

**Supplementary Figure S4:** Overview of the bioinformatics pipeline and the total number of sequences remaining after the different cleaning and processing steps. Sequences were considered “rare” if they occurred less than 5 times in the entire dataset.

