MU RRRC/MMRRC

Why we do it

There is abundant evidence that the gut microbiota plays an integral role in human health and disease susceptibility. There is also reason to believe that the mouse models we use are affected by the microbes harbored in the gastrointestinal tract. Of note, the composition of the gut microbiota in laboratory rats and mice can be affected by diet, genetic background, stress, housing density, and myriad other facets of animal husbandry. We believe that these and other factors capable of modulating the gut microbiota may be responsible for the subtle changes or even losses of model phenotype experienced by users of animal models. How does a scientist working with a limited budget monitor for such events? How does a scientist safeguard against such confounding factors affecting their studies?

The MU RRRC/MMRRC has leveraged the expertise of individuals in the fields of comparative medicine, microbiology, molecular biology, bioinformatics, and statistics to develop a robust and affordable next-generation sequencing pipeline to assist you in the characterization of the gut microbiota of your research animals. Using this technology, the MU RRRC/MMRRC has performed extensive surveys of the gut microbiota of multiple strains and stocks of rats, from various vendors, and at various ages, in order to provide you with knowledgeable guidance and consultation with your research. In order to provide the most consistent data, we perform all steps in the process from consultation on experimental design and extraction of fecal DNA to the annotation of sequence data and assistance with data interpretation.

How we do it

The process begins with extraction of fecal DNA. Working with a very small amount of starting material (1 rat fecal pellet or 1 to 2 mouse fecal pellets), we use a manual extraction process to extract and purify high-quality DNA. After assessing the yield and purity of DNA, polymerase chain reaction (PCR) is performed to amplify the hypervariable V4 region of the microbial 16S rDNA gene. Amplified DNA is then sequenced using the Illumina MiSeq platform and a set of 96 unique bar-coded primers. This results in the ability to sequence up to 96 samples simultaneously (reducing costs dramatically) at a depth of approximately 80,000-140,000 reads per sample. Sequence data is then processed through an informatics pipeline to remove chimeric sequences and bin sequences into representative operational taxonomic units (OTUs). Those OTUs are then BLASTed against a reference database of known microbial 16S rDNA sequences to annotate your data taxonomically.

What we offer

Service. Our goal is to assist the biomedical research community, as well as veterinary clinicians, with the resources to characterize the gut microbiota of research animals. At the front end, we are happy to provide consultation on experimental design, sample collection, or any other aspects of your project related to the gut microbiota. At the back end, we are happy to help users interpret and understand their data, especially in the context of their animal model. We can also offer guidance to researchers trouble-shooting changes in the phenotype of their animal models, and advice on ways to restore those models. Drawing on our expertise in assisted reproduction techniques, we are also able to offer services such as rederivation via embryo transfer to place your model on more than one background “enterotype”, in order to determine whether your model is susceptible to differences in the gut microbiota, or to optimize your model to suit your needs.

Data. The data provided by the RRRC/MMRRC comes in multiple formats including bar charts, two- and three-dimensional principal component analyses (PCAs), and rarefaction curves. Bar charts are provided at all taxonomic levels (phylum, class,…genus, OTU) with corresponding tables showing the relative abundance of the taxa detected (**Figure 1**). These data are amenable to transfer into Excel or similar programs for manipulation of data. Two-dimensional PCAs are provided as weighted and unweighted, discrete and

continuous analyses (**Figure 2**). Additionally, an unweighted three-dimensional PCA will be provided. Rarefaction curves are generated using multiple algorithms (Shannon, PD whole tree, Chao, and Observed Species)(**Figure 3**). Users also receive all raw data. These data are automatically stored at the MMRRC for one year and can be re-annotated to updated databases for a nominal fee.

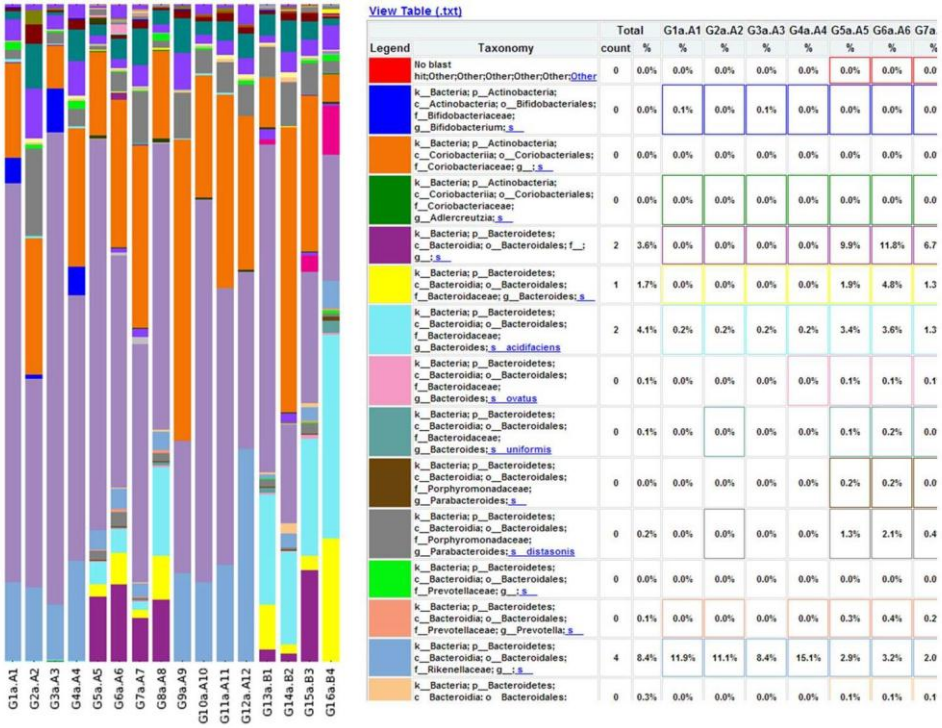

Figure 1. Example of sequencing data shown in bar chart format. Each column in the bar chart represents the sequence data from one animal. The table at left shows a small portion of the associated data indicating the abundance of each operational taxonomic unit detected in each sample.

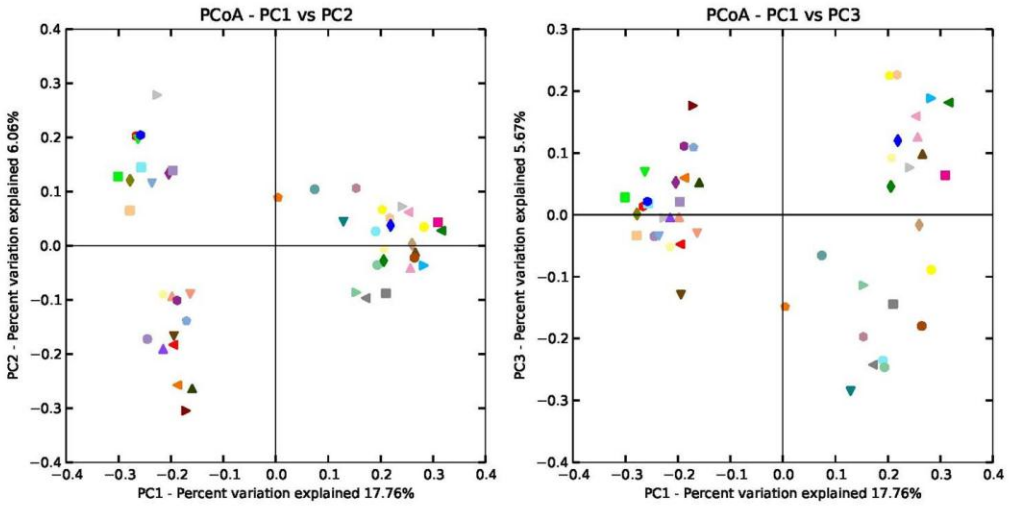

Figure 2. Example of sequencing data shown in PCA format. Each point represents one animal and distances between points indicate similarity between samples taking into account all OTUs. Principal component 1 (PC1) and PC2 are shown at left; PC1 and PC3 are show at right.

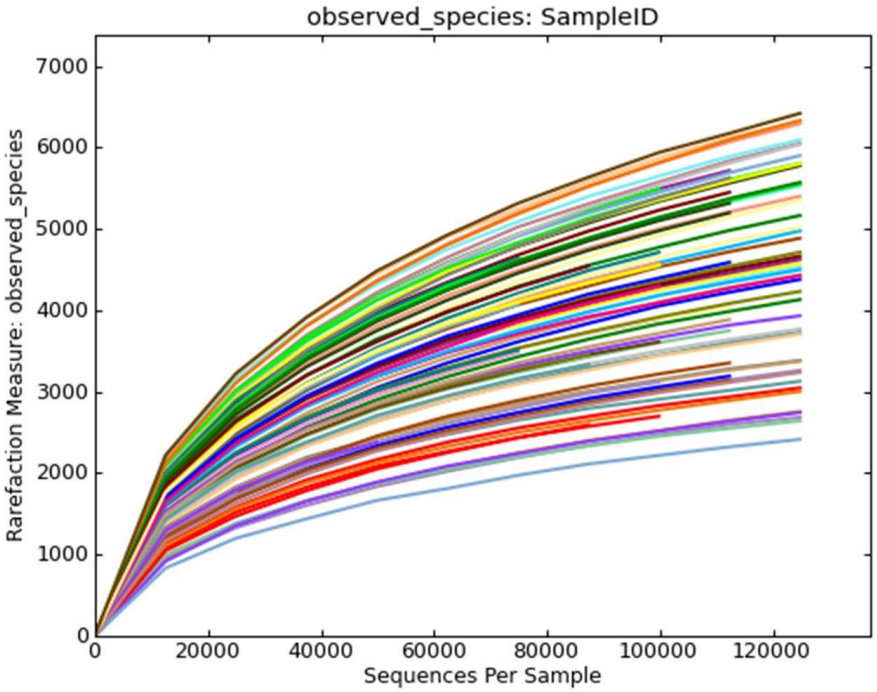

Figure 3. Example of rarefaction curves demonstrating the number of unique sequences detected when the metagenomic data for each sample is repeatedly sub-sampled. Each line depicts the relationship between increased sequencing depth and the number of OTUs detected.

Sample Submission

Prior to sample submission, please contact us to discuss your project so that we can ensure that your needs will be met.

www.rrrc.us

www.mmrrc.org

FAQs

1. **How should I collect fecal samples?**

   For rats and mice, we recommend placing the animal in an empty autoclaved cage (containing no bedding) and waiting until the animal defecates naturally. Using sterile toothpicks or syringe needles, spear the fecal pellet and place the sample directly in a sterile 1.5 mL Eppendorf tube, pre-labeled with any identifying information (e.g., animal ID, treatment group, etc.). Once samples have been collected, close tube and place in liquid N_2, or transfer immediately to a -80°C freezer. The main reason for prompt cooling of samples is to prohibit any proliferation of rapidly dividing bacteria and subsequent skewing of data.

2. **How "fresh" do the samples need to be?**

   Samples should be evacuated no more than a few minutes prior to sample collection. When collecting samples from a cage (even seemingly "fresh" samples), there is the chance that certain bacterial species with rapid doubling times will proliferate, resulting in falsely elevated levels. Additionally, there is nuclease activity in fecal material which can lead to degeneration of microbial DNA. Prompt collection and cooling of samples will minimize the nuclease activity and optimize the quality of DNA extracted from samples.

3. **Does it matter if the samples come in contact with urine or other biological fluids?**

   We have not performed experiments to answer this question definitively. To minimize any potential for a confounding effect, we recommend that samples that have come in contact with urine or other biological materials be discarded and fresh samples collected.

4. **How much/many samples should I send per animal?**

   Regarding the quantity of fecal material, we request a minimum of two normal fecal pellets per mouse sample, i.e., two fecal pellets per Eppendorf tube. For rats, we request one normal fecal pellet per sample. For other species, we ask that you contact us to discuss an appropriate amount for sample collection.

5. **What's an appropriate "*n*"?**

   Due to the inherent variability in species richness and composition of fecal samples, the answer to this question depends on the animals in question and the goals of the project. We strongly recommend that you contact us to discuss your study. You may also want to view the **Example Data** for an idea of the variability in the sequencing data generated from a set of fecal samples from sex-, age-, strain-, and vendor-matched animals.

6. **Does the order of fecal pellet matter?**

   No, we have not seen any appreciable difference in the microbial profile of sequential fecal pellets from the same rat or mouse.

7. **Does the time of day that I collect samples matter?**

   Again, we have not performed controlled experiments to answer this question definitively. GI motility and secretory activity are however both affected by circadian rhythms and we thus recommend collecting all samples at a consistent time of day to control for any potential effect.

8. **How should I ship the samples?**

   Samples should be shipped overnight in a container with sufficient dry ice to ensure that samples remain completely frozen during transport. Samples should be shipped to:

   Attn: Aaron Ericsson

University of Missouri
4011 Discovery Drive
Columbia, MO 65201

9. What is the expected turn-around time between sample submission and delivery of sequencing data?

Turn-around times will depend on our current sample load. Our goal is to return data to users within 4 to 6 weeks of sample submission.

10. What kind of data do you provide?

Along with the raw data, we perform several standard bioinformatics analyses. Sequence data will be annotated to a database of known microbial 16S rRNA gene sequences. Thus, the abundance of microbes at each taxonomic level (i.e., phylum, class, order, etc.) will be depicted in bar chart form with the relative abundance of each taxa listed below in a tabular format. Multiple principal component analyses (PCA) will be performed for the first three principal components. Two-dimensional PCAs depicting PC1 v. PC2, PC1 v. PC3, and PC2 v. PC3 will be generated using weighted and unweighted, and discrete and continuous algorithms. Additionally, a three-dimensional representation of PC1, PC2, and PC3 will be provided. Lastly, rarefaction curves will be generated to provide an estimate of the adequacy of sequencing depth, as well as overall species richness. Rarefaction curves will be generated using multiple algorithms (chao1, observed species, PD whole tree, and Shannon). The **Example Data** will give you an idea of the data provided.

11. Can you annotate sequence data to the species level?

The standard definition of species as it applies to eukaryotic organisms is problematic when considering prokaryotes. Our PCR primers and sequencing platform allow us to annotate many sequences to the species level, however, not all sequences will annotate to that level. Thus, we refer to our lowest level of annotation as “operational taxonomic unit” or OTU. This is an artificial grouping of DNA sequences by sequence identity into discrete OTUs that are within a pre-defined range of similarity, set during the informatics analysis. We use very stringent parameters in defining on OTU and routinely detect between 100 and 120 distinct OTUs in murine feces, binned from a total of 15,000-20,000 unique sequences. The **Example Data** shows the OTU-level annotation from a typical murine fecal sample.

12. Do you provide assistance with interpretation of metagenomics data?

We can provide assistance in the interpretation of sequencing data as either an additional fee-for-service or as a collaborative effort. Please contact us to discuss your project if you will require assistance with data interpretation.

13. Do you provide additional biostatistical analysis?

We do not routinely provide biostatistical analyses beyond the informatics described above. We are happy to provide consultation regarding what types of analyses would be of use to your projects and can provide *post hoc* biostatistical analyses (e.g., hierarchical cluster analysis or correlation to phenotypic data) for an additional fee.

14. How long do you maintain the sequencing data?

Raw sequence data will be maintained indefinitely.

15. Do you/can you archive fecal material?

We can archive fecal samples at -80°C for an additional fee. Please contact us prior to sample submission to discuss the number of samples and cost of sample storage.

16. Will I receive the raw data?

Yes, we will provide all raw sequence data.

17. How much data will I receive in total, i.e., how much computer storage space is required?

This will obviously depend on the number of samples submitted. We are able to sequence up to 96 samples per lane, resulting in the generation of approximately 1GB of data, including the informatics analyses (bar charts, PCAs, rarefaction curves). Smaller sample sets will require less storage space.

18. Will I require any specialized software to view or manipulate the data?

You will require current versions of Java and Google Chrome to view the informatics analyses (bar charts, PCAs, rarefaction curves). A current version of Excel is required for manipulation of the raw data and taxa abundances.

19. Can the raw data be re-annotated in the future as 16S rRNA gene databases are updated?

Yes, as long as the samples were submitted within the last year, we can re-annotate the sequence data at a nominal additional cost. Please contact us to discuss if you have questions regarding the cost and/or benefit of such analyses.

20. Do you work with samples from species other than mice and rats?

We have worked with fecal samples from other species. Please contact us to discuss your project goals before submitting samples from species other than mouse or rat.

21. Are all samples guaranteed to perform well, i.e., yield meaningful data?

Unfortunately, no. Occasionally, a handful of samples in a 96-well plate will not amplify well during amplification prior to sequencing. We recommend adjusting the experimental "*n*" to allow for such a possibility. Additionally, the presence of polysaccharides in fecal material may inhibit PCR amplification prior to sequencing. Thus, certain samples may be problematic, e.g., from mice that have ingested shipping gel pack material, DSS-treated mice, etc.

22. What is the typical depth of sequencing you provide?

We generally acquire between 100,000 and 120,000 reads per sample. If you require deeper sequencing, please contact us to discuss your project goals.

23. What region of the 16S rRNA gene do you amplify for sequencing?

The primers we use target the hypervariable V4 region of the 16S rDNA gene and result in an approximately 300 bp amplicon.

24. What sequencing platform do you use?

We use the Illumina MiSeq platform in conjunction with a set of 96 bar-coded primers.

25. What Informatics software and 16S rRNA sequence database do you use for sequence assembly and annotation?

Contiguous DNA sequences are assembled using FLASH software and representative OTUs are selected using Qiime v1.8. Selected OTUs are annotated to the Greengenes database of 16S rRNA gene sequences using BLAST.