**Objective:**To determine if the teat apex microbiome is different between cows that have mastitis (“cases”) versus those that don’t (“controls”). NOTE: our microbiome data was generated from shotgun metagenomic data using kraken2

**Study Design:**

**Animals:**Cows were sampled on 5 different dairy farms. Their teat apices were sampled weekly for several weeks prior to calving. At calving, their milk was collected and tested for mastitis to determine if they were a case or control (infected or un-infected).

**Samples:**839 samples collected by using gauze to swab the teat apices of dairy cows. Samples were extracted using PowerSoil Pro kit, and then subjected to shotgun metagenomic library preparation and sequenced on the NovaSeq 6000.

**Analysis:**Shotgun metagenomic data were run through kraken2 within the AMR++ pipeline.

**Tasks:**

1. Load AMR++ kraken2 output files and create a phyloseq object for analysis.
2. Summarize basic QC metrics:
   * Total raw reads, mean [range], median
   * Total trimmed reads, mean [range], median
   * Total nonhost reads, mean [range], median

Run statistical models to check for differences in average raw, trimmed and nonhost reads between cases and controls.

1. Create a stacked bar chart showing the proportion of host and non-host reads across all samples.
2. Calculate the total number of unique OTUs, species, genera, and phyla across all samples.
3. Calculate alpha diversity (richness, Shannon diversity) at the genus level, and create box plots to compare cases and controls.
   * Run statistical models to test if these values differ between cases and controls.
4. Assess if the microbiome composition (at the genus level) differs between cases and controls
5. Perform statistical tests to determine microbiome composition (at the genus level) is different between cases and controls.
   * Determine how much variation is partitioned to whether an animal was a case or control
   * Determine how much variation is partitioned to the farm where the animal lived
   * Determine how much variation is partitioned to batch effects
6. Plot the relative abundance of the microbiome at the phylum level, separated by cases and controls.
7. Conduct differential abundance analysis at the genus level, and identify genera that are significantly different between cases and controls.
8. Summarize key findings.