTGTAGAGGTGAAATTCTGAGATATTGCAAG
AACACCAGTGGCGAAGGCCGCTACTGCCCGTACTGACGCTGAGGTGCAAAGCGTGGGAGCAAACA
GGATTAGATAACCTGGTAGTCCACGCCGAAACGATGGATGCTAGCCGGTGGGAGCATGCTCTTCAGTG
GCGCAGCTAACGCTTAAGCATCCGCCCTGGGAGTACGGTCGCAAGGATAAACTCAAAGGAATTGACG
GGGGCCCGACAAGCGGTGGAGCATGTGGTTAACATCGAACGCGCAGAACCTTACAGCTTTGAC
ATGCCCGGTATGATGCCAGAGATGGCTTCTCCGCAAGGGCCGGAGCACAGGTGCTGCATGGCTGT
CGTCAGCTCGTGTGAGATGGTTAACATGCCGCAACGAGCGAACCCCTGCCCTAGTTGCCATC
ATTCAAGTGGGACTCTAGGGGACTGCCGTGATAAGCCCGAGGAAGGTGGGATGACGTCAAGTCCT
CATGCCCTACAGGCTGGGCTACACACGTGCTACAATGGCGGTGACAATGGGAAGCGAAAGGGCGACCT
GGAGCAAATCTCAAAAGCCGCTCAGTCGGATTGCACTCTGCAACTCGAGTGCATGAAGGTGGAATCG
CTAGTAATCGCAGATCAGCACGCTCGGGTAATACGTTCCGGCCTGTACACACCGCCGTCACACCA
TGGGAGTTGGTTTACCGAAGCGTTGCCAACCGCAAGGAGGCAGGCACCACGGTAGGGTCAGCGA
CTGGGGTG
```

Finally, I can chop out the 150 characters immediately following the primer sequence:

```
TACGAAGGGGCTAGCGTTCTCGGAATCACTGGCGTAAAGCGCACGTAGGCAGGATCTTAAGTCAGGG
GTGAAATCCGAGGCTAACCTCGGAACTGCCTTGATACTGGAGGTCTCGAGTCCGGAGAGGTGAGTG
GAACTGCGAG
```

I can then select **Feature** in the drop-down and enter my sequence (all caps, no spaces) and find all the samples where that exact ASV occurs:

Qjita Analysis Study Admin redbiom Qiimp Help More Info

Welcome jgsanders@ucsd.edu Log Out   

Redbiom only seal Create new analysis Create From Selected Samples See Previous Analyses 18th, 2018. Note that you will only be able to expand and add artifacts to analyses if you are signed into Qjita.

TACGAAGGGGGCTAGCGTTGGAAATCACTGGCGTAAAGCGCACGTAGCGGATCTTAAGTCAGGGGTGAAATCCGAGGCTAACCTCGGACTGCCTTGAT Feature 

Show 10 entries Filter results by column data:

Expand for analysis (artifact count)	Title	Study ID
 Artifacts: 1   Samples: 228	Patterns of microbiome transition along salinity gradients, and importance for fish host adaptive divergence across osmotic boundaries.	10308
 Artifacts: 1   Samples: 23	Microbial effects of livestock manure fertilization on freshwater aquaculture ponds rearing tilapia, <i>O. shiranus</i> , and African catfish, <i>C. gariepinus</i>	10353
 Artifacts: 1   Samples: 29	Most of the Dominant Members of Amphibian Skin Bacterial Communities Can Be Readily Cultured	10272
 Artifacts: 1   Samples: 3	Malaysia Pasoh Landuse Logged Forest	1714
 Artifacts: 1   Samples: 34	Bacterial Community Spatial and Temporal Variation in a North Temperate Bog Lake	1288
 Artifacts: 1   Samples: 381	Free-Living Amoeba Reservoirs of Pathogenic Leptospira	10442
 Artifacts: 1   Samples: 41	Subtle shifts in microbial communities occur alongside the release of carbon induced by drought and rewetting in contrasting peatland ecosystems	10278
 Artifacts: 1   Samples: 44	Longitudinal analysis of microbial interaction between humans and the indoor environment	2192
 Artifacts: 1   Samples: 5	Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils	1883
 Artifacts: 1   Samples: 56	Individuals diet diversity influences gut microbial diversity in two freshwater fish (threespine stickleback and Eurasian perch) - 1/4	2260

Found 61 artifacts with 1251 samples.

Open <https://qiita.ucsd.edu/redbiom/#> on this page in a new tab

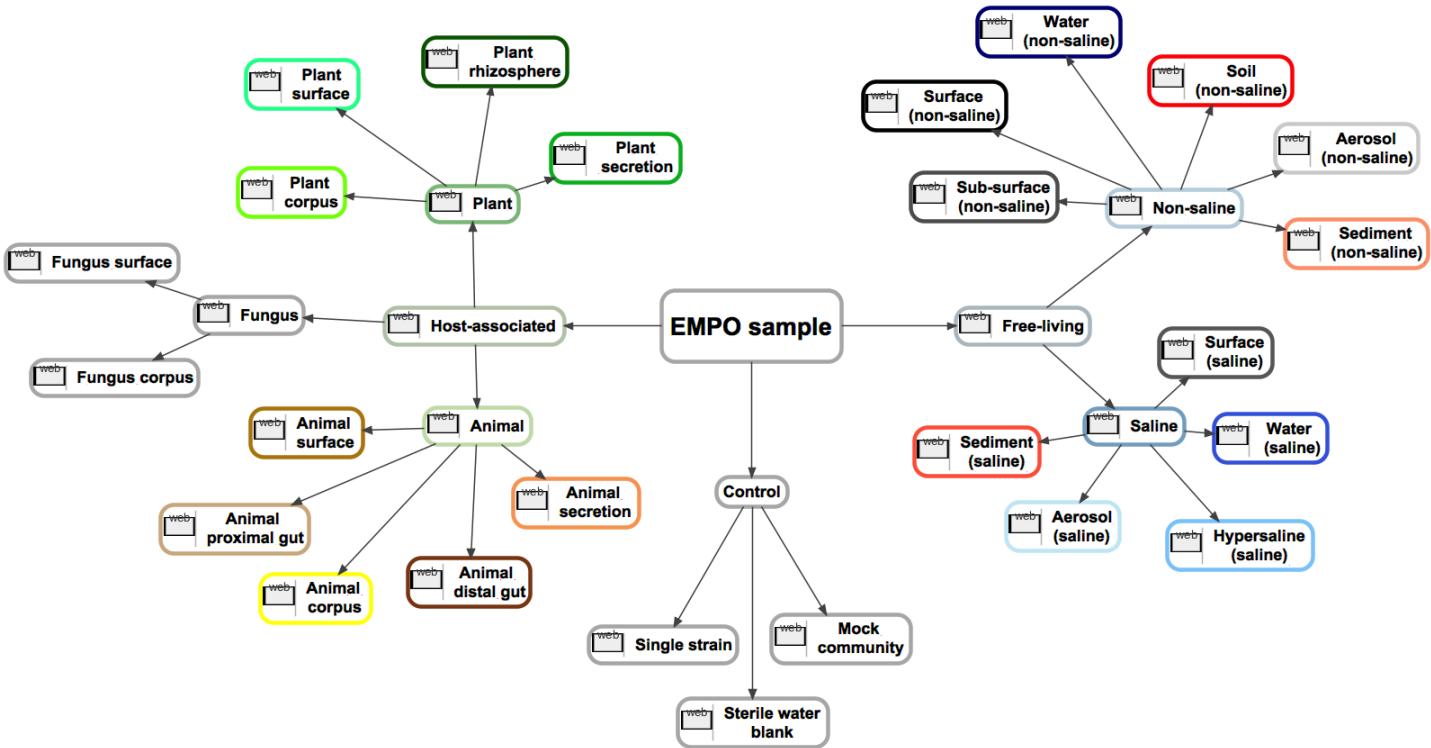
Showing 21 to 30 of 61 entries Previous 1 2 3 4 5 Next

One neat thing about this is that most of the studies that have large numbers of samples containing this feature are from freshwater environments, and the particular microbe we used to search was isolated from a freshwater bog. Cool!

## Finding samples by metadata

Finally, we can search the sample metadata associated with each sample. The syntax for this search is a little complex, so you should refer to the [redbiom documentation](#) for more specifics.

One group of metadata search columns that is really useful for this kind of search corresponds to the [Earth Microbiome Project Ontology](#). The EMPO a pretty easily-understood hierarchical organization of sample types we used in the EMP paper. All samples available in Qiita should have EMPO categories associated with them, under the `qiita_empo_1`, `qiita_empo_2`, and `qiita_empo_3` columns.



So for example, if we wanted to search for freshwater sediments, we could select `Metadata` from the drop-down and use the search query `where qita_empo_3 == 'Sediment (non-saline)'`:

The screenshot shows the Qitta web interface with a search query entered in the search bar: `where qita_empo_3 == 'Sediment (non-saline)'`. The results table displays 10 entries, each with a title, study ID, and options to expand for analysis.

Expand for analysis (artifact count)	Title	Study ID
Artifacts: 3   Samples: 347	Biodiversity and Functional Patterns of Microbial Assemblages in Postglacial Pond Sediment Profiles	1622
Artifacts: 5   Samples: 1	Rio de Janeiro Coastline	1039
Artifacts: 5   Samples: 5	Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau	1627
Artifacts: 5   Samples: 7	Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample - 5prime	722
Artifacts: 6   Samples: 138	Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils	1883
Artifacts: 6   Samples: 30	Global Coral Microbiome Project (GCMP)	10895
Artifacts: 7   Samples: 43	Human and environmental impacts on river sediment microbial communities	807
Artifacts: 7   Samples: 9	Patterns of microbiome transition along salinity gradients, and importance for fish host adaptive divergence across osmotic boundaries.	10308

Showing 1 to 8 of 8 entries

Found 44 artifacts with 580 samples.

We can restrict these to only samples that contain 'USA' somewhere in their metadata by using the query

```
usa where qiita_empo_3 == 'Sediment (non-saline)':
```

The screenshot shows the Qiita search interface. At the top, there's a navigation bar with links for Analysis, Study, Admin, redbiom, QIIME, Help, and More Info. On the right, it shows a welcome message for 'jgsanders@ucsd.edu' and links for Log Out and other user options. Below the header, a message states 'Redbiom only searches on public data. Last update: December 18th, 2018. Note that you will only be able to expand and add artifacts to analyses if you are signed into Qiita.' A 'Help and examples?' button is highlighted in blue. The main search bar contains the query 'usa where qiita\_empo\_3 == "Sediment (non-saline)"'. To the right of the search bar are buttons for 'Metadata' and a magnifying glass icon. Below the search bar, a table lists four study entries. The columns are 'Expand for analysis (artifact count)', 'Title', and 'Study ID'. The first entry has 5 artifacts and 7 samples. The second has 6 artifacts and 138 samples. The third has 6 artifacts and 16 samples. The fourth has 7 artifacts and 43 samples. A total of 24 artifacts with 204 samples were found. At the bottom left, it says 'Showing 1 to 4 of 4 entries'. At the bottom right, there are 'Previous' and 'Next' buttons, with the page number '1' in the center.

Expand for analysis (artifact count)	Title	Study ID
Artifacts: 5   Samples: 7	Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample - 5prime	722
Artifacts: 6   Samples: 138	Microbial diversity in arctic freshwaters is structured by inoculation of microbes from soils	1883
Artifacts: 6   Samples: 16	Global Coral Microbiome Project (GCMP)	10895
Artifacts: 7   Samples: 43	Human and environmental impacts on river sediment microbial communities	807

## Add data to a meta-analysis

Now that we've found some interesting samples, let's combine them into one dataset we can analyze.

Because Qiita processes sequence data using standardized parameters, and 'remembers' which parameters were used to produce each data output, it is possible to easily combine samples processed at different times into one meta-analysis.

[A complete tutorial](#) on performing analyses, as well as meta-analyses, is available at the CMI Workshop documentation.

## Selecting samples

Any time you do one of the above search types, you can select samples to add to a meta-analysis by clicking the green + icon in the column labeled 'Expand for analysis'.

Going back to my ant study from before, if I click the green + icon, I see a list of 10 different artifact types:

Expand for analysis (artifact count)	Title	Study ID	Samples	Shared With These Users	Principal Investigator	Publications	Status	QIITA EBI submission
10	Gut bacteria of Peruvian rainforest ants	10343	487	<a href="#">Modify</a> Owner: Jon Sanders	Jon Sanders		public	ERP014516 (submitted)
<b>Artifacts</b> <b>Processing method</b> <b>Data type</b>								
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: all.biom)   Trimming (length: 100)						16S (V4)
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: all.biom)   Trimming (length: 150)						16S (V4)
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: all.biom)   Trimming (length: 90)						16S (V4)
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: reference-hit.biom)   Trimming (length: 100)						16S (V4)
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: reference-hit.biom)   Trimming (length: 150)						16S (V4)
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: reference-hit.biom)   Trimming (length: 90)						16S (V4)
Add all	Per Artifact (1)	Pick closed-reference OTUs (reference-seq: /databases/gg/13_8/rep_set/97_ottus.fasta)   Split libraries FASTQ						16S (V4)
Add all	Per Artifact (1)	Pick closed-reference OTUs (reference-seq: /databases/gg/13_8/rep_set/97_ottus.fasta)   Trimming (length: 100)						16S (V4)
Add all	Per Artifact (1)	Pick closed-reference OTUs (reference-seq: /databases/gg/13_8/rep_set/97_ottus.fasta)   Trimming (length: 150)						16S (V4)
Add all	Per Artifact (1)	Pick closed-reference OTUs (reference-seq: /databases/gg/13_8/rep_set/97_ottus.fasta)   Trimming (length: 90)						16S (V4)

Each of these represents an output (BIOM table, or OTU table) of the raw data analyzed using a particular combination of parameters. You'll want to make sure you choose artifacts produced using the same parameter combinations for a metaanalysis, so as not to mix proverbial apples and oranges.

I'll choose the 'Deblur reference-hit | Trimming length: 100' parameter combination, which uses the Deblur ASV pipeline, selects only the sequences that look like 16S based on a reference database to remove obvious erroneous sequences, and trims the sequence lengths to 100. This represents a parameter combination that will be present across most studies, and so represents a good starting point for metaanalyses.

To add just this BIOM table to my analysis, I'll click the 'Per artifact' button to add just this one artifact, and then click the next button to confirm:

Expand for analysis (artifact count)	Title	Study ID	Samples	Shared With These Users	Principal Investigator	Publications	Status	QIITA EBI submission
10	Gut bacteria of Peruvian rainforest ants	10343	487	<a href="#">Modify</a> Owner: Jon Sanders	Jon Sanders		public	ERP014516 (submitted)
<b>Artifacts</b> <b>Processing method</b> <b>Data type</b>								
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: all.biom)   Trimming (length: 100)						16S (V4)
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: all.biom)   Trimming (length: 150)						16S (V4)
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: all.biom)   Trimming (length: 90)						16S (V4)
Add all	Per Artifact (1)	Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: reference-hit.biom)   Trimming (length: 100)						16S (V4)
<b>Name</b> <b>Samples in Prep Info</b> <b>Files</b>								
Add	deblur reference hit table (59290 - 2018-09-29 18:21:57)		484		reference-hit.biom			

Sometimes a study will have multiple BIOM artifacts using these same parameters, for example if it spans

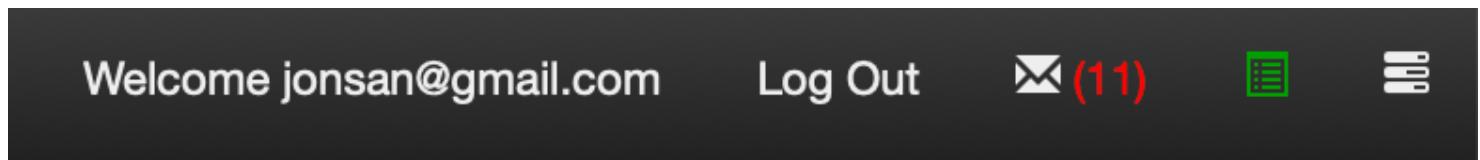
datasets sequenced on multiple sequence runs; in this case, there would be multiple such artifacts in the expanded table. In this case, you could click 'Add all' to add all of these BIOM artifacts from this parameter combination at once.

Now I'll search the table for 'honey bee' to find bee microbiome samples to compare, and add those:

Expand for analysis (artifact count)	Title	Study ID	Samples	Principal Investigator	Publications	Qiita EBI submission	
10	Microbiome of honey bees from Puerto Rico <small>EMP</small>	1064	391	MG Dominguez-Bello		ERP016607 (submitted)	
Artifacts						Processing method	Data type
Add all	Per Artifact (1)					Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: all.biom)   Trimming (length: 100)	16S (V4)
Add all	Per Artifact (1)					Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: all.biom)   Trimming (length: 150)	16S (V4)
Add all	Per Artifact (1)					Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: all.biom)   Trimming (length: 90)	16S (V4)
Add all	Per Artifact (1)					Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: reference-hit.biom)   Trimming (length: 100)	16S (V4)
Add all	Per Artifact (1)					Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: reference-hit.biom)   Trimming (length: 150)	16S (V4)
Add all	Per Artifact (1)					Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: reference-hit.biom)   Trimming (length: 90)	16S (V4)
Add all	Per Artifact (1)					Pick closed-reference OTUs (reference-seq: /databases/gg/13_8/rep_set/97_ottus.fasta)   Split libraries FASTQ	16S (V4)
Add all	Per Artifact (1)					Pick closed-reference OTUs (reference-seq: /databases/gg/13_8/rep_set/97_ottus.fasta)   Trimming (length: 100)	16S (V4)
Add all	Per Artifact (1)					Pick closed-reference OTUs (reference-seq: /databases/gg/13_8/rep_set/97_ottus.fasta)   Trimming (length: 150)	16S (V4)
Add all	Per Artifact (1)					Pick closed-reference OTUs (reference-seq: /databases/gg/13_8/rep_set/97_ottus.fasta)   Trimming (length: 90)	16S (V4)

## Creating the analysis

After I add an artifact, I will see the 'analysis' icon at the top right of the nav bar turn green:



If I click it, it will show me how many studies and samples I have selected:

an@gmail.com

Log Out

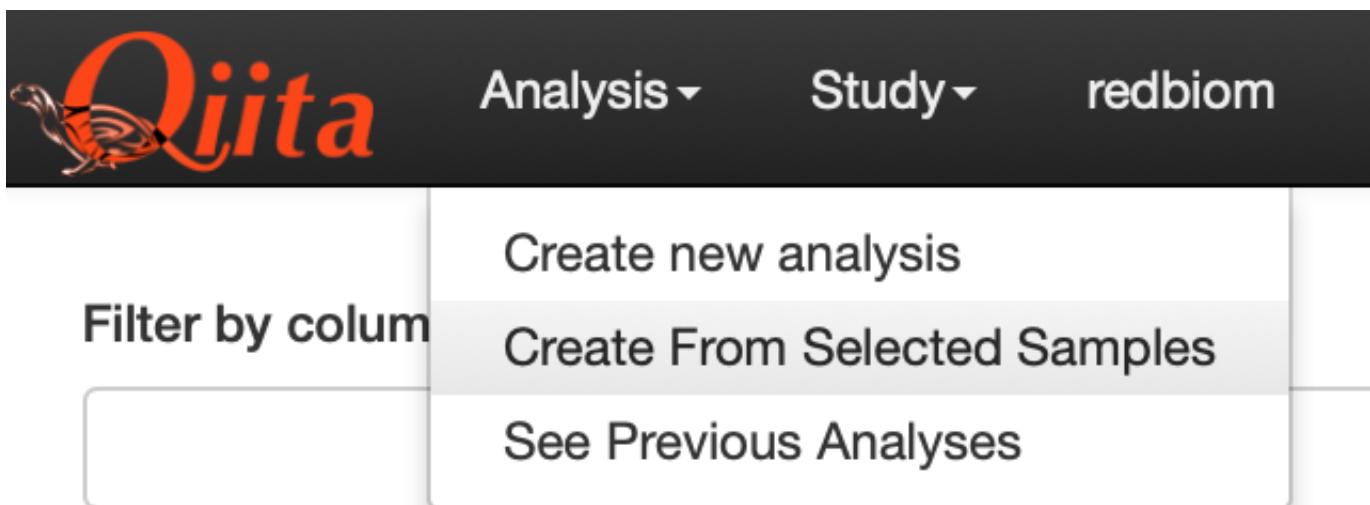
✉ (11)



## SELECTED SAMPLES

2 Studies  
Processed Data  
875 Samples

Next, I can click `Analysis / Create from selected samples` to create a new analysis with all of these samples:



This will bring me to a page that reviews all the different samples I've selected, where I can remove studies if desired (right column), and finally click the green `Create Analysis` button when I'm ready to go forward:

[Create Analysis](#)

[Clear Selected](#)

## Selected Samples

### Gut bacteria of Peruvian rainforest ants

Processed Data

ID	Datatype	Processed Date	Algorithm	Parameters	Samples selected from Prep Info	Show/Hide samples	Remove
59290 ⓘ	16S	2018-09-29 18:21:57.031840	deblur (Deblur)	Mean per nucleotide error rate: 0.005 Demultiplexed sequences: 7496 Insertion/deletion (indel) probability: 0.01 Reference phylogeny for SEPP: Greengenes_13.8 Minimum dataset-wide read threshold: 0 Positive filtering database: default Threads per sample: 1 Negative filtering database: default Error probabilities for each Hamming distance: 1, 0.06, 0.02, 0.02, 0.01, 0.005, 0.005, 0.005, 0.001, 0.001, 0.001, 0.0005 Indexed negative filtering database: default Indexed positive filtering database: default Sequence trim length (-1 for no trimming): -1 Minimum per-sample read threshold: 2 Jobs to start: 5 Maximum number of insertion/deletion (indel): 3	484	<a href="#">Show/Hide samples</a>	<a href="#">Remove</a>

### Microbiome of honey bees from Puerto Rico

Processed Data

ID	Datatype	Processed Date	Algorithm	Parameters	Samples selected from Prep Info	Show/Hide samples	Remove
56873 ⓘ	16S	2018-09-26 16:59:53.765504	deblur (Deblur)	Mean per nucleotide error rate: 0.005 Demultiplexed sequences: 5530	391	<a href="#">Show/Hide samples</a>	<a href="#">Remove</a>

It will then prompt me to give my analysis a name and description:

Help ▾ More Info ▾

### Create new analysis

**Analysis name**

**Description**

**Create analysis**

1840	deblur (Deblur)	Mean per nucleotide error rate: 0.005 Demultiplexed sequences: 7496 Insertion/deletion (indel) probability: 0.01	484
------	-----------------	--	-----

This sends Qiita the command to start processing the data into a single BIOM artifact for my meta-analysis. Initially, I see this splash page while the command is waiting to run on the server:

The screenshot shows the Qiita interface for an analysis titled "bees and ants - ID 22540 (Private)". At the top, there's a navigation bar with links for Analysis, Study, redbiom, Qiimp, Help, and More Info. Below the title, a button says "Make analysis public". A "Shared with:" section shows a user icon. A "Processing network" section has a "Hide" button. A note says "Click on the graph to navigate through it. Click circles for more information. This graph will refresh in 6 seconds or reload now". Below this are two status legends: "Job status (circles)" with categories success, running, error, in\_construction, queued, waiting, and deleting; and "Artifact status (triangles)" with categories artifact, type, outdated, and deprecated. A message "Hang tight, we are processing your request:" is displayed. A job log entry shows "Job: 808e3309-064f-464d-a881-0925ca1c4e42 Status: queued". At the bottom, there's a thank you message, email contact information, and links to terms and conditions.

bees and ants - ID 22540 (Private) [Make analysis public](#)

a comparison of microbiota from bees and ants

Shared with:

Processing network [Hide](#)

Click on the graph to navigate through it. Click circles for more information. This graph will refresh in 6 seconds or reload now

Job status (circles): success running error in\_construction queued waiting deleting

Artifact status (triangles): artifact type outdated deprecated

Hang tight, we are processing your request:

Job: 808e3309-064f-464d-a881-0925ca1c4e42 Status: queued

Thank you for using Qiita. [Citing Qiita?](#).  
Questions? [qiita.help@gmail.com](mailto:qiita.help@gmail.com); don't forget to add your study or analysis id.  
Read our [terms and conditions](#).

Once the server processes the data, I will see an interactive network diagram showing the artifacts available for analysis. In this case, there will actually be two: one BIOM containing all the sequences from both studies, and one containing just the sequences which have been successfully inserted into a microbial phylogenetic tree using the SEPP program.

I can click the artifact and see detailed information. Note that each 'artifact' also contains several files individually-downloadable files. In this case, I've clicked the top artifact icon and can see it contains the biom, a phylogenetic tree, and two html summary documents:

## bees and ants - ID 22540 (Private)

[Make analysis public](#)

### a comparison of microbiota from bees and ants

 Shared with:

Processing network

[Hide](#)

Click on the graph to navigate through it. Click circles for more information. This graph will refresh in 6 seconds or reload now

Job status (circles):       Artifact status (triangles): artifact   

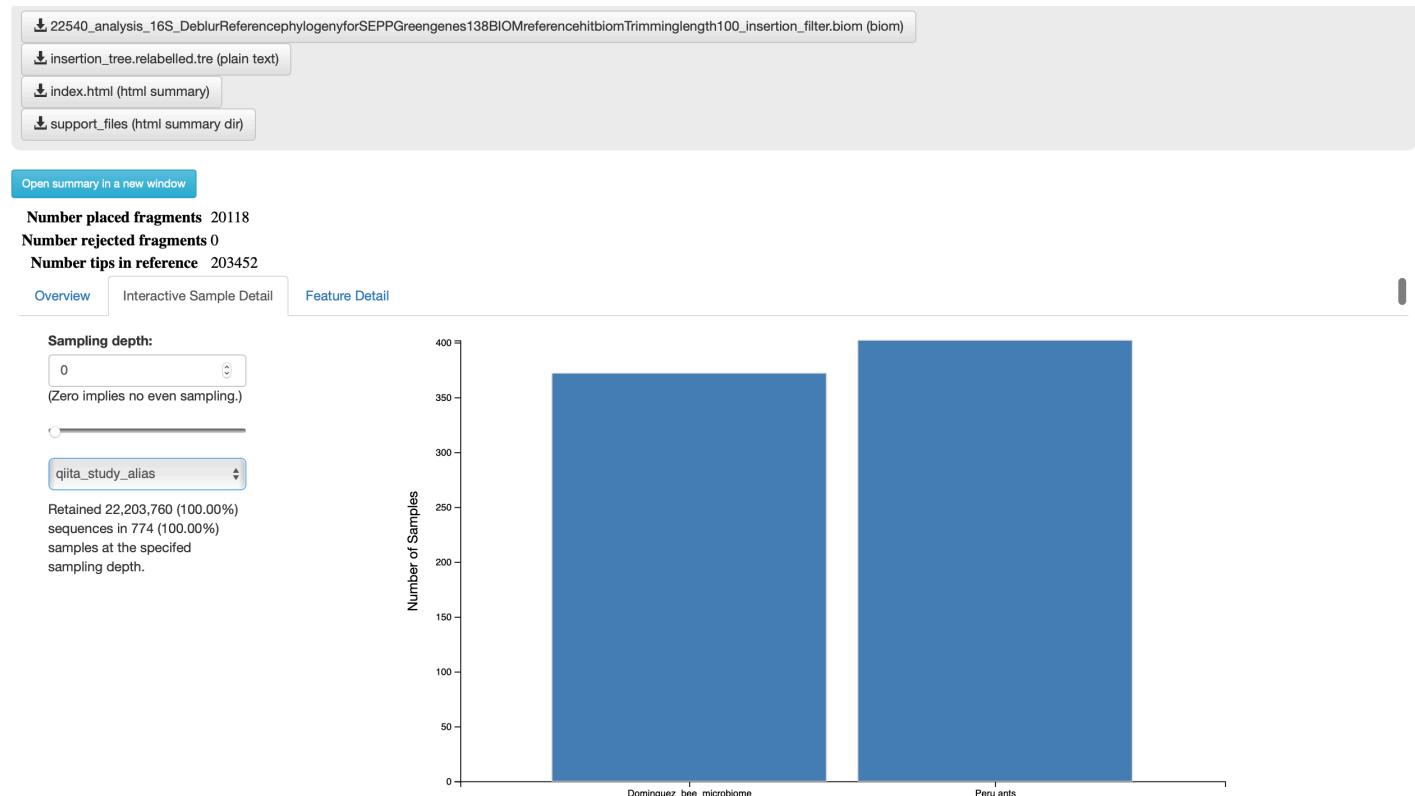
Hang tight, we are processing your request:

Job: 808e3309-064f-464d-a881-0925ca1c4e42 Status: queued

Thank you for using Qiita. [Citing Qiita?](#).Questions? [qiita.help@gmail.com](mailto:qiita.help@gmail.com); don't forget to add your study or analysis id.Read our [terms and conditions](#).

Clicking any of these will cause the file to download to my computer.

If I scroll down, I can also see an interactive summary of the different samples and sequence counts in this artifact. By clicking around, I am able to find that there's a metadata column called `qiita_study_alias` which shows me that I have just over 350 samples from each study:



## Add analysis steps

From here, it is possible to do a number of different analysis steps right on Qiita. For a more complete rundown of the options, check out the [CMI Workshop Tutorial](#).

Here, I'll just go ahead and add a few steps just to show how it works.

First, I'll click on the artifact I want (making sure it's the one that includes a phylogenetic tree) and click the **Process** button. This brings up a number of possible actions. I'll choose **Rarefy Table** to subset all my samples to the same number of sequences, and choose 5000 sequences per sample:

Choose command: Rarefy table

Required parameters:

The feature table to be rarefied.: dflt\_name (BIOM)

Optional parameters:

Parameter set: Default

Rarefy with replacement by sampling from the multinomial distribution instead of rarefying without replacement. (with replacement):

The total frequency that each sample should be rarefied to. Samples where the sum of frequencies is less than the sampling depth will be not be included in the resulting table unless subsampling is performed with replacement. (sampling depth):

5000

Add Command

Next, I'll add a **Beta diversity (phylogenetic)** step to calculate a beta diversity distance matrix between samples. I'll choose Unweighted UniFrac as my metric, and make sure to choose 'Artifact tree, if it exists' to make sure it uses the phylogenetic tree contained within my artifact:

.....

**Phylogenetic tree:**

Artifact tree, if exists



**The beta diversity metric to be computed. (metric):**

Unweighted UniFrac



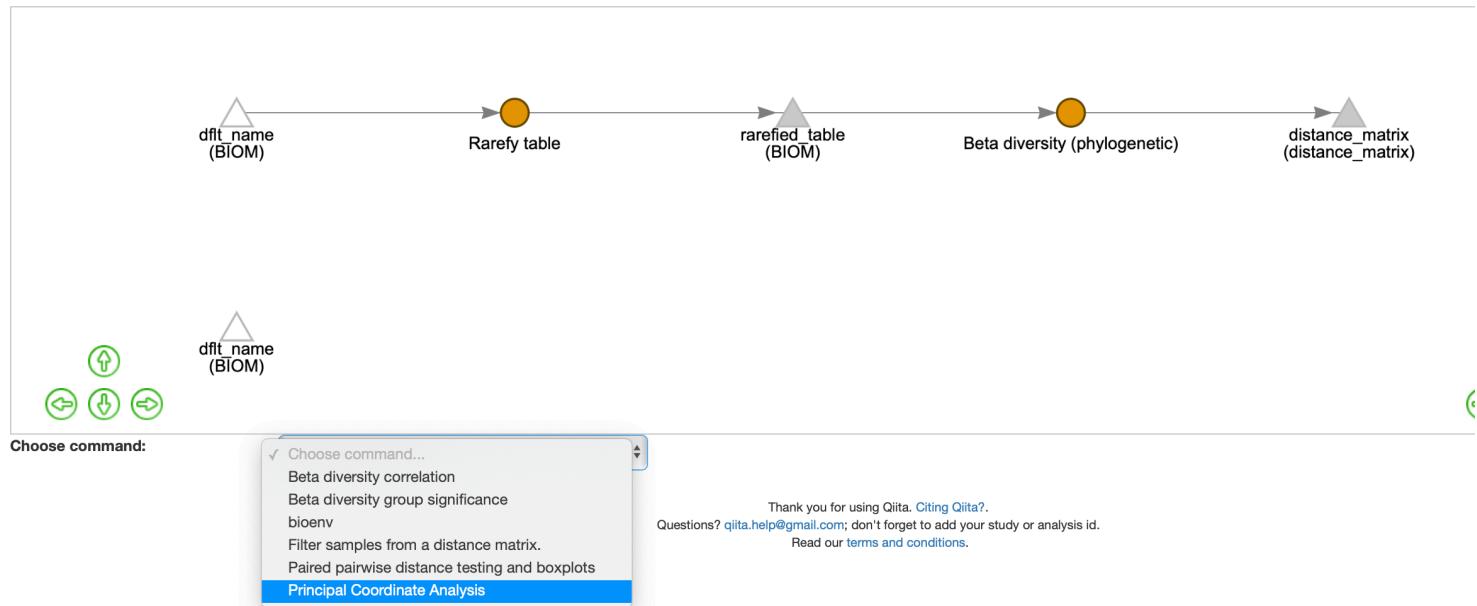
**The number of workers to use. (n jobs):**

1



Add Command

Finally, I'll add a step to generate a Principle Coordinates Analysis visualization of my beta-diversity distances:

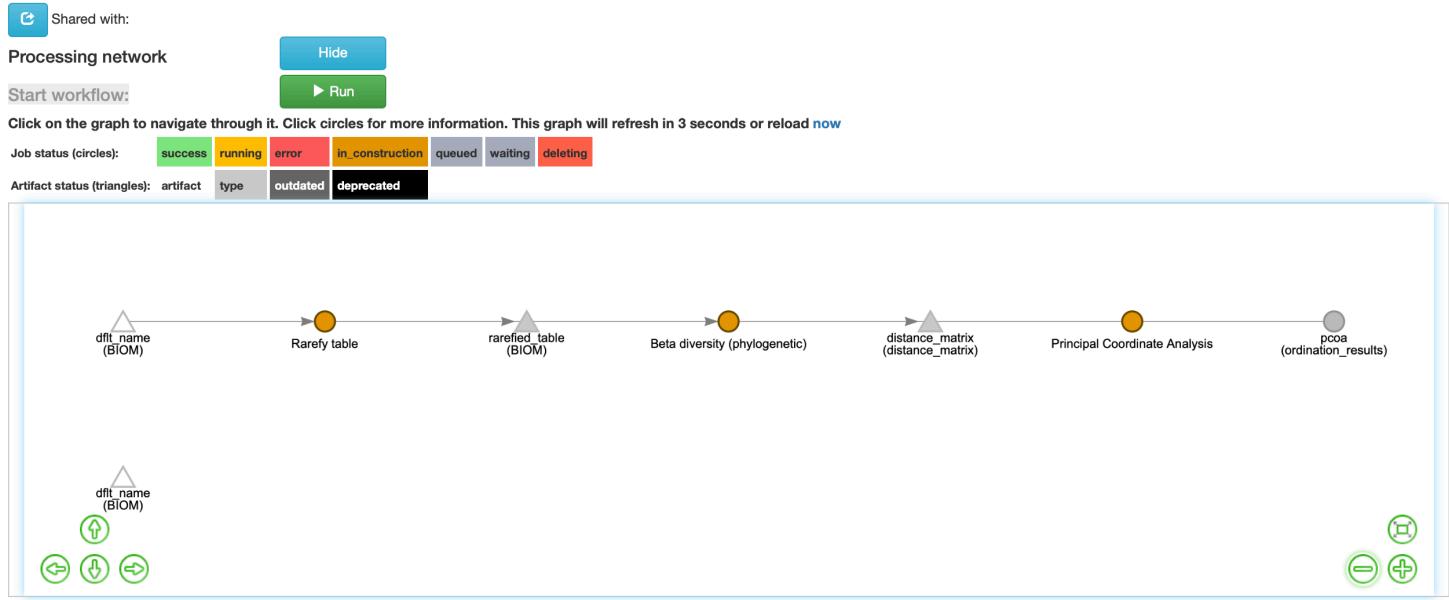


## Run analysis

Note that as each processing step is added, a processing step node (represented by a circle) and an output artifact (represented by a triangle) are each added to the network. Once I'm happy with the steps I've selected, I can click the green **Run** button at the top of the page to send the commands to the server:

## bees and ants - ID 22540 (Private) [Make analysis public](#)

### a comparison of microbiota from bees and ants



Once jobs are running, I can click the icon at top right of the nav bar to get a status report:

More Info ▾

## Active Jobs



successful jobs are not shown

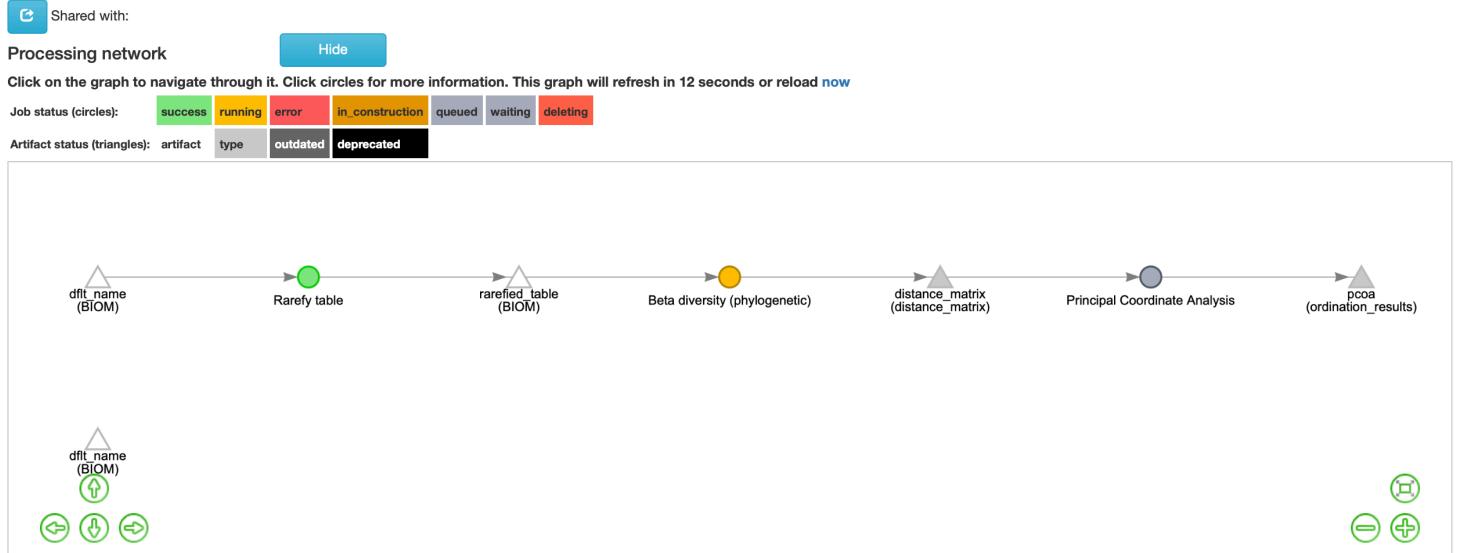
Search:

Heartbeat	Name	Status	Step
2019-03-27 18:50:31	Rarefy table	running	Step 4 of 4: Processing results
	Beta diversity (phylogenetic)	waiting	
	Principal Coordinate Analysis	waiting	

[Close](#)

As each step finishes, the corresponding command icon in the processing network will change color to indicate the outcome:

## a comparison of microbiota from bees and ants



Once each step completes, I am able to click on the resulting artifact to see a summary:

*distance\_matrix* (ID: 69630) Visibility: sandbox

Edit name

Process

Delete

Show processing information

Available files: Show/Hide

[distance-matrix.tsv \(plain text\)](#)

[index.html \(html summary\)](#)

[Open summary in a new window](#)

**Number of samples:** 576

**Minimum distance:** 0.0615

**Maximum distance:** 0.9948

**Mean distance:** 0.7513

**Median distance:** 0.7786

**Distance Matrix - hierarchical clustering**

