## 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, $j$ (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, $Y\_{ij}$, per species $j$ within replicate $i$, regardless of the experimenter and the day, have an underlying multinomial distribution with parameters equal to the expected relative abundances, $p\_{1},p\_{2},…,p\_{11}$. To account for overdispersion we assumed that the total number of reads, $N\_{i}$, per replicate $i$ to be Poisson distributed with mean λ. The likelihood of the data given this simple model is,

$$P\left(\left\{p\_{i}\right\}, λ\right)= P\left(\left\{p\_{i}\right\}, N\_{i}\right)P\left(λ\right)= \prod\_{i}^{}\frac{λ^{N\_{i}}e^{λ}}{N\_{i}!}\left(\frac{N\_{i}!}{\prod\_{j}^{}Y\_{ij}!}\right)\prod\_{j}^{}p\_{j}^{Y\_{ij}}$$

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 $\hat{p}\_{i}$ are estimated based on the observed data and are equal to $\frac{\sum\_{i}^{}Y\_{ij}}{\sum\_{i}^{}N\_{i}}$. 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.