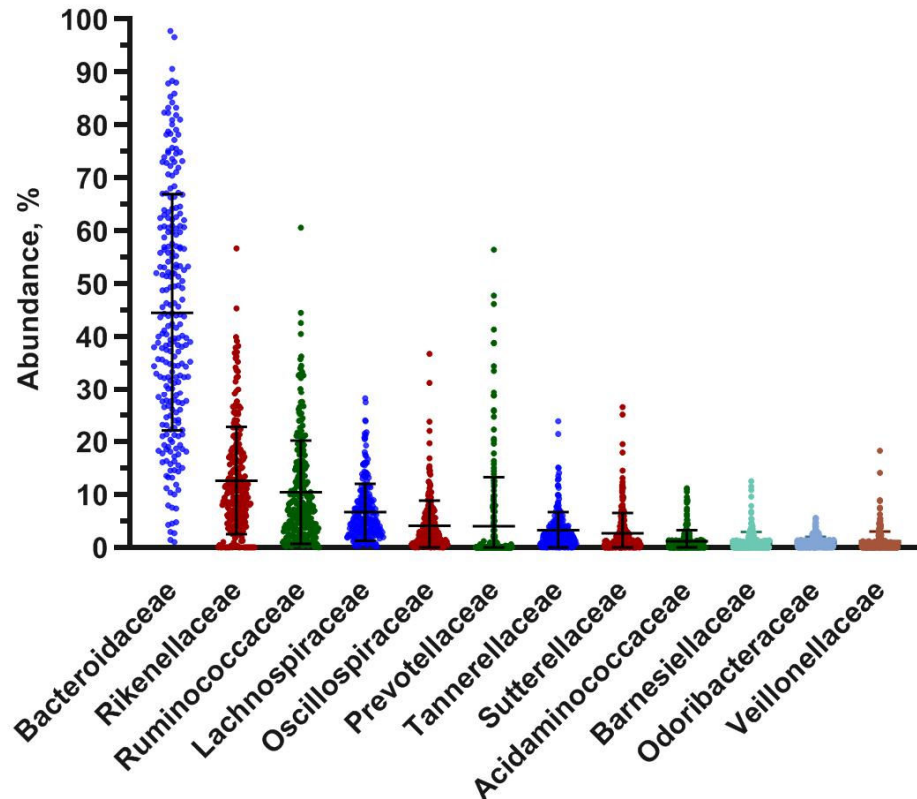
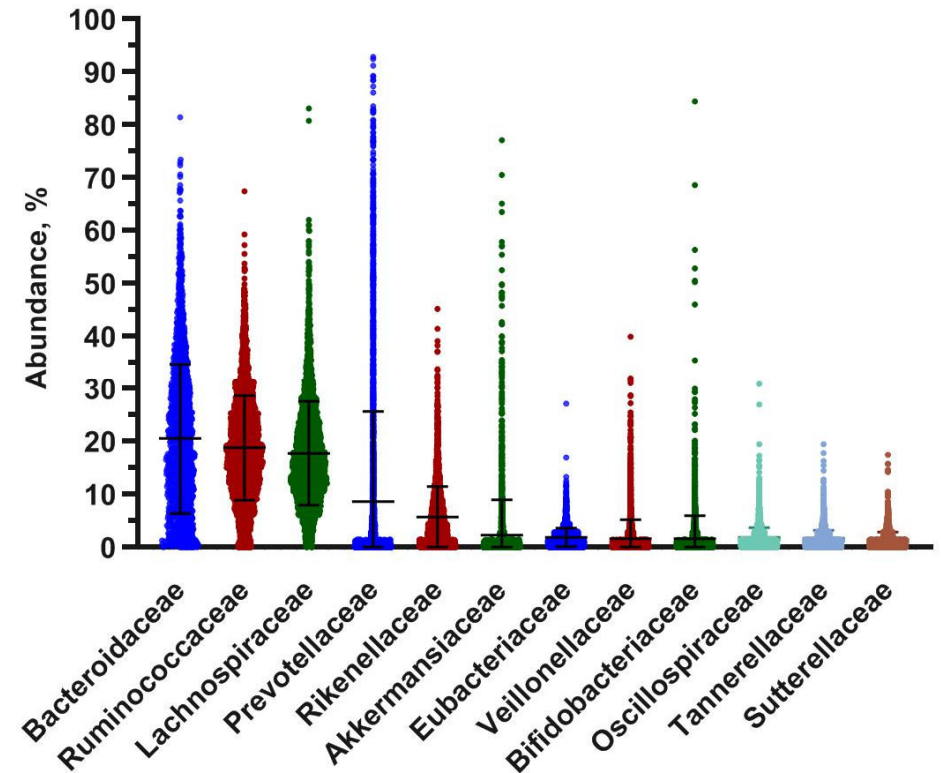


## HMP (245 samples)



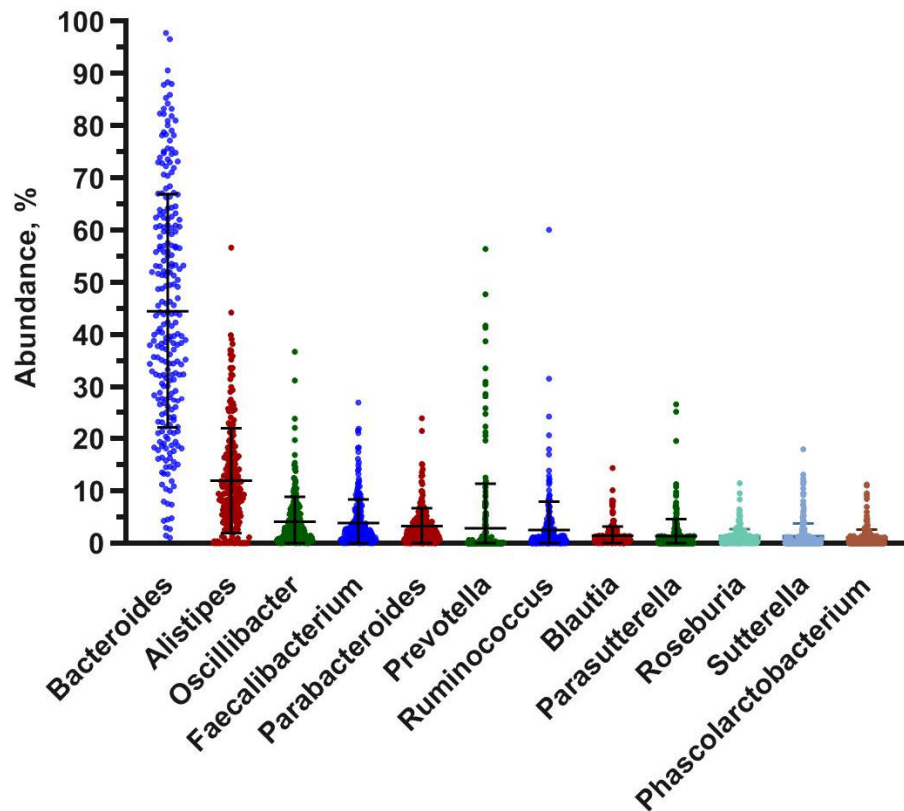
## AGP (2863 samples)



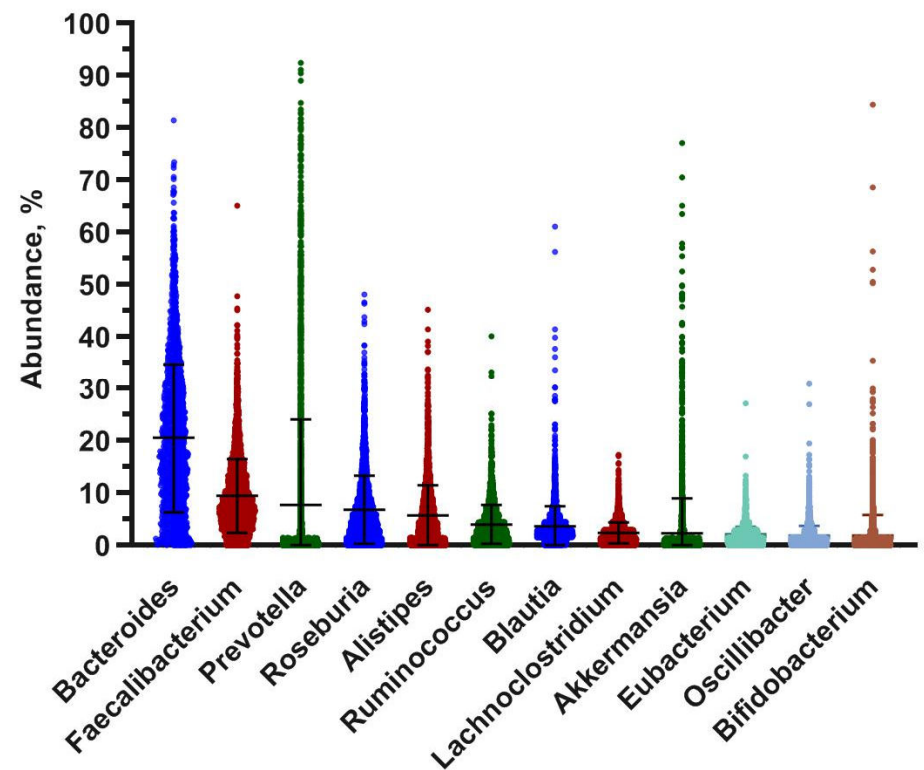
**Figure S4. (A) Abundance of major taxonomic families in HMP and AGP datasets.**

The obtained taxonomic profiles for HMP and AGP datasets differ substantially in the abundance for top contributing families. The HMP dataset is characterized by elevated average abundance of two families, the Bacteroidaceae (*Bacteroides* spp.) and Rikenellaceae (*Alistipes* spp.). In contrast, three major contributing families in the AGP dataset are Bacteroidaceae, Lachnospiraceae (*Blautia*, *Roseburia*, *Lachnoclostridium* spp.) and Ruminococcaceae (*Faecalibacterium*, *Ruminococcus* spp.).

## HMP (245 samples)



## AGP (2863 samples)



**Figure S4. (B) Abundance of major taxonomic genera in HMP and AGP datasets.**

The obtained taxonomic profiles for HMP and AGP datasets differ substantially in the abundance for top contributing genera. The HMP dataset is characterized by elevated average abundance of two genera, the Bacteroides and Alistipes. In contrast, three major contributing genera in the AGP dataset are Bacteroides, Prevotella, Blautia, Roseburia, Lachnospirillum, Faecalibacterium, and Ruminococcus.

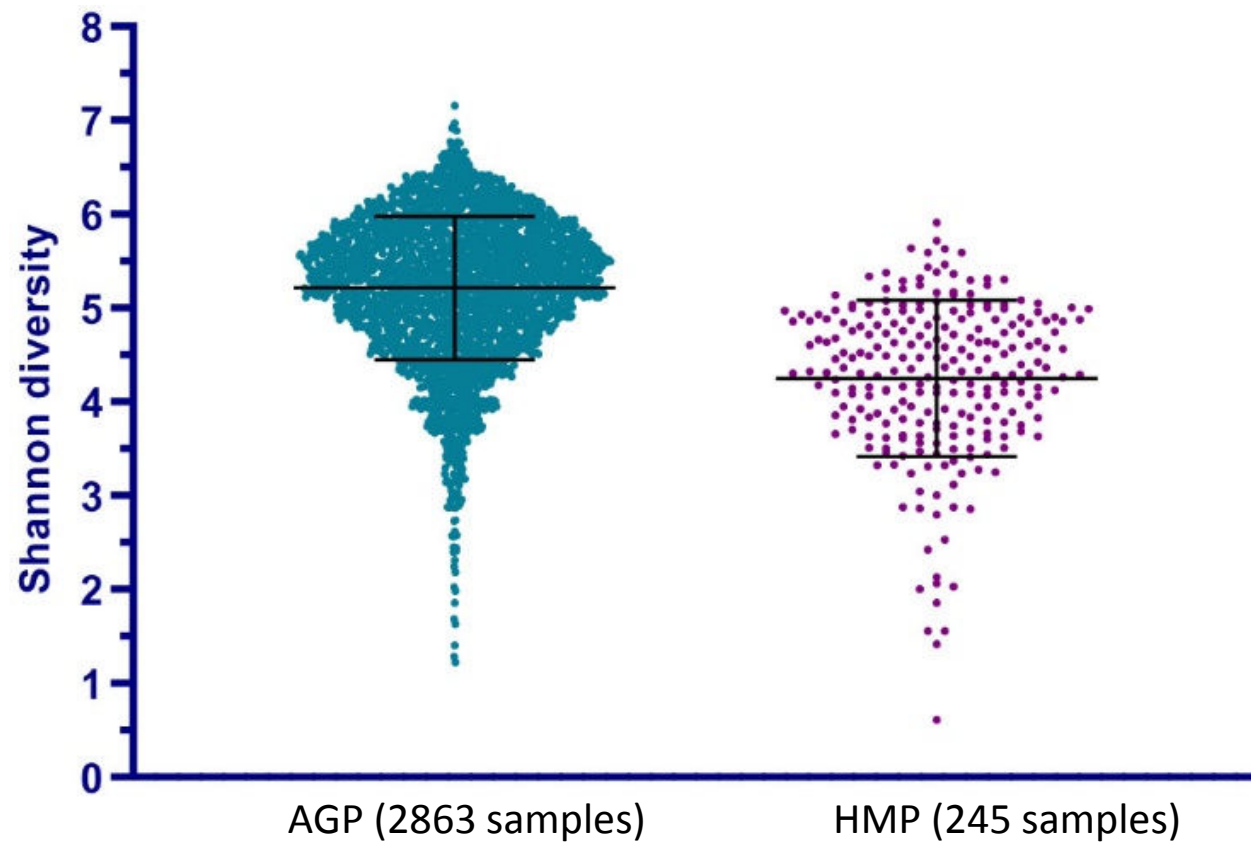


Figure S4. (C) Distribution of the Shannon diversity index in HMP and AGP datasets.