

PROJECT SUMMARY

Diet, Genetic Factors, and the Gut Microbiome in Crohn's Disease

Dr. Gary D. Wu, University Of Pennsylvania
Dr. James D. Lewis, University Of Pennsylvania
Dr. Frederic D. Bushman, University Of Pennsylvania

I. PROJECT ID NUMBER, PUBLICATION MORATORIUM INFORMATION, PROJECT DESCRIPTION:

This manuscript is part of a pilot effort on the part of NIH staff and the Nature publishing group to provide a more convenient archive for "marker papers" to be published. These "marker papers" are designed to provide the users of community resource data sets with information regarding the status and scope of individual community resource projects. For further information see editorial in September 2010 edition of Nature Genetics (Nature Genetics, 42, 729 (2010)), and the Nature Precedings HMP summary page.

Project IDs: 46313, 46879

Publication moratorium: 1 year

We propose to investigate the hypothesis that consistent changes in the human gut microbiome are associated with Crohn's disease, a form of inflammatory bowel disease, and that altered microbiota contributes to pathogenesis. Analysis of this problem is greatly complicated by the fact that multiple factors influence the composition of the gut microbiota, including diet, host genotype, and disease state. For example, data from us and others document a drastic impact of diet on the composition of the gut microbiome. No amount of sequencing will yield a useful picture of the role of the microbiota in disease if samples are confounded with uncontrolled variables. Our proposed project will take advantage of our experience with deep sequencing to characterize the composition of the gut microbiome, but also control diet, host genotype, and disease state.

Our initial study, known as Fecal Storage Methods (FSM), was designed to systematically evaluate methods for surveying bacterial communities in human feces using 454/Roche pyrosequencing of 16S rRNA gene sequences, generating a total of 797,276 tags. We analyzed fecal samples from 10 individuals and compared methods for fecal storage, DNA purification and sequence acquisition. To assess reproducibility, we compared samples one cm apart on a single stool specimen for each individual. To analyze storage methods, we compared 1) immediate freezing at -80°C, 2) storage on ice for 24 hours, 3) on ice for 48 hours, or 4) fresh samples. For DNA purification methods, we tested three commercial kits and prior bead beating in hot phenol. Six 16S rDNA primer sets were also compared. We plan to utilize these data to optimize protocols to collect, process and sequence bacterial 16S rDNA from fecal samples in our subsequent studies.

Additional studies that have been initiated include Controlled Feeding Experiments (CaFE) that address the effects of controlled diets on gut microbiome composition in normal human subjects sequestered in the Penn CTRC. CaFE 1 examines the effect of a low fat/high

carbohydrate vs. high fat/low carbohydrate diet on gut microbiome composition whereas CaFE 2 examines the effect of caloric intake. Data acquisition also includes metagenomic shotgun sequencing of DNA isolated on the first (day 1) and last (day 10) stool samples from subjects enrolled in the CaFE 1 study.

To determine the effect of diet on gut microbiome composition in the outpatient setting, we have also initiated a Cross-sectional Study of Diet and Stool Microbiome Composition (COMBO). Healthy volunteers age 2 to 50 are recruited who provide demographic data, a detailed diet history including measurement of usual diet with the Harvard Food Frequency Questionnaire (FFQ) and recent diet using three 24 hour diet recall interviews in the week immediately preceding sample collection.

II. DATA QUALITY:

We adopt the following quality control for the 16S rDNA 454 sequencing data: (1) the sequences need to show a perfect match to the bar code and 16SrRNA gene primer; (2) sequences need to be close in length to the run average; (3) the sequences have no more than one undetermined bases in the sequence read and (4) the sequences need to find at least a 75% match to a previously determined 16SrRNA sequence after alignment with NAST (most matches are much higher).

III. DATA ANALYSIS AND PUBLICATION PLANS:

To characterize microbiome samples, we purify DNA from mixed communities in stool or mucosal biopsies using optimized extraction methods based on our FSM study and use 454/Roche pyrosequencing for sequence acquisition. For first pass analysis, we use 454/Roche Titanium pyrosequencing of 16S rDNA gene segments. In some cases, we also characterize communities using shot-gun metagenomic analysis. The 16S reads are clustered into operational taxonomic units (OTUs) and aligned to curated databases of 16S sequences. Communities are characterized by standard measures of richness, diversity and evenness, and compared to each other using weighted or unweighted UniFrac. The UniFrac distance measures are used directly for comparison and association to metadata (e. g. diet, disease activity) or subjected to Principal Coordinate Analysis to visualize clustering or gradients. All the steps are automated within the QIIME software package. The inferred bacterial taxa are also visualized using our 16S Browser, which allows visualization of data as stacked bar graphs at user-configured taxonomic levels grouped by diverse forms of metadata. For shotgun metagenomic analysis, sequences are aligned to databases using BLAST or Phymm and PhymmBL, then, taxonomic information extracted using MEGAN. Gene types are summarized into functional groups based on KEGG or SEED pathways. Counts on gene types can be visualized and analyzed as heat maps separated out by bacterial Phylum.

Our first study, Fecal Storage Methods (FSM), evaluated the optimal method to collect, store, process, and sequence human fecal samples for 16S rDNA analysis. These results established the laboratory methodology for the rest of our projects. A manuscript entitled, "Sampling and pyrosequencing methods for characterizing bacterial communities in the human gut using 16S sequence tags" has been accepted for publication in BMC Microbiology 2010, 10:206.

Our next study, Cross-Sectional Study of Diet and Stool Microbiome (COMBO), provides evidence that there is an association between dietary intake, as determined by dietary questionnaire, and the composition of the gut microbiome in healthy subjects in the outpatient setting. In addition, certain demographic features were also associated with microbiome composition. We anticipate that a manuscript will be submitted for publication describing these results within the next year.

Our third study is a Controlled Feeding Experiment (CaFE). The results show that both dietary composition and caloric intake have a rapid and independent effect on gut microbiome composition, both bacterial and viral, supporting the rationale and feasibility of using the CaFE design to test the impact of dietary components in a controlled setting. We anticipate that a manuscript will be submitted for publication describing these results within the next year, perhaps with the results of the COMBO study described above.

In our fourth study, we have sequenced viral preparations from longitudinal stool samples. The results allow investigation of the composition of the gut virome, and specify considerable novel functionality. Longitudinal analysis specified that a dietary intervention was associated with a transition in the human virome to a new stable state.

In our fifth study, we will be examining the effect of an elemental diet on the composition of the gut microbiome in a study named, Pediatric Longitudinal Study of Elemental Diet and Stool Microbiome Composition (PLEASE). Elemental diet therapy is often effective in treating pediatric Crohn's disease, allowing us to investigate the microbiome changes associated with successful therapy, failed therapy, and relapse. Longitudinal studies, will allow us to specify microbial changes associated with successful or failed elemental diet therapy. This study has yet to be fully initiated so plans for publication cannot be determined at this time.

IV. DATA RELEASE PLAN:

We will comply with the resource sharing policy articulated for the HMP at <http://nihroadmap.nih.gov/hmp/datareleaseguidelines.asp>. Clinical metadata will be deposited into dbGaP at NCBI. Raw sequence data will then be deposited into SRA as it is generated, after filtering sequences for human DNA using BLAST.

V. CONTACT PERSONS:

Dr. Gary D. Wu, University Of Pennsylvania, 215-898-0158, gdwu@mail.med.upenn.edu
(Study design, sample collection and processing)

Dr. James D. Lewis, University Of Pennsylvania, 215-573-5137, jlewis@cceb.med.upenn.edu
(Study design, IRB issues, subject recruitment, and statistical analysis)

Dr. Frederic D. Bushman, University Of Pennsylvania, 215-573-8732,
bushman@mail.med.upenn.edu (Study design, DNA sequencing and data analysis)