




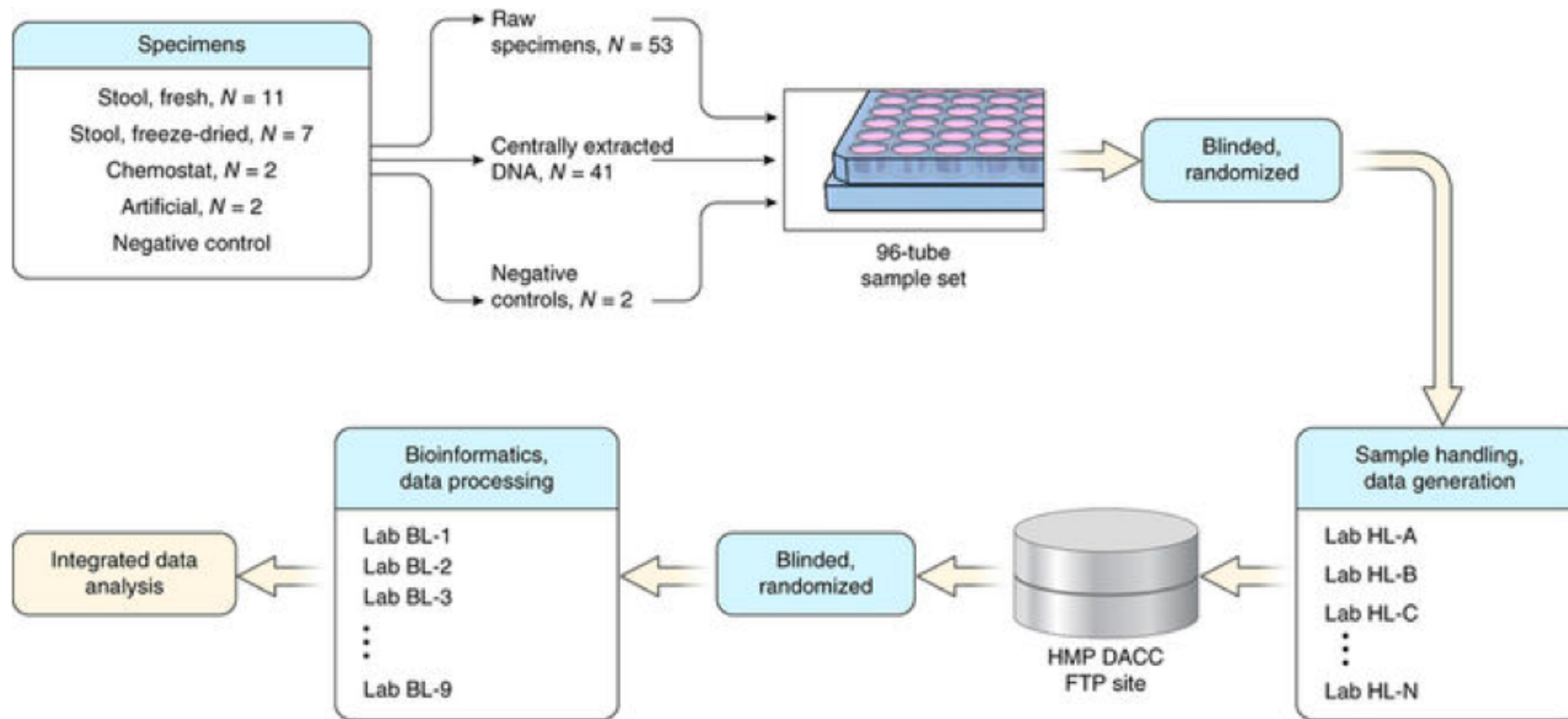


# Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium

Nature biotech  
Nov 2017

Rashmi Sinha<sup>1</sup>, Galeb Abu-Ali<sup>2,3</sup>, Emily Vogtmann<sup>1</sup> , Anthony A Fodor<sup>4</sup>, Boyu Ren<sup>2</sup>, Amnon Amir<sup>5</sup>, Emma Schwager<sup>2,3</sup> , Jonathan Crabtree<sup>6</sup>, Siyuan Ma<sup>2,3</sup>, The Microbiome Quality Control Project Consortium<sup>7</sup>, Christian C Abnet<sup>1</sup> , Rob Knight<sup>5,8</sup> , Owen White<sup>6</sup> & Curtis Huttenhower<sup>2,3</sup> 

In order for human microbiome studies to translate into actionable outcomes for health, meta-analysis of reproducible data from population-scale cohorts is needed. Achieving sufficient reproducibility in microbiome research has proven challenging. We report a baseline investigation of variability in taxonomic profiling for the Microbiome Quality Control (MBQC) project baseline study (MBQC-base). Blinded specimen sets from human stool, chemostats, and artificial microbial communities were sequenced by 15 laboratories and analyzed using nine bioinformatics protocols. **Variability depended most on biospecimen type and origin, followed by DNA extraction, sample handling environment, and bioinformatics.** Analysis of artificial community specimens revealed differences in extraction efficiency and bioinformatic classification. These results may guide researchers in experimental design choices for gut microbiome studies.



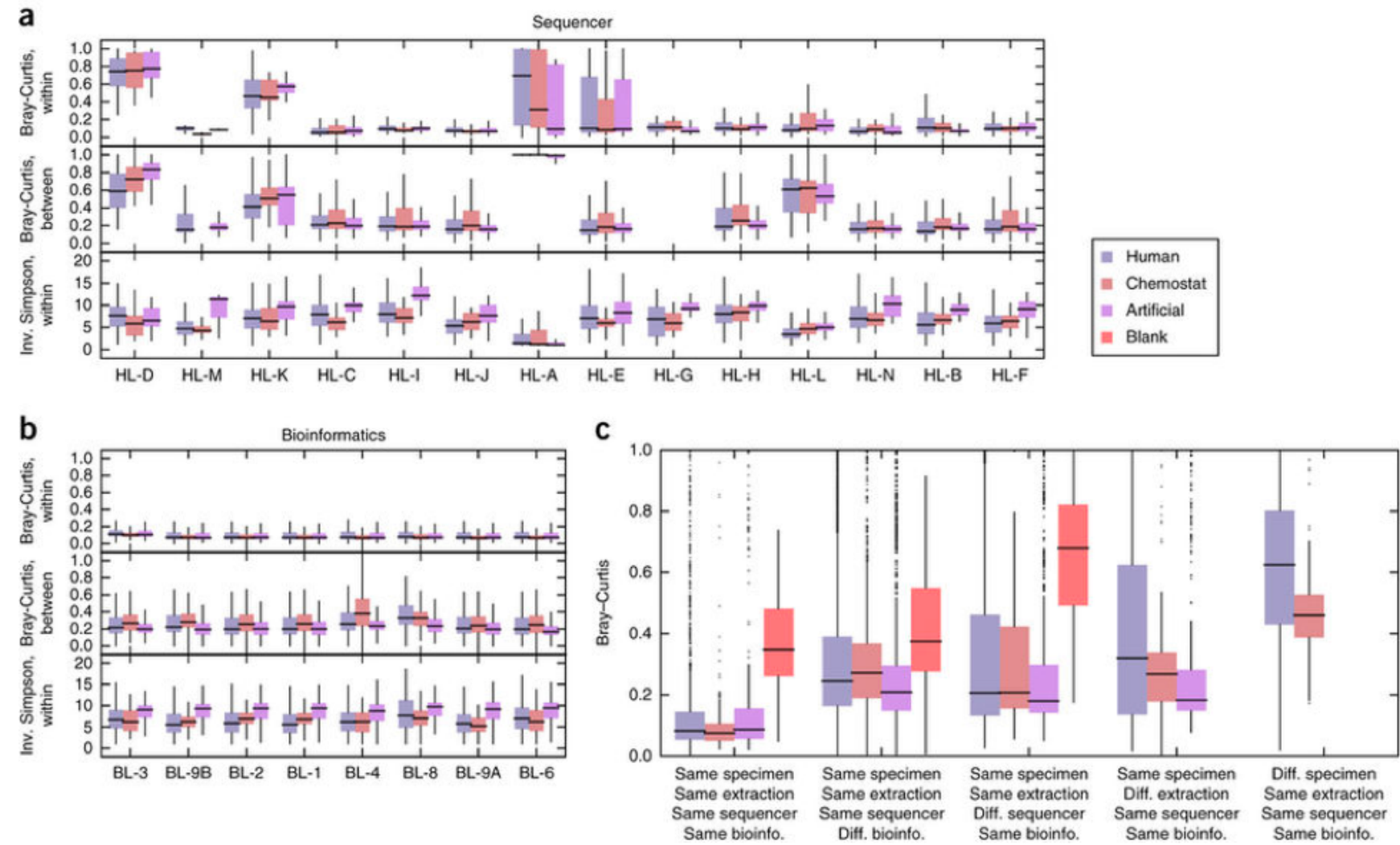
Study set up. Sent combo of raw specimens and aliquots from centrally-extracted DNA to different labs. Labs produced sequencing data. Data was randomized and sent to different bioinformatics labs.

Other notes: they didn't dictate which 16S region people sequenced. They did ask bioinformatics labs to call OTUs against the same GreenGenes database.

96 aliquots with 60% duplicates and 40% triplicates: 41 aliquots of centrally extracted DNA, 53 aliquots of raw specimens (frozen and freeze-dried feces, chemostat, and artificial communities), and two negative-control aliquots of storage buffer

a and b: c: difference between replicates for different comparisons

top = difference between replicates in the same lab  
middle = difference between replicates across different labs  
bottom = alpha diversity



Median alpha diversity moves around a bit, but not really affected (correlation across labs is high)

Bioinformatics labs all seemed similar (though paper claims that BL-8 was very different?)

Wet labs were all different, though disclaimer: huge differences tended to be caused by a few very dissimilar samples rather than consistent small differences across all samples.

Human samples are most variable, so biological variability is probably greater than technical variability.

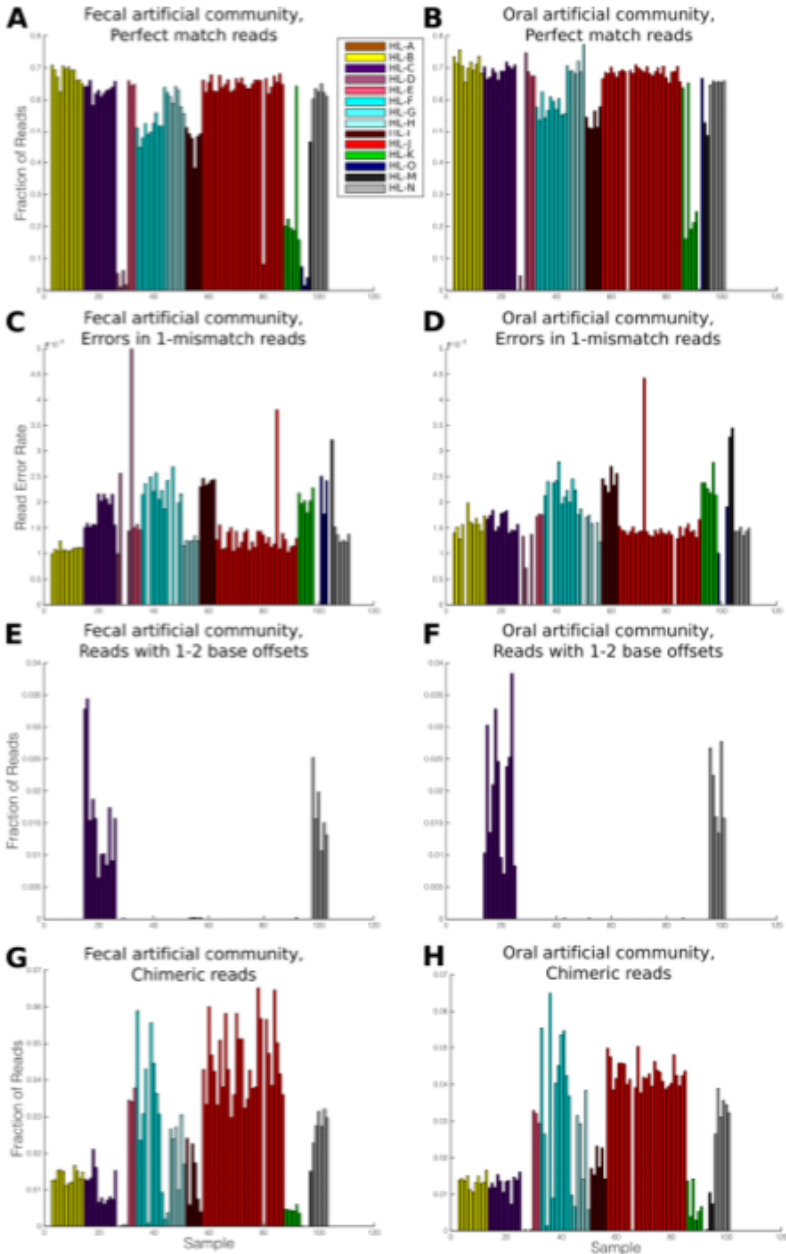
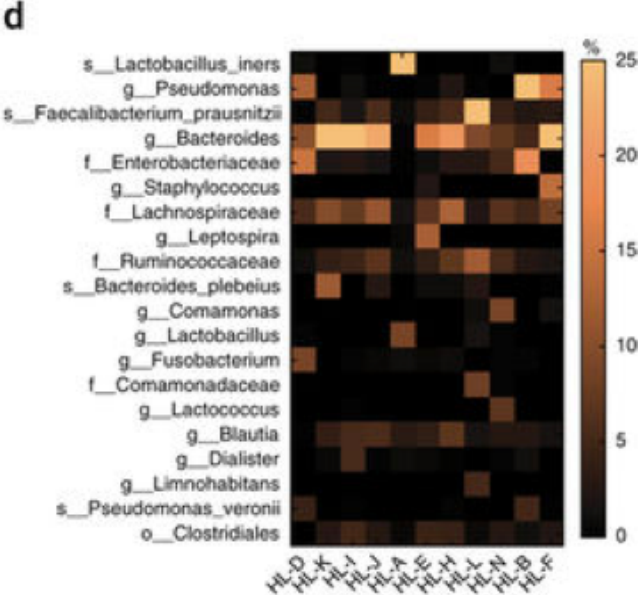
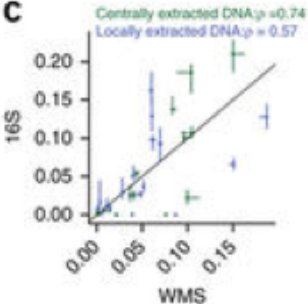
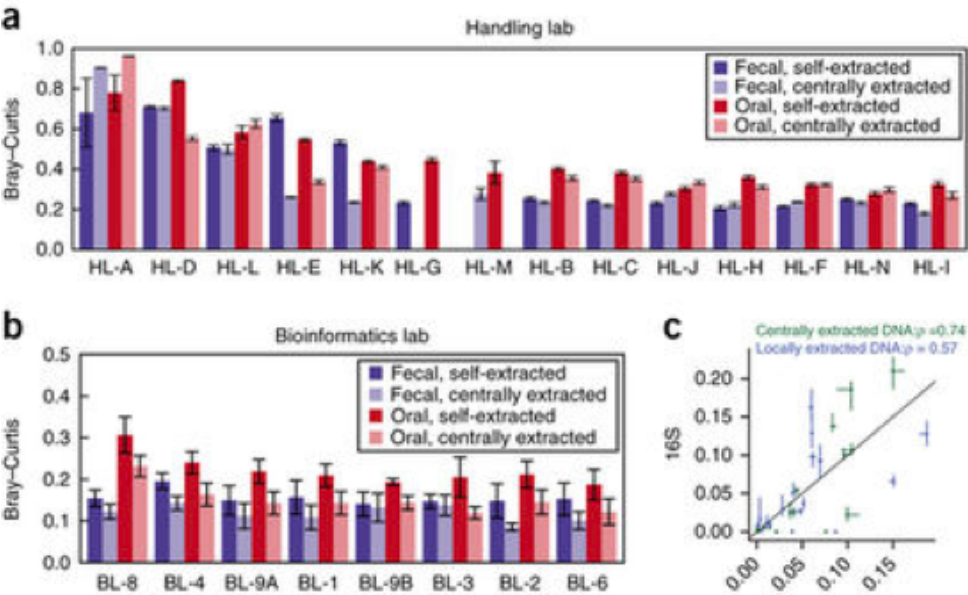
a and b: bray-curtis distance from two reference positive-control communities (20 fecal and 22 oral isolates)

c: meh, ignore

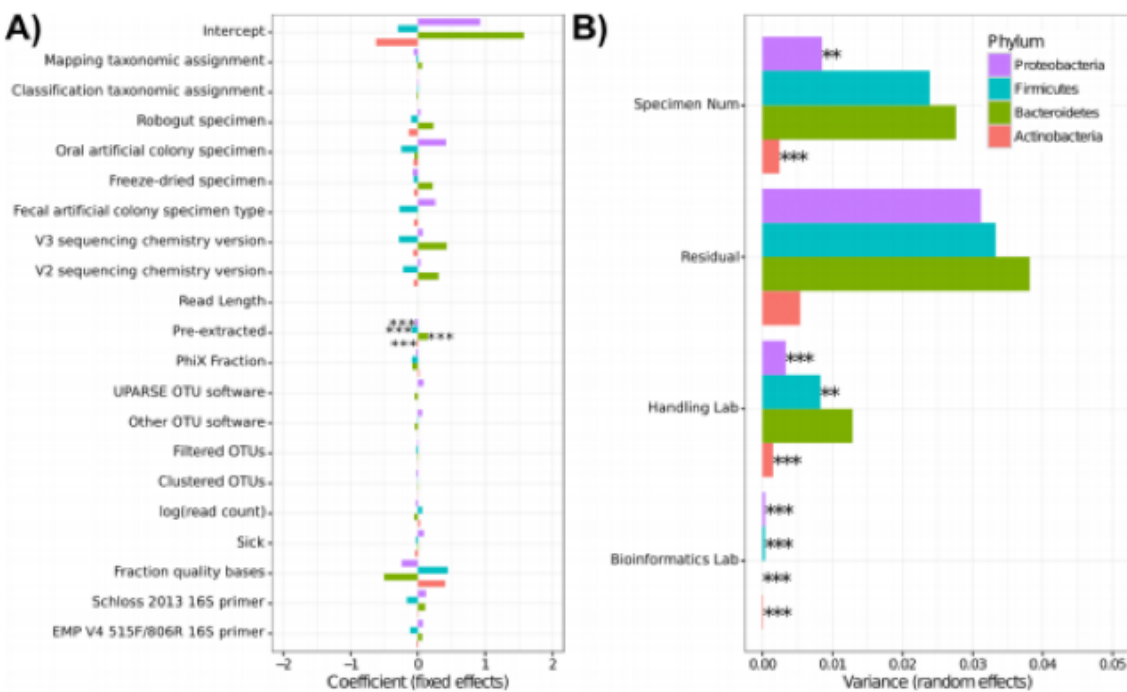
d. abundance heatmap for blank controls (Tris-HCL buffer), per lab

right (supp fig 13): percent of reads assigned to artificial community by handling lab

Again, take-away is that handling labs matter most.

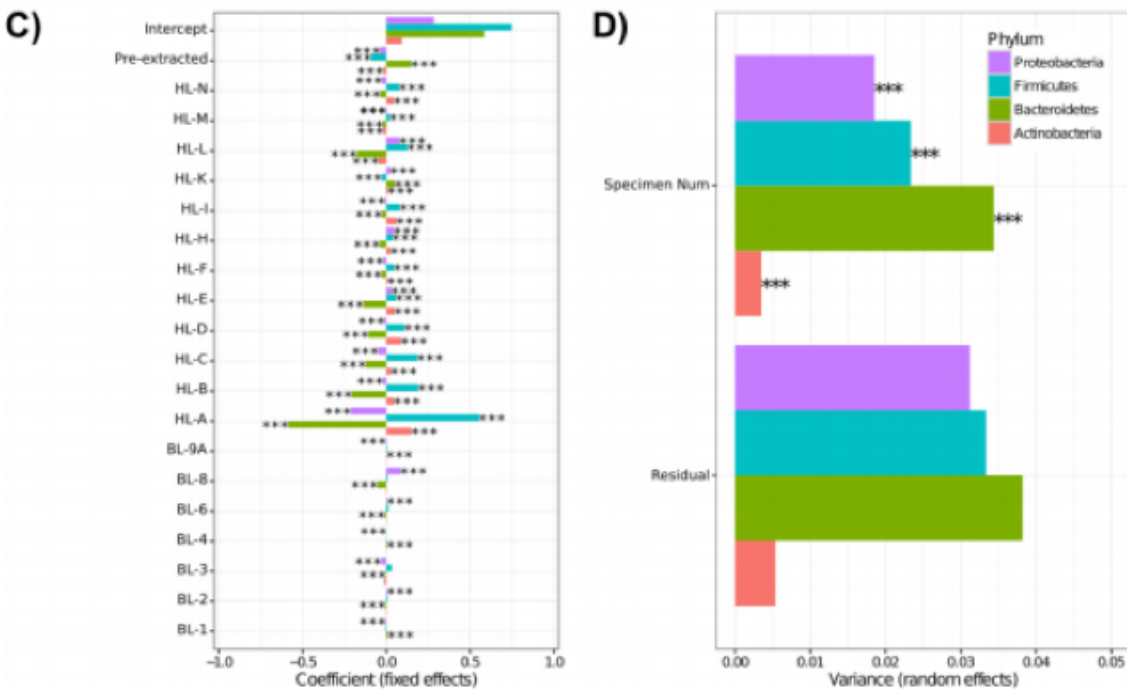




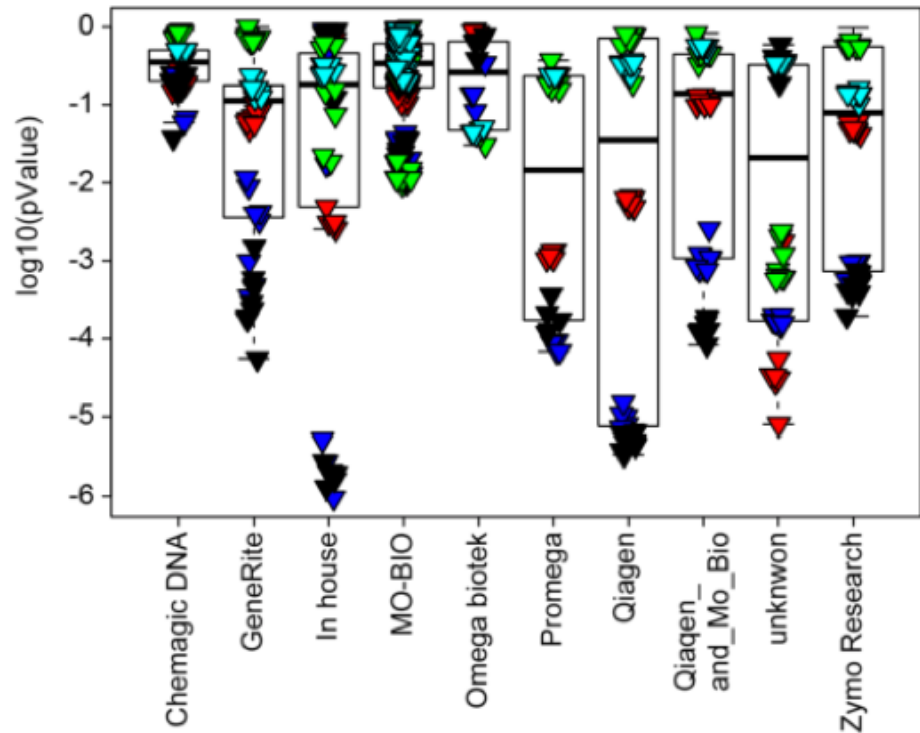
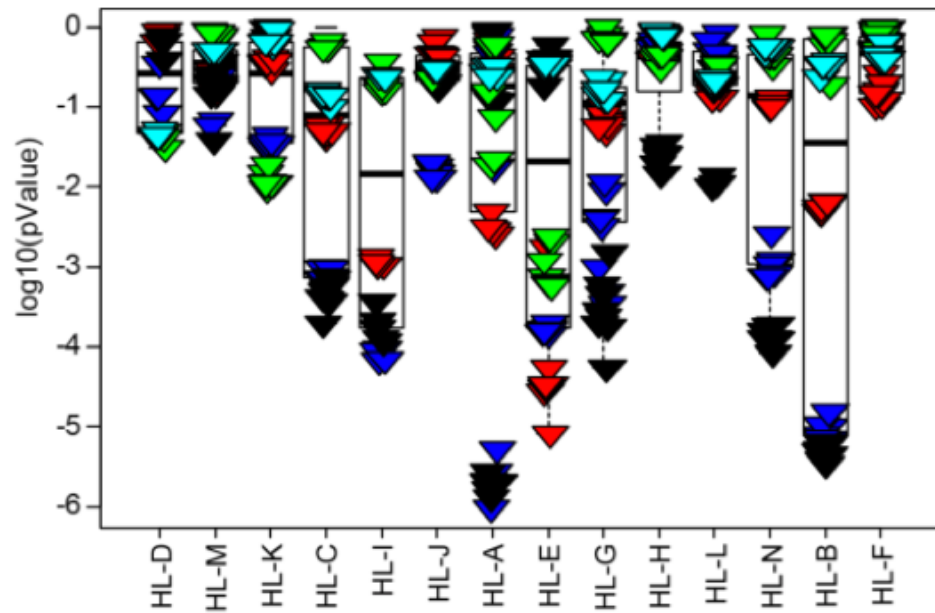


Different types of multivariate models, neither of which are particularly informative. Left panels show the coefficients for fixed effects, right panels show variation explained by random effects (colors are phyla).

Top panels are the “full” model, where all bioinformatics variables and labs are separated out, bottom panels are “simplified model” which just uses bioinformatics and handling labs as variables. Top doesn’t have enough power for significance, except for centrally-extracted vs. lab-extracted DNA. Bottom has everything significant.



Residual = “biological variation”



- Actinobacteria
- Bacteroidetes
- Firmicutes
- Fusobacteria
- Proteobacteria
- Verrucomicrobia

## DNA Extraction (In supplement)

Each triangle is a phylum from a bioinformatics lab. Shows paired p-value for "centrally extracted sample" vs. "lab-extracted sample." You can see that different labs/kits have very different extraction efficiencies for certain phyla.