redbiom - Notes on first impression

First impression on graphic interface
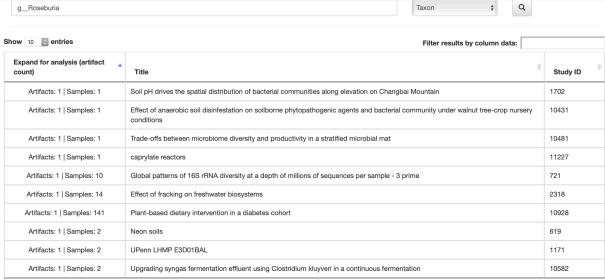- search
  - only one seq at a time
    - different input accommodation for seq and metadata
    - need data from actual usage cases
  - feature = seq (ambiguous label)
- no way to stop and re-search
- no progress status

Key Functionalities of CLT Missing from Graphic Interface
- only [search] is implemented
  - excluding searching for sample

- cannot put together the pipeline in advance
  - multiple cmd using pipe in CLT
  - have to wait for a long time for
- cannot have further specification for search
  - like context/features
- after the searching results are returned, no further actions seem applicable?
  - data are organized in a static table
  - Only three columns of info are displayed
  - Title? ambiguous lable

● Help and example lead to more confusions

Help and examples?

○ Not making sense, looks like making a search bar as space limited command line

○ Find all samples only belonging to the EMP in which the pH is under 7, for a variety of sample types:

■ soil:

```
soil where ph < 7 and emp_release1 == 'True'
```

■ ocean water:

```
water & ocean where ph > 7 and emp_release1 == 'True'
```

■ non-ocean water:

```
water - ocean where ph > 7 and emp_release1 == 'True'
```

○ not very interesting

○ Some other interesting examples:

```
feces & canine

(beer | cider | wine | alcohol)

where sample_type == 'stool'

usa where sample_type == 'stool' and host_taxid == 9606
```

CLT Observation (Following example on gitHub page)
– Help pages can be organized in a more intuitive way
    – listed top and secondary help menu (there are more levels down)

```
[(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom --help
Usage: redbiom [OPTIONS] COMMAND [ARGS]...

Options:
  --version  Show the version and exit.
  --help     Show this message and exit.

Commands:
  admin      Update database, etc.
  fetch      Sample data and metadata retrieval.
  search     Feature and sample search support.
  select     Select items based on metadata
  summarize  Summarize things.
[(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom admin --help
Usage: redbiom admin [OPTIONS] COMMAND [ARGS]...

  Update database, etc.

Options:
  --help  Show this message and exit.

Commands:
  coherency                   Assert coherency within contexts.
  create-context              Create context for sample data.
  load-sample-data            Load nonzero entries per sample.
  load-sample-metadata        Load sample metadata.
  load-sample-metadata-search Load sample metadata.
  scripts-read-only           Set scripts to read-only
  scripts-writable            Set scripts to allow write
```

```
[(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom fetch --help
Usage: redbiom fetch [OPTIONS] COMMAND [ARGS]...

  Sample data and metadata retrieval.

Options:
  --help  Show this message and exit.

Commands:
  features            Fetch sample data containing features.
  features-contained  Get features within a context.
  sample-metadata     Retreive sample metadata.
  samples             Fetch sample data.
  samples-contained   Get samples within a context.
  tags-contained      Get the observed tags within a context
[(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom search --help
Usage: redbiom search [OPTIONS] COMMAND [ARGS]...

  Feature and sample search support.

Options:
  --help  Show this message and exit.

Commands:
  features  Get samples containing features.
  metadata  Find samples or categories.
  samples   Get features present in samples.
  taxon     Find features associated with a taxon
```

```
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom select --help
Usage: redbiom select [OPTIONS] COMMAND [ARGS]...

  Select items based on metadata

Options:
  --help  Show this message and exit.

Commands:
  features-from-samples  Given samples, select the features associated.
  samples-from-metadata  Given samples, select based on metadata
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom summarize --help
Usage: redbiom summarize [OPTIONS] COMMAND [ARGS]...

  Summarize things.

Options:
  --help  Show this message and exit.

Commands:
  contexts           List names of available caches
  features           Summarize features over a metadata category.
  metadata           Get the known metadata categories and associated
                     sample...
  metadata-category  Summarize the values within a metadata category
  samples            Summarize samples over a metadata category.
  table              Summarize all features in a BIOM table.
  taxonomy           Summarize taxonomy at all levels.
```

- syntax presentation
  - long, linear, not structured

```
redbiom search metadata antibiotics | redbiom fetch samples --context <foo> --output my_table.biom
```

- can integrate with embedded cmd line functionalities
  - widely shared knowledge
  - linkage across tools
  - concise (5 steps in one line)

```
$ redbiom summarize contexts | cut -f 1,2,3 | grep 16S-v4 | grep Greengenes-illumina |  sort -k 2 -n
```

- limited by the format of cmd line
  - limited length
  - no use of fonts/colors/weights/position to construct a more functional presentation

```
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom summarize contexts | head -n 3
ContextName      SamplesWithData FeaturesWithData      Description
Pick_closed-reference_OTUs-SILVA-LS454-16S-V4-90nt-efd9ba      6067    8982    Pick closed-reference OTUs (reference-seq: |projects|qiita_data|reference|silva_119_Silva_119_rep_se
t97.fna) | Trimming (length: 90)
Deblur-Illumina-16S-V4-e40b83    1785    120797  Deblur (Reference phylogeny for SEPP: Greengenes_13.8, BIOM: reference-hit.biom) | Split libraries FASTQ
```

– lack feedback

```
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ export ctx=Pick_closed-reference_OTUs-Gree
ngenes-illumina-16S-v4-5c6506
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$
```

– unclear error message
  – format hard to read
  – little clues for potential solution

```
[(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom search metadata beer | redbiom fet]
ch samples --context $ctx --output example.biom
Traceback (most recent call last):
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/bin/redbiom", line 10, in <module>
    sys.exit(cli())
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/click/core.py",
line 764, in __call__
    return self.main(*args, **kwargs)
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/click/core.py",
line 717, in main
    rv = self.invoke(ctx)
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/click/core.py",
line 1137, in invoke
    return _process_result(sub_ctx.command.invoke(sub_ctx))
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/click/core.py",
line 1137, in invoke
    return _process_result(sub_ctx.command.invoke(sub_ctx))
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/click/core.py",
line 956, in invoke
    return ctx.invoke(self.callback, **ctx.params)
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/click/core.py",
line 555, in invoke
    return callback(*args, **kwargs)
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/redbiom/commands
/fetch.py", line 137, in fetch_samples_from_samples
    table, ambig = redbiom.fetch.data_from_samples(context, iterable)
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/redbiom/fetch.py
", line 295, in data_from_samples
    return _biom_from_samples(context, samples)
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/redbiom/fetch.py
", line 339, in _biom_from_samples
    redbiom._requests.valid(context, get)
  File "/Users/IreneYang/miniconda3/envs/qiime2-2019.7/lib/python3.6/site-packages/redbiom/_request
s.py", line 179, in valid
    raise ValueError("Unknown context: %s" % context)
ValueError: Unknown context: Pick_closed-reference_OTUs-Greengenes-illumina-16S-v4-5c6506
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$
```

```
requests.exceptions.ConnectionError: HTTPConnectionPool(host='qiita.ucsd.edu', port=7329): Max retr
ies exceeded with url: /HGETALL/state:contexts (Caused by NewConnectionError('<urllib3.connection.H
TTPConnection object at 0x10bb6fef0>: Failed to establish a new connection: [Errno 60] Operation ti
med out',))
```

– hard to know the current state
  – leave the work half-done over night

- not sure if previous data is still there
- have to go back 100 pages to know what step is next (if didn't take notes or if previous work is done by another person)
- hard to output or manipulate an intermedium file

- trouble shooting process

```
[(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom summarize contexts | cut -f 1,2,3 ]
| grep 16S-v4 | grep Greengenes-illumina |  sort -k 2 -n
[(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom summarize contexts                ]
ContextName      SamplesWithData FeaturesWithData       Description
Pick_closed-reference_OTUs-SILVA-LS454-16S-V4-90nt-efd9ba       6067    8982    Pick closed-referen
ce OTUs (reference-seq: |projects|qiita_data|reference|silva_119_Silva_119_rep_set97.fna) | Trimmin
g (length: 90)
Deblur-Illumina-16S-V4-e40b83   1785    120797  Deblur (Reference phylogeny for SEPP: Greengenes_13
.8, BIOM: reference-hit.biom) | Split libraries FASTQ
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V4-41ebc6    100     14434   Pick closed-referen
ce OTUs (reference-seq: |databases|gg|13_8|rep_set|97_otus.fasta) | Split libraries
Deblur-Illumina-16S-V4-250nt-3b87d8     105     11402   Deblur (Reference phylogeny for SEPP: Green
genes_13.8, BIOM: reference-hit.biom) | Trimming (length: 250)
Pick_closed-reference_OTUs-Greengenes-FASTA-16S-V3-100nt-a243a1 18       1706    Pick closed-referen
ce OTUs (reference-seq: |databases|gg|13_8|rep_set|97_otus.fasta) | Trimming (length: 100)
test    0       test context
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V9-100nt-a243a1      321     14922   Pick closed
-reference OTUs (reference-seq: |databases|gg|13_8|rep_set|97_otus.fasta) | Trimming (length: 100)
Pick_closed-reference_OTUs-SILVA-LS454-16S-V4-100nt-cb8fea      6067    9921    Pick closed-referen
ce OTUs (reference-seq: |projects|qiita_data|reference|silva_119_Silva_119_rep_set97.fna) | Trimmin
g (length: 100)
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V3-90nt-44feac       1135    14850   Pick closed
-reference OTUs (reference-seq: |databases|gg|13_8|rep_set|97_otus.fasta) | Trimming (length: 90)
Pick_closed-reference_OTUs-SILVA-not provided-16S-V3-54d83f     138     1014    Pick closed-referen
ce OTUs (reference-seq: |projects|qiita_data|reference|silva_119_Silva_119_rep_set97.fna) | Split l
ibraries FASTQ
Deblur-Illumina-16S-V9-100nt-fbc5b2     319     36485   Deblur (Reference phylogeny for SEPP: Green
genes_13.8, BIOM: reference-hit.biom) | Trimming (length: 100)
```

```
[(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom summarize conte]
xts | cut -f 1,2,3 | grep 16S-V4 | grep Greengenes-Illumina |  sort -k 2 -n
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V4-250nt-66a626          77      3
21
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V4-41ebc6        100     14434
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V4-125nt-65468f          16622   4
0899
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V4-150nt-bd7d4d          132702 7
1512
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V4-90nt-44feac           154430 7
3911
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V4-100nt-a243a1          154617 7
5633
Pick_closed-reference_OTUs-Greengenes-Illumina-16S-V4-5c6506        178259 83915
```

- weird behavior and conditional syntax not really making sense

```
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom search metadata
 "soil & europe where ph < 7" | wc -l
5521
      5818
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ 5521
-bash: 5521: command not found
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$
```

- format issue again

```
(qiime2-2019.7) Xuhans-MacBook-Pro-2:redbiome IreneYang$ redbiom search metadata
 "soil & europe where ph < 7" | redbiom summarize samples --category sample_type
  | head
XXQIITAXX        1489
soil     1186
fresh water      724
water    722
Soil     595
cheese   435
peat     192
biofilm 139
wetland soil     78
belly    41
```