

Step I: White Paper Application

Application Guidelines

1. *The application should be submitted electronically per requirements via the web site of any of the NIAID Genomic Sequencing Centers for Infectious Diseases. Include all attachments, if any, to the application.*
2. *There are no submission deadlines; white papers can be submitted at anytime.*
3. *GSC personnel at any of the three Centers can assist / guide you in preparing the white paper.*
4. *Investigators can expect to receive a response within 4-6 weeks after submission.*
5. *Upon approval of the white paper, the NIAID Project Officer will assign the project to a NIAID GSC to develop a management plan in conjunction with the participating scientists.*

White Paper Application

Project Title:

Impact of antibiotic administration on the establishment and development of infant gut flora

Authors:

David S. Newburg
Director, Program in Glycobiology
Pediatric Gastroenterology & Nutrition,
Massachusetts General Hospital

Ardyth L. Morrow
Professor and Director, Division of Biostatistics & Epidemiology
Cincinnati Children's Hospital Medical Center

Doyle V. Ward
Research Scientist
Broad Institute

Primary Investigators Contact:

Name	Doyle V. Ward
Position	Research Scientist
Institution	Broad Institute
Address	7 Cambridge Center, Cambridge
State	MA
ZIP Code	02142
Telephone	617-714-7906
Fax	
E-Mail	dward@broadinstitute.org

1. Executive Summary *(Please limit to 500 words.)*

The microbial composition of the gut likely contributes to a wide-range of health and disease states including intestinal inflammation, atopic disease, and possibly diseases of adulthood, such as heart disease and obesity. The early establishment of the gut microflora is suspected to have a particularly profound impact protecting the gut from infectious disease and on long-term subsequent health by predisposing individuals to atopic or autoimmune disease later in life. In contrast to the large-scale efforts of the Human Microbiome Project to characterize the microbial flora of healthy adults, relatively little has been accomplished to characterize the early establishment of the microbiota in infants.

The administration of antibiotics to infants has the capacity to profoundly affect the early acquisition, establishment, and natural maturation of commensal microflora. Prematurity is another major factor that is thought to influence 'normal' gut microbial composition in early life. A high number of pre-term infants receive antibiotic therapy prophylactically; the immediate and long-term effects of antibiotic treatment on infant and later health outcomes are unknown. We hypothesize that gestational age influences the ability to support 'normal' gut microbiota, and that antibiotic administration in the first few weeks of life can alter or delay the development of a 'normal' microbiota and may lead to adverse health outcomes. If we hope to promote the

healthy establishment of early gut colonization and thereby improve health outcomes, we need to establish what 'normal' colonization looks like across gestational ages and apply this knowledge as a benchmark to understand the impact of antibiotic treatment in early postnatal life.

In this project, we propose to examine the microbiome in four cohorts ranging from extremely preterm to term, with birth weights appropriate for gestational age: Group 1) 23 to 27 weeks and extremely low birth weight [ELBW, <1000 grams]; Group 2) 28 to 32 weeks and very low birth weight [VLBW, >1000 to <1500 grams]; Group 3) 33 to 36 weeks and low birth weight [LBW, >1500 to <2500 grams]; and Group 4) term (37 to 42 weeks) and normal [>2500 to <4000 gram] birth weight. The infant gut microbiome will be assessed longitudinally in study infants from the first postnatal week to 18 months of age, and will include untreated and antibiotic-treated infants.

Specific goals of the project

1. Establish a reference quality, sequence-based characterization of the early dynamic succession of the early microbiota in early and late preterm, near term and term infants who have not been treated with antibiotics by longitudinally sampling stool from 1 week to 18 months of age.
2. Determine the impact of antibiotic use on the early stages of colonization and the resilience of the early microbiota within each gestational age group.
3. Test the association between variation in the infant microbiome and infant health outcomes.

2. Justification

Background

Bacterial colonization of the infant. The human adult colon contains a microbial ecosystem of more than 400 bacterial species of primarily anaerobic genera, including Bacteroides, Eubacterium, Clostridium, Ruminococcus, and Faecalibacterium [1]. However, the precise composition of microbes varies across individuals, and in adults each of these unique microbial communities has a composition that remains stable for months [1-4]. In contrast, the infant gastrointestinal microbiota varies over time. In the first year, the infant intestinal tract goes from initially being sterile to an extremely dense colonization that becomes increasingly similar to that of the adult intestine [5]. The sterile gastrointestinal tract of the newborn is inoculated with microflora from the birth canal, close maternal contact, and the environment. Initial colonization is followed by waves of succession. The pioneering species include facultative anaerobes such as lactobacillae and bifidobacteria that initially establish the acidic anaerobic environment in which the subsequent, increasingly complex microbiota is established. A secondary wave of succession is noted at around 4 months by the increasing presence of bacteroides species, including *B. fragilis* and *B. thetaiotaomicron*, which become dominant species in the adult microflora.

The details of bacterial succession in early colonization of the gut are not known, nor are the factors that promote colonization by specific types of symbiotic bacteria. Many reports indicate that Bifidobacteria dominate the microbiota of breast-fed infants [5-8], but other reports indicate that this pattern occurs in only a small fraction of infants [9; 10]. A lower abundance of Bifidobacteria and a higher abundance of aerobic bacteria have been reported to exist in microbiota of formula-fed vs. breast-fed infants [8; 9; 11-13], but even this difference is not

universally observed [14; 15]. Vaginal delivery is often cited as a key factor that results in a healthy infant microbiota [6; 16; 17]. Microbiota of infants delivered by caesarean section differ from that of infants delivered vaginally in the natural progression of colonization and in composition [6; 18-21], but the clinical significance of these differences is unclear.

Preterm infants have microbiota distinct from that of healthy, full-term infants [20; 22-25]. The environment of the neonatal intensive care unit (NICU) and the frequent use of antibiotics are thought to contribute to aberrant gut colonization [26], which in turn is thought to contribute to risk of enteric diseases, including necrotizing enterocolitis (NEC). A variety of different organisms can become dominant in the unstable environment of the infant gut [20; 24; 25]. The neonatal microbiome of 20 premature infants was characterized by low diversity [25]. Greater understanding is needed of the origins and development of the gastrointestinal microbial ecosystem in order to prevent the morbidity and mortality associated with major diseases of preterm infants.

Antibiotic use and outcomes in premature infants. In premature infants, most of the antibiotic administered is empirical, based not on positive culture results but on clinical intuition or concern for potential risk of infection. In 5693 extremely low birth weight infants enrolled in 19 sites of the NICHD Neonatal Research Network, 71% of the premature infants who survived past the first week of life received some empirical antibiotic therapy, and 56% received it for 5 or more days [26]. In our study of 360 very low birth weight infants who survived the first 14 days without sepsis or necrotizing enterocolitis, 142 (39%) infants had received prolonged (≥ 5 days) antibiotic therapy in the first 14 days of life (Kuppala et al, unpublished). Unnecessary antimicrobial use can have serious, unintended consequences in premature infants, such as rapidly increasing drug resistance in sepsis cases [27; 28]. Broad-spectrum antibiotics use strongly increased risk of invasive fungal infection [29; 30]. We find that prolonged empirical antibiotic use in the first 14 days of life is significantly associated with subsequent risk of necrotizing enterocolitis, sepsis and death. We hypothesize that these adverse outcomes may be due to negative changes in the microbial community of the gastrointestinal tract, suggesting an urgent need to understand the impact of antibiotics on the microbiome.

Scope of work proposed

Study design. A cohort of 240 infants will be enrolled at Good Samaritan Hospital (GSH) and University Hospital (UH), which combined have about 10,000 deliveries per year. GSH and UH are major teaching hospitals in Cincinnati, OH, and each has a level III Neonatal Intensive Care Unit (NICU) staffed by faculty of the University of Cincinnati College of Medicine, Cincinnati Children’s Hospital Medical Center. Study infants will be enrolled over an 18 month period, stratified into eight groups. Seven groups will be infants who survive the follow-up period free of necrotizing enterocolitis (NEC). These infants will be defined by the combination of four gestational age

Table 1. *Stools collected by postnatal age, gestational age and antibiotic use of 210 infants who survive free of NEC.*

Postnatal age	Gestational age groups						
	23-27 wks		28-32 wks		33-36 wks		37-42 wks
Abx use at 5 days	NO	YES	NO	YES	NO	YES	NO
5-8 d	30	30	30	30	30	30	30
11-14 d	30	30	30	30	30	30	30
18-21 d	30	30	30	30	30	30	30
2-3 mo.	29	29	29	29	28	28	28
5-6 mo.	27	28	27	28	25	25	25
11-12 mo.	23	22	23	22	25	25	25
17-18 mo.	20	20	20	20	20	20	20

enterocolitis (NEC). These infants will be defined by the combination of four gestational age

groups and neonatal antibiotic use. The gestational age groups will be defined as 23-27 weeks, 28-32 weeks, 33-36 weeks, and 37-42 weeks. Neonatal antibiotic use will be defined as antibiotic use for the first 5 days of life or longer and will be compared to no antibiotic use. Cumulative antibiotic use will also be measured. The three preterm groups will include both antibiotic and no antibiotic use groups, whereas the term infants will include only a no antibiotic use group (table 1). The eighth study group will be all infants who develop NEC or die during the follow-up period.

All study infants will have data and stool sample collection performed at postnatal days 5-8, 11-14, 18-21, and months 2-3, 5-6, 11-12, and 17-18.

Subject enrollment: Infants will be enrolled in hospital by consent of the parents. Infants with known congenital anomalies who die or develop NEC in the first postnatal week will be excluded from study. Enrollment will not be considered final until the first three visits are completed, so that any study infants who drop out at or before 21 days will be replaced, thereby ensuring follow-up of 30 infants per group to at least 3 postnatal weeks. Drop-outs after that time will not be replaced, but is not expected to be high. We have conservatively estimated a minimum of 20 infants per group followed to 18 months of age. Among preterm infants (≤ 32 weeks gestational age), comprehensive data collection is being conducted in the NICUs of GSH and UH as part of the NICHD-sponsored Novel Biomarkers Study until discharge (pink region of table 1). Among late preterm infants (33-36 weeks GA), who are not part of the Novel Biomarkers Study protocol, infant enrollment and data collection will be performed in the same NICUs in the first visit 5-8 days, and as long the infants are in hospital thereafter (blue region of table 1). In term infants, infant enrollment will be achieved after delivery prior to discharge from hospital. In term and all other infants after discharge from hospital, study visits will be performed by trained staff at home, or if preferred, at their pediatrician's office (brown region of table 1). Among infants 23-27 weeks, 28-32 weeks, and 33-36 weeks GA, enrollment will be stratified on no antibiotic use versus antibiotic use of five or more days after birth, with 30 infants per group. These gestational age groups each have substantial antibiotic use, respectively, about 40%, 20%, and 5% receive antibiotics for 5 or more days, and the number of infants in each category for whom five or more days of antibiotic use occurs is sufficient to meet the target enrollment goals. Among term infants, however, far less than 1% receive antibiotics for 5 or more days after delivery. Thus, we will restrict enrollment of term infants only to those not given antibiotics after delivery as the standard to which the other groups are compared. To balance race/ethnicity across the gestational age and antibiotic use groups, we will enroll approximately one-third African-American and two-thirds non-African American infants in each study group. This balance may be important due to potential (but unknown) differences in bacterial colonization influenced by socio-cultural influences.

Of the anticipated 300 infants ≤ 32 weeks GA to be enrolled in the Novel Biomarkers Study in Cincinnati during the course of this project, approximately ten percent are expected to develop NEC or die. Samples from the estimated 30 infants who develop NEC or die will be included as a separate group for analysis.

Sample collection: Stool samples will be collected from study infants at 8 time points by trained nurses following a standardized protocol supervised by the research coordinator (B. Davidson) and study fellow (C. Greenwood). The 8 time points of collection are indicated in table 1, ranging from day 5 to month 18. Approximately 1400 to 1800 samples are expected to be collected and tested, depending upon the drop-out rate and need for replication to replace lost or doubtful samples. Substantial care will be taken in the collection and management of stool samples to ensure integrity and sufficient quantity for DNA extraction and testing. Stool samples will be collected from the study infant in the morning, and immediately placed in the study

refrigerator, labeled by the nurse. Laboratory staff will place samples in thioglycolate medium prior to cryogenic storage and preparation for shipping to Boston for DNA extraction, amplification, and testing. Stool sample collection from study infants will be performed in hospital or once discharged, at home, and will be accompanied by data collection regarding any antibiotic use or physician-diagnosed illness.

Data collection: The data to be collected regarding study mothers will include race, ethnicity, parity, and C-section or vaginal delivery. The data to be collected regarding study infants will include birth weight, gestational age at birth, age at sample collection, occurrence of death, NEC, sepsis, and antibiotic use.

3. Rationale for Strain Selection

4a. Approach to Data Production: Data Generation

Stool Extraction: Genomic DNA will be prepared from stool samples using protocols based on the methods established by the NIH Human Microbiome Project Jumpstart consortium. Extracted gDNA will be submitted to the Broad Institute for metagenomic sequencing and analysis.

454 Titanium sequencing of bacterial 16S rDNA: Bacterial 16S rDNA sequences will be produced by the Broad Institute using current production protocols established for the human microbiome project. The current protocol incorporates barcoded 454 Titanium adaptors, developed and validated by Broad, in the 16S rDNA amplification to facilitate multiplexing of samples. We will produce 5,000-10,000 passing quality 16S reads per sample. This project proposes to provide a minimum of 1400 infant stool samples over 2 to 3 years for metagenomic sequencing (240 subjects with 7 samples and additional samples for replicates and method validation for transport, DNA extraction and amplification of stools).

4b. Approach to Data Production: Data Analysis

Reads will be pre-processed prior to analysis. Preprocessing will include generation of read-based characteristics such as taxonomic classification and read quality and length parameters analogous to preprocessing of read metadata currently implemented at Broad and the HMP DACC for the HMP Jumpstart project.

Community analysis to be performed at Broad will include:

Estimation of species richness (number of OTUs). One metric to describe a microbiome is the number of OTUs found in that microbiome. An estimate of the total number of operational taxonomic units (OTUs) present in a community can be derived from metagenomic sequence data. We implement several estimators of species richness in EstimateS (2), including Chao1, ACE and ICE. In addition, we will use rarefaction curves to determine the lower bound of species richness, as rarefaction procedures do not estimate the existence of OTUs not discovered. We also fit the distributions of the abundant and rare species to the lognormal, Poisson, and gamma distributions (1). We will also use a sequence-based approach to determine the number of OTUs using CD-hit. Sequences will be clustered into OTUs of a certain relatedness (i.e., >97%). While CD-hit is not optimal, when compared to programs like DOTUR for clustering, DOTUR will be unable to handle the amount of sequence data generated in this project.

Estimation of species diversity (i.e., 'skew'). We will calculate Fisher's alpha diversity, Shannon's, and Simpson's diversity indices, along with the standard deviations associated with these means implemented in EstimateS at different taxonomic levels. This will indicate how evenly species are distributed within different classes of infants, which will enable statistical comparison of diversity among treatments, subjects, and samples.

Estimation of the effect of birthweight on species richness and diversity. Using the appropriate statistical tests, we will determine the effects of birthweight on species richness (both observed and estimated) and species diversity using t-tests or the appropriate non-parametric statistics if the data are non-normally distributed. These methods can also be used to assess the effects of other metadata such as date since birth. In addition, multifactorial ANOVA will be used to assess the relative importance of birthweight and age, as well as the strength of interaction between them.

The effect of birthweight on community composition. Taxonomic-based methods, at different levels of taxonomic hierarchy, will also be used. To differentiate community samples, we will perform principal components analysis (PCA) on both the frequencies of OTUs as well as their abundance. To examine the effect of rare OTUs, the data will also be transformed into presence/absence data and analyzed with PCA. Similar analyses will also be performed using canonical hierarchy analysis, which like, PCA, reduces the number of axes. Unlike PCA, however, canonical hierarchy analysis assigns a 95% confidence interval in this two-dimensional landscape and thus can quantitatively determine if community samples have different species composition. These analyses will be implemented in JMP 7.0 (SAS Institute). At higher taxonomic levels, the abundances and frequencies of OTUs will also be assessed with Bonferroni corrected G-tests.

To determine the influence of birthweight on community structure, we will use ANOVA models for OTUs of interest. Separate analyses will be conducted using either the frequencies of OTUs or the PCA values for the major axes (derived above). To ascertain the effects of various sources of variation on community diversity, ANOVA will be performed on the diversity indices described above.

To classify community samples based on sources of variation, discriminant analysis will be used. In this method, the likelihood that a community will be assigned to a given class (e.g., healthy or diseased) is estimated based on a set of variables known as predictors, which, in this case, are the individual frequencies of OTUs. One advantage of this method is that it can identify which specific community samples are 'typical' and 'atypical' members of a given class. This method also identifies those OTUs that significantly affect classification of community samples. This would identify particular OTUs associated with various birthweights or ages.

To compare the above infant microbiomes with matched infants who have been treated with a clinically relevant course of antibiotics, communities will be described using species richness, species diversity, PCA, and individual OTU frequencies. We will adopt the same analytical approaches, except that ANOVA and discriminant analyses will also incorporate the presence or absence of antibiotic therapy as an additional factor. Consequently, we will not only be able to determine the ecological effects of age and birthweight—and their interaction—but we will also determine how role antibiotic use might alter community structure, species richness, and species diversity.

6. Community Support and Collaborator Roles: Infant Gut Metagenomics Consortium

This project brings together diverse research expertise to address fundamental issues associated with the acquisition and maturation of infant gut flora and human health. Members will collaborate to share specific expertise and insights. Drs. Newburg and Morrow have been

working together on a program project whose funding for over 30 years is to understand the role on human milk components, especially the human milk glycans, on promoting infant health. The interrelationship between variable expression of glycans in target tissues, expression in milk, and disease risk has led to our strong interest on the interrelationship between these factors and establishment of the microbiome, and particularly how this may be influenced by other factors, such as antibiotic use. With the metagenomic capacities developed by Drs. Birren and Ward and associates at the Broad institute, the outstanding clinical research expertise, including exceptional quality sample acquisition and management at Cincinnati Children's Hospital, and the biochemistry and mucosal biology expertise in Dr. Newburg's laboratory, longstanding questions regarding intestinal colonization can now be addressed in a meaningful way.

Scientists participating in and affiliated with this project include.

David S. Newburg, PhD
Assoc. Professor - Glycobiology, microbiota
MGH/Harvard

Ardythe L. Morrow, PhD
Professor - Neonatal epidemiology, human milk, NEC & sepsis
Cincinnati Children's

Kurt R. Schibler, MD
Assoc. Professor – Neonatology/clinical research, NEC
Cincinnati Children's

Zhuoteng Yu, PhD
Research fellow - Microbiome analysis
MGH/Harvard

Ying Ying, MD
Research fellow - Cell biology/mucosal immunology
MGH/Harvard

Corryn Greenwood, MD
Clinical fellow – Neonatology, twinship and microbiome
Cincinnati Children's

Venkata Kuppala, MD
Clinical fellow – Neonatology, antibiotic use
Cincinnati Children's

Adrone LaForgia, MD
Clinical fellow - Neonatology fellow, risk of noroviruses
Cincinnati Children's

Jarek Meller, PhD
Associate Professor - Bioinformatics
Cincinnati Children's

Ted Denson, MD
Professor - Pediatric gastroenterology

Cincinnati Children's

Andrew South, MD
Assistant Professor - Neonatology/GI focus
Cincinnati Children's

Lisa J. Martin, PhD
Associate Professor – Genetic epidemiology/GWAS
Cincinnati Children's

Xi Jiang, PhD
Professor – Virology of enteric diseases
Cincinnati Children's

Beena Kamath, MD
Asst Prof. – Neonatology, maternal influences, infant outcomes
Cincinnati Children's

Mekibib Altaye, PhD
Assoc. Prof – Biostatistics/bioinformatics
Cincinnati Children's

Michael Kennedy, PhD
Professor – Metabonomics and the microbiome
Miami University

1. Availability & Information of Strains:

Metadata collection: The metadata to be collected regarding study mothers will include race, ethnicity, parity, and C-section or vaginal delivery. The data to be collected regarding study infants will include birth weight, gestational age at birth, age at sample collection, occurrence of death, NEC, sepsis, and antibiotic use.

No strains will be obtained in this study. Biological samples are expected to be consumed during the generation of sequence data and will not be available for deposition.

7. Compliance Requirements:

7a. Review NIAID's Reagent, Data & Software Release Policy:

NIAID supports rapid data and reagent release to the scientific community for all sequencing and genotyping projects funded by NIAID GSC. It is expected that projects will adhere to the data and reagent release policy described in the following web sites.

<http://www3.niaid.nih.gov/research/resources/mscs/data.htm>

<http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-013.html>

<Each Center to include their website that describes/points to the guidelines>

Once a white paper project is approved, NIAID GSC will develop with the collaborators a detailed data and reagent release plan to be reviewed and approved by NIAID.

Accept Decline

7b. Public Access to Reagents, Data, Software and Other Materials:

16S rDNA sequence data will be filtered to remove non-16S, contaminating reads prior to rapid deposition in the Short Read Archive at NCBI/NLM/NIH.

Metadata described in the above proposal will be submitted by the participating investigators/collaborators to the Broad GSCID at the time of sample submission and the GSCID will submit these data to NCBI in a mutually agreed upon format. Access to these data will also be available through the Broad GSCID website along with links to NCBI according to the time lines to be agreed upon by the Broad GSCID, NIAID, and the collaborators. We suggest that clinical metadata be embargoed for a period of 9 months following submission of trace data to NCBI.

Preprocessed read information of 16S rDNA sequence data will be made available as per timelines for genome annotation. This data will include pre-analysis, read-based characteristics such as taxonomic classification and read quality and length parameters analogous to preprocessing of read metadata currently implemented at Broad and the HMP DACC for the HMP Jumpstart project. This data will be made available on the Broad website and in an NIH supported host site, if available, as per NIAID request.

The Broad GSCID will make available results of analysis performed by Broad no later than upon acceptance of a manuscript or publication.

The Broad GSCID will make available relevant protocols used for sample processing and data generation.

Where applicable, the Broad GSCID will make available code for analysis tools developed for this project at the Broad's microbiome utilities site (<http://microbiomeutil.sourceforge.net/>).

7c. Research Compliance Requirements

Upon project approval, NIAID review of relevant IRB/IACUC documentation is required prior to commencement of work. Please contact the GSC Principal Investigator(s) to ensure necessary documentation are filed for / made available for timely start of the project.

Investigator Signature:

Investigator Name:

Date:

Blank Last Page