**Processing Pipeline for Parrot Fecal Microbiome Data**

**1. Set Up Environment**

* Load R libraries (e.g., tidyverse, phyloseq, vegan, etc.)
* Set up directories

**2. Load Metadata**

* Read main metadata (ParrotFecalSampleMetadata\_SerialIDs\_McKenzie.csv)
* Join with additional metadata (additional\_info\_seizures.csv)
* Format Sample IDs
* Remove problematic samples (S082895, S082806)

**3. Load Raw Sequence Data**

* Read 16S & 18S sequencing data: ESVs, reads, and sample data
* Separate reads by sequencing run (MiSeq and Nanopore)
* Clean column names (capitalization, suffixes)

**4. Filter Read Data**

* Remove samples with suffix .0 (zero reads)
* Remove problematic samples across datasets
* Separate read matrices by run and region:
  + otu16m\_df, otu16n\_df (16S MiSeq, Nanopore)
  + otu18m\_df, otu18n\_df (18S MiSeq, Nanopore + third run merge)
* Remove ESVs with total row sum = 0

**5. Standardize Sample Names**

* Remove suffixes from sample column names (e.g., .1, .2, .3)

**6. Convert to OTU Matrices**

* Convert filtered data frames to OTU matrices:
  + otu16m\_mat, otu16n\_mat, otu18m\_mat, otu18n\_mat

**7. Update Metadata**

* Calculate read depths for each OTU matrix
* Join read depth to metadata
* Format and standardize collection dates
* Calculate TimeSinceSeizure

**8. Generate Taxonomy Tables**

* Generate taxonomic assignments from reads\_16 and reads\_18
* Create full ScientificName fallback hierarchy
* Remove unclassified or unwanted taxa:
  + 16S: Remove Chloroplasts & Mitochondria
  + 18S: Remove Bacteria, Archaea, Host DNA, Birds

**9. Filter OTU Matrices**

* Remove unwanted ESVs from OTU matrices using taxonomy
* Create filtered OTU matrices:
  + \*\_mat\_filt, \*\_nohost\_mat\_filt

**10. Recalculate Filtered Read Depth**

* Recalculate and join filtered read depths into meta\_adj\_filt

**11. Coverage QC Analysis**

* Define analyze\_coverage() function
* Generate histograms, ECDFs, and summary stats for:
  + Raw and filtered OTU matrices
* Export to PDF: coverage\_plots\_no\_parrot.pdf

**12. Filter Based on Read Depth Thresholds**

* For each matrix, generate:
  + \_20 (≥20 reads)
  + \_500 (≥500 reads)
  + \_outlier (remove statistical outliers)
  + \_max\_9500 (≤9500 reads)

**13. Filter Based on OTU Prevalence**

* Remove ESVs not observed ≥2 times in:
  + ≥10% of samples (\_2count\_thresh\_10)
  + ≥30% of samples (optional, shown in comments)

**14. Generate Final Phyloseq Objects**

* Loop through all filtered OTU tables
* Subset metadata and taxonomy
* Create phyloseq objects using:
  + otu\_table, tax\_table, and sample\_data
* Match sequences using reads\_\*