Supplementary Appendix:
Improved detection of changes in species richness in high-diversity microbial communities

Amy Willis, John Bunge & Thea Whitman

Covariate structure & Type 1 error rate

To validate the claim that the rare species structure (high versus medium latent diversity) explains the higher Type 1 error rate for the Dethlefsen *et al.* (2008) dataset in comparison to the Whitman *et al.* (2016) dataset (Table 1), we investigated the effects of switching the continuous and discrete variables (Table 5). Under the reversal of the covariate structures, higher Type 1 error rates were again observed for the Dethlefsen *et al.* (2008) dataset compared to the Whitman *et al.* (2016) dataset. This supports our claim that the medium diversity structure drives the higher error rate because the gains in accounting for latent diversity are lower due to lower latent diversity.

It is worth noting that in the high diversity case (Whitman *et al.*, 2016), the improvements due to `betta` are even greater than previously, and remain consistently conservative. In contrast, it appears that accounting for error in estimation increases the Type 1 error probability for the Dethlefsen *et al.* (2008) dataset. However, it is important to note that the power (complement of Type 2 error rate) is far higher in this case, with power up to 18-fold greater in this case (see Section 3.2 and Table 2 of the manuscript).

Table 5: Empirical Type I error rates for significance levels of $\alpha = (0.01, 0.05, 0.10)$ for the true null hypothesis of $\beta_1 = 0$ for 5,000 homogeneous redraws from the Whitman *et al.* (2016) dataset, and 5,000 homogeneous redraws from the Dethlefsen *et al.* (2008) dataset, each partitioned into samples of 10 and 20 replicates.

Dataset	Model	Covariate	$n = 10$	$n = 20$
Whitman *et al.*	`betta (breakaway)`	Continuous	(0.014, 0.026, 0.038)	(0.004, 0.016, 0.024)
Whitman *et al.*	Regression on c	Continuous	(0.006, 0.048, 0.093)	(0.013, 0.051, 0.103)
Dethlefsen *et al.*	`betta (CatchAll)`	Discrete	(0.020, 0.072, 0.110)	(0.036, 0.069, 0.113)
Dethlefsen *et al.*	Regression on c	Discrete	(0.018, 0.058, 0.102)	(0.012, 0.048, 0.108)
Whitman *et al.*	`betta (breakaway)`	Discrete	(0.000, 0.002, 0.002)	(0.000, 0.004, 0.012)
Whitman *et al.*	Regression on c	Discrete	(0.002, 0.040, 0.094)	(0.012, 0.052, 0.088)
Dethlefsen *et al.*	`betta (CatchAll)`	Continuous	(0.108, 0.176, 0.212)	(0.084, 0.148, 0.188)
Dethlefsen *et al.*	Regression on c	Continuous	(0.012, 0.068, 0.118)	(0.012, 0.056, 0.116)

References

Dethlefsen, L., Huse, S., Sogin, M. L., and Relman, D. A. (2008). The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16s rRNA sequencing. *PLoS Biology*, **6**(11).

Whitman, T., Pepe-Ranney, C., Enders, A., Koechli, C., Campbell, A., Buckley, D. H., and Lehmann, J. (2016). Dynamics of microbial community composition and soil organic carbon mineralization in soil following addition of pyrogenic and fresh organic matter. *The ISME Journal*.