

Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples

In the format provided by the
authors and unedited

Microbial-enrichment method enables high-throughput metagenomic characterization from host-rich samples

Natalie J. Wu-Woods^{1,9}, Jacob T. Barlow^{1,9}, Florian Trigodet², Dustin G. Shaw^{2,3,4}, Anna E. Romano¹, Bana Jabri^{2,3,4}, A. Murat Eren^{5,6,7,8}, Rustem F. Ismagilov^{1,10}

¹ California Institute of Technology 1200 E. California Blvd., Pasadena, CA, 91125, USA

² Department of Medicine, University of Chicago, Chicago, IL 60637, USA

³ Committee on Immunology, University of Chicago, Chicago, IL, USA

⁴ Department of Pathology, University of Chicago, Chicago, IL, USA

⁵ Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA 02543, USA

⁶ Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, 26129 Oldenburg, Germany

⁷ Alfred-Wegener-Institute for Marine and Polar Research, 27570 Bremerhaven, Germany

⁸ Helmholtz Institute for Functional Marine Biodiversity, 26129 Oldenburg, Germany

⁹ These authors contributed equally to this work.

¹⁰ Correspondence to: rustem.admin@caltech.edu

Contents:

Figures S1–S9

Description of Tables S1–S11

Author Contact Information and ORCiDs

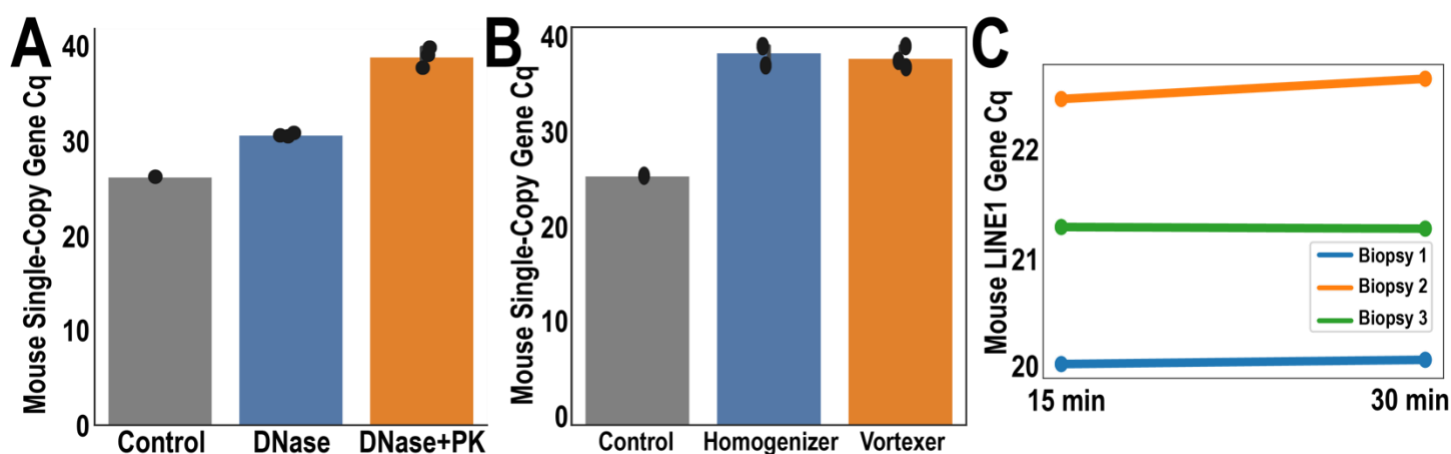


Fig S1: MEM protocol optimization. A) MEM protocol was tested with and without Proteinase K (PK) treatment on mammalian mouse cell culture. Host was quantified after MEM treatment and DNA extraction with single-copy mouse primers with qPCR. (N=3 technical replicates for DNase and DNase+PK condition, N=1 for control; error bars are 95% CI). B) MEM protocol was performed on mammalian mouse cell culture to compare the effectiveness of a homogenizer or vortexer in host cell disruption. A comparison between homogenizer (4.5m/s for 30 sec) vs vortexer adapter (1 min at max speed) was performed and remaining host DNA after MEM treatment was quantified using qPCR on single-copy mouse primers. We concluded homogenizers are not necessary and this may be important for field work for future environmental samples (N=3 technical replicates for Homogenizer and Vortexer condition, N=1 for control; error bars are 95% CI). C) MEM was performed on 3 rat biopsies split into two different reactions with 15-min or 30-min incubation at 37 °C. Host load was quantified through LINE1 transposon primers with qPCR. Minimal differences in host DNA removal with incubation of 15 or 30 minutes in nuclease. Final protocol used 15 minute incubation to minimize processing times (N=3 biological replicates from one rat).

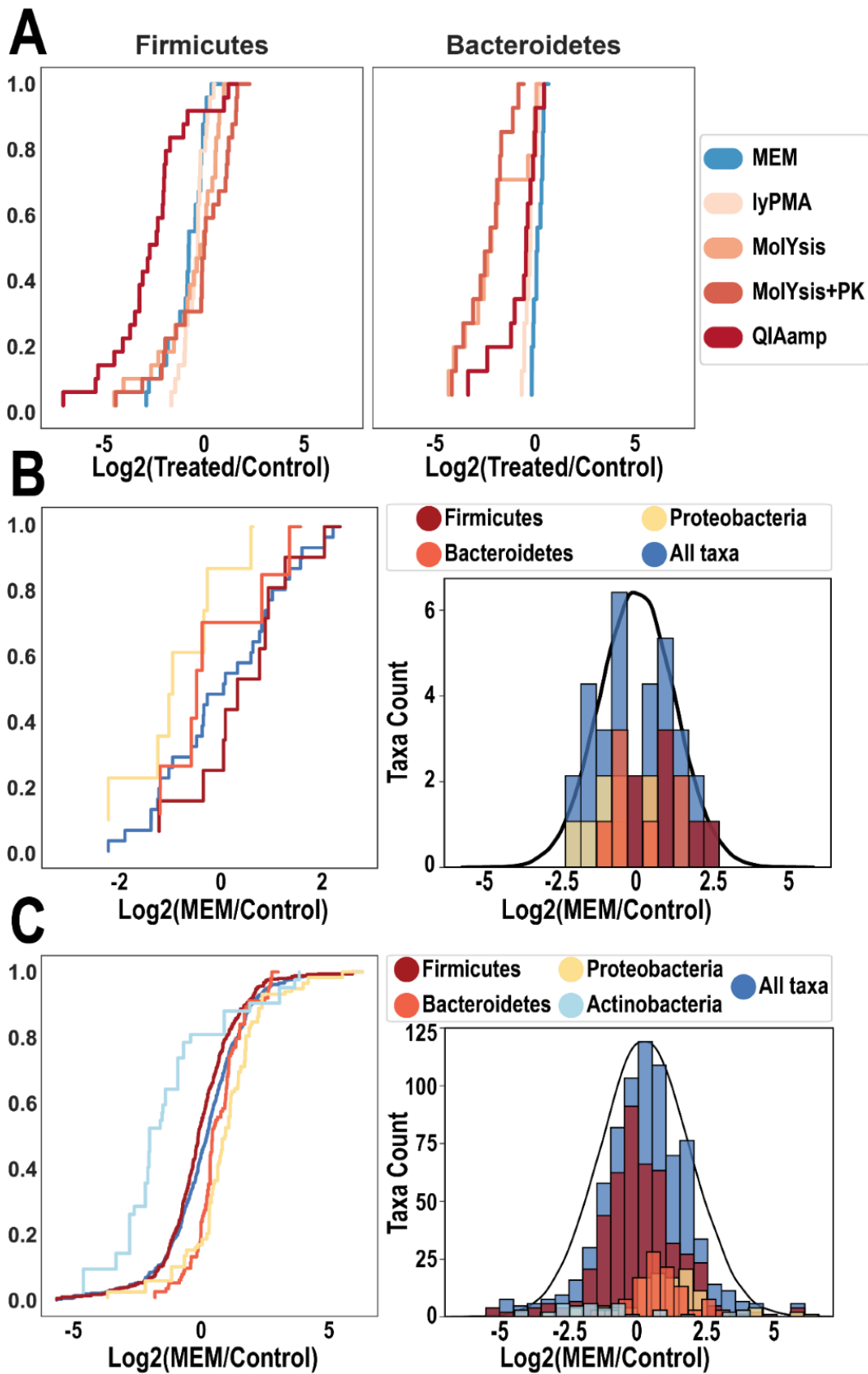


Fig S2: Impact of host depletion on specific bacterial phyla. Log2 fold-change between relative abundance of genera within each phylum in treated and control samples from 16S rRNA gene sequencing data for A) mouse stool with a variety of host depletion methods (N=3 biological replicates), B) human saliva on paired MEM-treated and untreated controls (N=3 technical replicates), and C) human biopsies on paired MEM-treated and untreated controls (N=4 biological replicates). The histograms in B-C) are overlaid with a normal distribution (black line).

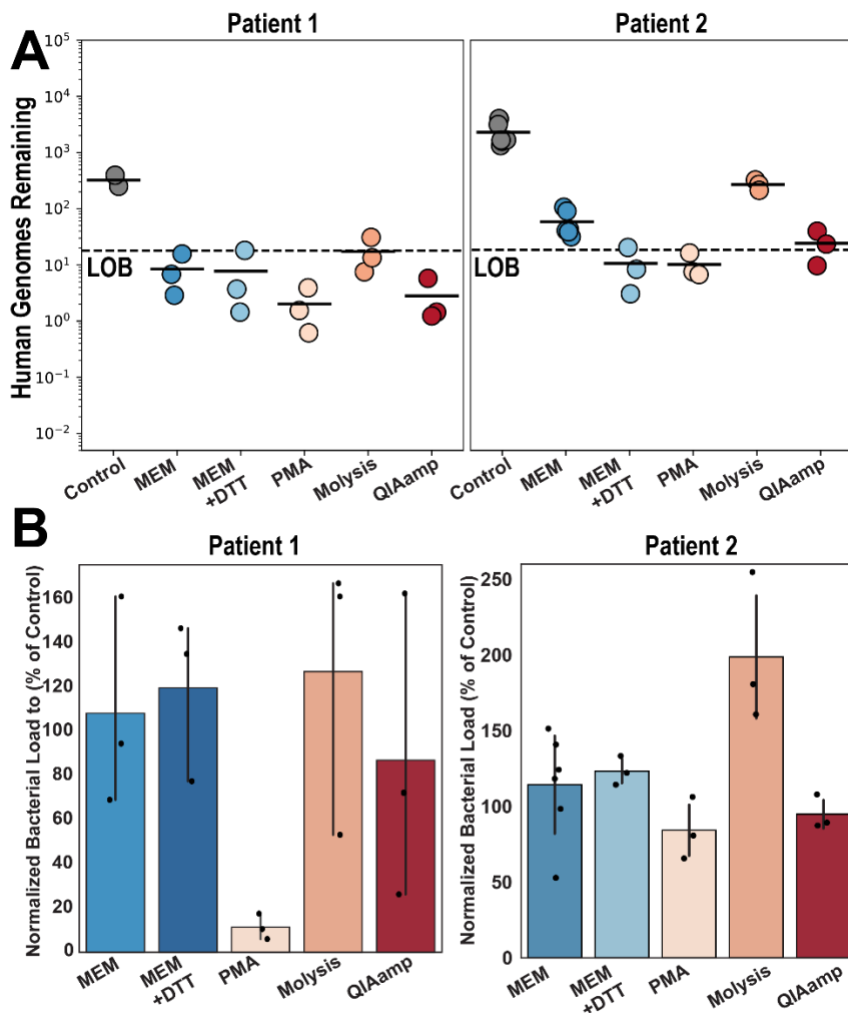


Figure S3: Saliva host depletion variation across participants. All commercial host depletion protocols and MEM were performed on two separate saliva donors a total of three times per donor. After host depletion, host load and bacterial load were quantified with qPCR of single-copy human primers and 16S rRNA gene primers. A) Host loads in participant 1 saliva was 10-fold lower compared to participant 2 (shown in Fig 1E) we tested. The addition of DTT did not have an effect on the participant with lower host load but caused higher reduction in host in the participant with higher initial host load. Dashed line indicates limit of blank (LOB) defined as $LoB = \text{mean}_{\text{blank}} + 1.645(SD_{\text{blank}})$ based on three processing blanks. B) Additionally, we see minimal loss of microbial DNA from PMA treatment in participant 2 saliva (right) but dramatic microbial losses in participant 1's saliva. This loss in low host saliva may be due to excess PMA that was incompletely inactivated. Disadvantage of lyPMA is it will need to be reoptimized for each sample's biomass. (Patient 1: N=3 biological replicates. Patient 2: N=3 biological replicates for lyPMA, MoYsis, and QIAamp. N=4 biological replicates for Control with N=2-3 technical replicates. N=4 biological replicates for MEM with N=3 technical replicates for one biological replicate. N=3 technical replicates for MEM+DTT; error bars are 95% CI).

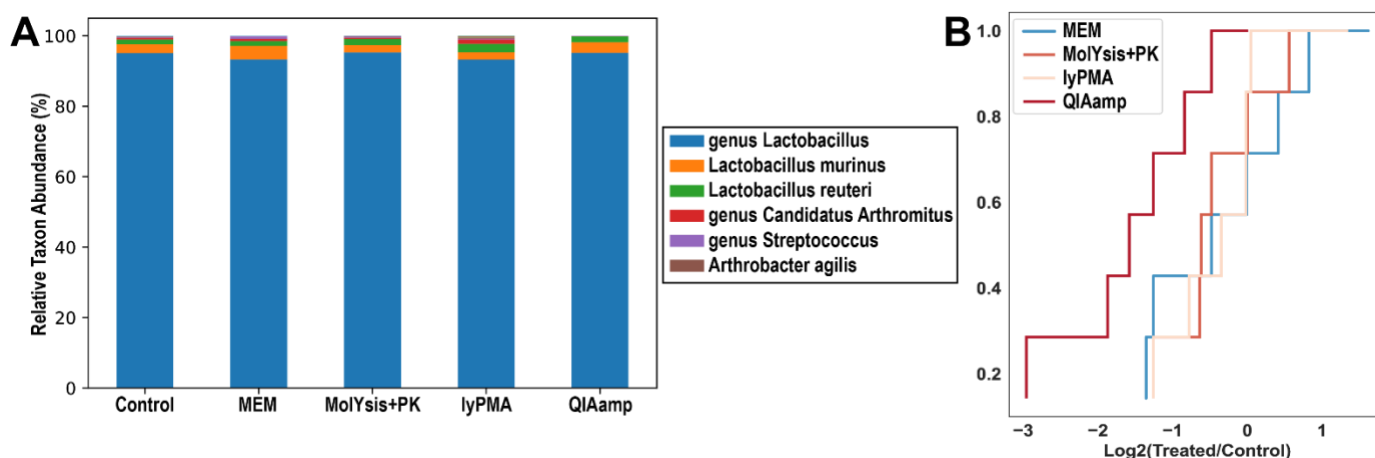


Figure S4: Characterization of mouse mucosal microbiomes following host-depletion protocols. Homogenized mouse intestinal mucosa scrapings were divided into each host depletion method with one sample immediately processed with nucleic acid extraction to serve as a control (N=3 technical replicates). Analysis of mouse mucosa scrapings through 16S rRNA sequencing revealed microbiome is heavily dominated by *Lactobacillus*. Attempted analysis of 16S rRNA sequencing data to characterize bacterial bias are difficult to interpret due to the low biodiversity. In the mucosal scrapings, the well-studied mucosal SFB, *Candidatus arthromitus*, was slightly enriched in MEM and the PMA method whereas it was depleted in the Molysis+PK and QIAamp methods.

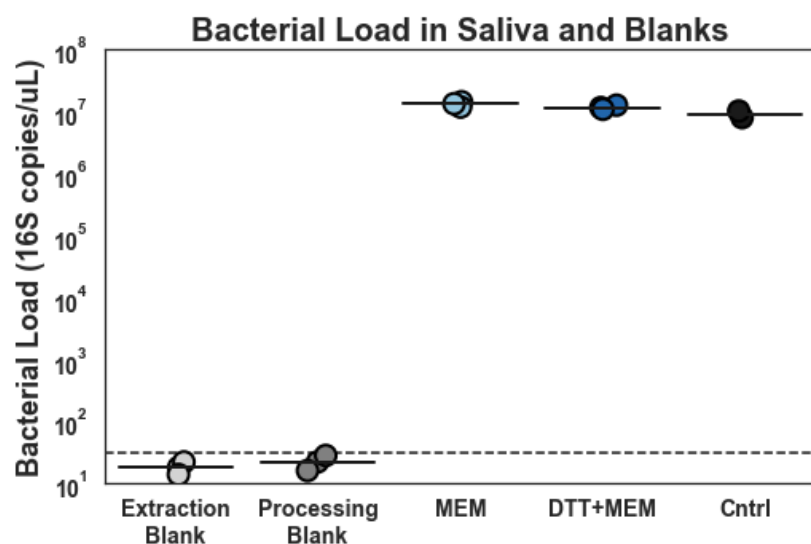


Figure S5: Bacterial load present in saliva and extraction/processing blanks. Bacterial load measurements for saliva samples that were shotgun sequenced compared to processing and extraction blanks. Dashed black line represents the measurement LOB. (N=3 technical replicates).

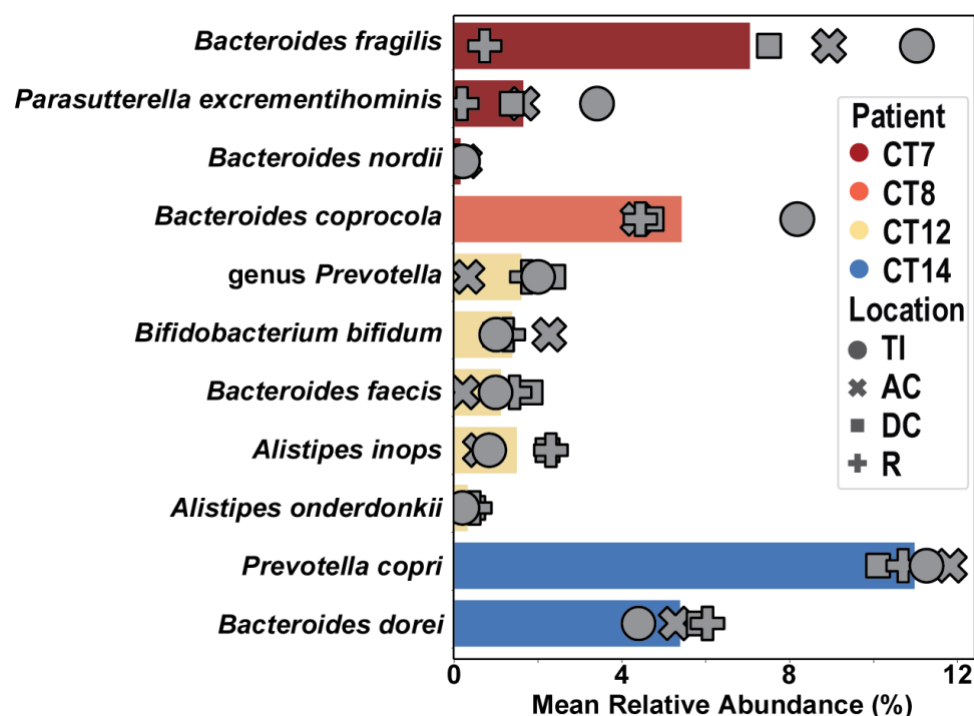


Fig S6: Abundance of participant unique species. From shotgun sequencing of 5 participants longitudinally, unique species were defined as present in only one of the five participants. Relative abundances of some of these unique species are shown with the bars representing the average abundance across all 12 biopsies from one participant.

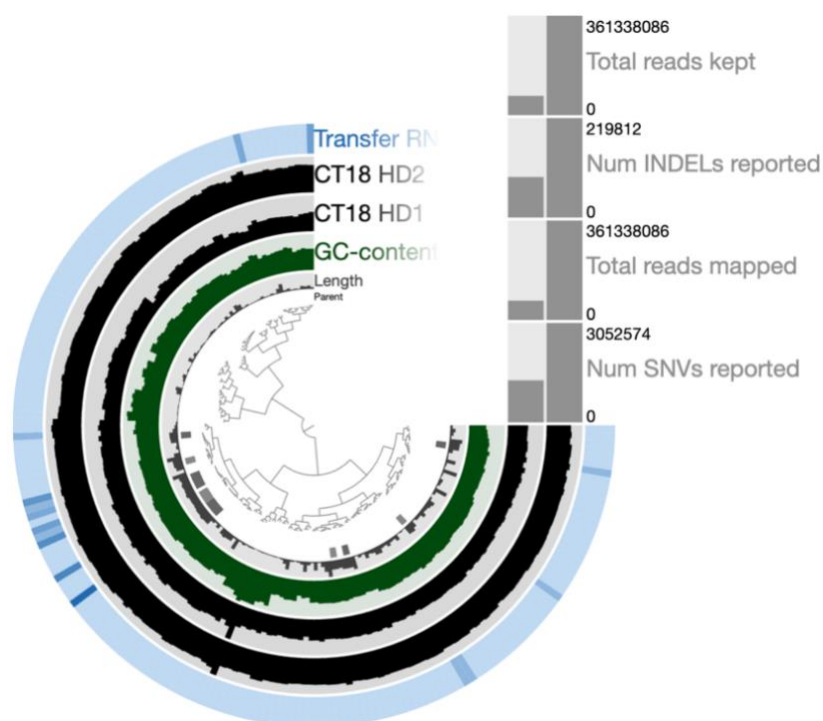


Fig S7: MAG of *Fusobacterium*. From two MEM-treated ascending biopsies from CT18, a MAG of *Fusobacterium* was constructed (completeness: 94%, redundancy: 1.4%).

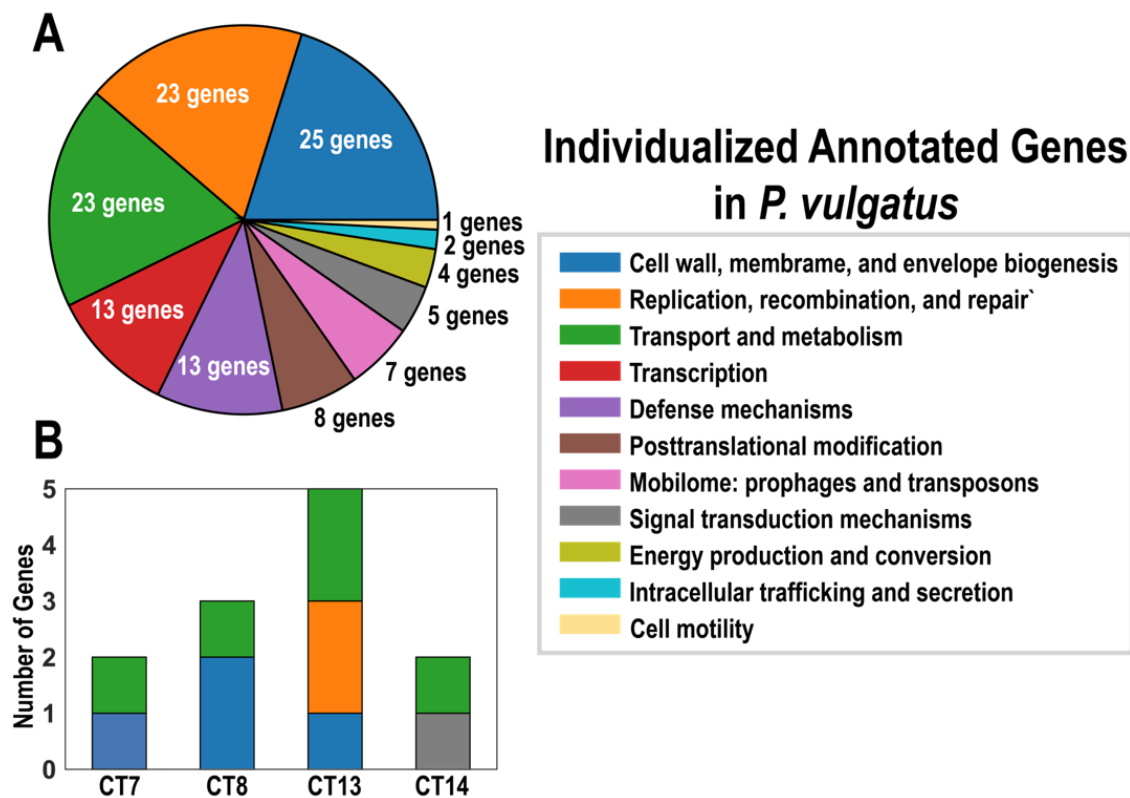


Figure S8: Genes driving participant unique strains of *Phocaeicola vulgatus*. A) 100 annotated genes found in only CT12 were sorted based on COG20 Category and the number of genes in each category are shown. B) The same analysis was repeated for genes found in only CT12 and one other participant (labeled).

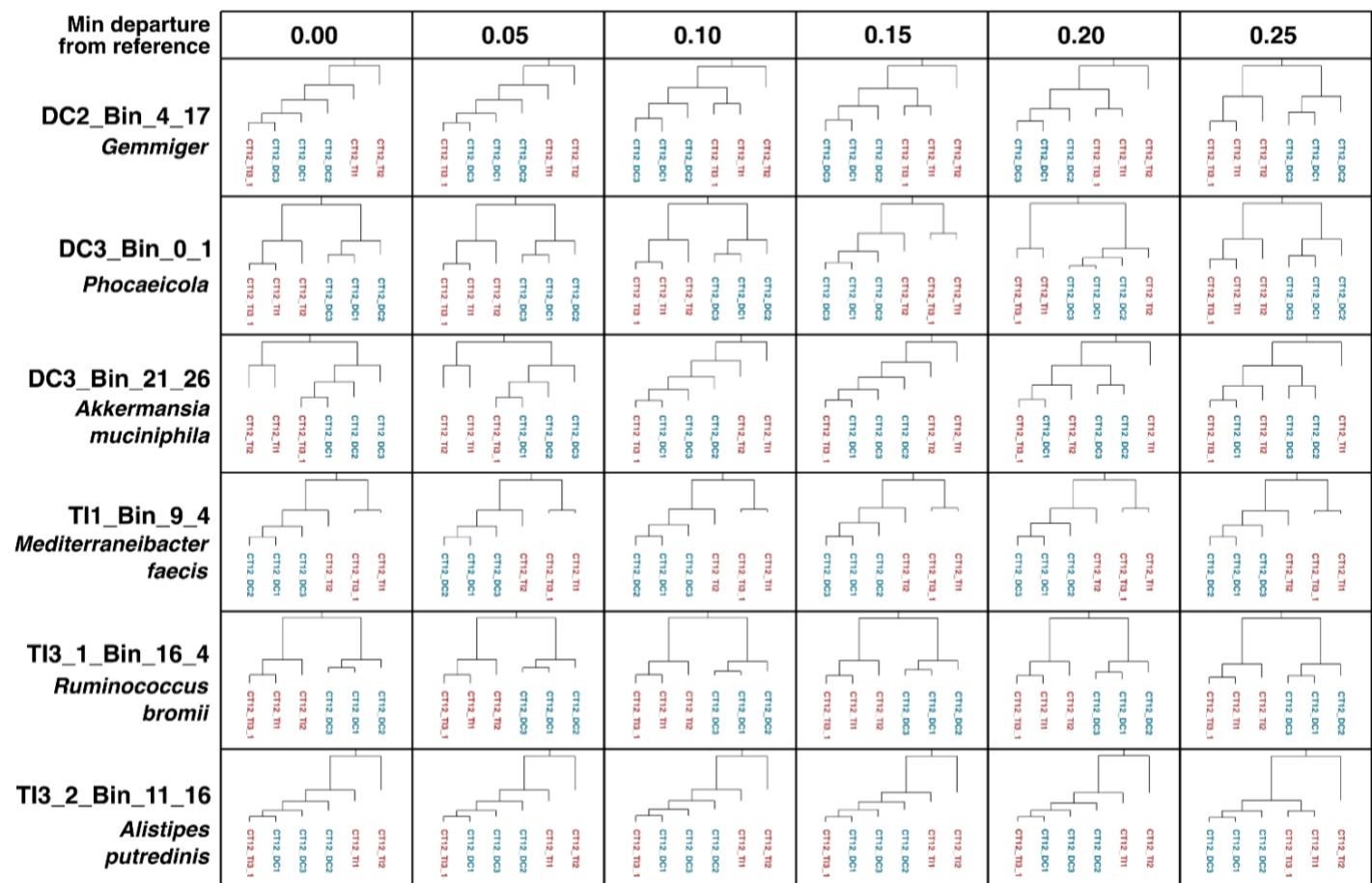


Figure S9: Fixation index across MAGs with varying deviation from reference. Six MAGs with greater than 50X mean coverage were selected for SNPs analysis. Fixation index analysis was performed on each MAG for various thresholds of minimum departure from reference nucleotide. Clustering of fixation index by location can be seen for some MAGs (red indicates terminal ileum samples vs blue are descending colon).

Natalie J. Wu-Woods, nwuwoods@caltech.edu, 0000-0001-5070-9091

Jacob T. Barlow, jtbarlow3@gmail.com, 0000-0002-1842-4835

Florian Trigodet, trigodet.florian@gmail.com, 0000-0002-4933-2896

Dustin Shaw, dgshaw@uchicago.edu, 0000-0003-3961-7588

Anna E. Romano, aromano@caltech.edu, 0000-0002-7148-0668

Bana Jabri, bjabri@bsd.uchicago.edu, 0000-0001-7427-4424

A. Murat Eren, meren@hifmb.de, 0000-0001-9013-4827

Rustem F. Ismagilov, rustem.admin@caltech.edu, 0000-0002-3680-4399

Detailed Author Contribution Statement:

Natalie Jamei Wu-Woods

1. Major contributor to MEM protocol development optimization
2. Performed MEM on all animal and human saliva samples
3. Performed library prep and all bulk and digital quantification experiments
4. Performed 16S rRNA gene and shotgun sequencing analysis
5. Performed all MAG construction
6. Constructed figures 1A-C, 1E-F, 3A, 4, 5, 6. Constructed SI figures 1-5, 7-8, and ED figures 1-5. Constructed tables 1-8.
7. Wrote and edited the manuscript
8. Wrote the supplemental information
9. Co-developed the IRB protocol with JTB to obtain healthy human saliva samples
10. Contributed to obtaining funding

Jacob T. Barlow

1. Designed and validated original MEM protocol
2. Major contributor to MEM protocol development optimization
3. Performed MEM on subset of stool samples and a range of validation samples
4. Designed MEM validation experiments (Fig 1 and 2)
5. Set up clinical sample collaboration with UChicago and Dustin Shaw
6. Designed sampling regime for all clinical samples collected from UChicago in collaboration with DS (Fig 3 and 4)
7. Developed pipeline for raw sequencing analysis of 16S data (Fig 1-4) and shotgun marker gene data (Fig 2,4)
8. Performed analysis for and generated figures 1D-G, 2, 3B-E, and the concept behind 4D
9. Constructed SI figure S6
10. Co-developed the IRB protocol for saliva samples with NW to obtain healthy human saliva samples
11. Wrote initial paper outline (only figures 1-4)
12. Helped edit paper

Florian Trigodet

1. Co-developed analysis plan for figures 5-6 with NW, RFI, and AME
2. Performed SNVs analysis utilized in figure 6
3. Constructed SI figure S9 and table S9-10
4. Co-designed figure 5D and 6A with NW
5. Provided feedback and advice on all shotgun sequencing analysis
6. Wrote and edited the manuscript

Dustin Shaw

1. Performed MEM on human intestinal samples
2. Designed sampling regime for all clinical samples collected from UChicago in collaboration with JTB (Fig 3 and 4)
3. Confirmed feasibility of MEM on samples obtained from the clinic

Anna E. Romano

1. Discussed various techniques for extraction modification of saliva specimens to increase DNA/RNA yield and remove inhibitors.
2. Discussed and researched shotgun sequencing library preparations and sequencing options.
3. Discussed previous research on RNA quality of different mouse SI tissues.
4. Revised methods.
5. Researched the use of MTM prime store as an alternative preservation solution compatible with Qiagen extraction kit.
6. Discussed, troubleshooted, and assisted in shotgun sequencing library prep along side NW. Contributed knowledge about library preparation of low input samples and bubble product formation.

R.F.I. contributed to the design and implementation of the study, to obtaining funding, and to editing the manuscript. A.M.E. oversaw the bioinformatic analysis, contributed to the design and implementation of the study, to obtaining funding, and to editing the manuscript. B.J. supervised the clinical work, contributed to the design and implementation of the study, to obtaining funding, and to editing the manuscript.