Genetic diversity of commensal *Blastocystis* gut protists reveals strain-specific changes in host-interfacing pathways

Abigail L. Lind[1], Ami S. Bhatt[2], Katherine S. Pollard[1]

1 Gladstone Institute of Data Science and Biotechnology , San Francisco, CA
2 Department of Genetics, Stanford University, Stanford, CA

The human gut microbiome is a microbial ecosystem containing bacteria, archaea, viruses, and microbial eukaryotes. The most common human gut microbial eukaryote is the commensal protist *Blastocystis*, with an estimated prevalence in industrialized countries at 25%. While the presence of *Blastocystis* is associated with a significant reduction in pro-inflammatory bacteria and with reduced gut inflammation, little is understood about its role in the gut microbiome and its fundamental biology. Genetic evidence suggests that *Blastocystis* comprise a group of genetically diverse subtypes, but we lack high quality genomic data for these subtypes and understanding of their functional similarities and differences. Here, we cultivate 6 *Blastocystis* strains spanning the genetic diversity of the genus and generate contiguous, annotated genomes using a combination of long-read DNA sequencing, Hi-C, and RNA-seq. These genomes range in size from 14-25 Mb and have protein-coding genes with unusual features, including a frequent lack of canonical stop codons and a regular intron length of exactly 30 base pairs. Through comparison with the genomes of closely related stramenopiles, we find a pattern of genome reduction and gene duplication in *Blastocystis*, as well as genomic organization patterns that likely arose during the transitions from a free-living lifestyle to an obligate within-host lifestyle and in transitions between host species. We find substantial strain and subtype-specific gene duplications, including those of likely host-interfacing genes such as those involved in cell-cell adhesion and cell surface glycan production. Together, these genomes and our analyses reveal the adaptations *Blastocystis* has undergone to thrive in the gut microbiome. These results identify substantial biological variability between subtypes of *Blastocystis* which are likely to drive differences in interactions with other gut microbiota and the host.

Longitudinal transmission patterns of diverse Clostridioides difficile lineages in a hospital: lessons and concerns.

Newcomer, Erin P1,2,3, Fishbein, Skye RS1,2, Hink, Tiffany4, Reske, Kimberly A4, Cass, Candice5, Dubberke, Erik R4, Dantas, Gautam1,2,3,6.

1 The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St Louis, United States.
2 Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, United States.
3 Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, United States.
4 Division of Infectious Diseases, Washington University School of Medicine, St. Louis, United States.
5 Department of Medicine, Washington University School of Medicine in St Louis, St Louis, MO USA.
6 Department of Molecular Microbiology, Washington University School of Medicine, St Louis, United States.

Background: Current methods for preventing Clostridioides difficile (Cd) infection (CDI) in healthcare environments include antibiotic stewardship programs and placing patients with CDI on contact precautions. However, a minority (19-30%) of CDI cases are verifiably transmitted from previous CDI cases, while asymptomatic Cd carriers (ACCs) outnumber patients with CDI 10-fold and can shed pathogenic Cd spores. These data suggest that ACCs could be an important vector in Cd transmission, yet little is known about the environmental contamination and resulting risk of CDI.

Methods: A prospective study, spanning 6 months in the hematopoietic cell transplant and leukemia wards at Barnes-Jewish Hospital in St. Louis, was conducted to investigate Cd transmission dynamics. Patients were screened for Cd carriage on admission and weekly, and environmental swabs from bedrails, sinks, and keyboards were collected at each timepoint, in addition to remnant stool specimens from patients tested for Cd during routine care. Multiplex PCR for toxin genes and whole-genome sequencing were conducted on cultured Cd isolates.

Results: This study collected 659 stool specimens and 495 rectal swabs from 384 patients and 400 environmental specimens, identifying a total of 22 CDI (EIA+) cases during the study period. In-silico MLST identified significant surface contamination by strains not carried by the patient occupying the room. Whole-genome SNP analysis revealed 5 lineages of Cd that spread between patients and involved ACCs carrying toxigenic strains.

Discussion: The involvement of ACCs in transmission networks associated with CDI indicates that ACCs may play a significant role in Cd transmission. Networks were found across diverse strains not typically considered hypervirulent, suggesting that strains that can asymptomatically colonize and cause CDI have a survival advantage in wards where only CDI patients are put on contact precautions. Future work will investigate genetic determinants of transmission associated with networks spanning multiple patients.

Metabolic profile of the prevalent human gut bacterium Eggerthella lenta

Cecilia Noecker1, Juan Sanchez2, Jordan E. Bisanz1, Veronica Escalante1, Margaret Alexander1, Kai Trepka1, Almut Heinken3, Yuanyuan Liu4, Dylan Dodd 4, Ines Thiele3, Brian DeFelice 2, and Peter J. Turnbaugh1,2*

1 Department of Microbiology & Immunology, University of California, San Francisco, San Francisco, CA 94143
2 Chan-Zuckerberg Biohub, San Francisco, CA 94158
3 School of Medicine, National University of Ireland, Galway, Ireland
4 Department of Pathology, Stanford University, Stanford, CA 94305
* Present address: Department of Biochemistry and Molecular Biology, Pennsylvania State University, State College, PA

Background: Human gut microbes perform diverse metabolic functions with consequences for host health. Yet the metabolic strategies relied upon for growth by many commensal gut microbes remain unclear. The prevalent and disease-linked Actinobacterium Eggerthella lenta does not metabolize sugars and instead displays several unusual metabolic traits: arginine dependence, a putative acetogenesis pathway, and numerous enzymes involved in xenobiotic metabolism and anaerobic respiration.

Methods: To obtain a comprehensive view of the metabolic network of E. lenta, we generated several complementary resources: novel chemically defined culture media, metabolomics profiles of strain isolates in vitro and in vivo, and a curated genome-scale metabolic reconstruction.

Results: We analyzed E. lenta growth and metabolism in defined media, finding that it uses arginine and acetate as energy and carbon sources in tandem and is auxotrophic for tryptophan and riboflavin. Untargeted metabolomics of culture supernatants revealed diverse metabolic byproducts, many of which were strain-variable and could be linked to accessory genes. E. lenta growth was strongly promoted by acetate, which we found to be incorporated into cell wall and nucleotide biosynthesis using stable isotope analysis. These characteristics were largely recapitulated in silico by flux balance analysis of an updated metabolic model of the E. lenta type strain. Finally, we compared the metabolic footprint of E. lenta in vitro and in gnotobiotic mice, finding consistent metabolite signatures across environments. This comparison led us to identify and characterize the host signaling metabolite agmatine as an alternative energy source for E. lenta that can replace the effect of arginine.

Discussion: These results elucidate the distinctive metabolic niche filled by E. lenta in the gut ecosystem and suggest potential novel impacts of this organism on human hosts and other community members. Our approach is a generalizable strategy to gain mechanistic and predictive insights into the metabolism of understudied microbes.

Association of diet and antimicrobial resistance in healthy US adults

Andrew Oliver1, Zhengyao Xue1,2, Yirui Tang1,2, Blythe Durbin-Johnson3, Zeynep Alkan1, Diana H. Taft2, Jinxin Liu4, Ian Korf 3, Kevin D. Laugero1,5, Charles B. Stephensen1,5, David A. Mills2,6, Mary E. Kable1,5, Danielle G. Lemay1,3,5

1 USDA-ARS Western Human Nutrition Research Center, 430 West Health Sciences Dr., Davis, CA, 95616
2 Department of Food Science & Technology, University of California, One Shields Ave., Davis, CA, 95616
3 Genome Center, University of California, One Shields Ave., Davis, CA, 95616 4 Laboratory of Gastrointestinal Microbiology, College of Animal Science and Technology, Nanjing Agricultural University, Nanjing 210095, China
5 Department of Nutrition, University of California, Davis, One Shields Ave. Davis, CA, 95616
6 Department of Viticulture and Enology, Robert Mondavi Institute for Wine and Food Science, University of California, Davis, One Shields Ave., Davis, CA, 95616

Background: Antimicrobial resistance (AMR) represents a significant source of morbidity and mortality worldwide. Since resistance to antibiotics is encoded in the microbiome, interventions aimed at altering the taxonomic composition of the gut might allow us to prophylactically engineer microbiomes that harbor less antibiotic resistant genes (ARGs). Diet is one method of intervention, yet little is known about the association between diet and antimicrobial resistance. Our primary hypothesis was that increased intake of dietary fiber would be associated with reduced ARG abundance in human fecal metagenomes. Beyond our directed hypotheses, we utilized machine learning approaches on a variety of diet, physiological, and lifestyle features to assess whether the abundance of antibiotic genes is correlated with variables outside the scope of our directed hypotheses.

Methods: We examined diet using the food frequency questionnaire (FFQ, habitual diet) and 24-hour dietary recalls (ASA24) coupled with analysis of the microbiome using shotgun metagenome sequencing in 290 healthy adult participants of the USDA Nutritional Phenotyping Study. Additionally, we applied machine learning to examine 387 dietary, physiological, and lifestyle features for associations with antimicrobial resistance.

Results: We found that aminoglycosides were the most abundant and prevalent mechanism of AMR in these healthy adults and that aminoglycoside-o-phosphotransferases (aph3-dprime) negatively correlated with total calories and soluble fiber intake. Individuals in the lowest quartile of ARGs (Low-ARG)

consumed significantly more fiber in their diets compared to Medium- and High-ARG individuals, which was concomitant with increased abundances of obligate anaerobes in their gut microbiota.

Conclusions: Our results indicate individuals with lower abundances of antibiotic resistance genes consumed more diverse diets that were richer in fiber and limited in animal protein. We suspect that increased fiber likely drives the composition of the gut towards a more obligate anaerobe state, reducing footholds for facultative anaerobes, which are known harbors of inflammation and antibiotic resistance.

Ultra-deep Sequencing of Hadza Hunter-Gatherers Recovers Vanishing Gut Microbes

Matthew R. Olm, Bryan D. Merrill, Matthew M. Carter, Dylan Dahan, Surya Tripathi, Sean P. Spencer, Brian Yu, Sunit Jain, Norma Neff, Aashish R. Jha, Erica D. Sonnenburg, Justin L. Sonnenburg

Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA; Chan Zuckerberg Biohub, San Francisco, CA, USA; Genetic Heritage Group, Program in Biology, New York University Abu Dhabi, Abu Dhabi, UAE; Center for Human Microbiome Studies, Stanford University School of Medicine, Stanford, CA, USA

The gut microbiome is a key modulator of immune and metabolic health. Human microbiome data is biased towards industrialized populations, providing limited understanding of the distinct and diverse non-industrialized microbiomes. Here, we performed ultra-deep metagenomic sequencing and strain cultivation on 351 fecal samples from the Hadza, hunter-gatherers in Tanzania, and comparative populations in Nepal and California. We recover 94,971 total genomes of bacteria, archaea, bacteriophages, and eukaryotes, 43% of which are absent from existing unified datasets. Analysis of in situ growth rates, genetic pN/pS signatures, high-resolution strain tracking, and 124 gut-resident species vanishing in industrialized populations reveals differentiating dynamics of the Hadza gut microbiome. Industrialized gut microbes are enriched in genes associated with oxidative stress, possibly a result of microbiome adaptation to inflammatory processes. This unparalleled view of the Hadza gut microbiome provides a valuable resource that expands our understanding of microbes capable of colonizing the human gut and clarifies the extensive perturbation brought on by the industrialized lifestyle.

Whole Genome sequencing and antibiotic resistance of isolates from the Alligator mississippiensis gut microbiome

Julissa Perez-Marron, Jenine Dong, Erica Meng, Paul Orwin

University of the Pacific

Multidrug resistant bacteria are a global threat to public health due in part to the lack of development of new antibiotics. One major avenue for the development of antibiotic resistance and the spread of resistant bacteria to humans is from agriculture. Although most studies of this process have focused on traditional farm animals (pigs, cows, chickens), the American alligator, Alligator mississippiensis is also raised on farms for food and leather. The gut microbiome of this animal is less well understood than its counterparts in agriculture. We cultured bacteria from the alligator colon, and identified 16 antibiotic resistant isolates using disk diffusion. Using 16S rRNA gene sequencing the 16 isolates were shown to fall into seven unique strains were identified among the selected isolates: Aeromonas sp., Citrobacter freundii, Citrobacter portuncales, Escheriera coli, Providencia revertii, Providencia sp., Pseudomonas aeruginosa and Pseudomonas sp. Through minimal inhibitory assays, it was determined that all strains were resistant to two or more of the seven antibiotics tested after 24 hours of exposure. All strains excluding Pseudomonas sp. exhibited resistance to both Ampicillin and Rifampin. Two of our isolates Pseudomonas sp. and Escherichia coli were resistant to six of the seven antibiotics tested. Genome assemblies were performed in Unicycler using Nanopore long reads and Illumina short read data with a hybrid approach. ABricate was used to identify the antibiotic resistance genes in the assembled genomes and to correlate the genes present with the observed phenotypes. The data presented here suggest that farmed alligators may be a significant and important reservoir of antibiotic resistant human pathogens.

Gut microbiome in endometriosis

Inmaculada Pérez-Prieto 1,2; Eva Vargas 3; Eduardo Salas-Espejo 1; Analuce Canha- Gouveia 1,4; Andres Salumets 5,6; Kreete Lüll 7,8; Elin Org 7; Signe Altmäe 1,2,9

1 Department of Biochemistry and Molecular Biology, Faculty of Sciences, University of Granada, Granada, Spain.
2 Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain.
3 Systems Biology Unit, Department of Experimental Biology, Faculty of Experimental Sciences, University of Jaen, Jaen, Spain.
4 Department of Physiology, Faculty of Veterinary, University of Murcia, Campus Mare Nostrum, IMIB-Arrixaca, Murcia, Spain.
5 Department of Obstetrics and Gynaecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia.
6 Competence Centre on Health Technologies, Tartu, Estonia. 7 Institute of Genomics, Estonian Genome Centre, University of Tartu, Tartu, Estonia.
8 Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia.
9 Division of Obstetrics and Gynecology, CLINTEC, Karolinska Institutet, Stockholm, Sweden.

Background: Endometriosis, defined as the presence of endometrial-like tissue outside of the uterine cavity, is one of the most common reproductive disorders affecting ~10% of women in reproductive age. There are different theories of the possible causes of endometriosis, however, its pathogenesis is not clear. Novel studies are indicating that the microbial composition may be involved in the etiology of endometriosis, nevertheless, the connection between microbes, its dysbiosis and the development of endometriosis remains unexplored. This study aims to analyze and compare the gut microbiome profile in women with and without endometriosis and determine microbial pathways potentially involved in the development of the disease.

Methods: This case-control study includes a subsample of 1000 women of the Estonian Biobank, a volunteer-based cohort of the Estonian adult population. 136 women with endometriosis and 864 control women who have not been diagnosed with endometriosis or any of its comorbidities were included. To characterize the microbiome composition and functionality, microbial DNA was extracted and sequenced by paired-end metagenomic shotgun sequencing on the Illumina Novaseq 6000 platform. Microbial functional pathways were annotated using the KEGG database (https://www.genome.jp/kegg/).

Expected results: The data processing is ongoing. The alpha- and beta-diversity and differential abundance analyses will be performed to assess the gut microbiome (species and KEGG orthologies [KO]) in both groups. The ANOVAlike Differential Expression tool (ALDEx2, v.1.18.0) will be used to identify species and KOs associated with endometriosis. Analysis will be adjusted for age, BMI, and stool consistency.

Discussion. To the best of our knowledge, this will be the biggest metagenome study in women with endometriosis. Our findings would provide better understanding of the endometriosis-associated microbiome in a large-scale cohort, which may lead to discovery of novel links between the microbiome and the etiology of endometriosis.

Building the World's Largest Phage-Host Interaction Atlas using Proximity Ligation Technology

Phase Genomics, Inc

Viruses, including bacteriophage and archaeal viruses, are the most abundant form of life on earth (10 31). They interact with all life and shape the global ecosystem through their impacts on community composition and horizontal gene transfer. However, phage-host relationships have proven challenging to identify without use of culture-based experiments to generate unambiguous evidence for a phage's presence in a given host. These experiments inherently require that all hosts are culturable, typically restricting the scope and microbial diversity that can be surveyed and limiting our understanding of potentially valuable phage-host relationships.

Proximity ligation sequencing is a powerful genomic method for associating viruses with their hosts directly in native microbial communities. Proximity ligation captures, in vivo, physical interactions between the host microbial genome and the genetic material of both lytic and lysogenic phage. Similar to culturing experiments, these linkages offer direct evidence that phage sequences were present within an intact host cell, thereby establishing a phage-host pair. However, unlike culturing experiments, proximity-ligation methods do not require the propagation of living bacterial cells and unlike single cell sequencing experiments, only capture phage-host interactions inside cells. The combination of intra-phage and phage-host signal enables us to simultaneously deconvolve viral genome bins (vMAGs) directly from metagenomes and to assign microbial hosts to large numbers of vMAGs without culturing.

Our application of this technology to hundreds of complex microbiome samples has yielded thousands of novel phage
and archaeal virus genomes with host assignments, as well as large numbers of new microbial genomes. Through
broad-scale application of proximity ligation sequencing, we are creating a global-scale database of highly diverse
phage-host interactions from samples from across the world. We will present published and unpublished work
highlighting the power of this approach in the field of metagenomic discovery.

“Microbes@Home”, a module to teach microbial diversity and bioinformatics to undergraduates.

Giorgia Pirino and Katherine L. Petrie

School of Biological Sciences, University of California, San Diego, LA Jolla, CA

The revised curriculum of the laboratory in microbiology at UCSD integrates a research component into the course, fostering critical thinking and a better understanding of the scientific process.
At the beginning of the pandemic, the course moved to a remote format, focusing more on data analysis. To keep a hands-on component and “student ownership”of experiments, we implemented the Microbes@Home module, in which students collected samples from surfaces in their household and shipped them to the lab, where 16S rDNA PCR amplicons were sequenced through the Illumina platform. Students then analyzed the microbial communities using the QIIME 2 pipeline via AWS, which provided a consistent computing environment for all students, regardless of their personal home computer.
Across 4 quarters, 221 samples were analyzed and over 7000 unique features were identified.
Each student analyzed the alpha and beta diversity of a subset of the samples and visualized the data through PCoA and taxonomy bar plots. Although students did not always find significant differences between surfaces in terms of alpha and beta diversity, they were usually able to discern distinct patterns in microbial composition at different sample sites via principle coordinates analysis.
The main challenge of the “Microbes@Home” module is the use of AWS for novices. If students have access to lab computers with QIIME installed, they do not need to use AWS, and the analysis can be scaled up to include more samples. The module can also be extended to the microbial communities of other environments accessible to undergraduate laboratories, such as soil samples and fermented food products.
The implementation of the Microbes@Home module engages students into the research process through hypothesis-driven experiments and discovery, while developing not only laboratory skills at the bench, but also bioinformatics skills at the command line.

Evaluation of taxonomic profiling methods for long-read shotgun metagenomic sequencing datasets

Daniel M. Portik 1, C. Titus Brown 2, and N. Tessa Pierce-Ward 2

1 Pacific Biosciences, 1305 O'Brien Dr, Menlo Park, California 93025 USA
2 Department of Population Health and Reproduction, University of California Davis, Davis, California USA

Background: Long-read shotgun metagenomic sequencing offers many advantages over short-read sequencing. Long reads contain higher information content which is useful for a variety of metagenomics analyses, including taxonomic profiling. New long-read specific tools for taxonomic profiling are being developed, yet there is a lack of consensus regarding their relative performance.

Methods: To address this, we performed a critical benchmarking study using both long-read and short-read methods. We applied these tools to several mock community datasets generated using PacBio HiFi or Oxford Nanopore Technology (ONT) sequencing. We evaluated method performance based on read utilization, detection metrics, and relative abundance estimates.

Results: We show that long-read methods generally outperformed short-read methods. Short-read methods (including Kraken2, Bracken, Centrifuge, and MetaPhlAn3) produced many false positives (particularly at lower abundances), required heavy filtering to achieve acceptable precision (at the cost of reduced recall), and produced inaccurate abundance estimates. By contrast, several long-read methods displayed near-perfect precision and suitable recall without any filtering required. These methods include BugSeq, MEGAN-LR using translation alignments (DIAMOND to NCBI nr) or nucleotide alignments (minimap2 to NCBI nt). Based on the abundance designs of the mock communities, we show that long-read methods detect all species down to the 0.1% abundance level with high precision in the PacBio HiFi datasets. Other methods (MetaMaps and MMseqs2) required filtering to reduce false positives and balance precision and recall. Read quality affected performance for methods relying on protein prediction or exact kmer matching, and these methods performed better with higher accuracy reads (PacBio HiFi reads). Finally, for a given mock community the long-read datasets produced better results than short-read datasets, demonstrating advantages for long-read metagenomic sequencing with regards to taxonomic profiling. Our critical assessment of methods provides recommendations for current research using long reads and establishes a baseline for future studies.

Enterococcus casseliflavus: Naturally vancomycin resistant motile species of Enterococcus

Janira Prichula1,2; Matheus M. S. M. Lima3; Adriana Seixas4; Michael S. Gilmore1,2

1 Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA 02114, USA; 2 Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA; 3 Bioinformatics Multidisciplinary Environment, Digital Metropolis Institute, Federal University of Rio Grande do Norte, Natal, RN, Brazil. 4 Department of Pharmacosciences, Federal University of Health Sciences of Porto Alegre, Porto Alegre, RS, Brazil.

Enterococcus casseliflavus is unusual among enterococci because it is motile, yellow-pigmented, and all isolates of the species harbor the VanC gene cluster, which reduces susceptibility to vancomycin, as a core chromosomal trait. The occurrence of this gene cluster long predates humans or human antibiotic use. Little is known of the nature of the environment to which this species adapted. To infer clues, we launched a study to systematically examine all available genomes in the NCBI database of E. casseliflavus (n=100) for insight into its ecology. Following quality curation 85 genomes were selected, being the largest group derived from human (29.1%), with the second largest group isolated from the environment (24.4%) but not a particular host. Average Nucleotide Identity (ANI) was determined for all 85 genomes. Surprisingly, of those, 82.35% (n= 70/85) possessed £ 95% ANI when compared to the prototype strain ATCC 12755 E. casseliflavus, with most of those belonging to the sister species E. flavescens based on >95% ANI with prototype strain ATCC 49996 E. flavescens. This reflects ongoing controversy about the relationship between E. casseliflavus and E. flavescens. Resistome analysis showed that all genomes harbor a vanC operon as expected. Interestingly, two genomes also harbor the vanD operon, which encodes resistance to high-level vancomycin. Additional antibiotic resistances were identified in 14 of the 85 genomes (12 of those 14 were derived from clinical or agricultural sites), including resistances conferred by ermB, tet(45), tet(C), tetM, and tetS, optrA, and poxtA genes. Pangenome analysis estimated a total of 18,938 genes, with 6.09% (1,155) being core to all strains. There is considerable confusion in the NCBI database with respect to species designation within the casseliflavus group, and the radiation of the species E. flavescens appears to be recent, and as for that within E. faecium, may be attributable to anthropogenic activity.

Isolating receptor dependent phages from the environment

Eleanor Rand1,2, Siân Owen1, Natalia Quiñones-Olvera1,2, Michael Baym1

1 Department of Biomedical Informatics and Laboratory of Systems Pharmacology, Harvard Medical School
2 Systems, Synthetic, and Quantitative Biology PhD Program, Harvard University

Bacteriophages (phages for short) are of growing interest both for their clinical uses as well as their evolutionary impacts. While there are a few well characterized phages that have been used across biology for decades, most phages in the environment remain uncharacterized. One of the most influential aspects used to determine a phage's use in the clinic or its ecological importance is the host cell component it requires as a receptor to begin infection. We developed an assay to quickly screen for and isolate environmental phages dependent on specific receptors. Our method expands upon the traditional plaque assay by plating multiple bacterial strains, each with one protein knockout and a unique fluorescent label, together on one plate with an environmental sample. Phages that display differential plaquing between the strains are likely to be dependent on a protein of interest. This method can be extended to any non-essential gene. Using this method, we have isolated phages dependent on proteins of interest in just a few days. Understanding the abundance and diversity of phages that interact with specific receptors could help inform the phages chosen for clinical use as well as contribute to our understanding of the ecological effects phages impose in the environment.

Isolating high quality nucleic acids for genome scanning and whole genome sequencing of “Haptophyte Genome Project”

Betsy Read, Hossein Khadivar

California State University San Marcos, Department of Biological Sciences

Haptophytes are a clade of marine phytoplankton. There are more than 300 extant species that range in size from between 2-40 μm. A number of haptophytes are coccolithophores that produces elegantly sculpted calcium carbonate cell coverings and significantly impact global carbon and sulfur cycling. Others exhibit mixotrophic tendencies, produce unusual very long fatty acids 〖(C〗_(36-42)) known as alkenones, form massive blooms, and synthesize harmful toxins. Although they extend deep into the tree of life, the evolutionary history of this important phylum of phytoplankton remains unresolved. For these many reasons the Department of Energy’s joint Genome Institute aims to sequence the genomes and transcriptomes of 35 members of the haptophyte phylum. The focus of this project was to grow 20 of these haptophytes and extract high quality nucleic acids for genome scanning to determine size and complexity, and eventually whole genome sequencing using PacBio and Illumina platforms. Haptophytes were grown in filtered-seawater supplemented with F/2, K/2 or K/8 vitamins and trace metals. DNA isolation was done during mid-late log phase. RNA was extracted to facilitate gene calling and was performed in the middle of the light and dark phases to maximize the number of genes expressed. Agarose gel electrophoresis was used to ascertain the integrity and size of the extracted DNA. High quality genomic DNA has been extracted from 20 species. Results from nucleic acid extractions and genome scanning will be presented, together with progress on RNA extractions.

Characterization of localized skin microbes recovered from patients undergoing immediate breast implant reconstruction to identify features correlated with surgical site infection development

*Olivia N. Rebeck 1, *Bejan Mahmud 1, David Zahrah 1, Meghan A. Wallace 2, Caitlin Johnson 2, Carey-Ann D. Burnham 2,3,4,5, Margaret A. Olsen 6, Gautam Dantas 1,2,3,7 * denotes equal contribution

1 The Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO, USA.
2 Department of Pathology and Immunology, Washington University School of Medicine in St. Louis, MO, USA.
3 Department of Molecular Microbiology and Microbial Pathogenesis, Washington University School of Medicine in St. Louis, MO, USA.
4 Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA
5 Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA.
6 Division of Infectious Diseases, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA. [7] Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO

Background: Breast cancer (BC) accounts for 15% of cancer-related deaths in women worldwide. Although mastectomy is a common and effective component of BC treatment, there is a high rate of surgical site infections (SSI) in patients undergoing mastectomy with immediate breast implant reconstruction (IBR). We hypothesize that the local skin microbiota is a reservoir for SSI-causing bacteria, such as Staphylococcus epidermidis and Cutibacterium acnes, and that differences in the community composition of the local skin microbiota can predispose a patient towards SSI development.

Methods: To test this hypothesis, we performed 16S rRNA gene sequencing on breast and axilla skin swabs and whole genome sequencing on isolates recovered from skin, implant, and seroma fluid samples of a cohort of patients undergoing mastectomy with IBR. This cohort consisted of 61 patients with risk factors for SSI development (e.g., obesity, previous SSIs) enrolled in two studies. Samples were collected across up to six timepoints (including pre- and post-mastectomy timepoints) from patients who developed infections, noninfectious wound complications, or no complications.

Results: To date, 192/432 metagenomic skin swabs have been sequenced. These metagenomes will undergo taxonomic identification via DADA2. We will analyze these sequencing data to assess differences in community composition. We have also sequenced 393 S. epidermidis and C. acnes isolates. We will analyze these sequencing data to track and compare isolates and characterize their microbial features, including virulence factors and antibiotic resistance genes. Together, we will use predictive modeling to determine which features of metagenomes and isolates are differentially associated with samples collected from patients who develop infections compared to samples collected from patients who did not.

Discussion: Robust characterization of these microbes can identify informative features that correlate to individual predisposition to SSI development, which can inform patient care by identifying patients at increased risk of SSI and improve antibiotic stewardship.

Plasmid evolution is guided by intracellular competition

Fernando Rossine, Michael Baym

Harvard Medical School

Bacterial plasmids carry historically variable loads of medically and ecologically relevant genes such as those responsible for antibiotic resistance or accessory metabolism. Yet, the evolutionary dynamics that drive this variation are poorly understood. Crucially, an arising plasmid variant must face challenges in two scales: bacteria hosting a new plasmid variant compete with bacteria hosting the ancestral type, but within each bacterium the new plasmid variant may directly compete for replication with the ancestral type. The extent to which intracellular competition for replication shapes plasmid evolutionary dynamics is unknown, as are the plasmid features that might favor its intracellular fixation. Here we propose that conflicts between plasmid replication and transcription limit the efficiency of plasmid replication, and therefore that low-transcription plasmids might be favored in the intracellular competition for replication. We develop a method to measure pairwise intracellular plasmid competition and show that plasmid transcription-replication conflicts can drive plasmid evolution regardless of cell-level costs or benefits. We start by setting up an experimental population in which each bacterium contains a well defined initial frequency of both plasmid variants. We achieve this by joining both plasmid variants into a single chimeric plasmid dimer that is then transformed into E. coli cells. Upon induction of a novel thermosensitive FLP-like recombinase, the dimers split into their component monomers, ensuring that each cell in the bacterial population contains both plasmid variants in equal frequencies. We find that when our plasmids differ in a few base pairs controlling the strength of a promoter, the weaker promoter plasmid increases in frequency as long as there is still co-occurrence of both plasmid types. More generally, these results indicate that intracellular competition might be a major deterrent to the plasmidial acquisition of new genes, as well as pose an alternative explanation for the longstanding observation of silenced plasmidial genes.

Longitudinal sequencing and variant detection of SARS-CoV-2 across Southern California wastewater

Jason A. Rothman; Katrine L. Whiteson

University of California, Irvine

Wastewater based epidemiology (WBE) is a useful method to detect pathogen prevalence and may serve to effectively monitor diseases at broad scales. WBE has been used throughout the COVID-19 pandemic to track disease burden through the quantification of SARS-CoV-2 RNA present in wastewater. Aside from case load estimation, WBE is being used to assay viral genomic diversity and the emergence of potential SARS-CoV-2 variants. Unlike sequencing SARS-CoV-2 in healthcare settings, WBE has been used to sequence virus without adding additional testing or supplies burden to the healthcare system.

Here, we present a study in which we sequenced RNA extracted from sewage influent samples obtained from eight wastewater treatment plants representing 16 million people in Southern California over April 2020 – August 2021. We sequenced SARS-CoV-2 with two methods: Illumina Respiratory Virus Enrichment and metatranscriptomic sequencing (N = 275), and QIAseq SARS-CoV-2 tiled amplicon sequencing (N = 96). We were able to build nearly full-length consensus genomes which largely mirrored emergent SARS-CoV-2 clades isolated from sequences derived from healthcare settings. Notably, in samples where RNA was too fragmented to assign clades, we identified single nucleotide variants (SNVs) of the virus that often corresponded to variants of interest and/or concern. Through our analyses, we also detected a large diversity of putatively novel SARS-CoV-2 SNVs, and SNVs of unknown potential function and prevalence. As the COVID-19 pandemic continues, and novel SARS-CoV-2 variants emerge, the ongoing monitoring of wastewater is an important tool in understanding local and population-level dynamics of the virus.

Our study shows the potential of WBE to detect SARS-CoV-2 variants throughout Southern California's wastewater and to track the diversity of viral SNVs and strains in urban and suburban locations. These results will aid in our ability to monitor the genomic potential of SARS-CoV-2 and help understand circulating SNVs to further combat COVID-19.

CRISPR-based killswitches for engineered microbes - stability and in vivo application

Austin G. Rottinghaus, Aura Ferreiro, Skye R.S. Fishbein, Gautam Dantas, Tae Seok Moon

Department of Energy, Environmental and Chemical Engineering, Washington University in St. Louis, St. Louis MO USA (AGR, TSM); The Edison Family Center for Genome Sciences & Systems Biology, Washington University in St. Louis, St. Louis MO USA (AF, SRSF, GD); Department of Biomedical Engineering, Washington University in St. Louis, St. Louis MO USA (AF, GD); Department of Pathology and Immunology, Washington University in St. Louis, St. Louis MO USA (SRSF, GD); Department of Molecular Microbiology, Washington University in St. Louis, St. Louis MO USA (GD); Division of Biology and Biomedical Sciences, Washington University in St. Louis, St. Louis MO USA (GD, TSM)

Background: Microbial biocontainment is an essential goal for engineering safe, next-generation living therapeutics. However, the genetic stability of biocontainment circuits such as genetically-encoded kill-switches is a challenge that must be addressed. Kill-switches are among the most difficult circuits to maintain due to the strong selection pressure they impart, often leading to evolution of escape mutant populations.

Methods: Here we engineer two CRISPR-based kill-switches in the probiotic Escherichia coli Nissle 1917, a single-input chemical-responsive switch and a 2-input chemical- and temperature-responsive switch. To maximize genetic stability, we employ parallel strategies including functional redundancy within the circuit, mitigation of the bacterial SOS response, antibiotic-independent plasmid maintenance, and innovatively, provision of intra-niche competition by a closely related strain. We test the efficacy of these kill-switch strains in vitro and in vivo, and characterize escape mutant populations through functional assays and sequencing of kill-switch circuit parts.

Results: We address two routes of DNA mutagenesis that contribute to kill-switch instability: i) stochastic inactivation of the kill-switch through introduction of functionally redundant Cas9 and gRNA expression cassettes, and ii) SOS-mediated DNA mutagenesis in response to Cas9-driven double-stranded genome breaks through knockouts of key SOS response elements. We further improve kill-switch stability through an antibiotic-independent plasmid maintenance system, and minimize the emergence of escape mutant populations through intra-niche competition provided by a related probiotic strain. Using these parallel strategies, we demonstrate that strains harboring these kill-switches can be selectively and efficiently killed inside the murine gut, or upon excretion in the case of the 2-input switch.

Discussion:
We developed CRISPR-based kill-switches in EcN towards a safe probiotic chassis for biomedical technologies, and which enable on-demand selective removal of engineered microbes from the gut. We explored diverse methods for improving the kill-switch stability and minimized key mechanisms of circuit inactivation, providing a template for future kill-switch development.

Comparative Genomics Of Diagnostic And Longitudinally-sampled Community-origin Staphylococcus Pseudintermedius Reveals The Absence Of Niche Adaptation But Widespread Strain Persistence Within Households

Sanjam S. Sawhney 1, Rhiannon Vargas 1, Meghan A. Wallace 1, Carol Muenks 1, Brian V. Lubbers2, Stephanie Fritz1, Carey-Ann D. Burnham*1, and Gautam Dantas*1

1 Washington University in St. Louis, St. Louis, MO
2 Kansas State University, Manhattan, KS; # authors contributed equally;
* corresponding authors

Background: Staphylococcus intermedius group (SIG) describes closely-related coagulase-positive staphylococcal species – Staphylococcus pseudintermedius (S.p.), delphini, and intermedius – associated with a range of host-species. Efforts investigating community MRSA spread have serendipitously identified widespread SIG colonization of household surfaces as well as human skin and fur of pet dogs and cats; additionally, clinical labs have reported increasing human infection resulting from SIG. In consideration of the growing host range and disease burden, we endeavored to interrogate the interrelatedness and potential niche adaptation of SIG across hosts and environments, and to assess the spatiotemporal dynamics of SIG.
Methods: We have assembled 572 SIG genomes collected from human diagnostic samples at the Barnes-Jewish Hospital (St. Louis, MO, n=181), animal diagnostic samples at Kansas State Veterinary Diagnostic Laboratory (n=100), and in community settings (St. Louis metropolitan area, n=291). All isolates were identified via MALDI-TOF (VITEK-MS) and underwent Illumina whole genome sequencing (WGS) for strain-level analysis.
Results: VITEK-MS had 100% sensitivity and 87.5% specificity at calling WGS-validated S.p. Remarkably, WGS did not identify niche adaptation between diagnostic and household S.p. by core, accessory, or gene family content, in contrast to reports on S. aureus; however, diagnostic S.p. had a higher ARG count ($p<0.001$) and isolates with the greatest ARG burden were often methicillin-resistant ($p<0.001$). Deeper analysis on strain persistence identified 72 strain-level clusters (≤45 pairwise SNPs at ≥98% coverage) comprising ~40% of isolates, primarily among intra-household isolates but also in diagnostic isolates from temporally-adjacent patients. Interestingly, a CRISPR-Cas operon and pathogenicity island were differentially present among diagnostic isolates of different strain backgrounds, while putative mobilizable elements differentiated certain isolates within a strain.
Conclusion: We argue against niche adaptation within S.p. but highlight widespread strain persistence within households, identifying genetic elements – often mobilizable – that differentiate isolates within the same strain and unite isolates of different strain backgrounds.

Preventing emergence of daptomycin resistance in Enterococcus faecium

Suelen Scarpa de Mello1, 2, Andrew G. Van Camp1, 2, Michael S. Gilmore 1, 2

1 Massachusetts Eye and Ear Infirmary, Boston, MA.
2 Harvard Medical School, Boston, MA.

Daptomycin is a last-line drug for treating multidrug-resistant E. faecium infections but is compromised by the emergence of resistant mutants. We discovered that defects in the gene lafB, which encodes a glycosyltransferase involved in lipoteichoic acid (LTA) biosynthesis, render E. faecium hypersusceptible to daptomycin. If lafB is dominant to other daptomycin resistance mutations, identifying a drug capable of blocking lafB would not only enhance daptomycin susceptibility by >10-fold but also block emergence of other daptomycin resistance mutations. To assess the dominance of a lafB mutation, we first generated an E. faecium strain with a clean deletion in lafB, which as expected was hypersusceptible to daptomycin. We next introduced mutant forms of liaR and cls known to confer increased daptomycin resistance to wild-type cells. However, they did not confer resistance in the lafB deletant, indicating that lafB mutation is in fact dominant to known daptomycin resistance mechanisms. Finally, we attempted to evolve daptomycin-resistant derivatives of the lafB-deleted strain. While controls evolved a 100-fold increase in resistance to daptomycin, little change was observed for the lafB-deleted strain. Little is known about the specific role of the disaccharide link between the diacylglycerol membrane anchor and polyphosphate domain of LTA, the formation of which is catalyzed by LafB, or why it varies. To identify variants related to lafB, we constructed a lafB gene tree. We observed that a subgroup of enterococci cluster with vagococci in possessing a variant form of the gene, indicating that these species likely produce an LTA with an altered disaccharide linkage. We are now comparing the properties of those cells to determine the potential contributions that this link makes to the physiology of the cell. We are also launching a small molecule screen for LafB inhibitors to identify compounds that sensitize E. faecium to daptomycin and prevent emergence of resistance.

Plant pathogenic Ralstonia species have multiple chemoreceptors for amino acids with overlapping, but divergent chemical specificities

Schomer R.A. 1, Avalos, J.K. 1,2, Lowe-Power, T.M. 2, Parales, RE. 1

1 University of California, Davis, Department of Microbiology & Molecular Genetics.
2 University of California, Davis, Department of Plant Pathology.

Background: Ralstonia solanacearum, a soil-borne plant pathogen, exploits motility and chemotaxis (directional movement along a chemical gradient) to locate host roots. Membrane-bound chemoreceptors drive chemotaxis and contain a conserved signal transduction domain and highly sequence divergent ligand binding domains (LBDs). We hypothesize that Ralstonia's repertoire of chemoreceptors and their respective chemical ligands drive plant host discovery and preference from the soil. Here, we target chemoreceptors that sense amino acids, a common component in root exudates.

Methods: To probe the chemotactic range of Ralstonia, we generated chimeras by fusing LBDs of 42 Ralstonia chemoreceptors with the transduction domain of a two-component system that controls gene transcription in E. coli. Using fluorescent reporter strains designed to measure transcriptional readout, we characterized the ligand ranges of LBDs from Ralstonia model strains with varying host ranges. A recently discovered LBD motif for amino acids allowed us to target predicted amino acid sensors.

Results: Across the genetically diverse Ralstonia solanacearum species complex, genomes contain two paralogous amino acid sensors, McpA and Rsc3307. We characterized the ligand range of McpA and Rsc3307 using the chimeric receptors. Interestingly, despite high sequence similarity between the Rsc3307 orthologs, the amino acid sensors tested from broad-host range Ralstonia solanacearum IBSBF1503 and narrow-host range Ralstonia syzygii PSI07 strains had variable ligand ranges. Using mcpA and rsc3307 mutants, their influence on the chemotactic range of Ralstonia and their effect on Ralstonia ability to infect plant hosts from the soil is being evaluated.

Discussion: To understand the role chemotaxis plays in host-microbe interaction, we must first identify which chemical signals bacteria detect. Our data indicate that subtle changes in LBD sequence can have significant influences on which chemicals a bacterium can detect. Variability in the chemical repertoire of a soil-borne plant pathogen could affect how effectively it senses host plants' roots or not.

Bloodstream Infection is Associated with Prior Gut Microbiome Colonization of the Same Strain in the Neonatal Intensive Care Unit

Drew J. Schwartz 1,2,3,4,5; N. Shalon 5; K. Wardenburg 5; C. Hall-Moore 1; B. Warner 1; P.I. Tarr 1,4; G. Dantas 4,5,6,7

1 Department of Pediatrics
2 Center for Women's Infectious Diseases Research
3 Department of Obstetrics and Gynecology
4 Department of Molecular Microbiology
5 The Edison Family Center for Genome Sciences and Systems Biology
6 Department of Pathology and Immunology
7 Department of Biomedical Engineering; Washington University School of Medicine

Background: Infants hospitalized in neonatal intensive care units (NICUs) have a high incidence of bacterial bloodstream infections (BSIs) leading to significant mortality. Common etiologies of BSI are pathogens that also colonize the infant gut. However, differentiating neonates who develop BSI with these organisms from those who are simply colonized remains challenging. We hypothesized that the same strain increases in abundance in the gut of affected neonates prior to BSI.

Methods: We performed shotgun metagenomic sequencing on 462 fecal samples from 19 infants with BSI and 37 controls in the same NICUs. We utilized MetaPhLan3 to infer taxonomy and InStrain to track the whole genome sequence of the BSI pathogen within the fecal metagenomes longitudinally. We used MaAsLin2 to determine significant associations between the microbiome and metadata.

Results: We found no differences in gut microbiome diversity between cases and controls. Infants who experience a BSI possess significantly higher relative abundance of the causative species in their gut relative to paired controls in the 2 weeks prior to bacteremia (median 21.4% vs. 0.4%, $p<0.05$). 10/19 infants produced at least 1 stool before BSI containing an organism with 0-1 genomic substitutions from the BSI isolate, establishing isogenicity between gut residing and invasive pathogen. The coverage (abundance) of the BSI isolate increased in the stool of affected neonates in the 10 days before infection to an average maximum of 0.48/1.0 (linear mixed effect model $p<0.05$). We observed the strongest associations between high gut microbiome abundance of E. faecalis, Klebsiella spp., E. coli, and S. marcescens of cases compared to all other samples ($p<0.01$).

Discussion: We show high relative abundance of the same strain that causes BSI in stool in hospitalized neonates. These data suggest the potential of gut microbial surveillance and management to identify blooms of gut pathogens to limit the incidence of bacteremia.

Arthropods are the Source of Genetically Diverse Enterococci

Bruna Sgardioli1 , Maria de Lurdes Enes Dapkevicius1, Paulo A.V. Borges1 , Ryan Whipple2, Abigail L. Manson3 , Ashlee M. Earl3, and Michael S. Gilmore2,4

1 University of the Azores, Angra do Herismo, Terceira, Portugal.
2 Harvard Medical School and Mass Eye and Ear, Boston, Massachusetts, USA.
3 Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.
4 Harvard Medical School, Boston, Massachusetts, USA.

Enterococci, leading causes of multidrug-resistant hospital infection, diverged from a common ancestor with vagococci (commonly found in the guts of fish) at the time animals colonized land and can be isolated from the guts of most land-dwelling animals – insects to humans. The first to colonize land were arthropods, and over 1,000,000 species of insects and invertebrates are believed to still exist. Preliminary work suggested that substantial enterococcal genetic diversity exists within insects and invertebrates. To test this, we collected insects and other invertebrates on the remote island of Terceira in the Azores. To maximize diversity and to detect the possible impact of human activity, insects were captured a) from the pristine wilderness at the top of the volcanic island, b) from agricultural pasturelands in the middle, and c) from the human-inhabited coastal region. From 125 insects, 125 presumptive enterococci were isolated and genomes sequenced. Twelve known enterococcal species were identified, the most common being E. faecalis, E. casseliflavus and E. mundtii. In addition, we obtained 13 with genomes more than 5% divergent (ANI) from a known species, indicative of novel species. Of candidate new species, 3 were also encountered in a parallel study in the lab. These 5 unique new species were obtained from Diptera, Annelid, Apidae, Syrphidae, Mythimna unipuncta, Popillia japonica, Forficula auricularia, Calosoma olivieri and Ocypus olens hosts. Given that we only have sampled 10’s of different insect species in total, we project that there are thousands of species of Enterococcus to be found possessing substantial genetic diversity. Enterococci are prolific genetic exchangers, and given the ubiquity of the species E. faecalis, which also colonizes and infects man, there is likely a very large pool of genes tuned to expression in the enterococcal background that pose a potential risk.

Hypermodification of dA by Mu Mom requires aminoacylated tRNA Gly

Rebekah M. B. Silva* 1, Anton Slyvka* 2, Yan-Jiun Lee 1, Chudi Guan 1, Sean Lund 1, Matthias Bochtler 2.3, and Peter R. Weigele

1 New England Biolabs
2 International Institute of Molecular and Cell Biology
3 Institute of Biochemistry and Biophysics, Polish Academy of Sciences

DNA hypermodification enzymes from bacteriophage install abundant and varied chemical groups on canonical bases that are protective against native host defenses (ex. Restriction endonucleases). Historically, much less has been understood about hypermodification of purines compared to pyrimidines. A classic example is the post-replicative hypermodification of adenine found in the phage Mu genome. The product of a single gene, mom (Modification of Mu), leads to an unusual carbamoylmethyl moiety found at the N6 position of adenine (6-NcmdA). Despite great effort, the group donor transferred by the Mom enzyme has eluded identification for over four decades. Mom has been predicted to belong to the Gcn5-Related N-Acetyltransferase superfamily (GNATs), most-commonly found to use acyl-CoA co-substrates for acylation of proteins, lipids, tRNA, and natural products. However, reconstitution of Mom activity in vitro has not been reported to date, even in the presence of putative co-substrates at high concentrations. Employing a suite of computational, analytical, biochemical, and biophysical tools, we have demonstrated that Gly-tRNA Gly is the co-substrate used by Mom to hypermodify dA, reconstituted the activity of Mom in vitro, and collected evidence for an on-base rearrangement of the glycyl group to yield the final structure of the Mu hypermodification. Using the predicted structure and an apo X-ray structure of Mom, we are gathering additional insights on the active site of Mom as well as the domains involved in co-substrate and substrate recognition. Furthermore, we have mined metagenome databases to find active Mom homologs and are currently aiming to harness the unique chemistry accomplished by Mom and Mom homologs to add new functionalities to DNA.

Genomic epidemiology and bacteriophage susceptibility over time of vancomycin-resistant Enterococcus faecium from bloodstream infections at a single hospital

Madison E. Stellfox, Hayley R. Nordstrom, Ryan K. Shields and Daria Van Tyne

University of Pittsburgh, Division of Infectious Diseases

Background: Enterococcus faecium is a gastrointestinal commensal and a difficult-to-treat nosocomial pathogen. Bacteriophage therapy is increasingly being studied as a potential treatment for E. faecium infections, and we recently treated a case of E. faecium bacteremia with two bacteriophages (9184 and Hi3). An important question regarding clinical use of bacteriophage therapy is scalability from personalized regimens to broader therapeutic tools. Therefore, we sequenced the genomes of 60 vancomycin-resistant E. faecium (VREfm) bloodstream isolates, contemporary with our clinical phage therapy experience, and characterized the lytic activity of both 9184 and Hi3 on all isolates.

Methods: 60 VREfm isolates from bloodstream infections in 58 unique patients were collected over two years and were sequenced on the Illumina platform. The earliest isolate was selected as a reference and was assembled with SPAdes and annotated with prokka. Variant calling was performed with snippy and a SNP-based, maximum likelihood phylogenic tree was constructed with RAxML. Pairwise SNP comparisons were used to assess genetic relatedness between isolates. Phage susceptibility of each isolate was determined in biologic duplicate with the soft agar overlay method.

Results: The 60 isolates spanned 12 different sequence types (ST), many of which are known to be hospital-associated. Using a 10 SNP cut-off, one third of the isolates were closely related to others in the dataset, suggesting patient-to-patient transmission. The probability of an isolate's phage susceptibility was stable over the study period. Approximately two-thirds of all isolates were susceptible to one or both phages, and susceptibility loosely clustered amongst isolates belonging to the same ST.

Discussion: Our findings suggest that a significant proportion of VREfm bloodstream infections are likely hospital-acquired. Differences in phage susceptibility between closely-related isolates will be investigated further. Overall, this study can serve as an initial step towards identifying genetic and phenotypic characteristics that could predict VREfm phage susceptibility.

Scalable pangenome-wide characterization of genetic variation in microbial species with Fusilli

Lucas R. van Dijk1,2, Abigail L. Manson1, Thomas Abeel2, Ashlee M. Earl1, Kiran V Garimella3

1 Infectious Disease & Microbiome Program, Broad Institute, Cambridge, MA 02142, USA
2 Delft Bioinformatics Lab, Delft University of Technology, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands
3 Data Sciences Platform, Broad Institute, Cambridge, MA 02142, USA

Identifying genetic variation among bacteria is critical for understanding their evolution, spread and functional potential. To date, most variant calling is done using a single reference, which often represents only a small fraction of a species pangenome, limiting the extent of variation that can be measured in an unrelated isolate. Graph-based approaches that accommodate additional species variation for interrogation are increasingly popular though most have limited applicability to bacteria. We developed Fusilli, a novel pangenome-wide variant calling tool designed specifically for diverse microbial species that combines multiple reference genomes in a Linked De Bruijn Graph and uses a multi-reference-assisted assembly algorithm to construct variant haplotypes. Fusilli works directly from raw reads, scales to hundreds of references and samples, and is able to identify more single nucleotide variants across a larger number of genes compared to single-reference tools, while also capturing large structural variants including unique genes. This tool will greatly improve our ability to analyze diverse bacterial genomes, including calling variants across an entire pangenome and rapidly detecting horizontally transferred and novel genes.

Characterizing the soil microbial community composition of MtDef4 defensin Arabidopsis variants

Rhiannon Vargas1, Maggie Newberry 1, Kimberley Sukhum 1,2, Gautam Dantas 1,2,3,4

1 The Edison Family Center for Genome Sciences and Systems Biology, Washington University in St. Louis School of Medicine, St. Louis, MO USA
2 Department of Pathology and Immunology, Washington University in St Louis School of Medicine, St Louis, MO, USA
3 Department of Molecular Microbiology, Washington University in St Louis School of Medicine, St. Louis, MO, USA 4 Department of Biomedical Engineering, Washington University in St Louis, MO, USA

Background: Plant defensins such as MtDef4 have antimicrobial properties and have the potential to be implemented in agricultural crops to decrease pathogen burden. Little is known about how plants with the MtDef4 gene influence plant growth and belowground microbial communities. Subcellular localization of transgenic defensins is shown to be a determinant of its antimicrobial activity but they are not well characterized with MtDef4. We predicted that the microbial community assembly on the roots and in the soil of Arabidopsis thaliana expressing MtDef-4 extracellularly (Ec) or in the endoplasmic reticulum (ER) will be different from the microbial communities of Col-0 wild-type (WT) plants.

Methods: A longitudinal study design was conducted with three Arabidopsis genotypes - Ec, ER, WT – and soil as a control. Growth measurements were collected along with three belowground compartments for each plant – endophytic root, rhizosphere soil, and bulk soil – at time of harvest. Plants were randomly assigned a harvest day starting at day 0 (post-germination), 24 hours, and week 1, 2, or 3. Each factor had a replicate of 10 (n=200). 16S rRNA and ITS sequencing were conducted to characterize the bacterial and fungal communities respectively.

Results: MtDef4 plants grew at a faster rate than WT plants. Bacterial alpha diversity and plant biomass were positively correlated. The composition of bacterial communities changed temporally but was generally consistent between genotypes, whereas subtle beta diversity differences were observed. Two phyla of interest that differ between genotypes overtime are Verrucomicrobia and Acidobacteria. Preliminary ITS results suggest that there may differences in Basidiomycota and Ascomycota phyla in the rhizosphere between genotypes.

Discussion: Differences between genotypes are shown when comparing growth rates, beta diversity overtime, and certain microbial phyla. Future directions include investigating significant bacterial and fungal correlations. This study will inform fundamental questions regarding the biological impacts of the MtDef4 plant-microbiome overtime.

Progress Towards Standardizing Metagenomics- Application of Metagenomic Reference Materials to Develop a Reproducible Microbial Lysis Methodology with Minimum Bias

Michael M. Weinstein 1,2, David C. Danko 3,4, Elaine Wolfe 1, Shuiquan Tang 1, Karen G. Jarvis 5, Christopher J Grim 5, Venu Lagishetty 6, Jonathan P. Jacobs 6, Jason Arnold 7, Ryan Kemp 1, and Christopher Mason 4,8

1 Department of Bioinformatics, Zymo Research Corp., Irvine, CA, USA.
2 Molecular, Cell, and Developmental Biology/Quantitative and Computational Biosciences Institute, University of California, Los Angeles, CA, USA.
3 Tri-Institutional Computational Biology and Medicine Program, Weill Cornell Medicine of Cornell University, New York, NY, USA.
4 Institute for Computational Biomedicine, Department of Physiology and Biophysics, Weill Cornell Medicine of Cornell University, New York, NY, USA.
5 Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, College Park, MD, USA. (6) Microbiome Core, Division of Digestive Diseases, Department of Medicine, University of California, Los Angeles, CA, USA.
7 Microbiome Core, School of Medicine, University of North Caroline, Chapel Hill, NC, USA.
8 The Feil Family Brain and Mind Research Institute, Weill Cornell Medicine of Cornell University, New York, NY, USA

Background: Citations involving microbiome analysis have undergone an exponential growth since approximately 2010 with little evidence of slowing down. Along with this rapid growth has come evidence of technical weaknesses causing a lack of reproducibility between different methods. This lack of reproducibility limits our ability to draw new conclusions from microbiome analyses and can call previous conclusions into question. To address these issues, leaders in the field have highlighted the importance of standardized protocols. However, the likelihood of a substantial portion of the microbiome research community standardizing around a single method is unlikely and will still leave unanswered the questions about the reality of a sample's nature.

Methods: To address these and similar challenges, we sought to compare the performance of many different microbiome lysis and extraction methods from both commercial and academic sources when extracting DNA from a mock microbial community standard of defined composition. In order to compare accuracy between different methods, we performed 16S rRNA gene sequencing as well as shotgun metagenomic analysis. Additionally, we present a new statistical measurement: The Measurement Integrity Quotient, providing a single, easy to understand number score that describes the accuracy of an observed composition from a known composition standard. Utilizing this method, a

comprehensive analysis of many variables potentially determining lysis efficiency were tested, including differences between thermal, enzymatic, and mechanical lysis.

Results/Discussion: Utilizing the data gathered from these analyses, we were able to determine which methods are likely to produce nearly complete microbial lysis. Additionally, we have developed and publicly released open-source bioinformatics packages for rapidly assessing the accuracy of a microbial extraction method relative to a defined composition microbial standard. The use of these data and analyses enables the selection of methods that generate minimal bias and facilitate benchmarking and quality control for many different microbiome pipelines.

Characterization of the gut metagenome in adults and children with asthma

Naomi G. Wilson, Leonard B. Bacharier M.D., Andrew L. Kau M.D. Ph.D.

Washington University School of Medicine in St Louis MO, Vanderbilt University Medical Center Nashville TN

Background: Asthma is a common allergic airway disease whose etiology remains elusive. The composition and function of the infant microbiota are linked to asthma risk, but the microbiota's role in older patients with established asthma is less clear.

Methods: Here, we performed metagenomic shotgun sequencing on human stool collected from patients with and without moderate-to-severe asthma to characterize the functional profiles of the gut microbiota in asthma beyond early childhood.

Results: Asthma contributes to beta diversity variation in gene content even when accounting for other important clinical demographics. Long-chain fatty acid metabolism gene-sets were associated with asthma and may be indicative of the involvement of these pathways' metabolites in airway smooth muscle and immune responses already associated with asthma. Antibiotic resistance gene (ARG) profiling revealed that people with asthma have a higher richness but a similar load of ARGs compared to healthy people. Further, we found an ARG enriched in asthma metagenomes that confers resistance to macrolides, which are often prescribed for asthma exacerbations, and most often encoded by Bacteroides species. Given our previous finding that the B. fragilis toxin, BFT, can alter lung inflammation in humanized gnotobiotic mice and is more prevalent in the asthma cohort compared to healthy, we calculated co-occurrence of the macrolide resistance marker and bft. We found that proportionally more MARS samples encode both genes compared to publicly available B. fragilis genomes, suggesting a possible relationship between B. fragilis toxin, antibiotic resistance in the asthmatic gut.

Discussion: Our results are the first to characterize the functional underpinnings of the gut microbiota in moderate-to-severe physician-diagnosed asthma and provide several avenues for future studies of gut-lung axis host-microbe interactions.

Automating microbial taxonomy workflows with PHANTASM: PHylogenomic ANalyses for the TAxonomy and Systematics of Microbes

Joseph S. Wirth, Eliot C. Bush

Harvey Mudd College

The nucleotide sequences of 16S ribosomal RNA (rRNA) genes have been used to inform the taxonomic placement of prokaryotes for several decades. Whole-genome approaches can better resolve evolutionary relationships of organisms, but these analyses often require computational proficiencies that are uncommon among microbiologists. PHANTASM is a new tool capable of automating these workflows. This tool was designed to work for a wide range of prokaryotes and is the first example of an automated reconciliation of NCBI's Taxonomy database with that of the List of Prokaryotic names with Standing in Nomenclature (LPSN). In this study, we describe the workflow of PHANTASM and provide several examples of results generated by it. The source code is freely-available on GitHub. In order to facilitate the ease-of-access for researchers, PHANTASM is also available as a Docker image. While other tools exist to facilitate starting points for these analyses, PHANTASM provides users with a greater degree of control and produces outputs that can be used to make publication-quality figures.

Translational Ramifications of Crowd-Sourced Genomics Data

David A Yarmosh, Juan G. Lopera, Nikhita P. Puthuveetil, Patrick Ford Combs, Amy L. Reese, Corina Tabron, Amanda E. Pierola, James Duncan, Samuel R. Greenfield, Robert marlow, Stephen King, Marco A. Riojas, John Bagnoli, Briana Benton, Jonathan L. Jacobs

American Type Culture Collection (ATCC), Manassas, Virginia, USA

Background: Public genomics databases serve a critical role in the life science research community. Despite existing guidelines which require metadata associated with a given genome assembly, other relevant data (e.g., sequencing platform, assembly method) are often incomplete or missing. Ultimately, this gap renders the assembly data itself questionable from the perspective of reliability, traceability, and accuracy. Previously, Yarmosh et al. illustrated the impact of poor data provenance by comparing several publicly available assemblies to assemblies which had complete traceability. While it was found that some public assemblies were labelled as derivatives of ATCC™ source material, there was a tendency toward fewer relative variants between these assemblies and their ATCC Standard Reference Genome (ASRG) counterparts. However, several of these assemblies still contained a large quantity of variants, including those inducing translational changes.

Methods: To better understand the consequences outlined in a previous study, the Prokaryotic Genome Annotation Pipeline was run on 190 public assemblies labeled as ATCC type material and their 127 corresponding ASRGs. Annotations from both sets were compared in terms of amino acid identity, gene count, gene identity, and the gain/loss of stop codons.

Results: Despite the claim of being assembled from type material, 25 of 190 public genomes contain premature stop codons and over 35,000 of these annotations have less than 50% reciprocal identity relative to their ASRGs.

Discussion: Public genomics databases are unable to curate their immense library of submitted assembly data to ensure the utmost quality. The de facto usage of these databases in modern research coupled with the repercussions of incomplete metadata underscores the urgent need for a more stringent curation process, such that future research and public health are not burdened by unreliable data. Toward this goal, ATCC's Enhanced Authentication Initiative aims to provide high quality reference genomic data directly from ATCC™ source material.

Comparison of long and short read metagenomic assemblers for recovery of low-abundance species

Sosie Yorki; Terrance Shea; Colin J. Worby; Abigail L. Manson; Ashlee M. Earl

Infectious Disease and Microbiome Platform, Broad Institute of MIT and Harvard

Recent technological and computational advances have made metagenomic assembly a viable approach to achieving high resolution views of complex microbial communities. In previous benchmarking studies, short-read (SR) metagenomic assemblers had the highest accuracy, long-read (LR) assemblers generated the longest sequences, and hybrid (HY) assemblers utilizing SR and LR balanced length and accuracy. However, until now, no benchmarking studies have specifically addressed the ability of assemblers to recover low-abundance organisms, which typify pathogens and other clinically relevant organisms in the gut. We generated synthetic long- and short-read metagenomic datasets by computationally spiking E. coli isolate reads into fecal backgrounds at fixed coverage levels, and compared the ability of different assemblers to generate E. coli scaffolds and recover antibiotic resistance genes. Then, we assessed how each assembler's performance changed when a competing E. coli strain was spiked in. Although SR assemblers recovered more resistance gene content with high accuracy, even at low coverages, the increased contiguity in LR assemblies allowed for greater association of E. coli genes with their taxonomic origin. HY assemblies identified resistance genes with high accuracy, but had lower contiguity than LR assemblies. LR and HY assemblers were able to recover most ARGs on E. coli-specific contigs and assemble 99% of an isolate plasmid, even with coverage as low as 4x. Our results highlight the trade-offs associated with each approach, and that the optimal tool is goal-dependent. For strain characterization and determining gene context, LR assembly is optimal, while for base accurate gene identification, SR assemblers outperform other options. HY assembly offers contiguity and base accuracy, though requires generating data on multiple platforms, and may suffer high misassembly rates when strain diversity exists.

Evolving the biofilm transcriptome of Mycobacterium tuberculosis

Madison A. Youngblom, Tracy M. Smith, Mohamed A. Mohamed, Caitlin S. Pepperell

Microbial Doctoral Training Program, University of Wisconsin-Madison; New York Genome Center; Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison; Department of Medicine, School of Medicine and Public Health, University of Wisconsin-Madison

Background: Tuberculosis (TB), caused by Mycobacterium tuberculosis (M. tb), is a leading cause of death due to infectious disease. TB is not often thought of as a biofilm infection, however M. tb readily forms a pellicle biofilm at air-liquid interfaces. Often underrecognized is the role that M. tb biofilms play in infection, including their contribution to drug and immune tolerance.

Methods: Here we used experimental evolution and transcriptomics to investigate adaptation of M. tb to this complex phenotype. We first selected a sample of clinical isolates and passaged them under selective pressure to grow as a biofilm. We then performed RNA-sequencing on the ancestral and evolved isolates to compare transcriptomes both across different strains as well as across evolutionary time.

Results: We found significant diversity of genes differentially expressed during biofilm growth among our ancestral isolates, with very few differentially expressed genes (DEGs) shared across isolates. After applying selective pressure to grow as a biofilm however, we observed convergence towards a shared biofilm transcriptome. Even more interesting is the observation that gene expression patterns of our strains converged in a genetic-background dependent manner; strains from two different sub-lineages of L4 evolved distinct biofilm transcriptomes.

Discussion: Elucidating strategies used by M. tb to adapt to new environments will help us to continue develop treatments that overcome M. tb's persistence. We have shown here that even closely related strains of M. tb have diverse transcriptomic landscapes, and respond differently to selective pressure, which has important implications for how we approach treatment of infections with different strains. We have also shown that it is possible to evolve bacterial transcriptomes in the laboratory, opening a new avenue for studying bacterial adaptation in vitro. Overall, this work highlights the complexity of bacterial adaptation, as well as the significant role played by epistasis in evolution to new environments.

Investigating the effects of temperate prophage elements on antibiotic resistance evolution.

Amy Zamora, Siˆan V. Owen, & Michael Baym

Laboratory of Systems Pharmacology and Department of Biomedical Informatics, Harvard Medical School, Boston, MA.

Pathogenic bacteria commonly experience exposure to antibiotics deployed to kill or inhibit growth. This pressure drives bacterial evolutionary pathways by selecting for genomic traits that enhance resilience. Bacterial viruses (bacteriophages) also act as a selective force on bacterial genomes. Importantly, lysogenic bacteriophages integrate into host genomes upon infection and are ubiquitously present as ‘prophages’. Bacteriophage integration is stable, but activation of the SOS response pathway can cause these integrated viruses to become lytic. In this process, bacteriophages exit the genome, replicate, and lyse their host to release new progeny. Several DNA-damaging antibiotics (e.g., Ciprofloxacin) trigger lysis through this process, known as “prophage induction”. It has been demonstrated that prophages in S.typhimurium genomes can increase bacterial susceptibility to antibiotics. This increased sensitivity could be due to phage-mediated death as a result of prophage induction, preventing growth of natural resistance mutants. Although this work elucidated how lysogenic phages affect antibiotic response, the interaction between prophages and antibiotic resistance evolution has yet to be rigorously investigated. To study this relationship, Ciprofloxacin resistant mutants were evolved in isogenic prophage (+) and (-) strains of S.typhimurium and challenged against a gradient of Ciprofloxacin to observe resistance development and extent. These mutants were collected and sequenced to understand mutational diversity at varying antibiotic concentrations and prophage effects on the mutational evolution pathways taken by bacterial hosts. Results yield that although the distribution of Ciprofloxacin mutants remains similar for both strains of S.typhimurium, the overall quantity of mutants evolved is increased in prophage (-) strains. Sequencing results further elucidate the effect which this mobile genetic element has on evolutionary pathways to resistance. Such results further characterize the specific role inducible prophages play in both the direct mechanisms they leverage to interact with antibiotics and the evolutionary mechanisms they exploit to influence bacterial survival and antibiotic resistance.

CRISPR spacer acquisition is a rare event in human gut microbiome

An-Ni Zhang1, Jeffry M Gaston2, Shijie Zhao1, Xiaoqiong Gu3, Eric J Alm1*

1Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, USA;
2 Google, Cambridge, USA;
3 Antimicrobial Resistance Interdisciplinary Research Group, Singapore-MIT Alliance 9 for Research and Technology, 1 Create Way, Singapore 138602, Singapore
* Corresponding author:
Address: Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, USA.

Host-parasite (host-virus) interactions are important for all cellular life, for example, human immunity and SARS-COV2. In bacteria and archaea, CRISPR systems actively acquire spacers to ensure continued defense against phages, which have been reported to acquire a new spacer within a few hours or days in an experimental setting. However, spacer acquisition in natural environments has often been to slow to observe, and limited literature reports suggest a much slower rate (new spacers only acquired over months or years). This high variance highlights the needs to improve our understanding of host-parasite interactions in natural environment.
By investigating temporal WGS datasets and metagenomes of human gut microbiome in healthy individuals, we found that spacer acquisition was a rare event in human gut microbiome, with an average rate of 1 spacer per 2,142-5,000 cell divisions, i.e. over 7-8 years. This low rate reflects only a small proportion of phage (0.4%-0.6% infection rate) challenged the CRISPR systems in human gut microbiome. The rare spacer acquisition in CRISPR suggest that CRISPR might not be the primary risk of effective phage therapy for human microbiome. The results of this study may inform future efforts involving phage therapy and pandemic defense.

Transcriptome Analysis of planktonic and biofilm cultures from several strains of Variovorax

Christina Zhang, Kalea Bringas, Paul Orwin

University of the Pacific

Variovorax is a genus in the Betaproteobacteria that exhibits diverse morphological, physiologic, and metabolic characteristics. This genus is thought to be important in the formation of the rhizosphere community. Previous studies have identified the static biofilm transcriptome of V. paradoxus EPS. Recent work on finishing the genomes of many Variovorax isolates has shown that the genomes are frequently multi-partite. We now wish to examine the roles of genes on these different replicons in biofilm formation, and to evaluate this phenotype more broadly. We evaluated two distinct biofilm growth phases (static and colony biofilms) along with two planktonic growth phases (logarithmic and stationary). A major goal of our study is to understand the common genetic programming that is responsible for changes in gene expression across these different bacterial isolates. Triplicate biological samples were imaged, tracked for growth, and RNA was extracted for the strains 4MfCol3.1, 110B, MF350, NFACC26, ATCC 17713(NBRC 15149), EPS, and VAI-C in 2.5 g/L YE media. 20-30 million paired end reads (2x150bp) were acquired from each sample and aligned to the genome. Prokka annotations were used to identify genes in the genomes, unless they had already been annotated by PGAP. The comparisons were used to identify genes up and down regulated >2-fold when comparing alternative conditions. Based on PCoA clustering, we observed substantial variation between the forms of biofilms, as well as the planktonic cultures. We also observe that very large sets of genes (1000-3000) are altered in expression during biofilm formation in all strains tested. We have also identified genes in all different replicons with differential biofilm regulation, and have identified genes commonly upregulated in biofilms by different strains. This work will form the basis of further comparative studies in biofilm formation across this important genus.