Laboratory 8: Working with functional data.

Objectives:

- Experiment with functional metagenomic data analysis;
- Perform gene set enrichment analysis using Fisher exact test.

1. Load the predicted metagenome data from for Nature 488, pp. 621-626: STAT.picrust.RData.
2. Study the provided data:
    I. genes variable is a phyloseq object containing the metagenomic predictions (enzymes);
    II. pathways variable identifies biological pathways in which the metagenomic predictions participate.
3. Use appropriate normalization technique to normalize the data.
4. Perform principal components analysis ordination of the data, plot the results.
5. Are these data multivariately associated with Location and/or Treatment? Which distance do you use for this analysis?
6. Perform univariate analysis of the metagenomic predictions with respect to the Location variable. What test do you use? Do not forget about adjustment for multiple comparisons! Report any positive results.
7. Fisher exact test, fisher.test, can be used to determine if a biological pathway is enriched for significant enzymes. Use this test on Level 1 pathways. Report any enriched pathways and perform the analysis at Level 2 for them.
8. Perform the same analysis with respect to Treatment variable within the cecal samples.
   Note: use coefficient of variation cv = s.d./mean to filter out enzymes that do not vary much. Keep only enzymes with cv>1.
   What statistical test do you use to compare enzyme abundance in control and antibiotic groups?

Bonus: Load whole metagenome sequence data from Human Microbiome Project and experiment with these data (http://hmpdacc.org/HMMRC/).