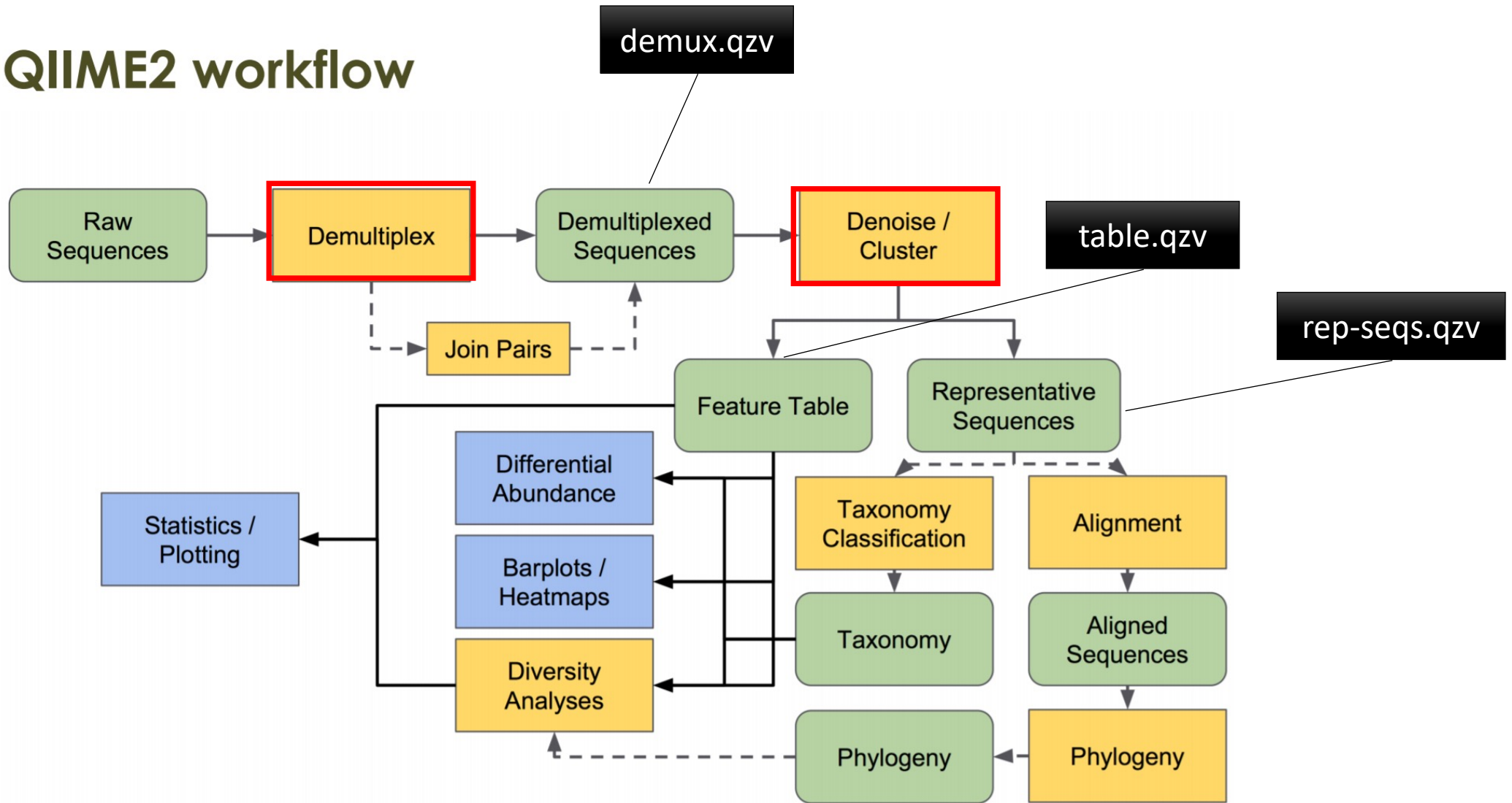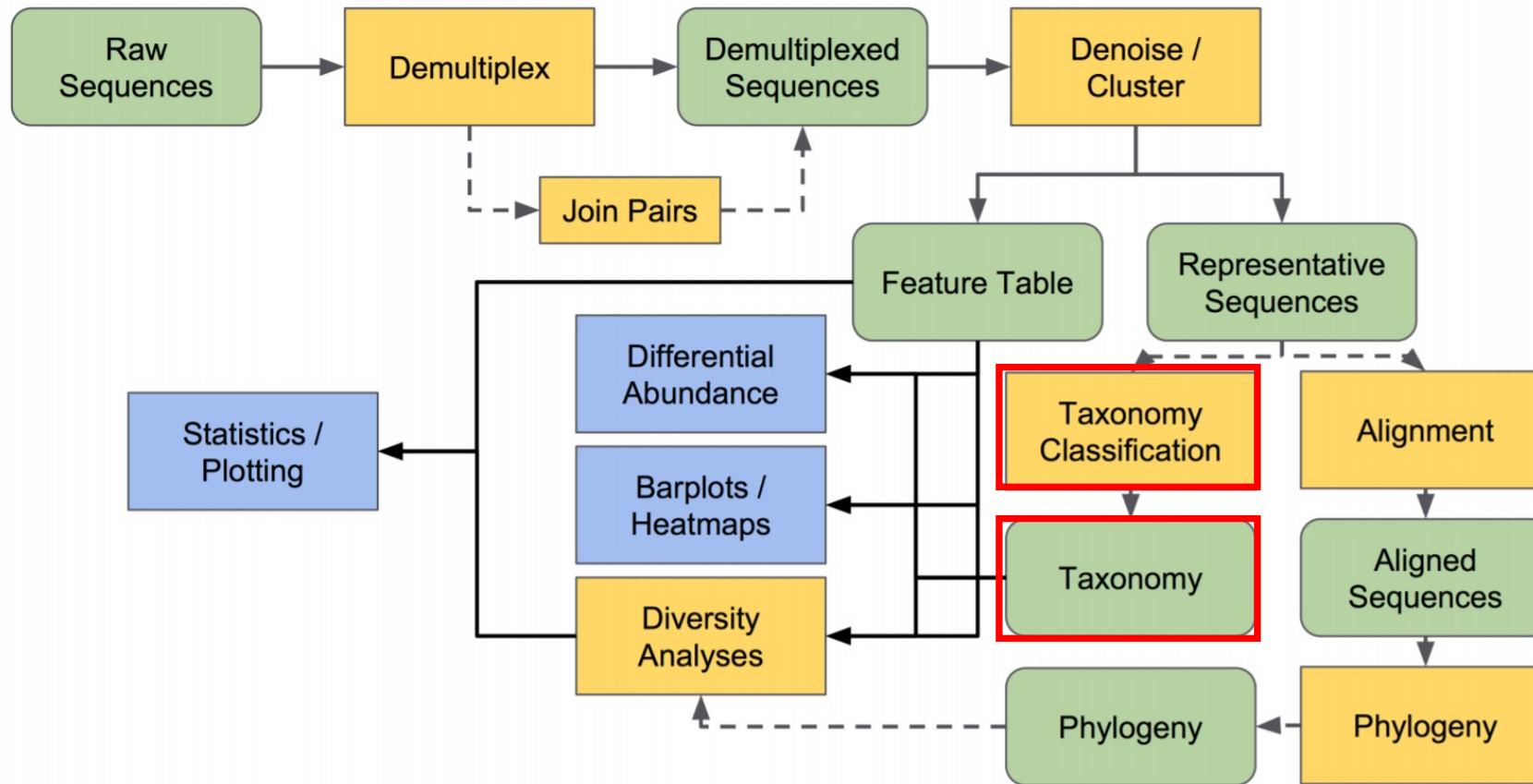# Module 6

Taxonomic analysis and data filtering

Module Outcomes

1. Train a classifier to do taxonomic analysis

2. Generate and interpret a taxa bar graph

3. Filter your ASV table based on taxonomy, frequency, or metadata
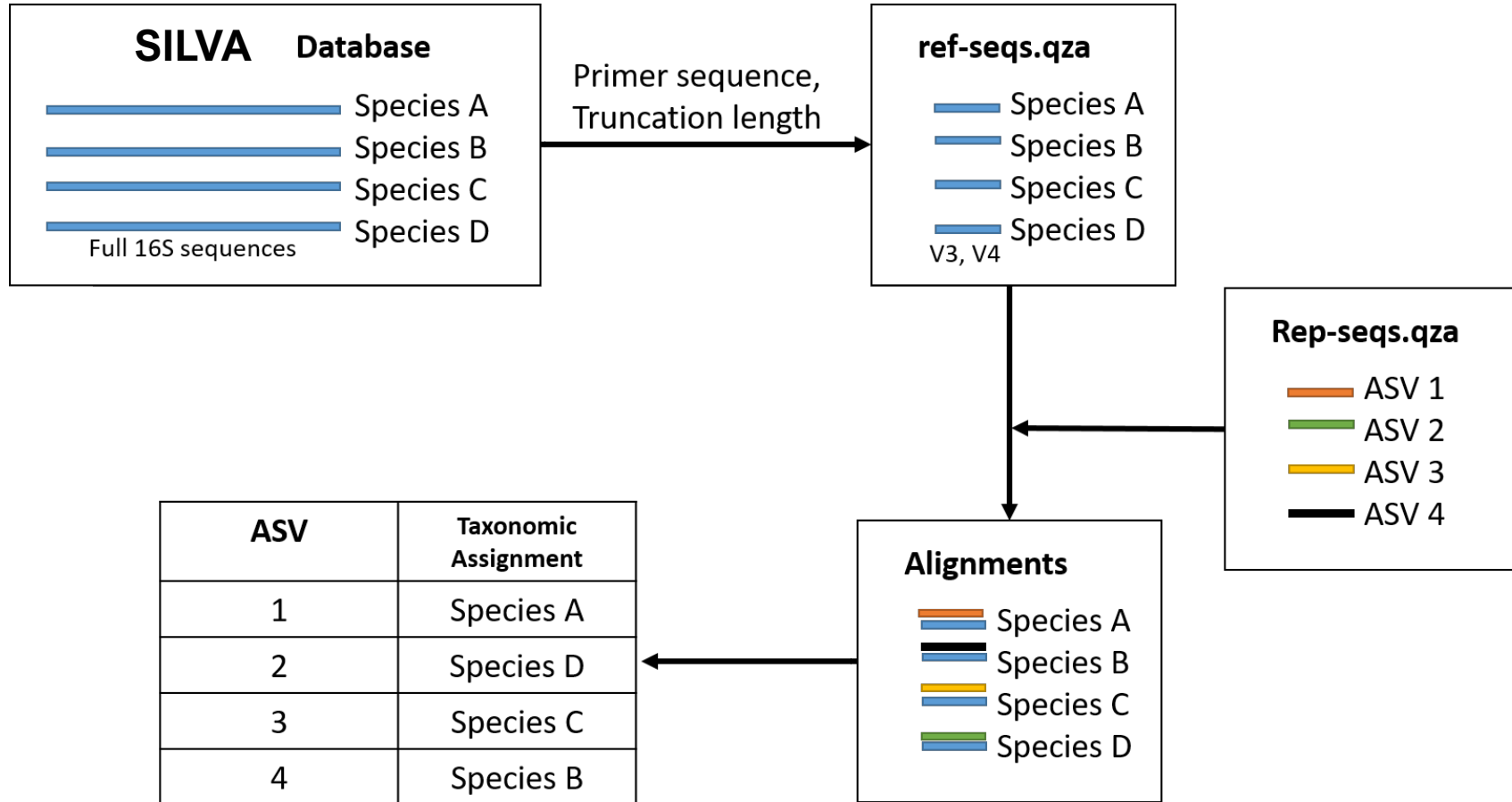
QIIME2 workflow

https://docs.qiime2.org/2018.8/tutorials/overview/#demultiplexing

QIIME2 workflow

https://docs.qiime2.org/2018.8/tutorials/overview/#demultiplexing

Training a classifier for Taxonomic analysis (visual)

Why do this instead of aligning to the whole 16S gene?

- Better alignments

- Assign more accurate taxonomic information
  - Used for diversity metrics

What do you need?

```
qiime feature-classifier extract-reads \
    --i-sequences ref-otus.qza \
    --p-f-primer GTGCCAGCMGCCGCGGTAA \
    --p-r-primer GGACTACHVGGGTWTCTAAT \
    --p-trunc-len 150 \
    --o-reads ref-seqs.qza\
```

Your imported database

Primer sequence

Truncation length from DADA2/Deblur step

https://docs.qiime2.org/2020.8/tutorials/feature-classifier/

Dissecting the code

```
qiime feature-classifier classify-sklearn \
    --i-classifier ref-seqs.qza \
    --i-reads rep-seqs.qza \
    --o-classification taxonomy.qza
```

Dissecting the code

```
qiime metadata tabulate \
   --m-input-file taxonomy.qza \
   --o-visualization taxonomy.qzv


qiime taxa barplot \
   --i-table table.qza \
   --i-taxonomy taxonomy.qza \
   --m-metadata-file metadata.txt \
   --o-visualization taxa-bar-plots.qzv
```

Taxonomy-based filtering

- Removing mitochondria or chloroplast sequences

```
qiime taxa filter-table \
--i-table table.qza \
--i-taxonomy taxonomy.qza \
--p-exclude mitochondria,chloroplast \
--o-filtered-table table-no-mitochondria-no-chloroplast.qza
```

Feature-based filtering

- Remove rare ASVs (<0.005 or 0.01% of total reads)
  - Unlikely true biological variants
  - Often not biologically informative (represented in too few samples)
  - Negatively affects downstream analyses
- Moving pictures tutorial: Total reads153,884 × 0.00005 ≈ 8 reads

```
qiime feature-table filter-
features \
--i-table <input table> \
--p-<filter metric> <filter
value> \
--o-filtered-table
<filtered table name>
```

```
qiime feature-table filter-
features \
--i-table table.qza \
--p-min-frequency 8 \
--o-filtered-table feature-
frequency-filtered-
table.qza
```

Metadata-based filtering ⚠️

```
qiime feature-table filter-samples \
--i-table <input table> \
--m-metadata-file <metadata file> \
--p-where "<filter condition>" \
-o-filtered-table <filtered table>
```
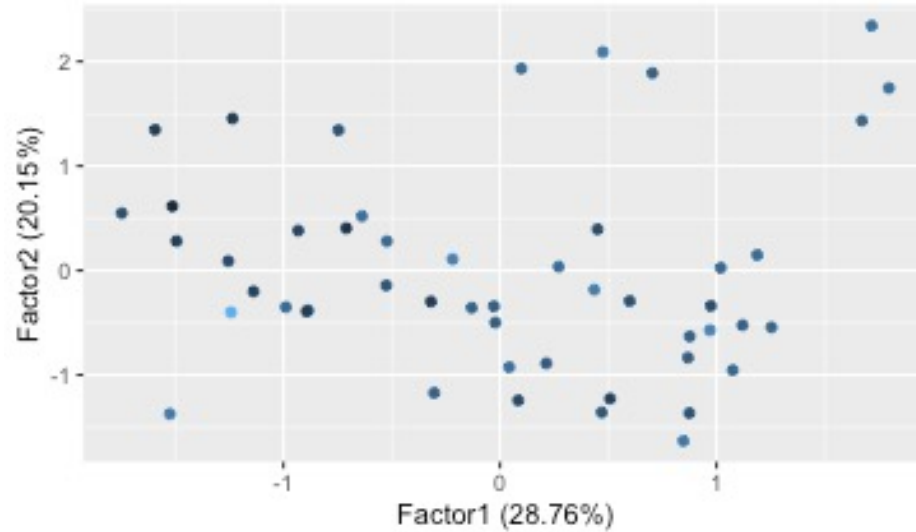
- Example:
  ```
  qiime feature-table filter-samples \
  --i-table table.qza \
  --m-metadata-file /mnt/datasets/moving_pictures/sample-metadata.tsv \
  --p-where "[subject]='subject-1'" \
  --o-filtered-table subject-1-filtered-table.qza
  ```
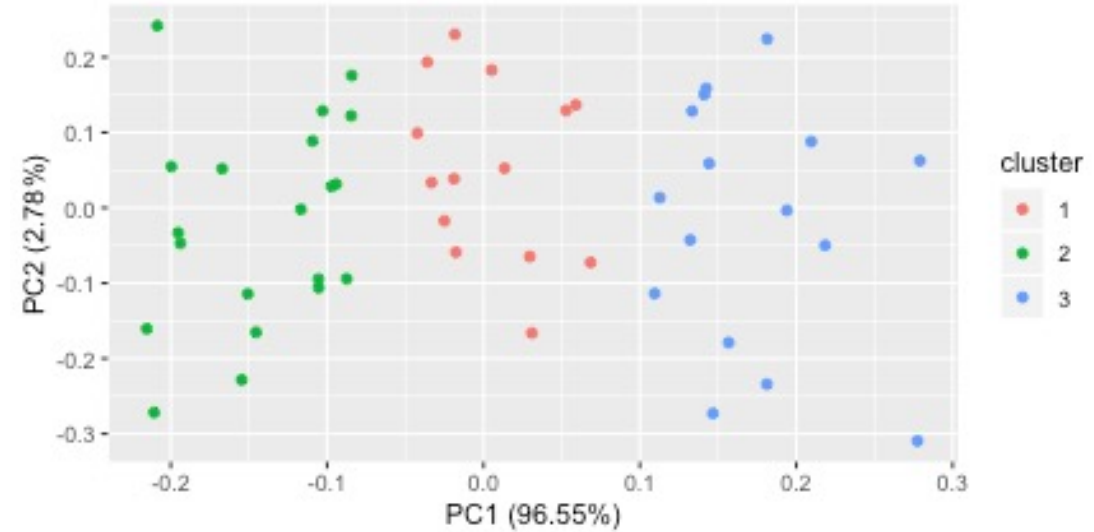 - column name in brackets.
  - `=` checks for matches.
  - Filter value in single quotes.

Why is it important to filter based on metadata?

Before filtering

Qiime2 Tutorial on Filtering Data

- https://docs.qiime2.org/2020.11/tutorials/filtering/