Meeting #3 Agenda
Hayley to take minutes

1. Confirm that we are working with variable region 4 (for training classifier).
   a. We read in a UJEMi paper which used V4, and proceeded through the QIIME pipeline assuming V4 was sequenced.
2. At the data filtering step in the QIIME2 pipeline, what metadata categories should we filter for?
   a. We just filtered out mitochondrial and chloroplast DNA.
   b. We were going to filter out the infants that aren't at 4 months (to pick one time point for analysis and to ensure we have enough samples and they are not confounded by birth mode as much as the earlier time points), but we couldn't figure out which column in the metadata to use.
   c. Is there anything else we need to filter for? (ex: birth mode?)
   d. Is this where we filter out to control for confounding variables, or is that done at a later step?
3. How are we planning on controlling for confounding variables and at what stage of the analysis is this done?
   a. Birth mode, age, etc.
   b. Is this done later in the analysis or initially?
4. How do we use code to associate BMI, mode of feeding, and infant microbial diversity all together?
   a. Are we generating plots to answer the question?
   b. How do we associate three variables together in QIIME and R?
5. Review of QIIME2 pipeline results
   a. Review of resulting plots
   b. Do we need to apply extra filters to our data (more than filtering out mitochondrial and chloroplast sequences?)
6. Confirm sampling depth and rarefaction curve
   a. Mom bmi as metadata column? (doesn't seem to have barcode or other sample ID but table shows 1 sample per BMI for the most part)
   b. Sampling depth ~14000? (jump from 3717 to 14097)

Meeting Minutes: 2023-10-11

Training Classifier: Variable Region 4 probably right

Filtering:
- Just 4 months: remove other ages (don't need mom's sequences: just BMI in metadata)
  - Month = 4 in R
- Don't filter birth mode
- Make sure all babies have mom BMI for each sample (do in R)

Age category = age
Host Subject ID groups mom & baby (CH in table)

Proposal
- Justify controls (don't need to say how)
    - Get stats regardless, just look at stuff of interest
- Say which tests in R (not just "use R")
- Redundancy is fine
- Specify what's done in qiime and what's done in R (once per column where switch happens)
- Specify which diversity metrics (don't need >1 alpha)
    - Numerical, continuous
- Specify stats (for alpha and beta)
    - Shannon has 2 numericals; spearman correlations work (continuous data)
    - Permanova for beta diversity
    - Dseq for pie crust
- Beta diversity for BMI should work in R
- Using linear models
- Background
    - Do literature review definitely within UJEMI but also within the broader literature
    - Specifically mention knowledge gap you are trying to fill
    - Need other UJEMI papers, BMI effect on microbiome by delivery method
    - Highlight what hasn't been done
- Aims
    - Broad to specific
    - Logical flow
    - Justification for BMI in general
    - Why controlling for method of delivery
- Email by Thursday before deadline if needed

Notebook
- Include troubleshooting
- Streamlined important code
- Run again to make sure if works (for UJEMI+)

Presenting results
- One plot per mode of feeding
- Can overlay slopes to visualize difference
- Compare side by side or try merging together (interaction)
    - Harder to get significant results for interaction but good to try
- Try pie crust (relevant, meaningful)
    - Creates dataset; apply stats to it
- Dseq - differential abundance (stat tool)
    - Basically a t test

- Any stat test (have to interpret if difference is meaningful)
- Just choose one for proposal
- Look at differences in general
- Keep experimental aims broad
- Make sure we have mom **and** baby data at 4 months
- Histogram distribution of mom bmis to decide if we should categorize them into low med high or keep continuous. Perhaps remove outliers?
    - If data looks like normal distribution, consider keeping continuous
    - Stats get tricky for >2 groups
    - In final proposal, we could divide into three bmi groups just for the purposes of the graph but all the analyses will have been run with bmi being continuous
- Babies will have less diverse biomes
- Make column for mom bmi for baby

Rarefaction:
- 14097 is good
- Mom BMI fine to use for proposal