

# 16S AMPLICON DATA REPORTING FORMAT AND KBASE INTEGRATION

**PAMELA WEISENHORN**

Assistant Computational Ecologist

Biosciences

Argonne National Lab



Argonne National Laboratory is a  
U.S. Department of Energy laboratory  
managed by UChicago Argonne, LLC.

May 25, 2021  
ESS-DIVE Community Workshop

# COORDINATED BETWEEN PROJECTS



# OVERVIEW

- Details of reporting format – ***Seeking community feedback***
- Application of the reporting format in KBase

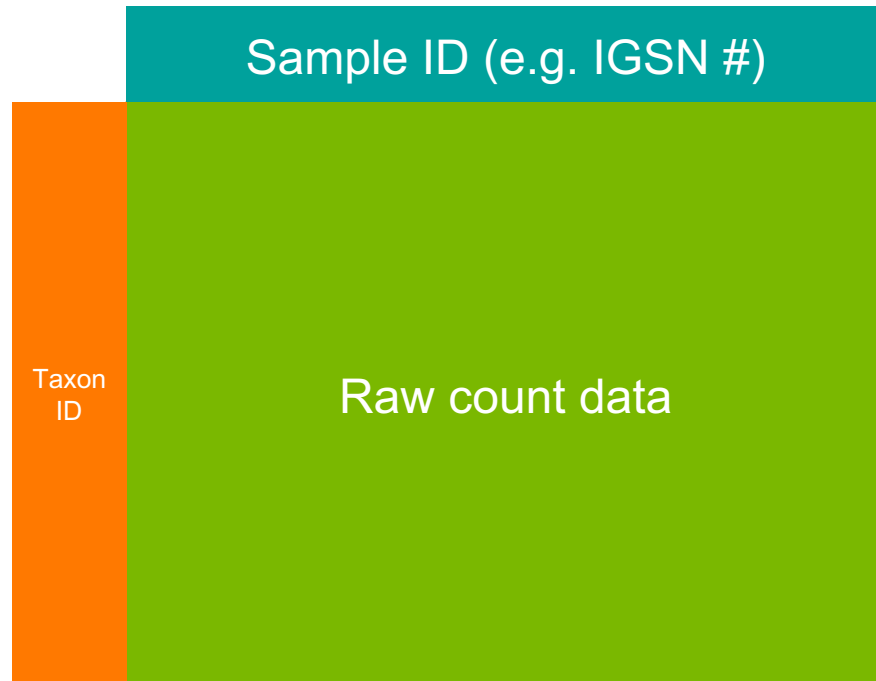
# AMPLICON REPORTING FORMAT

# MULTIPLE FILES

- Amplicon / Taxon Table (tsv)
- Consensus sequences (fa)
- Key Metadata (tsv)
- Taxonomy (tsv) – optional

# AMPLICON / TAXON TABLE

- Tab-delimited text file
- Rows are taxon IDs
- Columns are sample IDs
- ***Raw count data***



# EXAMPLE TEMPLATE

	b1rfm3	c1rsm1	b1rfm1	c2rfm3	c1rfm3	c1rpw1	b1rbm3	c1rbm2	c1rpw3	c1rbm4
OTU_253;size=251;	88	0	46	1	1	0	0	0	0	0
OTU_11;size=6816;	0	211	0	0	0	17	115	93	0	152
OTU_300;size=264;	26	0	95	24	4	9	0	0	0	0
OTU_2078;size=44;	5	0	2	13	1	0	0	0	1	0
OTU_1807;size=54;	7	0	3	1	4	4	0	0	0	0
OTU_35;size=1487;	7	0	18	5	62	827	78	14	0	0
OTU_218;size=1013;	49	4	19	78	82	6	119	17	3	9
OTU_1275;size=151;	0	0	0	0	0	1	14	11	0	5
OTU_2903;size=75;	0	0	1	2	2	1	0	0	5	0
OTU_2999;size=71;	0	0	0	0	0	0	8	4	0	9

# CONSENSUS SEQUENCES

- Names must match Taxon IDs in Taxon Abundance table
- One sequence per Taxon ID
- Link between assertions of taxonomic identity and sequence data
- FASTA file

```
>OTU_1;size=19095;
```

```
tacagaggtctcaagcgttggtcggattcattgggcgta  
aagggtgcgtaggcggcgcggaagtcggatgtga  
aatctcggagcttaactccgaaattgcattcgatactg  
ccgtgcttgaggactggagaggagactggaatttacg  
gtgtagcggtgaaatgcgtagatatcgtaaggaa
```

```
>OTU_2;size=7025;
```

```
gacggaggatgcaagtgttatccggaatcactgggc  
gtaaagcgtctgtaggtggtttaataagtcaactgttaa  
atcttgaggcttaacttcaaaatcgagtcgaaactatt  
agactagagtatagtaggggttaaagggaatttccagt  
ggagcggtgaaatgcgtagatatattggaaa
```



# REQUIRED METADATA

## SEQUENCING

- Target gene (required)
- Target region (required)
- Forward primer (required)
- Reverse primer (required)
- Sequencing platform (required)

## PROCESSING

- Quality filtering cutoff (required)
- Clustering cutoff (required)

# OPTIONAL OR OMITTED METADATA

## SEQUENCING

- Extraction kit or protocol (optional)
- Sequencing run (optional)
- Sequencing kit (optional)
- Sequencing center
- Chemistry version

## PROCESSING

- Clustering method (optional)
- Denoising method (optional)
- Sequence depth cutoff for samples
- Barcode error threshold

Note: Data in taxon table should be ***raw counts***

# TAXONOMY

- Taxonomic string
- Taxonomic database
- Version of database
- Assignment algorithm

# KBASE WALK-THROUGH

<https://appdev.kbase.us/narrative/ws.57083.obj.19>



Argonne National Laboratory is a  
U.S. Department of Energy laboratory  
managed by UChicago Argonne, LLC.

