

Control Number: 2022-A-81-NGS

Topic 1: Microbial Chatter: Microbial ecology in health and disease

Publishing Title: Noroviruses are one of the leading causes of acute gastroenteritis (AGE)

Author Block: **M. Seabolt**¹, J. Cannon¹, R. Xu¹, A. Montmayeur¹, S. Suh¹, M. Diez-Valcarce¹, F. Bucardo², S. Becker-Dreps³, J. Vinjé¹;
¹Centers for Disease Control and Prevention, Atlanta, GA, ²National Autonomous University of Nicaragua-León, León, NICARAGUA, ³University of North Carolina Chapel Hill, Chapel Hill, NC.

Abstract Body: Noroviruses are one of the leading causes of acute gastroenteritis (AGE), characterized by diarrheal illness and/or vomiting and accounting for approximately one-fifth of AGE cases worldwide. Despite its ubiquity, little is known about the effects of norovirus-associated AGE on development of gut microbiomes in infants, which is hypothesized to play a role in the development of robust immune responses. Using a shotgun metagenomics-based approach to detect and quantify changes in the taxonomic and functional diversity incurred by AGE-induced disturbance, we sequenced stools at time points before, during, and after the first norovirus-associated AGE occurring in five children enrolled in a birth cohort study conducted in León, Nicaragua. Amongst the children, the taxonomic and functional diversity of microbiomes increased substantially during the first symptomatic norovirus episode, with a dominance of Gammaproteobacteria. After recovery from norovirus AGE, there was a return of many of the probiotic microbial taxa typical of breast-feeding infants profiled prior to the norovirus AGE episode. Most notably, our findings demonstrate a temporary disturbance to the microbiome during norovirus-associated AGE that did not alter the development of healthy microbiomes amongst the immunologically naïve children in this birth cohort. The metagenomic approach presented here should be broadly applicable to other AGE-associated pathogens to provide foundational data important for developing probiotic treatments and vaccines.

---PAGEBREAK---

Control Number: 2022-A-114-NGS

Topic 1: Microbial Chatter: Microbial ecology in health and disease

Publishing Title: Leveraging Metagenomic Data from SRA for finding *Candida auris* reservoirs

Author Block: **U. Bagal**¹, R. Welsh¹, J. Sexton¹, V. Nayak¹, M. Jorge¹, A. Morgulis², S. Shiryev², R. Agarwala², N. Chow¹;
¹Centers for Disease Control and Prevention, Atlanta, GA, ²National Center for Biotechnology Information, Bethesda, MD.

Abstract Body: **Objectives:** *Candida auris* is a multidrug-resistant pathogenic fungus, capable of causing invasive infections with high mortality. Despite intense efforts to understand how this pathogen rapidly emerged and spread to healthcare facilities worldwide, the environmental reservoir of *C. auris* remains unknown. Environmental surveillance aimed at identifying origins for emerging pathogens requires extensive sampling that is often not feasible. However, as metagenomic studies expand to encompass more environments, it is now possible to flip the traditional paradigm of collecting and screening individual samples for specific pathogen sequences and instead leverage big data repositories to identify all

potential environments and sample types that contain a pathogen of interest. Here, we present a collaborative effort between CDC and NCBI to identify fungal pathogen sequences in public metagenomic datasets available to date. **Methods** The system is currently designed to run at CDC as a bash pipeline for searching NCBI's Sequence Read Archive (SRA) database housed in the cloud. It downloads only whole genome sequence metagenomic datasets. Reads are aligned to a curated reference assembly database comprising *C. auris* and related species using SRPRISM. The pipeline computes genome coverage for each genome in the reference set using padding. **Results** We successfully created a sequence-based pipeline that scans the SRA metagenomic data for the emerging, multidrug-resistant *Candida auris*. To date, ~300,000 SRA read sets from 2010 onwards have been scanned to produce an output of *C. auris* positive hits with varying genome coverage. Moving forward, we plan to associate data (location, time, environment) from positive *C. auris* hits and integrate it with epidemiologic data to inform public health intervention strategies aimed at breaking chains of transmission from the environment into human populations, as well as monitor for spillback events into the environment. **Conclusion** Next steps will be evaluating positive hits for false positives and adapting the pipeline into a real-time monitoring system. Making use of available sequence data represents nearly endless opportunities for discoveries around innovations in public health, diagnostics, and therapeutics research and development. Before this project, there were no measures in place to monitor and detect *C. auris* reintroductions into human populations. We expect this pioneering collaboration to benefit public health by identifying reservoirs and new routes of *C. auris* exposure, elucidating the origin of *C. auris*, and providing the foundation for using this tool for other emerging pathogens.

---PAGEBREAK---

Control Number: 2022-A-137-NGS

Topic 1: Microbial Chatter: Microbial ecology in health and disease

Publishing Title: Inter-Species Gene Exchange and Selection in the Core Genome of *streptomyces* from Diverse Bat Species

Author Block: M. Montoya-Giraldo, T. Garcia-Aroca, C. P. Andam;

University at Albany, Albany, NY.

Abstract Body: Members of the bacterial genus *Streptomyces* (phylum Actinobacteria) are best known as major bacterial producers of antibiotics and other useful compounds commonly used in human medicine, animal health, and agriculture. They are ubiquitous in nature, often inhabiting diverse soil environments. They are also often associated with animals, yet their diversity in these habitats and the evolutionary processes that shape such diversity are often poorly understood. Here, we investigated the contributions of homologous recombination and selection on the core genome of 73 previously published *Streptomyces* genomes from bats. Of the 1141 core genes, calculation of the pairwise homoplasmy index revealed 924 genes that did not show evidence of recombination and 217 genes that have experienced recombination. We also estimated the ratio of non-synonymous to synonymous polymorphisms (dN/dS) of each core gene and found 116 that show evidence of positive selection. Combining the two analyses, we found six frequently recombining genes that also exhibit positive selection: *ettA3*, *gcvT1*, *guaA*, *lpdA*, *nrdZ*, *sdaA*. We also identified 207 frequently recombining genes that also exhibit negative (purifying) selection. We discuss the functions of these genes in the context of host-microbe interactions and natural drug

discovery efforts. Overall, these results indicate that genes acquired through recombination can be maintained in recipient organisms through selection, thereby facilitating the adaptation to their bat hosts.

---PAGEBREAK---

Control Number: 2022-A-146-NGS

Topic 1: Microbial Chatter: Microbial ecology in health and disease

Publishing Title: Associations among milk microbiota, glycans, lipids, and somatic cells from lactating Holstein cows

Author Block: L. Coates¹, S. Durham², D. Storms¹, A. Magnuson³, D. Van Hekken⁴, B. Plumier⁴, J. Finley⁵, N. Fukagawa⁶, P. Tomasula⁴, D. Lemay¹, M. Picklo⁷, D. Barile², K. Kalscheur⁸, M. Kable¹; ¹USDA-ARS Western Human Nutrition Research Center, Davis, CA, ²UC Davis, Food Science and Technology, Davis, CA, ³USDA ARS Grand Forks Human Nutrition Research Center, Grand Forks, ND, ⁴USDA-ARS Dairy and Functional Foods Research, Wyndmoor, PA, ⁵USDA-ARS George Washington Carver Center, Beltsville, MD, ⁶USDA-ARS Beltsville Human Nutrition Research Center, Beltsville, MD, ⁷USDA-ARS Grand Forks Human Nutrition Research Center, Grand Forks, ND, ⁸USDA-ARS U.S. Dairy Forage Research Center, Madison, WI.

We aimed to determine the relationships among milk microbiota, milk components (lactose, oligosaccharides (MO) and fatty acids (MFA)), and somatic cell counts (SCC), in lactating Holstein cows. Raw milk samples were collected from cows at three timepoints, ranging from early to late lactation. Milk microbial community was determined by 16S rRNA amplicon sequencing, lactose abundance by Fourier transform infrared spectroscopy, MO by nano-liquid chromatography quadrupole time-of-flight tandem mass spectrometry, MFA by fatty acid methyl esters analysis by gas chromatography with flame ionization detection, and SCC (as a measure of mammary inflammation) by flow cytometry. Data were analyzed using (generalized) linear mixed effects modeling and repeated measures correlation. Microbial diversity was not correlated with MO diversity or MFA diversity. However, many MO were positively correlated with potentially pathogenic genera (e.g. *Corynebacterium*, *Enterococcus*, *Pseudomonas*), while numerous MO were negatively correlated with the symbiont *Bifidobacterium*. The neutral, non-fucosylated MO composed of eight hexoses had a positive relationship with SCC, while lactose had a negative relationship with SCC. No significant relationships between microbial taxa, or microbial α -diversity, and SCC were identified. Lactose had negative relationships with several potentially pathogenic genera: *Corynebacterium*, *Pseudomonas*, and *Enterococcus*. Unsaturated MFA and short chain MFA had mostly negative relationships with potentially pathogenic genera, including *Corynebacterium*, *Pseudomonas*, *Acinetobacter*, and an unknown *Enterobacteriaceae* genus, but numerous positive relationships with symbionts *Bifidobacterium* and *Bacteroides*. MO and MFA are likely not dominant factors in the structuring of the milk microbial community. Instead, distinct microbes have unique relationships with distinct milk components. Mammary inflammation may impact these components to promote a more beneficial milk microbiota. The associations of economically and nutritionally important milk components with milk microbes and mammary health suggests that holistic optimization of milk traits is needed.

---PAGEBREAK---

Control Number: 2022-A-147-NGS
Topic 1: Microbial Chatter: Microbial ecology in health and disease
Publishing Title: Machine Learning Reveals Metagenomic Biomarkers for Human Norovirus Gastroenteritis

Author Block: C. Coarfa, A. Koirala, S. L. Grimm, F. H. Neill, S. J. Javornik Cregeen, K. L. Hoffman, J. Petrosino, M. K. Estes, R. L. Atmar, S. Ramani; Baylor College of Medicine, Houston, TX.

Human noroviruses (HuNoV) are a leading cause of acute gastroenteritis across all ages in the United States. Genetic differences in histo-blood group antigen (HBGA) expression drives susceptibility to different HuNoV strains. While antibodies that block HuNoV binding to HBGAs correlate with protection from illness, whether other host and microbial factors influence HuNoV gastroenteritis is unknown. We used samples from a human experimental infection study to determine whether whole genome shotgun (WGS) metagenomic profiles, including taxonomy and enriched pathways, can predict the risk of developing gastroenteritis in genetically susceptible individuals.

Abstract Body: We profiled 604 stool samples from 55 healthy adults who were challenged with the prototype HuNoV Norwalk virus (GI.1) or placebo using WGS metagenomics. Samples collected pre-challenge and at multiple visits (day 2-56) post-challenge were tested. Of 41 genetically susceptible persons, 20 were infected and 13 developed symptoms of acute gastroenteritis (diarrhea and/or vomiting). Machine learning classification was performed pre- and post-challenge on a combined set of clades and pathways via RandomForest (RF) and Support Vector Machines (SVM) to build models predicting acute gastroenteritis. We used cross-validation, dividing our samples 80% for training and 20% for testing, using the Area Under the Curve (AUC) metric. Fifty iterations were performed for each model and median AUC was reported. Each sample was considered independently.

WGS analysis identified 603 clades and 533 pathways. In pre-challenge profiles of individuals who received high dose GI.1 inoculum (4800 or 48 genome equivalents, GE), RF models achieved AUC=0.99; *Dorea formicigenerans* and *Gemmiger formicilis* were more abundant in persons with acute gastroenteritis. In those who received low doses of GI.1 inoculum (4.8 or 0.48 GE), SVM achieved AUC=0.99, with higher abundance of *Streptococcus vestibularis* and *Denitrobacterium detoxificans* in persons with acute gastroenteritis. In post-infection profiles, RF and SVM achieved AUC=0.99 for both high and low dose groups, revealing higher abundance of *Eubacterium hallii*, *D. formicigenerans* and *G. formicilis* in the high dose group and higher abundance of *S. vestibularis* and *Dialister invisus* in the low dose group. Despite a relatively low sample size and considering each sample independently, we built effective models to predict acute gastroenteritis with GI.1 HuNoV. Identifying key microbial species associated with gastroenteritis will broaden our understanding of HuNoV pathophysiology and aid the design of potential therapies.

---PAGEBREAK---

Lower airway dysbiosis drives lung inflammation in early COPD

Imran Sulaiman^{1,2#}, Benjamin G. Wu^{1,2,4#}, **Matthew Chung**³, Bradley Isaacs^{1,2}, Jun-Chieh J. Tsay^{1,2,4}, Meredith Holub^{1,2,5}, Clea R Barnett^{1,2}, Benjamin Kwok^{1,2}, Matthias Kugler^{1,2}, Jake Natalini^{1,2}, Shivani

Singh^{1,2}, Yonghua Li^{1,2}, Rosemary Schluger^{1,2}, Joseph Carpenito^{1,2}, Destiny Collazo^{1,2}, Luisanny Perez^{1,2}, Beno W. Oppenheimer^{1,2}, Kenneth I Berger^{1,2}, Roberta M Goldring^{1,2}, Sergei B. Koralov⁶, Michael D. Weiden^{1,2}, Jeanine D'Armiento⁷, Jose C. Clemente⁸, Elodie Ghedin³, and Leopoldo N. Segal^{1,2,9*}

¹ Division of Pulmonary and Critical Care Medicine, New York University Grossman School of Medicine, NYU Langone Health, New York, NY

² Department of Medicine, New York University Grossman School of Medicine, NYU Langone Health, New York, NY

³ Systems Genomics Section, Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD

⁴ Division of Pulmonary and Critical Care Medicine, VA New York Harbor Healthcare System, New York, NY

⁵ Division of Pulmonary and Critical Care Medicine, Hartford Health Care, Hartford, CT.

⁶ Department of Pathology, NYU Langone Health, New York, New York.

⁷ Department of Physiology and Cellular Biophysics, Columbia University School of Medicine

⁸ Department of Genetics and Genomic Sciences and Immunology Institute, Icahn School of Medicine at Mount Sinai, NY.

⁹ Laura and Isaac Perlmutter Cancer Center, New York University School of Medicine, NYU Langone Health, New York, NY.

Chronic obstructive pulmonary disease (COPD) is a common condition associated with significant morbidity and mortality as well as increased health care costs. Smoke exposure has been shown to be a causative factor in its development. However, some patients with a heavy smoking history may not develop physiological evidence of COPD. In this study we sought to investigate whether lower airway dysbiosis occurs in early-stage COPD and contributes to the pathogenesis of the disease.

To explore this, we recruited 57 patients with a greater than 20 pack year smoking history to perform an in-depth microbiome analysis. Of the 57 patients, 26 had physiological evidence of COPD and the remaining were grouped as smoker controls. All patients had a bronchoscopy performed where lower airway (BAL), upper airway (UA) and bronchoscopy control (BKG) samples were collected. Samples were analyzed by 16S rRNA gene sequencing, whole genome sequencing, RNA metatranscriptome and host RNA transcriptome sequencing. In addition, a mouse model was used to evaluate the different contributions that smoke and dysbiosis have on lower airway inflammatory injury.

---PAGEBREAK---

Control Number: 2022-A-182-NGS

Topic 1: Microbial Chatter: Microbial ecology in health and disease

Publishing Title: Field-forward threat sensing through raw nanopore electrical signal: a case study in arthropod-borne pathogen detection on Saint Catherines Island

Author: J. Russell, A. Shteyman, C. Price;

Block: MRIGlobal, Gaithersburg, MD.

The ability to field genome-sensing technologies in distributed, austere locations changes the nature of the public health practitioner's relationship with extant biological information present in the environment. Determining the presence of active pathogens (naturally-occurring or synthetic/weaponized) is a key advantage for biosurveillance and biodefense applications. Accessing direct sequence-level information, as opposed to indirect detection of sequence-level signatures via amplification technologies (e.g., PCR), confers distinct bio-intelligence benefits. These benefits are most acutely realized when the active biological threat is an unknown, emergent entity without any comparable reference in genome repositories. Synthetically-engineered chimeric biothreats are poised to become a more pressing biodefense concern, underscoring the limitations to pre-designed assay technology and the need for direct sequence-level data sensing. Currently, nanopore-based sequencing platforms are the most developed technologies that could meet this need in a tactical or field-forward environment, outside of traditional laboratories and more proximal to the point-of-sampling. However, the maturity of these platforms for the rigors of tactical scenarios are still evolving to the required levels. Here, we field-test Oxford Nanopore's MinION sequencing device, embedded in a ruggedized tactical laboratory platform (Mercury Lab, Figure 1), for the purposes of un-biased sequence-level surveillance and detection of pathogen signatures directly on site, from arthropod samples found in the testing environment. The field site -- St. Catherines Island -- is a barrier island off the coast of the U.S. state of Georgia, 30 miles south of Savannah. The island is home to the only free-ranging troops of ring-tailed lemurs outside of Madagascar, making the island a unique environment to surveil for arthropod-borne pathogens of human concern since non-human primates are embedded in the ecosystem. We collected mosquito and tick samples, in addition to lemur serum and whole blood from several potential rodent reservoirs. Arthropod samples were processed for nucleic acid extraction and sequencing library preparation in the field on Mercury Lab, and the samples were sequenced in the field on the MinION device. Instead of converting the raw electrical signal to nucleotide sequences, we leveraged existing open-source algorithms designed to search for sequences-of-interest directly from the raw electrical signal. We report on the utility of this approach, gained advantages in operational time and size, weight, and power (SWAP) requirements, and detection results compared to base-called data. We comment on current gaps and proposed approaches for moving these technologies closer to the tactical reality and rigors of field-forward biosurveillance environments.

Abstract Body:

---PAGEBREAK---