**Summary of procedures to select samples for AMKAJ microbiome analyses**

1. Select observations: Exclude those with low sequence reads
   1. 45 Meatal samples – exclude 2 with total sequence read <10,000; 43 for analysis
2. Delete genus level taxa with zero sequence reads
3. Filter taxa

**Step 1: Select observations for analysis**

Figures 1 and 2. Total sequence reads per observation for 45 meatal samples.



Low sequence reads identified below and to be excluded from analyses, leaving 43 meatal samples.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mehta1042.209960 | 209960 | 31-Aug-17 | AMKAJ-MEA | 6021 |
| Mehta1038.209632 | 209632 | 29-Aug-17 | AMKAJ-MEA | 877 |

**Step 2: Delete taxa with zero reads**

1. For MEA

* 2,209 genus level taxa from the larger dataset. 1,609 (72.8%) with 0 sequence reads

**Step 3: Filter taxa**

* Among the 600 genus level taxa detected (i.e., at least one sequence read), there are a total of 1,758,443 sequence reads across the 43 observations.
* 0.1% = 1,758 sequence reads across all observations = 44 taxa
* Ranking bacteria by total sequence reads across observations
  + Mycoplasma – 30th
  + Chlamydia – 189th
  + Neisseria – 105th
  + Treponema – 198th
* Will conduct statistical analyses with top 45 genus level taxa that account for at least 0.1% of taxa across all observations and add in the putative pathogens that aren’t in the top 45 (i.e., Chlamydia, Neisseria, Treponema)
* Excluding the 553 genus level taxa (92.2% of all genus taxa with at least one sequence read) that don’t achieve 0.1% (i.e., 1,758 sequence reads across all observations) [while including chlamydia, Neisseria, and treponema], excludes 48,716 sequence reads (2.77% of total sequence reads)