## Appendix S2

## A Poisson-Binomial mixture model to account for overdispersion in microbiome sampling

In testing representation our goal was to evaluate if the observed number of 16S rRNA reads per species, (one of eleven), within a sample are similar to those expected while accounting for experimental variation between replicates due to 454 pyrosequencing. To do so we used a likelihood ratio test, accounting for overdispersion using a Poisson-Binomial mixture model. This likelihood ratio test assumes two models: a null simple model and an alternative more complex model. The null model assumes that the observed reads, , per species within replicate , regardless of the experimenter and the day, have an underlying multinomial distribution with parameters equal to the expected relative abundances, . To account for overdispersion we assumed that the total number of reads, , per replicate to be Poisson distributed with mean λ. The likelihood of the data given this simple model is,

The likelihood of the alternative model is similar to the null except that it assumes that the replicates have a multinomial distribution different than what is expected and that the parameters of this model are estimated based on the observed data and are equal to . This alternative model still includes the Poisson model to account for sample variation. The likelihood ratio test we implemented compares the null and the alternative models using a strategy equivalent to that presented by Schütte et al. 2009 [61] and based on the bootstrap.