Metagenome Preparations and Analysis

Bryan Yoo[1], Jessica Griffiths[1], Sarkis Mazmanian[1]

[1]California Institute of Technology

Nov 02, 2021

1

dx.doi.org/10.17504/protocols.io.bzqep5te

Mazmanian Lab

Jessica Griffiths

Protocol for metagenome preparation and analysis used in Yoo et al 2021

DOI

dx.doi.org/10.17504/protocols.io.bzqep5te

Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian 2021. Metagenome Preparations and Analysis. **protocols.io** https://dx.doi.org/10.17504/protocols.io.bzqep5te

Center for Environmental and Microbial Interactions (CEMI)
Grant ID: n/a

Emerald Founation
Grant ID: n/a

Heritage Medical Research Institute
Grant ID: n/a

National Institutes of Health
Grant ID: GM007616 and DK078938

Department of Defense
Grant ID: PD160030

Aligning Science Across Parkinson's
Grant ID: ASAP-000375

protocol ,

Nov 02, 2021

Nov 02, 2021

Fecal Collection

1 AAV-PHP.S:hSYN1-hM3Dq-mRuby2 (10^{12}VGs) was delivered systemically to TH-Cre and ChAT-Cre mice. 3-4 week after infection, C21 (3mg/kg) was administered daily for 10 consecutive days. Fecal pellets were collected in sterile containers one day before initial C21 dose, and between doses thereafter.

Fecal sample DNA extraction and Library Preparation

2 DNA was extracted with the Qiagen MagAttract PowerSoil DNA kit as previously described (Marotz et al 2017). Our standard protocol is optimized for an input quantity of 1ng DNA per reaction.

3 Prior to library preparation, input DNA is transferred to a 384-well plate and quantified using a PicoGreen fluorescence assay (ThermoFisher, Inc).

4 Input DNA is then normalized to 1ng in a volume of 3.5µL of molecular-grade water using an Echo 550 acoustic liquidhandling robot (Labcyte, Inc).

5 Enzyme mixes for fragmentation, end repair and A-tailing, ligation, and PCR are prepared and added in approximately 1:8 scale volumes using a Mosquito HV micropipetting robot (TTP Labtech).

6 Fragmentation is performed at 37°C for 20min, followed by end repair and A-tailing at 65°C for 30min.

7 Sequencing adapters and barcode indices are added in two steps, following the iTru adapter protocol (**Glenn et al., 2019**). Universal adapter “stub” adapter molecules and ligase mix are first added to the end-repaired DNA using the Mosquito HV robot and ligation performed at 20°C for 1h. Unligated adapters and adapter dimers are then removed using AMPure XP magnetic beads and a BlueCat purification robot (BlueCat Bio). 7.5 µL magnetic bead solution is added to the total adapter-ligated sample volume, washed twice with 70% EtOH, and then resuspended in 7µL molecular-grade water.

8 Next, individual i7 and i5 are added to the adapter-ligated samples using the Echo 550 robot. Because this liquid handler individually addresses wells, and we use the full set of 384 unique

Citation: Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian Metagenome Preparations and Analysis
https://dx.doi.org/10.17504/protocols.io.bzqep5te

error-correcting i7 and i5 indices, we are able to generate each plate of 384 libraries without repeating any barcodes, eliminating the problem of sequence misassignment due to barcode swapping (**Costello et al., 2018**;**Sinha et al., 2017**).

9 To ensure that libraries generated on different plates can be pooled if necessary, and to safeguard against the possibility of contamination due to sample carryover between runs, we also iterate the assignment of i7 to i5 indices each run, such that each unique i7:i5 index combination is only repeated once every 147,456 libraries.

10 4.5µL of eluted bead-washed ligated samples is added to 5.5µL of PCR master mix and PCR-amplified for 15cycles.

11 The amplified and indexed libraries are then purified again using magnetic beads and the BlueCat robot, resuspended in 10µL water, and 9µL of final purified library transferred to a 384-well plate using the Mosquito HV liquid andling robot for library quantitation, sequencing, and storage.

12 384 samples are then normalized based on a PicoGreen fluorescence assay.

Shallow shotgun metagenome sequencing and diversity analysis

13 The Illumina data for each HiSeq lane was uploaded to Qiita, a tool with standardized pipelines for processing and analyzing metagenomic data (**Gonzalez et al., 2018**).

14 Adapter sequences were removed from the reads using the Atropos v.1.1.15 (**Didion et al., 2017**) command (from the qp-shogun 0.1.5 pipeline) and the trimmed sequences were downloaded from Qiita.

15 The reads for each sample were filtered of any potential mouse contamination using Bowtie2 v.2-2.2.3 (**Langmead and Salzberg, 2012**).

16 The filtered reads were then aligned to the Web of Life (WoL) reference phylogeny (**Zhu et al., 2019**) with Bowtie2 using an adapted SHOGUN pipeline (**Hillmann et al., 2018**). The WoL contains 10,575 bacterial and archaeal genomes with each genome representing an OTU. Sequencing reads that did not map to a single reference genome as well as reads that mapped to multiple genomes were not included in the analysis.

17 If an OTU had a relative abundance less than 0.01% in a given sample, the OTU was not included for that sample. Additionally, OTUs with fewer than 5 assigned reads were not

Citation: Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian Metagenome Preparations and Analysis
https://dx.doi.org/10.17504/protocols.io.bzqep5te

considered.

18 The samples were rarefied to a depth of 12,750 reads and those with fewer than the rarefaction depth were excluded. T

19 he QIIME2 v.2019.7 (**Bolyen et al., 2019**) DEICODE plugin was used to calculate the Aitchison distances, a compositional beta diversity metric, and perform Robust Aitchison PCA to create biplots that visualize relationships between features and samples (**Martino et al., 2019**).

20 The QIIME2 diversity plugin was used to calculate the other alpha- and beta-diversity metrics used in this study.

Metagenomic based functional profiling

21 The filtered reads were also analyzed using HUMAnN2 v2.8.1 (**Franzosa et al., 2018**) to establish functional profiles for the samples. HUMAnN2 is a pipeline that begins by using MetaPhlAn 2 to compile custom databases of reference genomes based on the species detected in a sample (**Truong et al., 2015**).

22 HUMAnN2 then maps the filtered onto these custom databases and the reads that do not map to any of the references are then subjected to a translated search against UniProt Reference Clusters or UniRef (**Suzek et al., 2007**). Here, the UniRef90 database was used for the translated search and installed according to the HUMAnN2 documentation. The results from both the search performed using the custom reference genome database and the search against the UniRef90 database were combined and the gene families identified in each sample were reported in units of read per kilobase (RPKs) to account for gene length.

23 HUMAnN2 also compared the gene families found in a sample with the MetaCyc pathways database (**Caspi et al., 2018**) and output a table reporting the pathway abundances found in each sample.

24 After rarefying gene family tables to a depth of 166,000 RPKs and using a depth of 22,600 for pathway abundances, the QIIME2 diversity and DEICODE plugins were used to calculate alpha- and beta-diversity metrics.

Citation: Bryan Yoo, Jessica Griffiths, Sarkis Mazmanian Metagenome Preparations and Analysis
https://dx.doi.org/10.17504/protocols.io.bzqep5te