**DNA isolation and microbial composition analysis**

DNA was extracted from samples using bead beating and modified protocols from commercially available kits, as previously described [1]. Briefly, 200 μL of each sample was added to 300 μL of SLX buffer (Omega Bio-Tek E.Z.N.A.® Stool DNA kit; Norcross, Georgia), 10 μL of 20 mg/mL proteinase K and 200 mg of glass beads, with bead-beating for 3 min. Samples were incubated at 70 °C for 10 min, 95 °C for 5 min and on ice for 2 min. The E.Z.N.A. protocol was followed to the inhibitor removal step, and DNA was extracted using Maxwell®16 DNA Purification Kit cartridges (Promega: Madison, Wisconsin), according to the Maxwell kit instructions.

You will need to change this isolation protocol to your specific protocol.

Primer sequences, barcodes used, Perl scripts and full protocols are in release 1 at <https://github.com/ggloor/miseq_bin/Illumina_SOP.pdf>. In brief, one microliter of DNA sample (1-5 ng) was used for PCR amplification using the Earth Microbiome universal primers (named 515F, 806R) specific for the V4 variable region of the rRNA gene [2]. Amplification was carried out using Promega GoTaq hot start colorless master mix reagent (Promega: Madison, WI) for 25 cycles with an annealing temperature of 52°C. Inline barcodes, similar to those described in [3] were used, with the modification that the barcodes were 8 nucleotides long and were preceded by 4 randomly synthesized nucleotides. Sequencing was carried out on the Illumina MiSeq platform at the London Regional Genomics Centre, with the 600 cycle v3 chemistry kit (Illumina, California). Paired-end sequencing was carried out with a 2x220 cycle profile with 5% PhiX-174 spike in. Data was exported as raw fastq files. The forward and reverse reads were overlapped with PandaSeq [4] disallowing overlapping segments that contained ambiguous bases (N characters). (Of note, if you use the userearch overlap method then use this sentence). The forward and reverse reads were overlapped by USEARCH (version number), keeping only overlapped reads that contained 1 or fewer sequencing errors.

Custom Perl scripts were used to separate and enumerate the reads from each sample as described [3]. Barcodes were chosen to contain a minimum edit distance of 4, and sequences were demulitplexed requiring perfect barcode matching to be assigned to a sample. USEARCH [5,14] version 7 for MacOSX was used to identify and remove chimeric sequences and to cluster sequences de-novo at a 97% identity threshold into operational taxonomic units (OTUs). Singleton OTUs, and OTUs that occurred at an abundance of less than 0.1% in any sample were discarded prior to analysis. OTU seed sequences were taxonomically assigned with the classify.seqs function of mothur 1.3.4 [6] using the Silva v119 database [7]. For analysis, OTUs were filtered to keep only those that occurred at a frequency greater than 1% in any sample, although this can be set by the user.

OTU frequencies are relative abundances. Such data are constrained by a constant sum, have an inherent internal correlation structure, and subsets of the data can give different results than the full set [8,9,10]. These problems were minimized by transforming the data into log-ratios between the abundance of each OTU and the geometric mean abundance of OTUs in the sample [8]. This transformation, termed the centered-log ratio (CLR), converts the data into unconstrained data where the distances between OTU abundance are linearly related, less affected by the subset problem, and directly comparable between samples [8,9,10]. Exploratory analysis of the data was performed in R [11] according to the protocols contained in the compositions R package described in [12]. When necessary, zero replacement of the sparse data was performed using the CZM approach from the zCompositions R package [13].

**References:**

1: McDonald JA, Fuentes S, Schroeter K, Heikamp-deJong I, Khursigara CM, de Vos WM, Allen-Vercoe E. **Simulating distal gut mucosal and luminal communities using packed-column biofilm reactors and an in vitro chemostat model.** J Microbiol Methods. 2015 Jan;108:36-44)

2: [Caporaso JG](http://www.ncbi.nlm.nih.gov/pubmed/?term=Caporaso%20JG%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Lauber CL](http://www.ncbi.nlm.nih.gov/pubmed/?term=Lauber%20CL%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Walters WA](http://www.ncbi.nlm.nih.gov/pubmed/?term=Walters%20WA%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Berg-Lyons D](http://www.ncbi.nlm.nih.gov/pubmed/?term=Berg-Lyons%20D%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Huntley J](http://www.ncbi.nlm.nih.gov/pubmed/?term=Huntley%20J%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Fierer N](http://www.ncbi.nlm.nih.gov/pubmed/?term=Fierer%20N%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Owens SM](http://www.ncbi.nlm.nih.gov/pubmed/?term=Owens%20SM%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Betley J](http://www.ncbi.nlm.nih.gov/pubmed/?term=Betley%20J%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Fraser L](http://www.ncbi.nlm.nih.gov/pubmed/?term=Fraser%20L%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Bauer M](http://www.ncbi.nlm.nih.gov/pubmed/?term=Bauer%20M%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Gormley N](http://www.ncbi.nlm.nih.gov/pubmed/?term=Gormley%20N%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Gilbert JA](http://www.ncbi.nlm.nih.gov/pubmed/?term=Gilbert%20JA%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Smith G](http://www.ncbi.nlm.nih.gov/pubmed/?term=Smith%20G%5BAuthor%5D&cauthor=true&cauthor_uid=22402401), [Knight R](http://www.ncbi.nlm.nih.gov/pubmed/?term=Knight%20R%5BAuthor%5D&cauthor=true&cauthor_uid=22402401). **Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms.** ISME J. 2012 Aug;6(8):1621-4

3: **Gloor** GB, Hummelen R, Macklaim JM, Dickson RJ, Fernandes AD, MacPhee R, Reid G. **Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products.** PLoS One. 2010 Oct 26;5(10):e15406

4: [Masella AP](http://www.ncbi.nlm.nih.gov/pubmed/?term=Masella%20AP%5BAuthor%5D&cauthor=true&cauthor_uid=22333067), [Bartram AK](http://www.ncbi.nlm.nih.gov/pubmed/?term=Bartram%20AK%5BAuthor%5D&cauthor=true&cauthor_uid=22333067), [Truszkowski JM](http://www.ncbi.nlm.nih.gov/pubmed/?term=Truszkowski%20JM%5BAuthor%5D&cauthor=true&cauthor_uid=22333067), [Brown DG](http://www.ncbi.nlm.nih.gov/pubmed/?term=Brown%20DG%5BAuthor%5D&cauthor=true&cauthor_uid=22333067), [Neufeld JD](http://www.ncbi.nlm.nih.gov/pubmed/?term=Neufeld%20JD%5BAuthor%5D&cauthor=true&cauthor_uid=22333067). **PANDAseq: paired-end assembler for illumina sequences.** BMC Bioinformatics. 2012 Feb 14;13:31

5: Edgar, RC. **Search and clustering orders of magnitude faster than BLAST.** Bioinformatics. 2010 Oct 1;26(19):2460-1

6: [Schloss PD](http://www.ncbi.nlm.nih.gov/pubmed/?term=Schloss%20PD%5BAuthor%5D&cauthor=true&cauthor_uid=19801464)1, [Westcott SL](http://www.ncbi.nlm.nih.gov/pubmed/?term=Westcott%20SL%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Ryabin T](http://www.ncbi.nlm.nih.gov/pubmed/?term=Ryabin%20T%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Hall JR](http://www.ncbi.nlm.nih.gov/pubmed/?term=Hall%20JR%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Hartmann M](http://www.ncbi.nlm.nih.gov/pubmed/?term=Hartmann%20M%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Hollister EB](http://www.ncbi.nlm.nih.gov/pubmed/?term=Hollister%20EB%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Lesniewski RA](http://www.ncbi.nlm.nih.gov/pubmed/?term=Lesniewski%20RA%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Oakley BB](http://www.ncbi.nlm.nih.gov/pubmed/?term=Oakley%20BB%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Parks DH](http://www.ncbi.nlm.nih.gov/pubmed/?term=Parks%20DH%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Robinson CJ](http://www.ncbi.nlm.nih.gov/pubmed/?term=Robinson%20CJ%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Sahl JW](http://www.ncbi.nlm.nih.gov/pubmed/?term=Sahl%20JW%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Stres B](http://www.ncbi.nlm.nih.gov/pubmed/?term=Stres%20B%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Thallinger GG](http://www.ncbi.nlm.nih.gov/pubmed/?term=Thallinger%20GG%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Van Horn DJ](http://www.ncbi.nlm.nih.gov/pubmed/?term=Van%20Horn%20DJ%5BAuthor%5D&cauthor=true&cauthor_uid=19801464), [Weber CF](http://www.ncbi.nlm.nih.gov/pubmed/?term=Weber%20CF%5BAuthor%5D&cauthor=true&cauthor_uid=19801464). **Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.** Appl Environ Microbiol. 2009 Dec;75(23):7537-41

7: [Yilmaz P](http://www.ncbi.nlm.nih.gov/pubmed/?term=Yilmaz%20P%5BAuthor%5D&cauthor=true&cauthor_uid=24293649)1, [Parfrey LW](http://www.ncbi.nlm.nih.gov/pubmed/?term=Parfrey%20LW%5BAuthor%5D&cauthor=true&cauthor_uid=24293649), [Yarza P](http://www.ncbi.nlm.nih.gov/pubmed/?term=Yarza%20P%5BAuthor%5D&cauthor=true&cauthor_uid=24293649), [Gerken J](http://www.ncbi.nlm.nih.gov/pubmed/?term=Gerken%20J%5BAuthor%5D&cauthor=true&cauthor_uid=24293649), [Pruesse E](http://www.ncbi.nlm.nih.gov/pubmed/?term=Pruesse%20E%5BAuthor%5D&cauthor=true&cauthor_uid=24293649), [Quast C](http://www.ncbi.nlm.nih.gov/pubmed/?term=Quast%20C%5BAuthor%5D&cauthor=true&cauthor_uid=24293649), [Schweer T](http://www.ncbi.nlm.nih.gov/pubmed/?term=Schweer%20T%5BAuthor%5D&cauthor=true&cauthor_uid=24293649), [Peplies J](http://www.ncbi.nlm.nih.gov/pubmed/?term=Peplies%20J%5BAuthor%5D&cauthor=true&cauthor_uid=24293649), [Ludwig W](http://www.ncbi.nlm.nih.gov/pubmed/?term=Ludwig%20W%5BAuthor%5D&cauthor=true&cauthor_uid=24293649), [Glöckner FO](http://www.ncbi.nlm.nih.gov/pubmed/?term=Gl%C3%B6ckner%20FO%5BAuthor%5D&cauthor=true&cauthor_uid=24293649). **The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks.** Nucleic Acids Res. 2014 Jan;42(Database issue):D643-8

8: Aitchison, JR. 1986 **The Statistical Analysis of Compositional Data**, Chapman and Hall, New Jersey ISBN: 1-930665-78-4

9: [Lovell D](http://www.ncbi.nlm.nih.gov/pubmed/?term=Lovell%20D%5BAuthor%5D&cauthor=true&cauthor_uid=25775355), [Pawlowsky-Glahn V](http://www.ncbi.nlm.nih.gov/pubmed/?term=Pawlowsky-Glahn%20V%5BAuthor%5D&cauthor=true&cauthor_uid=25775355), [Egozcue JJ](http://www.ncbi.nlm.nih.gov/pubmed/?term=Egozcue%20JJ%5BAuthor%5D&cauthor=true&cauthor_uid=25775355), [Marguerat S](http://www.ncbi.nlm.nih.gov/pubmed/?term=Marguerat%20S%5BAuthor%5D&cauthor=true&cauthor_uid=25775355), [Bähler J](http://www.ncbi.nlm.nih.gov/pubmed/?term=B%C3%A4hler%20J%5BAuthor%5D&cauthor=true&cauthor_uid=25775355). **Proportionality: a valid alternative to correlation for relative data.** PLoS Comput Biol. 2015 Mar 16;11(3):e1004075

10: [Fernandes AD](http://www.ncbi.nlm.nih.gov/pubmed/?term=Fernandes%20AD%5BAuthor%5D&cauthor=true&cauthor_uid=24910773), [Reid JN](http://www.ncbi.nlm.nih.gov/pubmed/?term=Reid%20JN%5BAuthor%5D&cauthor=true&cauthor_uid=24910773), [Macklaim JM](http://www.ncbi.nlm.nih.gov/pubmed/?term=Macklaim%20JM%5BAuthor%5D&cauthor=true&cauthor_uid=24910773), [McMurrough TA](http://www.ncbi.nlm.nih.gov/pubmed/?term=McMurrough%20TA%5BAuthor%5D&cauthor=true&cauthor_uid=24910773), [Edgell DR](http://www.ncbi.nlm.nih.gov/pubmed/?term=Edgell%20DR%5BAuthor%5D&cauthor=true&cauthor_uid=24910773), [Gloor GB](http://www.ncbi.nlm.nih.gov/pubmed/?term=Gloor%20GB%5BAuthor%5D&cauthor=true&cauthor_uid=24910773). **Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis.** Microbiome. 2014 May 5;2:15.

11: R Core Team (2015). **R: A language and environment for statistical**

**computing.** R Foundation for Statistical Computing, Vienna, Austria.

URL <https://www.R-project.org/>.)

12: van den Boogaart, KG and Tolosana-Delgado, R. **Analyzing compositional data with R**. 2013, Springer, New York ISBN 978-3-642-36809-7

13: Palarea-Albaladejo, Javier, and Josep Antoni Martín-Fernández. **zCompositions—R package for multivariate imputation of left-censored data under a compositional approach**. Chemometrics and Intelligent Laboratory Systems, 2015, 143:85-96

14: Edgar,RC, Haas,BJ, Clemente,JC, Quince,C, Knight,R (2011) **UCHIME improves sensitivity and speed of chimera detection**, *Bioinformatics* doi: 10.1093/bioinformatics/btr381