Control Number: 2022-A-17-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: A flexible Galaxy-based platform for the analysis of microbial WGS data in public health

Author Block: **B. Bogaerts**, J. Van Braekel, R. Winand, N. H. Roosens, S. C. De Keersmaecker, K. Vanneste; Sciensano, Brussels, BELGIUM.

Abstract Body: In recent years, the use of whole-genome sequencing (WGS) has increased substantially in the public health sector. Sciensano (the Belgian Institute for Health) has set up a public Galaxy instance enabling scientists to perform complete characterization of their microbial WGS data. Besides support for community-endorsed staple tools available from the Galaxy ToolShed, several optimized bioinformatics push-button pipelines, resulting in a summarizing analysis report, were developed in-house for priority pathogens, such as *Escherichia coli*, *Mycobacterium tuberculosis*, and *Neisseria meningitidis*. These workflows were validated extensively and described in peer-reviewed publications. The same underlying architecture was used to construct additional (unvalidated) workflows for other bacterial pathogens, including *Shigella spp*., *Enterococcus faecalis*, *Enterococcus faecium*, *Listeria monocytogenes*, and *Staphylococcus aureus*. Individual assays such as, for example, antimicrobial resistance gene detection or sequence typing, are also provided as stand-alone tools, with the option to use different detection methods (i.e., BLAST+, SRST2, KMA) and various automatically updated databases (e.g., ResFinder, PointFinder, PubMLST, EnteroBase, ⋯). The Galaxy instance also contains tools for phylogenomic investigation, supporting both core-genome MLST and SNP-based approaches. Most of these tools and workflows generate interactive HTML reports that can be accessed within Galaxy. An extensive set of video tutorials is provided to train scientists to use these tools efficiently. This Galaxy-based platform has contributed to the successful integration of WGS into the routine and research activities of Sciensano. The instance is freely available for non-profit and academic usage at https://galaxy.sciensano.be/.

---PAGEBREAK---

Control Number: 2022-A-19-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: NGS-Based Agnostic Diagnostics: BARDA’s Vision for Pandemic Preparedness

Author Block: **D. Mittar**, K. Caravelli, S. Selimovic, S. Patel;
BARDA\DRIVe, Washington DC, DC.

Abstract Body: The availability of widespread diagnostic capability is critical for an effective pandemic response. Today’s diagnostic tools are designed to specifically detect a single pathogen or a few pathogens, which must already be known. Although these targeted tests are critical for routine diagnosis and management of infectious diseases, they take time to develop and validate. When a new pathogen emerges in the future, a diagnostic capability that can identify any and every pathogen—an agnostic diagnostic—will be needed to inform the public health response and individual care paths on day one. Next generation sequencing (NGS) and specifically metagenomic NGS (mNGS) could provide a promising agnostic

diagnostic capability for infectious diseases. Already in use worldwide for genomic surveillance, a clinical application of mNGS would be informative—especially in case of future pandemics—to individuals in directing their care pathway and to public health authorities in understanding disease prevalence before validated, targeted diagnostics are widely available. The Biomedical Advanced Research and Development Authority's (BARDA) Division of Research, Innovation, and Ventures (DRIVe) recognizes the potential of mNGS technology as an agnostic diagnostic tool. To advance this capability toward commercialization, DRIVe has funded five industry and academic partners to develop, verify, and validate mNGS-based agnostic diagnostic assays for respiratory RNA viruses. DRIVe's partners are addressing several challenges associated with clinical implementation of this technology, including lowering the sample-to-result time to under 24 hours; reducing interference from host RNA; performing analytical validation of their platforms using both contrived and clinical respiratory samples; developing protocols for performing high-throughput testing, and pursuing appropriate regulatory strategies (particularly for use in public health emergencies). Clinical mNGS has the potential to revolutionize diagnostics as a pathogen-agnostic tool for infectious disease management and pandemic preparedness. DRIVe seeks to engage the mNGS community regarding this vision, to better understand nuanced challenges and channel interest toward implementation and regulatory approval.

---PAGEBREAK---

Control Number: 2022-A-23-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Whole Genome Sequencing Reveals Diverse Modes of Carbapenemase Gene Dissemination in Indian Klebsiella pneumoniae clinical isolates

Author Block: **V. Shamanna**[1], A. Underwood[2], S. Argimón[3], A. Prasanna[1], G. Nagaraj[1], V. Govindan[1], N. S[1], M. Bhaskaran[1], D. M. Aanensen[3], K. Ravikumar[1];
[1]Central Research Laboratory, Bengaluru, INDIA, [2]Broken String Biosciences LTD, BioData Innovation Centre, Cambridge, London, UNITED KINGDOM, [3]Centre for Genomic Pathogen Surveillance, Big Data Institute, University of Oxford, Old Road Campus, Oxford, London, UNITED KINGDOM.

Abstract Body: Mobile genetic elements and plasmids harboring multiple drug resistance genes are a major concern in Klebsiella pneumoniae causing nosocomial infections. The dynamics underlying the dissemination of these genes remain poorly resolved in the Indian context. Hence, using whole-genome sequencing (WGS), we described the mobile genetic elements (MGE) harboring carbapenemases genes and studied the evolutionary dynamics of horizontal gene transfer (HGT) of these genes within the population.

A total of 1072 K. pneumoniae clinical isolates collected from 21 centers across India from the year 2013-2021 were sequenced on the Illumina platform. Antimicrobial resistance (AMR) determinants, mobile genetic elements (MGEs), and plasmids were identified using AMRfinder, MobileElementFinder, and Plasmidfinder respectively. The sequence type analysis was performed using pubMLST. Additionally, subsets of isolates were sequenced using Nanopore to determine carbapenemase-encoding plasmids' complete structure.

The 1072 isolates, belonged to 105 different STs and ST231 (n=225), ST147(n=170) & ST395 (n=83) were the major lineages. The AMR analysis showed 791 strains (74%) as carbapenemase-producing K. pneumoniae (CP-Kp) with 54% (n=427) producing OXA48-like

carbapenemases; 17% (n= 138) NDM producers and the remaining 29% (n= 226) are co-producers of NDM and OXA. The plasmid analysis revealed carbapenemase gene dissemination through 104 different plasmids. blaOXA-48-like genes such as OXA-181 and OXA-232 have spread primarily via the ColKP3 plasmid transmitted predominantly by stable association with clonal lineage ST147 and ST231 respectively. Whereas, blaNDM genes have spread via IncF type of plasmids across numerous lineages. Further, the complete structure of plasmids carrying the NDM and OXA genes was deduced through the hybrid assembly. Mobile genetic element analysis in genomic surveillance systems provides a more comprehensive understanding of AMR spread. Identifying the key molecular markers of resistance and plasmids harboring them improves strategies to control AMR dissemination.

---PAGEBREAK---

Control Number: 2022-A-25-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Utility of Whole Genome Sequencing in Diagnosis of Mono- and Polyresistant Tuberculosis

Author Block:
M. Dohál[1], I. Porvazník[2], M. Škereňová[3], J. Mokrý[4];
[1]Department of Pharmacology and Biomedical Centre Martin, Jessenius Faculty of Medicine in Martin, Comenius University, Martin, SLOVAKIA, [2]National Institute of Tuberculosis, Lung Diseases and Thoracic Surgery, Vyšné Hágy, SLOVAKIA, [3]Department of Molecular Medicine and Biomedical Centre Martin, Jessenius Faculty of Medicine in Martin, Martin, SLOVAKIA, [4]Department of Pharmacology, Jessenius Faculty of Medicine in Martin, Comenius University, Martin, SLOVAKIA.

Abstract Body:
Tuberculosis (TB), especially drug-resistant forms, is still a global medical emergency. The global prevalence of mono- and polyresistant strains is almost 17%. The detection and rapid diagnosis of these cases is very important, as incorrect treatment and management can result in the development of multidrug resistant TB. Whole genome sequencing (WGS) is an increasingly preferred method in the diagnostics and monitoring of the transmission dynamics of the resistant forms of TB. The study was aimed to compare WGS data and results of phenotypic susceptibility testing (pDST) and to determine the degree of false positivity and negativity, respectively. This cohort study performed WGS on 66 *Mycobacterium tuberculosis* isolates, including 47 monoresistant (71.2%) and 19 polyresistant (28.8%) strains characterized by pDST on solid and liquid media (predominantly resistant to isoniazid - INH, streptomycin - STM and pyrazinamide - PZA; we excluded monoresistant strains to rifampicin, as these strains were analyzed in several studies). The data showed the variable sensitivity and specificity of WGS in the identification of gene variants encoding drug resistance: INH 88.9% and 97.44%; STM 41.2% and 100.0%; PZA 12.5% and 96.0%. The low sensitivity of WGS in STM resistance can be explained by the poor bacterial growth on the solid medium containing STM in the critical concentration (CC) recommended by the WHO (4mg/L). We also performed pDST above the CC (10mg/L) and did not observe any growth. In addition, variant calling analysis revealed 2 novel single nucleotide polymorphisms and 2 novel deletions in *gidB* gene, which role in resistance to STM should be explored. PZA resistance was encoded predominantly by a mutation in the *pncA* gene. However, many phenotypically resistant strains did not harbor these mutations, indicating a high rate of false-positive pDST results. Moreover, 11 isolates

(16.7%) phylogenetically belonged to the highly virulent Beijing lineage 2.2.1 and 55 isolates (83.3%) belonged to the Euro-American lineage In this study, we emphasized the need to apply WGS in clinical settings to establish an appropriate treatment regimen for the patient and thus prevent the development and spread of resistant TB.This research was funded by grant APVV-18-0084 and VEGA-1/0093/22.

---PAGEBREAK---

Control Number: 2022-A-28-NGS
Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology
Topic 2:
Publishing Title: Surveillance of SARS CoV-2 genomic variants in wastewater
Author Block: **s. sherchan**;
Morgan State University, Baltimore, MD.
Abstract Body: Genomic surveillance of wastewater sampling sites are increasingly exploited to map the prevalence and spread of SARS CoV-2 variants in Nepal. We launched Environmental surveillance project back in March 2021, and successfully demonstrated the feasibility of environmental surveillance as supplementary epidemiological tool with clinical surveillance. For effective public health measures, the understanding of the circulating variants of the virus is very important. Environmental genomic surveillance is expected to predict upcoming variant(s) with possible next wave of COVID-19. The main aim of genomic surveillance project in Nepal is to inform public health authorities and relevant stakeholders regarding the circulating SARS CoV-2 variants. The project also aims to understand the feasibility to utilize the wastewater based epidemiology for minimizing the infectious diseases in low income countries like Nepal. So far, key variant assignment is based on mutations in spike protein of the virus. Therefore we approached spike gene amplicon sequencing with NGS strategy on Illumina sequencing instrument. Out of 72 sewage and hospital wastewater attempted, significant number of reads were obtained only 49 samples with higher reads in hospital wastewater samples. We assigned variants in the sequences using “kallisto” bioinformatics pipeline. Most frequent mutations were SNPs - T477K, D614 and L452, indicating mutations in ACE-2 receptors. Among assigned eleven variants the dominant one was Delta variant until mid-December 2021, then after replaced by Omicron which correlates with clinical scenario. Nepal had Omicron wave with massive positivity rates in January 2022. Though further analysis with whole genome is continued, our results genomic surveillance provide usefulness of wastewater testing to predict genomic variants of SARS CoV-2 in poor resource setting countries. Our results are suggestive that genomic surveillance of SARS CoV-2 variants in wastewater is feasible and can be exploited as supplementary to clinical surveillance for effective public health interventions to minimize the spread of the diseases.

---PAGEBREAK---

Control Number: 2022-A-33-NGS
Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology
Topic 2:

Publishing Title: Detecting within-host diversity of *Staphylococcus aureus* during colonization: "Pool" genomes vs. Individual genomes

Author Block: **V. Raghuram**[1], J. J. Gunoskey[2], N. F. Jacko[2], K. S. Hofstetter[1], T. D. Read[1], M. Z. David[2]; [1]Emory University, Atlanta, GA, [2]University of Pennsylvania, Philadelphia, PA.

Abstract Body: As pathogenic bacteria go through cycles of growth and adaptation within a host, the genetic makeup of the initial population may change with time. A straightforward approach to tracking a microbial population within an infected or colonized host would be to isolate and characterize the genome of a single colony obtained from a culture sample. However, this method may not capture the complete genetic diversity in the population. Alternatively, the “pool” of all colonies obtained from a culture can be investigated, but this bears the disadvantage of having a non-homogeneous sample, making it difficult to perform specific experiments. To compare the genetic diversity detected between 8 single-colony isolates (SCIs) and all pooled colonies, we periodically sampled 3 specific body-sites of *Staphylococcus aureus* colonization on subjects presenting with a methicillin-resistant *S. aureus* skin and soft-tissue infection. *S. aureus* is a ubiquitous pathogen and one of the leading causes of healthcare associated infections worldwide. From 86 subjects, we obtained 257 pools and 2056 SCIs. Using whole genome sequencing (WGS), we compared the total nucleotide diversity in each pool to the corresponding SCIs. We found that in most cases, the S. aureus colonizing isolates did not stem from multiple clonal lineages; the genetic diversity observed within a patient was relatively low, suggesting single strain origins. However, few pools contained greater genetic diversity than what we observed from the genomes of 8 SCIs combined, suggesting that the pool may contain clinically relevant genes with phenotypes that may be missed when only examining SCIs. This was corroborated by our analysis of the number of antimicrobial resistance (AMR) genes in the pools vs SCIs. The median number of AMR genes detected in the pools was one greater than found in genomes of SCIs. In rare cases pools had >16 AMR genes, none of which were found even from genomes of 8 SCIs combined.

---PAGEBREAK---

Control Number: 2022-A-44-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Single cell microbial sequencing for global surveillance of AMR

Author Block: **B. AVOT**;
DTU, Lyngby, DENMARK.

Abstract Body: **Background:** Antimicrobial resistance (AMR) in pathogenic bacteria was the cause for 35,000 deaths in the US according to the CDC and has been qualified by the WHO as one of the biggest threats for public health. There is therefore a need for improved surveillance programs that include NGS technologies. As per now most studies regarding AMR have relied on clinical isolates or classical environmental shotgun sequencing. This study shows the potential of single cell sequencing of microbial communities and suggest how it could be a more specific alternative to the aforementioned shotgun metagenomics in global surveillance programs. This study aims to develop a method to infer a taxonomic assignment and identify AMR genes for each sequenced microbe. It also has for goal to develop a workflow that will uncover the dynamics of AMR through plasmid dissemination.

Methods: A pilot study was conducted on three environmental samples: two sewage samples from Tchad and Bangladesh and one fecal sample collected in a Danish pig farm. All samples were deep and shallow sequenced (> 100,000 reads and > 10,000 reads respectively) to show the analyses enabled by each depth. To assign taxonomy as well as identify AMR genes KMA (an ultra-fast k-mer mapping tool) was used against the GTDB-Tk microbial marker genes database and ResFinder respectively. Furthermore, a workflow is introduced to assemble the deep sequenced droplets, assess genome quality, and predict plasmid contigs.
Results: Successful association between taxon with AMR genes was achieved for two thirds of the sequenced microbes. However, a large proportion of the microbes remains unassigned as a result of the bias of microbial databases towards genomes sequenced in high income countries and computational time concerns. Assemblies of deep sequenced droplets and the genome assessment process confirmed that higher sequencing depth enables enhanced genome qualities. Finally, plasmid contig prediction emphasized the need for more accurate machine learning predictor tools.
Conclusions: Single cell metagenomics was proved to be a very powerful approach that enables a more accurate overview of microbial communities and fills in some gaps left by shotgun sequencing. Shallow sequencing was proved to be a cost-efficient way to grasp the taxonomic and AMR composition of a microbial community; deep sequencing showed potential to study in silico the mechanisms of AMR.

---PAGEBREAK---

Control Number: 2022-A-54-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Prediction of antimicrobial resistance and associated whole genome analysis for Salmonella Heidelberg isolated from Minnesota in 10 years

Author Block: **J. Haan**[1], D. Boxrud[2];
[1]Minnesota Department of Health, St. Paul, MN, [2]Centers for Disease Control and Prevention, Atlanta, GA.

Abstract Body: Title : Prediction of antimicrobial resistance and associated whole genome analysis for Salmonella Heidelberg isolates from Minnesota in 10 years BackgroundSalmonella Heidelberg has been one of the most common serotypes associated with Salmonellosis foodborne illness in US. Understanding of ARM in foodborne pathogens is a critical to treat clinical cases in terms of using an appropriate drug to shorten the period of hospital stay. To have better insights of changes on the resistance profiles over time and their associated genomic features, 146 Salmonella Heidelberg collected between 2005 and 2014 were used in this study.
Methods139 clinical isolates 7 non-clinical isolates were performed for E-test for minimum inhibitory concentration (MIC) and WGS. 12 antibiotics were used to screen phenotypic AMR according to CLIA guideline. The same isolates were sequenced using Nextera XT and Illumina V2 chemistry on Miseq platform. The detection of AMR gene and sequencing analysis were performed using Bionumerics 7.6 software.
ResultsPhenotypically, 65.9% of isolates were resistant to Streptomycin followed by Ampicillin (35.2%) and Tetracycline (31.5%). None of S. Heidelberg was resistant to Ciprofloxicin. 33.6% of isolates were multidrug resistance. There was no correlation between

the level of resistance and the collection year. 29 resistance genes that are belonging to 8 different drug classes were detect from WGS analysis. The most common resistant genes detected were aph(3")-lb (28.7%), aph6-ld (26.7%) and blaCMY (16.4%). Genotype-phenotype concordance was 92%. Phylogenetic relationship was assessed by cgMLST indicating 6 distinct sub-clusters are associated with their resistome profile. ConclusionWGS as a sole diagnostic tool to detect resistance has some limitations due to the discrepant results when compared with phenotypic resistance data. However, our study proposes that WGS data acquired from routine surveillance system would provide an efficient way to monitor the level of antimicrobial resistance with potential applications in clinical settings.

---PAGEBREAK---

Control Number: 2022-A-56-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Transcriptomic Signature for the identification of Drug-Susceptible *Mycobacterium tuberculosis*

Author Block: **H. Poonawala**[1], S. Kuchibhotla[2], K. Zhang[3], M. R. Farhat[3];
[1]Tufts Medical Center, Boston, MA, [2]Harvard College, Cambridge, MA, [3]Harvard Medical School, Boston, MA.

Abstract Body: Background: Transcriptomic-based diagnostics identify antibiotic-induced differentially expressed genes (DEGs) in susceptible organisms that are absent in drug-resistant organisms and are independent of resistance mechanism. We hypothesized that antibiotic-induced DEGs in *Mycobacterium tuberculosis* (MTb) can be used as a marker of drug-susceptibility. Methods: We searched for MTb microarray and RNASeq datasets on NCBI's Gene Expression Omnibus. We analyzed each microarray (using Limma) and RNASeq (using bwa, samtools, featurecounts, and DESeq2) dataset with false-discovery rate of 0.1 to generate transcription profiles (TPs) of DEGs. We ranked DEGs using log-2-fold ratios and identified the top 10 DEGs seen in at least half the TPs for a specific antibiotic generate a 10-gene signatures for each antibiotic. We assessed the performance of gene signatures against antibiotic and control TPs. Results: We identified 29 datasets and generated 196 TPs from 582 isolates of which 158 TPs (101 antibiotic and 57 controls) were analyzed further. After excluding TPs from antibiotics with concentration < 0.5 WHO critical concentration, there were 101 antibiotic TPs (49 microarray and 52 RNASeq) across 18 antibiotics with a median of 5.6 TPs per antibiotic, ranging from 17 TPs for isoniazid to 1 for kanamycin, amikacin, levofloxacin, cycloserine and rifapentine. The majority (77%) of TPs were from strains cultured between 3 and 24 hours, with significant heterogeneity in the concentration of antibiotic used. Antibiotic gene signatures showed similar genes upregulated for isoniazid, ethionamide, and ethambutol (*kasA, kasB, fbpC, iniA,* and *iniB*). Transcriptional regulators such as *whiB7, lexA,* Rv3074 were noted to be upregulated for aminoglycosides, fluoroquinolones, and newer drugs like linezolid, pretominid and delaminid. Conclusion: A transcriptomic-based assay may allow the rapid detection of drug-susceptible *Mycobacterium tuberculosis*.

---PAGEBREAK---

Control Number: 2022-A-57-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: A Unified molecular epidemiology approach for the utilization of SARS-CoV-2 whole genome sequencing, a state public health perspective

Author Block: **X. Wang**, J. Garfin, N. Lehnertz, S. Seys, S. Meyer, K. Como-Sabetti, S. Vetter, R. Lynfield; Minnesota Department of Health, Saint Paul, MN.

Abstract Body: The COVID-19 pandemic marks the first time that whole genome sequencing (WGS) has been utilized as a real-time analysis tool to guide public health pandemic response and policy. The Minnesota Department of Health (MDH) has continuously sequenced severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) since its introduction into Minnesota. Initially, molecular epidemiology-based approaches were used to study the introduction and baseline genomic diversity of SARS-CoV-2 importations into Minnesota, with strategies and techniques shifting with the changing priorities of the ongoing pandemic. Since March 2020, the MDH Public Health Laboratory has sequenced more than 35,000 SARS-CoV-2 genomes, established the Minnesota Surveillance of SARS-CoV-2 (MN-SOS) consortium to coordinate sequencing efforts and logistics with other sequencing and testing sites in Minnesota, collaborated closely with MDH infectious disease epidemiologists to investigate SARS-CoV-2 transmission dynamics in congregate settings, tracked and identified emerging variants including tracking the spread of the Alpha variant in a community, the nation's first Gamma variant, the first domestically-transmitted Omicron variant, and reported SARS-CoV-2 lineage prevalence in clinical samples to inform public health policy. SARS-CoV-2 sequencing has proven to be a vital tool in the larger COVID-19 pandemic response; particularly in characterizing disease severity, vaccine breakthrough infections, and characterizing viruses introduced into Minnesota through foreign travel or specific events that may have led to further spread of disease. From a state public health perspective, a robust NGS infrastructure and unified molecular epidemiology approach are key components for current and future pandemic responses.

---PAGEBREAK---

Control Number: 2022-A-64-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: What Organisms Are Found in Culture-negative Periprosthetic Joint Infections? Insights with Next Generation Sequencing

Author Block: **C. M. Baker**, K. Goswami, S. Tarabichi, E. Chisari, J. Parvizi; Rothman Orthopaedics, Philadelphia, PA.

Abstract Body: **BACKGROUND**: There is a dire need to address the issue of culture negative periprosthetic joint infection (PJI). Without knowledge about the infective organism(s) in these cases, the antimicrobial treatment is usually done blindly leading to issues with antimicrobial stewardship, as broad spectrum and usually multiple antimicrobials are administered, as well the risk for failure as rare organisms may not be covered. This study sought to describe the organism profile in culture-negative PJI, using next-generation sequencing (NGS) as the molecular census instrument.

METHODS: In this prospective study samples were collected from consecutive patients

undergoing revision total hip and knee arthroplasties at a single academic center. Patients meeting International Consensus Meeting (ICM) criteria for PJI were included. Intraoperative samples (synovial fluid, deep tissue and swabs) were obtained and sent for both routine culture and NGS. Patients for whom NGS was positive and standard culture was negative were included in our analysis. Organisms above an abundance threshold for 5% were reported as “common”.

RESULTS: The overall cohort included 100 patients who met the ICM criteria for PJI and were culture-negative. A pathogen could be identified by NGS in 51 (51.0%) of these culture-negative patients. Twenty species were identified as common based on a study-wide incidence threshold of 5%. NGS revealed a polymicrobial infection in 71% (n=36/100) of culture-negative PJI cases, with a mean of 4.6 ± 4.8 and median of 3 different species per patient. *Staphylococcus epidermidis*, *Escherichia coli*, and *Staphylococcus aureus* were the most frequently detected
organism. *Staphylococcus*, *Acinetobacter*, *Streptococcus*, *Escherichia*,
and *Cutibacterium* ranked highest in terms of study-wide mean incidence at the genus level.

CONCLUSIONS: NGS provides a more comprehensive picture of the microbial profile of infection that is often missed by traditional culture. Examining the profile of PJI in a prospective cohort using NGS, this study describes the most prevalent opportunistically pathogenic organisms identified in culture-negative PJIs, furthermore, that the majority of PJIs were polymicrobial.

FIGURE 1: Frequency distributions for the number of species detected per case. Distribution only for species characterized as common on the basis of having a study-wide incidence of at least 5%.

FIGURE 2: Stacked bar plot of the number of cases detected (incidence) by microbial species and genera. Species are categorized by their respective genus for each bar. Genera are organized in the panels based on incidence rates. Only species that were identified as common (having a minimum study-wide incidence of at least 5%) were included for plotting.

---PAGEBREAK---

Control Number: 2022-A-66-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: In-scilico discovery of uncategorized anti-microbial resistance genes

Author Block: **M. Hallgren**;
Technical University of Denmark, National Food Institute, Kongens Lyngby, DENMARK.

Abstract Body: Modern day methods for identification of anti-microbial resistance (AMR) genes in DNA sequencing reads primarily relies on one of two approaches: Genotype screening using alignment-based methods or pretrained Hidden Markov Models. Alignment-based approaches generally suffer from a high number of false negatives due to underrepresentation of genes in curated databases. Contrary to this, probabilistic approaches usually tend to heavily overestimate and therefore suffer from a high number of false positives. This often leads to inaccuracies in results because identified variants are incorrectly reported.To improve alignment-based efforts, we investigate publicly available data to search for uncategorized resistance genes to increase the size of our genomic databases. This is done using an extremely loosely tuned mapping pipeline using KMA

mapping which is designed to output large amounts of assembled consensus sequences of observed gene variants. We then investigate each unique gene variant which has multiple occurrences across different samples to map variable genomic positions and check if published literature for the corresponding isolate have observed phenotypical resistance behavior. Only high-quality trimmed Illumina paired-end samples were used, because highly precise local assemblies of unaligned regions were desired. In our preliminary findings, which includes screening of 13.000 clinical isolate sequence samples, we have identified 48 novel gene variants belonging to the three AMR classes of beta-lactams, aminoglycosides and fosfomycins.In future work (which is currently on-going at this time of writing), we plan to screen a much larger number of public sequencing samples which have association with AMR genes. Hopefully, this will contribute to an expansion of known AMR gene variants, which in turn will lead to more precise genotyping when using alignment-based screening tools.

---PAGEBREAK---

Control Number: 2022-A-72-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: (this abstract has been withdrawn) Genetic Diversity of HIV-1 Subtypes in African Populations A Case Study of Sub-Saharan Africa

Author Block: **B. M. Karumbo**;
Pwani University, Kilifi, KENYA.

Abstract Body: Effective control of the HIV-1 infection pandemic remains elusive despite advances inantiretroviral therapy that have revolutionized HIV-1 disease management. Increased HIV-1infection rates and genetic diversity in Sub-Saharan African countries pose a challenge to HIV-1clinical management. This study evaluated the HIV-1 genetic diversity and its implications forHIV-1 disease spread which have an impact on the effectiveness of therapies. Whole-genomesequences of HIV-1 were retrieved from Genbank corresponding to 10 African countries withhigh HIV-1 prevalence. Sequence alignment that included reference sequences retrieved from theLos Alamos database was conducted. The alignment file was viewed and curated in Aliview.Molecular genetic diversity analysis was inferred using the maximum-likelihood methodimplemented in iqtree. The clustering pattern of the HIV-1 sequences from the Sub-SaharanAfrican countries under study showed homogeneous and heterogeneous clustering. All theZambian sequences clustered with HIV-1 subtype-C reference sequence suggesting the localdistribution or adaptation of subtype-C. Sequences from nine countries showed heterogeneousclustering along with different subtypes suggesting genetic exchange and gene flow between thecountries that can be attributed to the cross-border movement. Sequences from Kenya andNigeria clustered with almost all the HIV-1 subtypes suggesting high HIV-1 genetic diversitycompared to other countries. These results can be related to the presence of subtype-specificpolymorphisms and interaction during border movements.

---PAGEBREAK---

Control Number: 2022-A-76-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Molecular Epidemiology of the First NDM-1-Producing *Klebsiella pneumoniae* Outbreak in Minnesota

Author Block: **S. B. Namugenyi**, J. L. Dale, J. Garfin, M. Plumb, B. VonBank, S. O'Malley, R. Lynfield, P. Snippes Vagnone, X. Wang;
Minnesota Department of Health, Saint Paul, MN.

Abstract Body: New Delhi-metallo-β-lactamase (NDM) is an enzyme, found in gram negative bacteria. NDM is capable of hydrolyzing penicillin, cephalosporin, monobactam, and carbapenem classes of antibiotics, which contributes to bacterial resistance to these antimicrobials. Though uncommon in the United States, NDM-producing Enterobacterales are more common in healthcare settings in other countries and are an emerging public health threat. In Minnesota, fewer than 10 cases were reported annually from 2012-2018 most of whom received recent healthcare abroad. Between December 2018 and May 2019, the first outbreak of NDM-producing *Klebsiella pneumoniae* occurred in Minnesota among eleven cases with an epidemiologic link to one long-term care facility, and without patient history of international travel. *Klebsiella pneumoniae* isolates were obtained from 9 out of 11 cases with one case also harboring an NDM-producing *Escherichia coli*. Whole genome single nucleotide polymorphisms (wgSNPs) analysis demonstrated relatedness between all nine *K. pneumoniae* isolates with SNP differences ranging from 1-18. WGS analysis also determined that all *K. pneumoniae* isolates harbored the $bla_{\text{NDM-1}}$ gene, which encodes for NDM. Detailed epidemiological and WGS data point to clonal spread of NDM-producing *K. pneumoniae* in this outbreak. However, since NDM is typically found on a plasmid, there is potential for horizontal gene transfer of antibiotic resistance between different bacterial genera. We did find two different genera of NDM-1-producing organisms (*K. pneumoniae* and *E. coli*). Long read (Oxford Nanopore) sequencing was performed on the isolates of the two organisms. Further analysis was done to pinpoint the location of $bla_{\text{NDM-1}}$ gene and genetic relatedness of those plasmids in order to understand antimicrobial resistant transmission mechanisms in a healthcare facility outbreak setting. WGS, combined with high fidelity short reads high throughput sequencing and long reads sequencing, can inform antimicrobial resistant bacteria outbreak response and facilitate infection control interventions.

---PAGEBREAK---

Control Number: 2022-A-78-NGS

Topic 1: Bridging Silos: Exploring mechanisms for collecting and sharing microbial genomic data to foster interoperability

Topic 2:

Publishing Title: Identification of candidate antimicrobial resistance determinants in Enterococcus using machine learning

Author Block: **J. Kim**[1], T. A. McAllister[2], R. G. Beiko[1];
[1]Dalhousie University, Halifax, NS, CANADA, [2]Agriculture and Agri-Food Canada, Lethbridge, AB, CANADA.

Abstract Body: Machine learning (ML) is increasingly being applied in antimicrobial resistance (AMR) genomics investigations to better predict the resistance potential of bacteria. Many investigations have published ML models for high-priority pathogens like *Salmonella* and *Mycobacterium tuberculosis*. However, the question of which features

the ML models are using to predict is not as comprehensively explored. We examined different candidate feature sets and feature-selection approaches to elucidate potential AMR determinants of *Enterococcus faecium* and *Enterococcus faecalis*.
Vancomycin-resistant strains of *E. faecium* have a collection of well-known resistance-conferring genes, such as the Van operon. Experimentally validated AMR genes were identified through the Resistance Gene Identifier software of the Comprehensive Antibiotic Resistance Database. A gradient-boosting model with AMR-specific feature sets resulted in 98.0% accuracy in a set of 308 genomes, and the top features used by the model included the well-known *vanHAXSR* genes of the VanA operon. Six genomes were misclassified, but two of these were found to have incorrect resistance labels, while the other four lacked one or more of the key *van* genes. The VanA operon typically localizes to plasmids; when we used plasmids predicted using the MOB-suite package as features, the classification accuracy remained high (93.8%), suggesting that mobile genetic elements can stand in for specific genes.
We then expanded our modeling approach to include all identified in the pangenome of 600+ enterococci species. Genes were annotated with Prokka and compiled with Roary, and the accuracy of vancomycin-resistance prediction was similar to the RGI-based features, at 97.7%. AMR genes such as *vanHAX* were also ranked as the most important features. Other sequences previously not associated with vancomycin resistance were also selected as relevant features, including putative bleomycin- and fosfomycin-resistance genes.
Predicted resistance to other antimicrobials had lower associated accuracy. *E. faecalis* erythromycin (87.8%) and doxycycline resistance (90.8%) prediction models with pangenome features did not provide such a straightforward outcome as vancomycin resistance in *E. faecium*. The top selected features did not overlap with the previously known AMR genes for erythromycin & doxycycline resistance. Our results revealed that several top features were undefined hypothetical proteins that may be worth further computational and experimental follow-ups. We propose that ML can help fast-track the confirmation of new AMR determinants. Our workflow allows the development of precise hypotheses that can be prioritized for laboratory testing.

---PAGEBREAK---

Control Number: 2022-A-84-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Clinical Utility of Whole Genome Sequencing of AMR Pathogens and its benefits to Infection Prevention and Control and Antibiotic Stewardship

Author Block: **A. C. Materna**, S. Beisken, J. Weinberger, P. Májek, L. Lueftinger, I. Ferreira, T. Weinmeier; Ares Genetics GmbH, WIEN, AUSTRIA.

Abstract Body: Over five years, the estimated number of annual deaths that might have been prevented by effective antimicrobial therapy has grown by 81% to approximately 1.27M deaths [1,2]. Healthcare-associated infections (HAIs) with antimicrobial resistant (AMR) pathogens are widespread and are a common cause of complications among hospitalised patients. AMR can negatively impact patient outcomes, and the economic burden associated with HAI infections is enormous, due to longer hospital stays, higher treatment costs, and a reduced availability of intensive care unit beds [3]. Whole Genome Sequencing (WGS) coupled with bioinformatic solutions can comprehensively characterize AMR pathogens and is

increasingly adopted for epidemiological analysis. In healthcare settings, it can identify AMR outbreaks, distinguish between hospital-acquired and community-acquired infections, and thus help to curb or even prevent transmissions. Broad-scale adoption of WGS in healthcare settings has not yet occurred.This presentation explores the utility of WGS of AMR pathogens in healthcare settings with a focus on the benefits to patients and healthcare economics. We survey advances in WGS and bioinformatics solutions that facilitate a broad-scale adoption. Finally, we introduce a bioinformatic approach for antibiotic susceptibility testing from WGS data (WGS-AST) for the augmentation of therapeutic guidance and improve antibiotic stewardship.

References:
[1] Collaborators, Antimicrobial Resistance *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* (2022) doi:10.1016/s0140-6736(21)02724-0.
[2] O'Neill, J., Antimicrobial resistance: tackling a crisis for the health and wealth of nations. Wellcome Collection. (2014) Attribution 4.0 International (CC BY 4.0)
[3] 1.Gordon, L. G. *et al.* Budget impact analysis of routinely using whole-genomic sequencing of six multidrug-resistant bacterial pathogens in Queensland, Australia. *Bmj Open* 11, e041968 (2021).

---PAGEBREAK---

Control Number: 2022-A-88-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Lack of Host Sex Effect on*Brugia pahangi*Gene Expression

Author Block: **C. Holt**[1], S. Baumberger[2], R. E. Bromley[1], B. C. Sparklin[1], J. Mattick[1], S. Ott[1], L. Sadzewicz[1], L. J. Tallon[1], J. M. Foster[3], M. L. Michalski[2], J. C. Dunning Hotopp[1];
[1]University of Maryland, Baltimore, Baltimore, MD, [2]University of Wisconsin, Oshkosh, Oshkosh, WI, [3]New England Biolabs, Ipswich, MA.

Abstract Body: **Background:** Lymphatic filariasis is a mosquito-borne neglected tropical disease that occurs in regions of South America, Africa, and Southeast Asia that currently afflicts 67 million-120 million people globally. Lymphatic filariasis is caused by three nematodes: *Wuchereria bancrofti*, *Brugia malayi*, and *Brugia timori*. The disproportionate infection of males with lymphatic filariasis has been long observed in human populations as well as in the laboratory models. The purpose of this study is to determine if the sex of the vertebrate host affects the gene expression of the nematode. **Methods:** We used RNA-Sequencing of *Brugia pahangi* to study the effect of host-sex on nematode gene expression by comparing worms from male gerbils to worms from female gerbils. *Brugia pahangi* is closely related to *Brugia malayi* and has been used to study lymphatic filariasis and the effect of host-sex on lymphatic filariasis in animal models. After injecting *B. pahangi* larvae into the intraperitoneal cavity of 7 male and 7 female gerbils, and injecting *B. pahangi* larvae into 22 male and 32 female gerbils subcutaneously, 3 gerbils from each category were selected with adult male nematodes, adult female nematodes, and microfilariae isolated, using male/female litter mate pairs when possible. **Results:** As expected and consistent with previous results, life stage-specific differentially expressed genes could be identified. However, samples did not cluster by host-sex and differentially expressed genes for host-sex

were not identified. This suggests that at these time points, there is no detectable differential expression in worms from male gerbils relative to female gerbils **Conclusions:** Despite studies showing a preferential infection in males in both human populations and in the laboratory as well as differing immune response in male gerbils that may be associated with host hormones, host sex did not influence the gene expression of *B. pahangi* adult male nematodes and adult female nematodes. A statistically significant difference between gerbil models in terms of numbers of worms recovered was not observed.

---PAGEBREAK---

Control Number: 2022-A-99-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Ribavirin therapeutic as a mutagen of Crimean-Congo haemorrhagic fever virus (CCHFV) genome in clinical samples

Author Block: **N. Wand**[1], J. D'Addiego[1], N. Elaldi[2], K. Osman[1], B. Koksal Bagci[3], E. Kennedy[1], A. Nur Pektas[4], E. Hart[1], G. Slack[1], R. Hewson[1];
[1]UK Health Security Agency, Porton, Salisbury, UNITED KINGDOM, [2]Cumhuriyet University Faculty of Medicine, Sivas, TURKEY, [3]Cumhuriyet University, Faculty of Health Sciences, Sivas, TURKEY, [4]Cumhuriyet University, Advanced Technology Application and Research Centre, Sivas, TURKEY.

Abstract Body: Crimean-Congo haemorrhagic fever virus (CCHFV) is a negative-sense single-stranded segmented RNA virus with a genome composed of L, M and S segments. RNA viruses have been observed to exploit high rates of genetic variation to secure their survival and continued transmission. The high levels of CCHFV mutagenesis fuels the rise and fall of viral quasispecies within its host, creating an ever-changing complex population of minor variants. This existence on the edge of lethal level mutation rates of CCHFV can be used to develop new therapeutic strategies that may tip the balance of the viral population causing reduction of fitness and decreasing viremia. The only current WHO-recommended therapeutic for treating CCHFV infection is ribavirin, and it has been proposed to act as a mutagen. However, its efficacy against CCHF is controversial and definitive data underpinned by case-controlled trials have not been conducted. In this study, we used whole genome sequencing to investigate whether ribavirin could increase the rate at which CCHFV variants arose in the viral population in clinical samples. Patients positively diagnosed with CCHF were either administered ribavirin treatment or received supportive care only, based on the decision of attending physicians. The study was a retrospective analysis and patient samples were collected over the course of normal procedures at Cumhuriyet University Hospital. Serum samples were collected on the first day of hospitalisation, the following day when treatment was initiated and two days post treatment initiation. CCHFV genome sequences in all clinical samples were processed using Sequence-Independent Single-Primer Amplification (SISPA) method and sequenced on Illumina MiSeq instrument. Both high- (≥10%) and low-frequency (2-9%) variants rates were determined and compared between ribavirin-treated and control group patients.We observed a limited mutagenic effect of ribavirin on the CCHFV genome from ribavirin-treated patients with a significant increase only in low-frequency C-to-T transitions, compared to the control group-derived CCHFV populations. This mutagenic effect, however, was not observed to affect viremia,

since the viral load decrease in ribavirin-treated patients was not significantly different from that in patients who received supportive care only.

---PAGEBREAK---

Control Number: 2022-A-101-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Genotyping of isolates from children with tuberculosis in a tertiary care hospital in Mexico

Author Block: **D. Manzano**, J. A. Ferrer, I. V. Diaz, A. G. Bravo, P. Saltigeral, M. Macias;
INSTITUTO NACIONAL DE PEDIATRIA, MEXICO, MEXICO.

Abstract Body:
Introduction: The World Health Organization (WHO) published this year that children and adolescents (under 15 years of age) account for about 11% of all people with tuberculosis (TB) in the world. This means that about 1.1 million children fall ill with TB each year, almost half of them under the age of five. The main problem in diagnosing young children is due to the paucibacillary nature of TB at this age. As a result, smear microscopy is not useful and cultures take 4-6 weeks. Today, a wide range of molecular tests have been developed for TB detection, such as next-generation sequencing (NGS), which is available to clinical laboratories, provides rapid diagnosis and has increased its use as a diagnostic test for TB. **Objective:** The use of next-generation sequencing (NGS) molecular testing for rapid diagnosis of isolates from children with tuberculosis in a tertiary care children's hospital for genus- and species-level detection, as well as sensitivity to first- and second-line antibiotics for tuberculosis.
Methods: All strains of Mycobacterium tuberculosis complex (MTBC) were isolated from patients under 16 years of age at the National Pediatrics Institute, Mexico. Matrix-assisted laser desorption/ionization-assisted laser desorption/ionization mass spectrometry (MALDI-TOF MS) (Bruker) was used for their identification. For the differentiation between MTBC species and the detection of resistance to first- and second-line antibiotics in children with TB, the Deeplex®-MycTB assay was used following the manufacturer's instructions.
Results: We analyzed from December 2021 to June 2022 five children with a mean age of 9.4 years, corresponding 3/5 (60%) to males and 2/5 (40%) to females. The diagnosis and identification of the MTBC species detected in each patient is: Patient 1 had a kidney transplant and extrapulmonary tuberculosis (lymph node and intestinal) detecting *Mycobacterium bovis*. Patient 2 had hydrocephalus and venticulitis detecting *Mycobacterium bovis BCG*. Patient 3 had miliary tuberculosis in which *Mycobacterium tuberculosis* was detected. Patient 4 had miliary tuberculosis in which *Mycobacterium bovis* was detected and patient 5 had bone and vertebral tuberculosis with *Mycobacterium tuberculosis complex*. It is important to mention that all patients were treated with isoniazid, pyrazinamide, ethambutol, rifampicin and, in the cases of *M. bovis* and *M. bovis BCG*, levofloxacin was added to the treatment, presenting improvement of the infection.
Conclusion: This year we added NGS molecular tests to the microbiological Workflow of tuberculosis, which has made it possible to determine whether infections by the *Mycobacterium tuberculosis complex* are *M. bovis* or *M. bovis BCG*, since they are intrinsically resistant to pyrazinamide, allowing the addition of an appropriate antituberculosis treatment.

---PAGEBREAK---

Control Number: 2022-A-103-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: β-Lactamase Alleles in *Acinetobacter* and *Pseudomonas*: How do they stack up?

Author Block: **A. R. Mack**, A. M. Hujer, R. A. Bonomo;
Case Western Reserve University & Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, OH.

Abstract Body:

Background

Pseudomonas aeruginosa and *Acinetobacter baumannii* are important, non-fermenting, Gram-negative pathogens responsible for a wide variety of nosocomial and community acquired infections. Carbapenem resistant *A. baumannii* and multidrug resistant *P. aeruginosa* are recognized as particularly high-risk organisms. Both encode a chromosomal class C β-lactamase (*bla*$_{ADC}$ or *bla*$_{PDC}$) and class D β-lactamase (*bla*$_{OXA}$) and can acquire an array of additional β-lactamases.

Surveillance studies reveal the diversity and distribution of β-lactamase alleles but can be difficult and expensive to conduct. Herein, we apply a novel approach, using publicly available data derived primarily from whole genome sequences to explore these questions and to compare the alleles present in these two organisms.

Methods

Details of β-lactamase alleles present in 20,341 *P. aeruginosa* and 17,164 *A. baumannii* isolates were obtained from the National Center for Biotechnology Information's Pathogen Detection Project databases. Analysis of frequency and distribution of alleles was conducted in RStudio.

Results

The *A. baumannii* dataset contains 151 of 241 (62.7%) distinct *bla*$_{ADC}$ alleles and 189 of 1067 (17.7%) distinct *bla*$_{OXA}$ alleles and the *P. aeruginosa* dataset contains 228 of 498 (40.0%) distinct *bla*$_{PDC}$ alleles and 90 of 1067 (8.4%) distinct *bla*$_{OXA}$ alleles. Additionally, 662 novel *bla*$_{ADC}$ alleles, 355 novel *bla*$_{PDC}$ alleles, and 530 *bla*$_{OXA}$ alleles were found. The ten most frequent *bla*$_{ADC}$ and *bla*$_{PDC}$ alleles represent 12,000 (77.3%) and 15,237 (77.6%) alleles respectively, with the two most frequent ADCs accounting for 8,801 (56.7%) of total alleles. The dominant *bla*$_{OXA}$ family is the chromosomal allele in both datasets (*bla*$_{OXA\text{-}51\text{-}family}$ in *A. baumannii* and *bla*$_{OXA\text{-}50\text{-}family}$ in *P. aeruginosa*), representing 54.4% and 89.4% of total *bla*$_{OXA}$ alleles, respectively. Interestingly, none of the acquired *bla*$_{OXA}$ families occurring in 1% or more of isolates overlap between the organisms.

Conclusions

In both organisms, very few distinct alleles represent the majority of alleles present, while the lack of overlap in acquired alleles represents just how unique the closely related bacteria are in terms of their β-lactamase profiles. The greater number of distinct novel alleles versus distinct assigned alleles demonstrates just how understudied β-lactamase diversity remains even in these well-known organisms.

Understanding the diversity and distribution of these alleles is critical on multiple levels: understanding the epidemiology and spread of resistant bacteria around the world, prioritizing alleles in basic research, and choosing target enzymes for the development of

new antibiotics and inhibitors to overcome resistance. We propose using information in NCBI Pathogen Detection Project databases as a novel and useful approach to help answer these questions, and one which has advantages over standard surveillance studies in terms of scalability, logistics, and cost.

---PAGEBREAK---

Control Number: 2022-A-104-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Wastewater Based Genomic Surveillance of SARS CoV 2 in Karachi City, Pakistan

Author Block: **W. Khan**, S. Kanwar, F. Kabir, F. Aziz, N. Ansari, U. Mehmood, F. Jehan, M. I. Nisar;
The Aga Khan University, Karachi, PAKISTAN.

Abstract Body: Wastewater-based surveillance has been used increasingly in many different parts of the world since the early phase of COVID-19 pandemic for monitoring the prevalence and temporal trends of SARS-CoV-2 strains in communities. Detection of viral RNA particles in wastewater sample can signal the initiation of an outbreak within a defined catchment area earlier than clinical surveillance. Here, we describe the feasibility of using a sewage network to describe the trend of SARS-CoV-2 in limited resource settings, such as, the biggest metropolitan city of Pakistan, Karachi with undefined infrastructure of sewage connections. We have performed genomic analysis on sewage samples collected from 4 sites to describe the abundance of circulating COVID-19 variants. Sample collection and filtration was done on-site using all aseptic precautions. SARS-CoV-2 was detected trough RT-qPCR for either of N1, N2, or E gene. Library preparation was done using using ARTIC-NEB V3 and V4 protocols and sequencing was performed using Illumina iSeq100 and MiSeq platforms. Percentage abundance of SARS-CoV-2 lineages was identified using KALLISTO pipeline which was then compared with clinical samples obtained from Nextstrain Pakistan build. A consensus genome of wastewater sequences was generated using web-based CZ ID consensus genome pipeline. Quality of genomes was assessed using NextClade and passed sequences were submitted to Global Initiative on Sharing All Influenza Data (GISAID) and NCBI's GenBank. A total of 123 COVID-19 positive samples were sequenced collected from June 2021 to January 2022, 120 samples passed all quality control, filter checks and abundance analysis. Based on sample collection month, 8 reference sets were built to acquire actual abundance of SARS-CoV-2 variants circulating at that time period. Consensus genome of 115 samples was successfully generated. In current study, the Omicron variant was recovered 33 days earlier than the first clinically reported case. However, sequencing of wastewater samples differs from clinical surveillance because of 1) low concentration of viral RNA, 2) presence of multiple variants in the same sample, and 3) presence of contaminants. It is still unknown how well this method alone can identify SARS-CoV-2 genetic diversity. Nevertheless, our findings suggest that wastewater-based surveillance can signal an impending outbreak in areas with low disease prevalence, low levels of testing, or delays in reporting. To the best of our knowledge this is the first study from Pakistan to demonstrate this and builds on existing evidence from high income countries on the use of sewage surveillance as a disease early warning system for SARS-CoV-2.

---PAGEBREAK---

Control Number: 2022-A-107-NGS

Topic 1: Bridging Silos: Exploring mechanisms for collecting and sharing microbial genomic data to foster interoperability

Topic 2:

Publishing Title: Characterizing Consequences of X:Autosome Fusions in Onchocercidae Through Secondary Analysis of RNA-seq Data

Author Block: **K. A. Hackbarth**[1], E. S. Haag[2], J. C. Dunning Hotopp[3];
[1]University of Maryland, Maryland, MD, [2]University of Maryland, College Park, MD, [3]University of Maryland School of Medicine, Baltimore, MD.

Abstract Body: Through secondary analysis of publicly available next-generation sequencing data, research questions separate from the initial aims of the data generation can be addressed. Combining datasets extends the potential for new discoveries. *Brugia malayi* and *Onchocerca volvulus* are parasitic nematodes consequential for public health. In each species, a different autosome fused with the X chromosome to form a neo-X. As a result, their ancestrally XX/XO systems became XX/XY, where the Y represents a degenerated version of the unattached autosomal homolog chromosome. The new association of a large region of the genome with the sex chromosomes introduces evolutionary pressures due to a change in gene dosage. These independent fusion events offer an opportunity to investigate and compare the consequences of X:autosome fusions for these parasites' genomes. Here, we combine and reanalyze RNA-seq data from *B. malayi* and *O. volvulus* to show that the fused regions exhibit a lesser degree of dosage compensation and show different patterns of gene conservation when compared to the ancestral X regions.

---PAGEBREAK---

Control Number: 2022-A-109-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Molecular Surveillance of Monkeypox in the District of Columbia

Author Block: S. V. Nguyen[1], E. L. Vaughn[1], **J. H. Doss**[1], M. Adjei[1], C. Williams[1], M. Mann[1], A. Cabello[1], S. Byfield[1], L. Sealey[1], S. Merid[1], R. Blackwell[1], E. Zelaya[1], M. Doucette[2], L. Gagne[2], D. Payne[1], J. A. Kiehlbauch[1], J. R. Hauser[1];
[1]DC Department of Forensic Sciences Public Health Laboratory, Washington, DC, [2]Massachusetts Department of Public Health, Boston, MA.

Abstract Body: The 2022 global outbreak of West African clade monkeypox is the largest monkeypox outbreak to occur in non-endemic regions. Monkeypox is an emerging zoonotic virus in the *Orthopoxvirus* genus, and it has epidemic potential due to the vacant ecological niche left by the eradicated *Orthopoxvirus* smallpox. As of July 18, 2022, the District of Columbia has reported the most monkeypox cases per capita in the United States, with at least 122 confirmed positive patients. Ongoing challenges for molecular surveillance of monkeypox by public health laboratories include nonspecific real-time PCR assays that are not able to distinguish between monkeypox and other non-variola *Orthopoxviruses,* difficulties in sequencing and bioinformatic analyses, and a lack of biosafety level 3 (BSL-3) laboratory infrastructure for sequencing. Methods undertaken to ensure extracted nucleic acids are

rendered non-infectious may result in lower quality DNA samples and may not be conducive for next-generation sequencing. Early efforts at monkeypox sequencing at the District of Columbia Public Health Laboratory (DCPHL) include a metagenomic approach utilizing Illumina and Oxford Nanopore technologies. Here we provide an overview of methods undertaken at the DCPHL for next-generation sequencing and genomic analysis for epidemiological surveillance of monkeypox in the District.

---PAGEBREAK---

Control Number: 2022-A-130-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Genome Wide Association Studies (GWAS) to predict Cryptic Carbapenem Resistance Mechanisms in *Klebsiella pneumoniae* Detected in Italy

Author Block: A. Cornacchia[1], **A. Chiaverini**[1], A. Janowicz[1], G. Centorotola[1], M. Saletti[1], S. Chiatamone Ranieri[2], A. Di Pasquale[1], C. Cammà[1], F. Pomilio[1];
[1]Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale", Teramo, ITALY, [2]Operative Unit of Clinical Pathology and Microbiology, ASL of Teramo, Teramo, ITALY.

Abstract Body: *Klebsiella pneumoniae* (*Kp*) is one of the most common causes of hospital-acquired infections and it is recognized by WHO as a “critical” priority pathogen for the development of novel antimicrobial strategies. In particular, the rapid emergence of carbapenem-resistant *Kp* in diverse environmental niches, even outside of the clinical setting, poses a challenge for detection and the real-time monitoring of novel AMR trends using molecular and WGS-based methods. The aim of our study was therefore, to identify emerging resistance determinants responsible for the phenotypic carbapenem resistance observed in strains circulating in Italy by performing a genome wide association study (GWAS).In this study, we collected 320 *Kp* strains from foods, environmental samples, animals and clinical cases isolated between 2018-2020 in Italy. The AMR profile of all isolates was assessed using both phenotypic and WGS-based methods. High quality draft assemblies were obtained using Illumina platform. Carbapenemase genes were detected querying the Pathogenwatch publicly available database. We identified 14 strains resistant to carbapenems, which did not carry any known genetic determinants explaining their AMR phenotype. The genome annotation of all *Kp* genomes was carried out producing GFF-files, including sequences and annotations, which were used to extract the pangenome.Then, the GWAS was used to identify genes significantly associated with the inconsistent carbapenem phenotypic resistance. In order to control for the confounding population stratification, a pairwise comparison based on linear mixed model was performed using a reference SNP-based phylogenetic tree. The preliminary data analysis resulted in identification of potential genetic markers for carbapenems, including a genomic cassette of nine genes involved in peptidoglycan synthesis and MOP efflux pump. These preliminary results confirmed the potential of GWAS to identify genetic variants that could be associated with antibiotic resistance traits in *Kp*. In light of the emergent diffusion of carbapenem-resistant *Kp* in Italy, there is an urgent need to understand cryptic and complex resistance mechanisms associated and subsequently to improve antimicrobial therapy. However, precautions need to be taken to prevent the detection of spurious associations using the GWAS approach, but

the implementation of such advanced methods with increased AMR surveillance will potentially improve *Kp* infection treatment and patient outcomes.

---PAGEBREAK---

Control Number: 2022-A-132-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Whole Genome Sequencing Study on *Listeria monocytogenes* Clinical Strains: 9 Years of Surveillance in Umbria and Marche Regions (Italy)

Author Block: F. Guidi[1], **M. Torresi**[1], G. Centorotola[1], A. Chiaverini[1], E. Rocchegiani[2], F. Pomilio[1], G. Blasi[2]; [1]Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "G. Caporale", Teramo, ITALY, [2]Istituto Zooprofilattico Sperimentale dell'Umbria e delle Marche "Togo Rosati", Perugia, ITALY.

Abstract Body: Whole Genome Sequencing (WGS) revolutionised the surveillance of listeriosis with unprecedented resolution in outbreak detection and improved support for epidemiological investigations. After a severe listeriosis outbreak occurred in Central Italy between 2015 and 2016, the isolate-based surveillance of *Listeria monocytogenes* (*Lm*) was intensified in the affected regions (Umbria and Marche), leading to a large amount of WGS data. Our aim was to conduct a retrospective study on 129 clinical *Lm* collected since 2014, in order to identify the main circulating clones, detect genomic clusters and define virulence and antimicrobial resistance (AMR) profiles. WGS data of *Lm* were analysed using Multi Locus Sequence Typing (MLST), core genome MLST (cgMLST) according to the Institute Pasteur scheme (chewBBACA's allele-calling algorithm) and SNPs analysis (CFSAN). All the genomes were screened for virulence and AMR genes on the BIGSdb platform. Strains grouped into 31 Sequence Types (STs) and 28 Clonal Complexes (CCs). CC7 (24.8%) and CC1 (20.2%) were the most frequent. CgMLST and SNPs analysis identified several clusters; the largest one belonged to CC7 and represented the outbreak occurred during 2015-2016 and re-emerged in 2018. Other clusters belonged to CC1, CC2, CC3, CC5, CC6, CC8, CC29, CC155 and CC224; most of them sporadically recurred over time (2-7 years). Recurrence of sporadic infections caused by a same cluster indicated its persistent circulation probably at food environment level. In the NCBI Pathogen Detection Database CC7 outbreak strains were identified as SNP cluster PDS000005985.16, while a CC4 genome responsible for a sporadic case associated with respiratory symptoms, grouped in the multi-country cluster PDS000043761.4 together with 3 clinical strains (UK, USA and Germany) and an environmental one (UK). All the CC1, CC3, CC4, CC6, CC77 and CC224 strains carried the additional Listeria Pathogenic Island 3 (LIPI-3); the CC4 ones also presented the LIPI-4 also carried by CC88, CC296 and CC517. The CC19 isolate carried the highest number of internalins. All the *Lm* presented the AMR genes *mdrM*, *mprF*, *norB*, *sul* and *fosX; tetM* and *tetS* were detected in the CC517 strain. None of these genes encoded for resistance against antibiotics used as treatment of choice for human listeriosis. Our study demonstrated how a well-structured isolate-based surveillance system supported by the WGS technology together with the global data-sharing represent a powerful tool for local and global surveillance of listeriosis.

---PAGEBREAK---

Control Number: 2022-A-135-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Genomic Epidemiology of SARS-CoV-2 Transmission among University Students in Western Pennsylvania

Author Block: **V. Rangachar Srinivasa**, M. P. Griffith, K. D. Waggle, L. Zhu, J. W. Marsh, D. Van Tyne, L. H. Harrison, E. Martin;
University of Pittsburgh, Pittsburgh, PA.

Abstract Body: **Background** SARS-CoV-2 control on college campuses is challenging given communal living and student social dynamics. The aim of this study was to understand and reconstruct patterns of SARS-CoV-2 spread on campus to develop targeted infection prevention strategies by using an integrated approach comprising both epidemiological and viral whole-genome sequencing (WGS) data from three different University of Pittsburgh campuses. **Methods** SARS-CoV-2 nasal swabs were collected from students for symptomatic testing and asymptomatic surveillance from August 2020 through April 2021. Samples collected from 308 students underwent WGS. Contact tracing information collected from students was used to identify transmission clusters. Genomes were considered related if they had ≤ 2 single nucleotide polymorphisms (SNPs) between them. **Results** We identified 31 Pangolin lineages of SARS-CoV2, the majority belonging to B.1.1.7 (Alpha) and B.1.2 lineages. Contact tracing identified 142 (46%) students clustering with each other; WGS identified 53 putative transmission clusters involving 216 (70%) students. WGS identified transmissions that were missed by contact tracing. However, 84 (27%) cases could not be traced by either WGS or contact tracing. Clusters were most frequently linked to students residing in the same dormitory, off-campus roommates, friends, or athletic activities. **Conclusion** The majority of SARS-CoV-2 positive samples clustered by WGS, indicating significant transmission across campuses. The combination of WGS and contact tracing maximized the identification of SARS-CoV-2 transmission on campus. This approach can be used as a strategy to mitigate transmission among students.

---PAGEBREAK---

Control Number: 2022-A-139-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Genomic Sequencing of *Bacillus cereus* Group Strains Isolated in Conjunction with 2020 Welder Anthrax Cases in the United States

Author Block: **L. M. Carroll**[1], C. K. Marston[2], C. B. Kolton[2], C. A. Gulvik[2], J. E. Gee[2], Z. P. Weiner[2], J. Kovac[3];
[1]EMBL, Heidelberg, GERMANY, [2]Centers for Disease Control and Prevention, Atlanta, GA, [3]Pennsylvania State University, University Park, PA.

Abstract Body: Anthrax-causing members of the *Bacillus cereus* group pose a serious risk to public health. While most anthrax-causing strains resemble *B. anthracis* phenotypically, rare cases of anthrax-like illness caused by strains resembling "*B. cereus*" have been reported. Here, whole-genome sequencing was used to characterize three *B. cereus* group strains isolated in conjunction with two 2020 welder anthrax cases in the United States. Two strains were isolated from male welders in their thirties with severe anthrax pneumonia (Patients F and G from Louisiana and Texas, respectively); one strain was isolated from soil collected from

Patient F's worksite in Louisiana. All three strains possessed anthrax toxin-encoding *cya, lef,* and *pagA*; however, all strains resembled "*B. cereus*" phenotypically per the Bacteriological Analytical Manual. Following pre-processing and assembly (e.g., via SKESA v2.4.0), annotation (via Prokka v1.13), and quality control (e.g., via FastQC v0.11.9, QUAST v5.0.2), the three genomes sequenced here were compared to all publicly available, high-quality *B. cereus* group genomes (*n* = 2,890 total genomes) using average nucleotide identity (ANI), multi-locus sequence typing, and pangenomic methods (via BTyper3 v3.2.0 and Panaroo v1.2.8). Notably, strains associated with each case effectively belonged to separate species at the conventional 95 ANI prokaryotic species threshold. Two sequence type 78 (ST78) strains affiliated with Patient F's case in Louisiana were nearly identical (1 core SNP difference), were most closely related to *B. tropicus*, and possessed genes encoding the Bps exopolysaccharide capsule, as well as enterotoxins hemolysin BL (Hbl) and cytotoxin K (CytK). Comparatively, a ST108 strain isolated from Patient G in Texas was most closely related to *B. anthracis*; however, like other anthrax-causing strains most closely related to *B. anthracis*, this strain did not possess Bps-, Hbl-, or CytK-encoding genes. Overall, all anthrax-causing *B. cereus* group strains that did not belong to the clonal *B. anthracis* lineage typically associated with anthrax illness belonged to one of two lineages: ST78 and the ST365 clonal complex (CC; *n* = 8 and 5 total genomes with anthrax toxin genes, respectively). Anthrax-causing ST78 and ST365 CC were both identified in the U.S. Gulf Coast and were each additionally detected in China and Côte d'Ivoire, respectively. Overall, results presented here provide insight into the evolution of anthrax-causing members of the *B. cereus* group.

---PAGEBREAK---

Control Number: 2022-A-148-NGS

Topic 1: Bridging Silos: Exploring mechanisms for collecting and sharing microbial genomic data to foster interoperability

Topic 2:

Publishing Title: Status of public *Candida auris* whole genome assemblies

Author Block: **A. R. Gener**[1], P. Hemarajata[2];
[1]Association of Public Health Laboratories; Los Angeles County Public Health Laboratories, Los Angeles, CA, [2]Los Angeles County Public Health Laboratories, Los Angeles, CA.

Abstract Body: **Background:** Since the discovery of *Candida auris* (*C. auris*) in 2009, the sequencing platforms and associated analytical tools available to ascertain *C. auris*' genomic content have improved. Reference genome choice has changed from the B8441 scaffold assembly, with the current reference genome B11205 lacking obvious features like a mitochondrial genome. The US Centers for Disease Control and Prevention is working with US state and local public health laboratories to standardize *C. auris* whole genome sequencing and analysis to track its emergence and spread. The current recommended approach relies on reference guided assembly, which can bias the assemblies made. To be able to compare existing *C. auris* datasets and to efficiently interrogate future samples, the status of available whole genome assemblies bore review. We hypothesized that genomes would vary in contiguity across periods of collection.
Methods: *C. auris* assemblies in NCBI were assessed prior to 15 July 2022. For mitogenome analysis, available read data were mapped to the B8441 mitochondrial reference genome (GenBank:MT849287.1) and consensuses were called.

Results: Genome assemblies available in NCBI (n=112) varied by assembly length, assembly level, annotation level, chromosome number, presence of chromosome ends, presence of mitochondrial chromosome. In addition, we noticed disparity between available reference genomes and the community resource candidagenome.org. Drug resistance gene lists were not standardized. Most complete assemblies lacked associated read data for validation or reanalysis. Clade II and V mitogenomes had structural variants which were previously incorrectly annotated as introns in COX1. The clade V assembly exhibited a possible replication signature at a feature formerly annotated as COB intron. Mitochondrial assemblies were made when reads were available.
Conclusions: Available public *C. auris* assemblies varied drastically. True variability between assemblies from different clades with different lengths co-occurred with artefactual variability. For example, "complete" genomes often lacked mitochondrial chromosomes. We made mitochondrial assemblies from available public data to complete assemblies. There is no consensus on reference genomes for different clades, which may impact downstream analyses of virulence and drug resistance. To date, there remains no explicit standards for *C. auris* genome assembly. It is important for stakeholders to communicate and harmonize efforts to maximize efficiency and utility of *C. auris* surveillance.

---PAGEBREAK---

Control Number: 2022-A-150-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Evaluating the detection of *Mycobacterium tuberculosis* heteroresistance by Whole Genome Sequencing

Author Block: S. Danchuk[1], **O. Solomon**[1], T. Kohl[2], S. Niemann[2], D. van Soolingen[3], J. van Ingen[4], M. Behr[1]; [1]McGill University, Montreal, QC, CANADA, [2]Molecular and Experimental Mycobacteriology, Research Center Borstel, Borstel, GERMANY, [3]National Institute for Public Health and Environment, Bilthoven, NETHERLANDS, [4]Radboud University Medical Centre, Nijmegen, NETHERLANDS.

Abstract Body: Background: The rise of drug-resistant tuberculosis (DR-TB) has necessitated the development of diagnostic tools to better inform treatment. *M. tuberculosis* (*M.tb*) is slow growing, acting as the rate limiting step for phenotypic drug-susceptibility testing (DST), the current gold standard. As such, TB diagnostics have turned to more rapid techniques, using molecular approaches to look for resistance-associated mutations. With these newer assays, patients simultaneously infected with resistant and susceptible strains of *M.tb* (heteroresistance = HR) present a unique challenge. Methods and Results: Previously our lab developed DR strains of *M. bovis* BCG as a tool for quality control (QC) in clinical and reference settings. To determine the ability of phenotypic and genotypic tests to detect HR, we combined DR- and WT-strains in varying proportions (50:50; 90:10; 99:1). Importantly, we sought to detect diagnostic outcomes if DR-strains made up the minority population of a sample. Phenotypic DST was conducted using the agar proportion method. Genotypic testing compared the GeneXpert MTB/Rif assay and whole-genome sequencing (WGS), done according to protocols established in participating labs. Phenotypic DST reliably detected HR at 10% and even at 1%. As has been previously reported, GeneXpert was unable to detect HR at 10% or even 50%; only at >60% Rifampin-resistance was resistance detected. With WGS, several results were observed. All three participating labs detected majority

populations (90%, 99%) of first-line resistant strains and 50% resistance. At 10% HR, minority DR populations were reported 2/3 times. WGS was unable to detect 1% HR. Significance: Engineered BCG strains are a safe QC tool for diagnostic efficacy in detection of DR *M.tb* isolates. Here, we show that they can be mixed at specified proportions for QC testing of molecular diagnostics, including targeted and WGS testing. While labs using WGS diagnostic pipelines could detect some HR mixes, phenotypic DST remains the gold standard, able to effectively detect DR populations down to 1%. As more labs turn to WGS instead of phenotypic DST for *M. tuberculosis* resistance testing, the use of engineered BCG strains or other combinations of isolates is encouraged, to ensure that external proficiency testing is offered.

---PAGEBREAK---

Control Number: 2022-A-153-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Genomic epidemiology and population structure of *Campylobacter jejuni* in Peru

Author Block: W. Quino, J. Caro-Castro, D. Flores-Leon, **R. G. Gavilan**;
Instituto Nacional de Salud, Lima, PERU.

Abstract Body: *Campylobacter jejuni* is the principal cause of human bacterial gastroenteritis worldwide and has a high impact on global public health. Also, it is considered the most frequent factor associated with Guillain-Barré syndrome (GBS). In fact, a recent large outbreak of GBS associated with *C. jejuni* strains was confirmed in Peru using molecular techniques. Due to this, it is necessary to analyze the circulating strains of *C. jejuni* using Whole Genome Sequencing (WGS) to understand the relationship between these foodborne pathogens and emerging diseases such as GBS in Peru. The aim of this study was to apply WGS to determine the genetic diversity of the populations of *C. jejuni* strains in Peru. A total of 106 *C. jejuni* strains (102 clinical strains and 4 chicken strains) recovered from 2010 to 2020 were sequenced using Illumina Miseq platform. Also, 62 genomes of *C. jejuni* sequenced by previous studies in Peru were included. MLST analysis identified a high genetic diversity among those strains, represented by 50 sequence types (STs), several of them within 21 clonal complexes (CC). All *C. jejuni* strains linked to GBS were isolated from 2019 to 2020, and characterized as ST-2993. Phylogeny analysis showed that Peruvian *C. jejuni* strains were divided into 2 clades (I and II) with 5 populations (A to E).The majority of STs belonged to the clade I; while ST-2993 belonged to the clade II, whose strains encodes for lipooligosaccharides (LOS) locus genes related to molecular mimicry with gangliosides in peripheral nerves; besides, a close relationship between human and chicken *C. jejuni* ST-2993 indicated this avian as one of the probable reservoirs for human infections. On the other hand, genetic markers associated with antibiotic resistance to quinolones, tetracycline (*tetO, tetW/N/W*), beta-lactamases (*blaOXA-61*), macrolides (A2075G in 23S rRNA) were found in 94.5, 21.7, 66.7, 6.2, 69.8, and 18.6% of strains, respectively. Also, the *cmeABC* multidrug efflux operon was present in 78.3% of strains. In conclusion, WGS provided important information about genetic diversity and virulence factors as determinants of antimicrobial resistance inside *C. jejuni* genome. Furthermore, *C. jejuni* ST-2993 was not detected until the GBS outbreak, probably circulating silently in reservoirs such as chickens. All this evidence shows the necessity to reinforce the epidemiological

surveillance of *C. jejuni* from different sources to improve the prevention and control of this emerging pathogen in our country.

---PAGEBREAK---

Control Number: 2022-A-156-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Nanopore adaptive sampling enriches for target antibiotic resistance genes in soil microbial communities

Author Block: **D. Wrenn**, D. M. Drown;
University of Alaska Fairbanks, Fairbanks, AK.

Abstract Body: Antibiotic resistance is an ongoing public health threat. The surveillance of environmental reservoirs of antibiotic resistance genes (ARG) can both inform and enhance the way that we combat resistant pathogens. While genomic analyses of ARG can be invaluable in the detection and monitoring of resistant organisms, sequencing is costly and time intensive. The Oxford Nanopore Technologies (ONT) MinION sequencer allows for a streamlined library preparation that reduces cost and the time elapsed between sample collection and sequence analysis. The objective of this research is the development of DART (Detection of Antibiotic Resistance Toolbox), a novel toolbox including an environmental antibiotic resistance gene panel to detect antibiotic resistant microbes within soil microbiomes. This research tests the ability of adaptive sampling technology to enrich ARG implemented with a combination of low cost GPUs and the MinION sequencer. Prior research has demonstrated that adaptive sampling can effectively enrich for large sequencing targets, such as whole genomes, in high quality specimens. DART tests the ability of adaptive sampling to enrich for comparatively smaller targets (e.g., ARG) in both high quality and environmental specimens. Using a mock community of environmentally relevant Alaskan soil microbial isolates, we enriched for ARG with adaptive sampling sequencing. Adaptive sampling increased the target composition for the majority of the ARG panel. We also applied this technology to rapidly detecting antibiotic resistant organisms within a soil microbiome in the field. Overall, nanopore sequencing and adaptive sampling have demonstrated the ability to enrich antibiotic resistance genes. DART could serve as a viable tool for the detection of antibiotic resistant pathogens in the field as well as in clinical settings. Future research is warranted to further optimize these protocols as well as test its effectiveness in environmental samples.

---PAGEBREAK---

Control Number: 2022-A-161-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: AusTrakka - integrated pathogen genomics across Australia and New Zealand

Author Block: **C. Sloggett**, S. Carswell, V. Phu, T. Hoang, T. Seemann, B. Howden;
Microbial Diagnostic Unit Public Health Laboratory at the Peter Doherty Institute and University of Melbourne, Melbourne, AUSTRALIA.

Abstract Body:

AusTrakka is Australia's national pathogen genomics surveillance platform. The AusTrakka project was initiated to address a long-standing need for integrated genomic surveillance of pathogens across Australian jurisdictions. Its development under the Communicable Diseases Genomics Network (CDGN) was accelerated to support management of the SARS-CoV-2 outbreak, and in 2020 AusTrakka was endorsed as the national SARS-CoV-2 genomics data sharing platform by the Chief Health Officers of all states and territories in Australia. AusTrakka has formed a core part of the pandemic response in Australia and New Zealand, providing national reporting for COVID-19 surveillance and multi-jurisdictional investigations into priority public health pathogens.

AusTrakka enables Australian and New Zealand public health laboratories to securely submit sequence data for pathogens along with epidemiological and sequence metadata. An integrated phylogenetic analysis is produced by a national team of bioinformaticians and the made available in the form of a searchable, metadata-annotated phylogenetic tree, together with a database of available sample metadata. AusTrakka has been conceived and developed under a highly consultative model, allowing member PHLs to confidently share data within the trusted network of users.

Since its deployment, AusTrakka has hosted over 120,000 genomes, covering eight pathogens, including SARS-CoV-2, Salmonella, Listeria, JEV and MPXV. Analyses and reports, including phylogenetic trees, sample metadata tables, and report summaries for SARS-CoV-2 variants of concern, have been made available to users in real-time upon uploading of new data.

In 2021-22, AusTrakka has been redesigned and re-implemented. This will support larger data volumes and more data types, allow for more flexible sharing and restriction of data, will provide an API interface and command-line tools to users, and will allow for extension with further genomics visualisations.

The development of AusTrakka continues to be a highly consultative process, leveraging Australia's genomic and bioinformatics expertise to bridge the gap between raw genomic sequence data, bioinformatic analysis and epidemiology for real-world application of public health pathogen genomics. We will discuss the engagement, design and implementation challenges around a platform like AusTrakka, its utility for cross-jurisdictional genomic pathogen surveillance, and the ongoing plans for development.

---PAGEBREAK---

Control Number: 2022-A-162-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Genomic Analysis of Novel Dog-Bite-Associated, Fastidious, Gram-negative Bacilli from Multiple Patient Specimens

Author Block:

B. M. Liu[1], M. A. Fisher[2];

[1]Univ of Utah / ARUP; Children's National Hospital / George Washington University, Washington, DC, [2]Univ of Utah / ARUP, SLC, UT.

Abstract Body:

Background: As a leading cause of nonfatal ED visits, animal bites are a significant burden on the health care system, totaling more than $50 million per year in U.S. healthcare costs. Recent epidemiology studies have shown an uptick of dog bites in pediatric populations in the U.S. during the COVID-19 pandemic. In 2020, we received a non-hemolytic, slowly growing, facultatively anaerobic Gram-negative rod (OL1) from a dog bite wound for

identification. Partial 16S rRNA gene sequencing showed OL1 was 95.9% identical to *Ottowia pentelensis* in the family *Comamonadaceae*. Most published *Ottowia* sp. have been isolated from industrial and food sources, with no reports from patient samples. These findings suggest this clinical isolate likely represents a novel *Ottowia* species. Here we aimed to characterize this and other *Ottowia*-like (OL) isolates using genomic approaches.
Methods: Partial 16S rRNA gene sequencing and BLAST searches were performed to determine initial taxonomic status and identify related isolates. Paired-end genomic DNA libraries (Nextera DNA Flex Library Prep, Illumina) of the OL isolates were sequenced as 150 nt reads by Illumina NovaSeq. De novo assembly, annotation, functional prediction, and phylogenetic analyses were performed with Geneious and the suite of analytical tools on the PATRIC web service.
Results: Seven additional isolates from hand wounds/abscesses, including 3 known to be from dog bites, collected in the past decade were identified from our historical sequence database. BLAST search showed they had ⩾ 99.8% identity with OL1. Clustal W alignment of 16S sequences of OL1 and other OL isolates revealed just a single polymorphic position in the 5' third of the gene. OL isolates were phylogenetically distant from all known *Ottowia* spp. as confirmed by genomic phylogenetic analysis and 16S rRNA gene sequencing. Among the >3500 predicted coding genes, several that may confer antimicrobial resistance (AMR) were identified. For example, a resistance nodulation cell division (RND) efflux system and Auxin efflux carrier family proteins were identified, which are predicted to function as multi-drug efflux pumps. Also, genes encoding class C-like Beta-lactamases, and more than five metal-dependent hydrolases of the beta-lactamase superfamily I were found. These findings are consistent with elevated MICs observed in phenotypic antimicrobial susceptibility testing.
Conclusions: These OL isolates are fastidious, Gram-negative bacilli from clinical wound specimens, and are associated with dog bites. Whole genome sequencing analyses confirmed 16S rRNA gene sequence results suggesting these isolates may represent a novel species within the family *Comamonadaceae*. The identification of AMR genes may be predictive of phenotypic resistance.

---PAGEBREAK---

Control Number: 2022-A-163-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: SARS-CoV-2 Evolution Within Immunocompromised Individuals: An Origin of Variants?

Author Block: **D. Lemmer**[1], P. Montfort[1], W. T. Porter[1], C. Hepp[1], S. Dadwal[2], D. Engelthaler[1];
[1]TGen North, Flagstaff, AZ, [2]City of Hope, Duarte, CA.

Abstract Body: Long-term SARS-CoV-2 infection within immunocompromised patients has been proposed as a mechanism that allows for many mutations to emerge within a single infection and within a short period of time, driving the emergence of new SARS-CoV-2 variants. Although this is a common hypothesis, data supporting this theory is limited and challenging to collect due to the need for repeated sampling within long-term infected patients. Thus, previously published articles often focus on a single patient or a couple of patients, limiting the conclusions that can be drawn. To explore these dynamics further, we combine mutational profiles identified within previously published long-term infections and SARS-CoV-2 whole

genome sequence (WGS) data from 13 longitudinally sampled immunocompromised patients. WGS sequencing was performed on nasopharyngeal samples that were collected from the patients between 3 and 11 times over the course of their infection, depending on the patient. Each sample was sequenced using New England Biolab VarSkip Library prep and analyzed using TGen's Amplicon Sequencing Analysis Pipeline (ASAP) to generate consensus genomes and report mutations. Single Molecule Overlapping Read (SMOR) analysis was performed to reduce the effects of sequencing error by only accepting base calls where read 1 and read 2 overlap and agree, to allow confident calls of low-level SNPs. In addition, Synthetic DNA spike-ins (SDSIs) were added to each sample to quantify any potential sample cross-contamination which would reduce confidence in low-level variants. Pangolin was used to determine the lineage of each patient sample to verify whether there were any changes to the lineage in a patient, which could indicate a re-infection from a separate strain. Overall, several patterns emerged within these patients including common deletions within the spike gene and convergence of several mutations known to enhance the virus's ability for immune escape (i.e., E484K/E484Q). These data highlight the selective advantages that specific mutations can have on the SARS-CoV-2 infection. Overall, these data support the hypothesis that infections within an immunocompromised individual can pose an evolutionary advantage for the SARS-CoV-2 virus, allowing gain of both novel and independently recurrent mutations that confer some advantage to the virus.

---PAGEBREAK---

Control Number: 2022-A-170-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: A Novel Approach to Characterizing and Sequencing Influenza A from Rapid Tests

Author Block: D. Erickson[1], D. Lemmer[1], Z. Barrand[1], K. Simmons[1], P. Hawkinson[1], B. Brock[1], **K. H. Sheridan**[2], M. Valentine[1], H. Centner[1], D. Engelthaler[1], C. Hepp[1];
[1]TGen North, Flagstaff, AZ, [2]TGen North, Ashburn, AZ.

Abstract Body: Presently, H1N1 and H3N2 are the most frequently detected subtypes of Influenza A, but given precautions taken during the pandemic, we were interested in determining if different subtypes would arise. Additionally, we hypothesized that precautions taken during the omicron surge may have substantially limited influenza introductions over a short period of time within a Northern Arizona community that resides in congregate living settings. In January and February of 2022, we collected residual rapid test swabs from samples that were confirmed to be positive for Influenza A from this community. The swabs were evaluated for viral load of Influenza A with ReverseTranscription-qPCR (RT-qPCR). To determine which subtypes were circulating, multiple samples with high viral loads were selected for whole-genome sequencing with a DNA bait capture kit. This approach allowed for enrichment of full Influenza A genomes from three of the residual rapid test samples, all of which aligned to the H3N2 subtype. To feasibly sequence and characterize H3N2 genomes from the remaining samples, we designed a tiled amplicon primer scheme based on the DNA bait-capture output to span the entirety of contemporary H3N2 genomes. While tiled amplicon sequencing has frequently been used to enrich small portions of large pathogen genomes and entire unsegmented viral genomes, to our knowledge, it has not been used to amplify whole segmented viral genomes. Application of the primer scheme resulted in near

complete H3N2 genomes from the remaining nasal swabs. However, it failed to amplify an H3N2 positive control from a sample collected in 2012, indicating the necessity of closely related references for development of tiled amplicon sequencing schemes. A phylogenetic tree revealed that at least four introductions occurred within the sampled population, and only one of these clades indicated a rapid outbreak over a short time-frame. This study demonstrates the utility of DNA baits to capture distantly related viruses that can be leveraged for more feasible amplicon approaches.

---PAGEBREAK---

Control Number: 2022-A-173-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Disease surveillance and outbreak investigations using AusTrakka, a national genomic surveillance platform

Author Block:
C. Lam[1], M. Wilmot[2], P. Andersson[2], T. Hoang[3], S. Nguyen[4], K. Horan[3], A. Arnott[5], L. Leong[6], T. Seeman[2], A. V. Jennison[7], V. Sintchenko[1], B. P. Howden[2];
[1]Centre for Infectious Diseases and Microbiology- Public Health–Institute of Clinical Pathology and Medical Research, Westmead, NSW, AUSTRALIA, [2]Microbiological Diagnostic Unit Public Health Laboratory, The University of Melbourne, at The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, AUSTRALIA, [3]Microbiological Diagnostic Unit Public Health Laboratory, The Department of Microbiology and Immunology, The University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, AUSTRALIA, [4]Forensic and Scientific Services, Queensland Health, Brisbane, QLD, AUSTRALIA, [5]Centre for Infectious Diseases and Microbiology Laboratory Services, NSW Health Pathology–Institute of Clinical Pathology and Medical Research, Westmead, NSW, AUSTRALIA, [6]Public Health Laboratory, Microbiology and Infectious Diseases, SA Pathology, Adelaide, SA, AUSTRALIA, [7]Public Health Microbiology, Queensland Public Health and Infectious Diseases Reference Genomics (Q-PHIRE Genomics), Forensic and Scientific Services, Brisbane, QLD, AUSTRALIA.

Abstract Body:
As the number of SARS-CoV-2 cases has continued to rise globally, the appetite and audience for genomic data and relevant analyses has grown over the past 3 years. In Australia, the national genomics surveillance platform AusTrakka, was quickly established to meet this demand and is governed by the Communicable Diseases Genomic Network (CDGN). Participating jurisdictions are able to share genomic data on a centralized platform which is then curated and analysed by a team of bioinformaticians and genomic epidemiologists. To date, over 135,000 SARS-CoV-2 genomes have been shared since January 2020 by jurisdictions across Australia and New Zealand, and providing a comprehensive national overview of the changing SARS-COV-2 viral population within a global context. AusTrakka also has the capacity to generate and visualize phylogenetic relationships between sequences, which together with epidemiological data, can indicate potential inter-jurisdictional COVID-19 disease transmission or emerging clusters of cases. Regular interpretation and reporting of SARS-CoV-2 genomic trends are provided for jurisdictional and national stakeholders to inform public health and policy decisions. These have included the early detection and monitoring of novel variants and variants of concern (VOCs), recombinant strains, and potential drug resistance within viral populations. In addition to SARS-CoV-2 genomic surveillance, AusTrakka has also provided a platform for the

investigation of multi-jurisdictional foodborne disease outbreaks including Salmonella *spp* and Vibrio parahemolyticus. Having a centralized platform and standardized bioinformatic pipeline ensured consistency in repeated analyses over the course of each outbreak. With each outbreak, state public health units and national disease surveillance programs were engaged and consulted throughout the investigation, ensuring the genomic analyses were relevant to end users. For both ongoing genomic surveillance as well as outbreak investigations, AusTrakka has provided a versatile integrated genomic surveillance and analysis platform which can be used to complement and better inform public health responses.

---PAGEBREAK---

Control Number: 2022-A-176-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: The Utility of Next-Generation Sequencing in Revision Total Joint Arthroplasty

Author Block: **C. Baker**, S. Tarabichi, E. Chisari, K. Goswami, G. S. Goh, J. Parvizi;
Rothman Orthopaedics, Philadelphia, PA.

Abstract Body: BACKGROUND:Next Generation Sequencing (NGS) has emerged as a promising diagnostic tool for periprosthetic joint infection (PJI). Despite several studies demonstrating promising results, a common criticism of the technique has been that it overestimates the presence of infection. This study aimed to explore: (1) The ability of NGS to identify pathogens in CN-PJI; and (2) determine whether organisms detected by NGS, as part of a prospective observational study, had any role in later failure of patients undergoing revision surgery. METHODS:In this prospective study, samples were collected from 452 consecutive patients undergoing revision total hip and knee arthroplasties. Of these, 121 patients had PJI, as defined by the International Consensus Meeting (ICM) criteria, and of these 23 were culture-negative (19%). Synovial fluid, deep tissue and swabs were obtained at the time of surgery and sent for NGS and culture/MALDI-TOF. Treatment failure was assessed using the Delphi criteria. In cases of re-operation, organisms present were confirmed by culture. Concordance of the infecting pathogen(s) at failure with the NGS analysis at the initial stage CN-PJI procedure was determined. RESULTS:Among 23 patients with CN-PJI, NGS identified organisms in 11 cases. Twelve cases were both culture and NGS-negative. Five CN-PJIs failed by re-operation with infection recurrence confirmed on culture. In 3 of these 5 cases (60%), NGS reported organisms at the initial CN-PJI procedure. The remaining cases failed with a new organism. NGS detected several organisms in CN-PJI cases. Additionally, 1/3 (33%) of aseptic revisions failed from an organism identified via NGS at the initial procedure. DISCUSSION:CN-PJI is often associated with polymicrobial metagenomic organism profile. Furthermore, most failures by infection recurrence were due to an organism previously detected by NGS. Our findings suggest some cases of PJI may be polymicrobial and escape detection using conventional culture.

---PAGEBREAK---

Control Number: 2022-A-177-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Genetic Diversity of USA300 Strains: Y2K to the Present

Author Block: **E. A. Felton**[1], S. J. Kennedy[1], J. K. Jackson[1], E. Cella[2], T. Azarian[2], L. N. Shaw[1];
[1]University of South Florida, Tampa, FL, [2]University of Central Florida, Orlando, FL.

Abstract Body: *Staphylococcus aureus* is an adaptable and dangerous human pathogen, causing infection in every single niche of the human body. The clonal lineage USA300 has become the predominant *S. aureus* clone within the United States since it first emerged at the turn of the century. Originally identified as causing wound infections, the strain has more recently evolved towards eliciting blood stream infections. To gain insight into factors that have influenced changes in USA300 pathogenesis over the last twenty years, we performed whole genome sequencing (WGS) of two distinct clinical isolate libraries - one isolated ~ 2000 (n=69) and the other from ~2019 (n=49). WGS was performed using the Hackflex library preparation, a method of optimization for Illumina DNA sequencing of microbial genomes. Hackflex takes advantage of Illumina tagmentation chemistry and relatively small genome size of most microbes, to allow users to dramatically increase the number of samples for the same amount of reagent. The genomes of both libraries were analyzed for single nucleotide polymorphisms (SNPs) against the well characterized FPR3757 genome (isolated in 2000) using Snippy (v4.6.0). The resulting analysis reveals, as one might expect, increased sequence variability in USA300 strains over time. The older isolates display far less divergence from the reference strain in terms of overall numbers of SNPs (1965) and their distribution throughout the genome. Conversely, the more recent isolates showed significantly more SNPs throughout their genomes (5723). When reviewing nSNPs in circa 2000 strains, we observed variants in the proteases *sspB* and *sspA*, enterotoxins *seq* and *sek*, and the surface associated virulence factors *clfAB* and *sdrDE.* Interestingly, the more recent strains had nSNPs in all major regulators of virulence (except *codY*), as well as variants in 9/10 secreted proteases, alpha toxin, *lukSF-PV*, and most MSCRAMM genes. Importantly, we show a number of these nSNPs phenotypically influence processes such as hemolysis, proteolysis, and biofilm formation. Collectively, this study provides new insight into the evolution of the USA300 strain, and its phenogenomic adaptation over the last 20 years.

---PAGEBREAK---

Control Number: 2022-A-181-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Evaluation of the ONT MinION Mk1C platform vs Illumina MiSeq for the epidemiological surveillance of SARS-CoV-2 in the Northern Metropolitan Area of Barcelona, Spain

Author Block: **A. C. Pelegrin**[1], D. Panisello Yagüe[1], L. Soler[1], A. E. Bordoy[1], S. Martínez-Puchol[1], G. Clara[1], S. González-Gómez[1], A. Paris de León[1], A. Blanco[1], M. Noguera[2], M. Armengol[3], O. Francino[4], I. Blanco[5], V. Saludes[1], P. Cardona[1], E. Martró[1];
[1]Microbiology Department, Laboratori Clínic Metropolitana Nord. Germans Trias i Pujol University Hospital, Germans Trias i Pujol Research Institute (IGTP), Badalona, Spain, Badalona, SPAIN, [2]IrsiCaixa AIDS Research Institute, Germans Trias i Pujol Research Institute (IGTP), Badalona, SPAIN, [3]Translational Genomics Unit, IGTP, Badalona, SPAIN, [4]Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra, SPAIN, [5]Clinical Genetics

Department, Laboratori Clínic Metropolitana Nord, Germans Trias i Pujol University Hospital, Badalona, SPAIN.

BackgroundSince the early days of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic, whole genome sequencing (WGS) has been a key element, stepping up as the gold standard, high-resolution epidemiological marker method to trace the evolution of SARS-CoV-2 for global surveillance. It has been openly discussed that third-generation sequencing platforms confer certain economical and practical advantages over their second-generation counterparts, which could be useful to continue SARS-CoV-2 epidemiological surveillance efforts. We aimed to evaluate the technical (number of single nucleotide polymorphisms (SNPs), coverage, lineage assignment) and practical (turn-around time (TAT), costs) differences between the Oxford Nanopore Technologies (ONT) Mk1C sequencing device compared to its Illumina MiSeq counterpart.

MethodsNinety-four nasopharyngeal swab samples selected for routine surveillance purposes from patients with a positive real time RT-PCR and a Ct value <30, and two controls (negative and positive) were processed in parallel on both sequencing platforms during the 14th week of 2022. Samples were kept at 4 C until nucleic acid extraction (< 2 days). RT-PCR and library preparation were executed as follows: LunaScript reverse transcriptase and ARTIC 4.1 tiled PCR followed by Illumina DNA prep kit for the Illumina MiSeq sequencing (v2, 2x150 cycles), and Midnight RT PCR Expansion followed by Rapid Barcoding Kit 96 kit for ONT Mk1C with an R9.4 flow cell run overnight. Results from both sequencing runs were analyzed using the nf-core/viralrecon pipeline (v2.4.1). TAT and total sample analysis cost were registered.

Results & Discussion Three out of the 94 samples could not be sequenced by either of the two methods, probably due to RNA degradation. Mean genome coverage was 99% for both sequencing platforms while median depth of coverage was 1028X and 543X for Illumina and ONT devices, respectively. Four samples were discordant for the following amoniacid substitutions not detected by the ONT method: ORF1a:A2944V (1), S:T19I (3), ORF3a:V77I (1). Despite this, lineage assignments were fully concordant (48% BA.2, 23% BA.2.9, 14% BA.2.3, 8% BA.2.3.2, 7% others). ONT had a 9-fold shorter TAT library preparation step (20 min) and an approximate 3-fold cheaper sequencing process. Compared to the Illumina MiSeq, the ONT Mk1C sequencing methodology is a rapid, reliable and affordable alternative to implement SARS-CoV-2 epidemiological surveillance.

Abstract Body:

---PAGEBREAK---

Control Number: 2022-A-185-NGS

Topic 1: Bridging Silos: Exploring mechanisms for collecting and sharing microbial genomic data to foster interoperability

Topic 2:

Publishing Title: Management of Wastewater Data and Dashboard Development for SARS-CoV-2 Variant Tracking

C. H. Bias[1], T. Kayikcioglu[2], J. Amirzadegan[1], P. Ramachandran[2], R. Timme[2], M. Balkey[2]; [1]Oak Ridge Institute for Science and Education/Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, MD, [2]Center for Food Safety and Applied Nutrition, Food and Drug Administration, College Park, MD.

Author Block:

As wastewater surveillance rapidly gains status as a means to assess and predict the course of the SARS-CoV-2 (SC2) pandemic, it is crucial to develop standard methods to collect,

Abstract Body:

curate, and visualize sequence data and associated metadata. With years of experience managing a laboratory network sequencing enteric pathogens, collecting standard metadata for environmental samples, and making data public in real-time, the FDA GenomeTrakr program stepped into this role in March 2021. GenomeTrakr laboratories perform targeted sequencing of SC2 from wastewater samples and collect a suite of standard metadata for public release, fully describing and submitting the sampling and laboratory methods employed with custom versions of the NCBI BioSample "SARS-CoV-2: wastewater surveillance" package and the SRA metadata template. We built an interactive dashboard in Tableau Desktop using five data sources: 1) NCBI Run Selector's integrated sample and sequencing metadata; 2) BioSample metadata for wastewater samples without sequencing records, queried via NCBI Entrez; 3) Geographic and BioProject information for each laboratory; 4) Analysis results from CFSAN's Wastewater Analysis Pipeline (C-WAP) showing relative abundance of SC2 variants in each sample; 5) Quality control metrics for each sequencing record from C-WAP. The dashboard is uploaded to Tableau Public to embed in the FDA.gov "Wastewater Surveillance for SARS-CoV-2 Variants" webpage, which garnered 3936 unique page views by July 25, 2022. Our current dashboard consists of three visualizations: an interactive map of laboratories and sample contributions, a Gantt chart plotting timing of sample collection and sequencing by lab, and a bar graph of lineage abundance estimations against time. Displayed data spans from the week of September 12, 2021 to present and contains 1392 samples and 1521 sequencing runs (on July 25, 2022). The incoming data is added once available in NCBI. This project shows the strength of implementing a standard data structure within NCBI for submitting data collected for public health application and encourages transparency in data sharing standards. Our approach for documenting methods as machine-readable metadata enables automated downstream analyses (e.g., variant analysis and dashboard monitoring). Additionally, because participant labs can see their data visualized within days of public release, metadata gaps can be identified quickly and corrected. Most importantly, this project shows the promise of interoperability with similar sequencing surveillance efforts among collaborating labs adhering to the same standards.

---PAGEBREAK---

Control Number:	2022-A-188-NGS
Topic 1:	Bridging Silos: Exploring mechanisms for collecting and sharing microbial genomic data to foster interoperability
Topic 2:	
Publishing Title:	The Mercury Workflow Series and Terra_2_NCBI: Bioinformatics Solutions to Facilitate Genomic Data Submission to NCBI & GISAID for Public Health Laboratories
Author Block:	**S. M. Wright**[1], F. J. Ambrosio[1], K. G. Libuit[1], D. Park[2], J. Sevinsky[1]; [1]Theiagen Genomics, Highlands Ranch, CO, [2]Broad Institute, Cambridge, MA.
Abstract Body:	Submission of pathogen genomic data to internationally accessible repositories such as NCBI and GISAID is essential to public health and improving our understanding of infectious disease dynamics. Some benefits of data sharing include helping local epidemiologists understand introduction events and the rise of novel variants; globally, data sharing can inform stronger vaccine and diagnostic assay development. Yet, for public health laboratories (PHLs) that lack the bioinformatics capabilities to programmatically prepare and submit genomic data to NCBI or GISAID, the process of

sharing critical genomic data can be cumbersome and include many manual steps prone to error. For some PHLs, this challenge prohibits data sharing of any kind.
To address this challenge, the Mercury Workflow Series and Terra_2_NCBI bioinformatics resources were developed to assist PHLs in submitting pathogen genomic data to internationally accessible data repositories. These workflows were made accessible through Terra, a bioinformatics web-application that provides a graphical user interface to run bioinformatic pipelines, enabling easy adoption for PHLs.
The Mercury Workflow Series consists of open-source Workflow Description Language (WDL) workflows developed to programmatically prepare NCBI and GISAID submission files for SARS-CoV-2 amplicon sequences from clinical samples. The Mercury workflows ingest SARS-CoV-2 read, assembly, and sample metadata files and format these data for submission per the Public Health Alliance for Genomic Epidemiology (PHA4GE)'s SARS-CoV-2 Contextual Data Specifications.
Terra_2_NCBI is a separate WDL workflow that facilitates the submission of genomic data to BioSample, SRA, and GenBank directly from the Terra platform for any organism. After creation of a BioProject and submission group, the user of the Terra_2_NCBI workflow will not need to interact with the NCBI web portal, saving valuable time and energy as submission is performed for the user through the NCBI FTP interface.
Mercury workflows have been adopted by nearly 50 US city, county, and state PHLs, and helped facilitate hundreds-of-thousands of SARS-CoV-2 submissions to NCBI and GISAID. The recently released Terra_2_NCBI workflow enabled the rapid data sharing from over a hundred fungal genomes to NCBI. Some of these submissions originated from PHLs that never previously submitted data to public repositories.
Programmatic data submission to NCBI greatly reduces the burden on PHLs. These workflows ease the crucial yet time-consuming process of data sharing, allowing scientists to focus not on operational duties, but on the science.

---PAGEBREAK---

Control Number: 2022-A-194-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Modularizing SARS-CoV-2 Bioinformatics Solutions for Adaptability to Other Pathogens of Public Health Concern

Author Block: **F. J. Ambrosio**[1], C. J. Kapsak[1], N. Hull[2], K. G. Libuit[1], J. R. Sevinsky[1];
[1]Theiagen Genomics, Highlands Ranch, CO, [2]APHL, Silver Spring, MD.

Abstract Body: Bioinformatics workflows for the analysis of viral genomes have become increasingly valuable tools in the public health laboratory analysis arsenal. The SARS-CoV-2 pandemic spurred innovation and collaboration throughout the field of public health pathogen genomics resulting in rapid development and dissemination of effective techniques to analyze viral genomic material and highlighting the value of the tiled amplicon approach to viral whole genome sequencing. The tiled amplicon approach has been used in a variety of sequencing assays, yet typically a custom bioinformatics workflow must be developed to address organism-specific considerations. This custom workflow development process inexorably leads to redundancy of effort, a lack of standardization, and issues in ongoing tool maintenance.
Often, a laboratory scientist is in search of a pipeline that is functionally equivalent to one

which already exists, but the ability to repurpose this pipeline is restricted to those laboratories that have an experienced bioinformatician to modify the pipeline components to suit their needs. With a more modular approach, pipelines would not need to be developed from scratch, reducing the need for duplication of efforts, and allowing bioinformaticians to focus their time on the organism-specific components of the workflow. Moreover, maintaining a single modular workflow requires far less effort than a suite of unique organism-specific workflows. Finally, a modular pipeline would provide a framework of quality control metrics, intermediate files and output reports which will facilitate comparison of results between collaborators.
The TheiaCoV_Illumina_PE workflow was developed specifically for SARS-CoV-2 sequencing data produced using the ARTIC Network primer scheme. This workflow has been modularized to allow for non-SARS-COV-2 datasets, models and analytical modules to be triggered or excluded from the analysis by a user with no programming background. By making the components of the TheiaCoV_Illumina_PE workflow modular, traditional microbiologists are empowered to operate workflows as they see fit, and our bioinformatics collaborators are enabled to focus on developing particular components of the workflow for new organisms rather than entire organism-specific workflows.
The effectiveness of this approach has been demonstrated by the adaptation of the TheiaCoV_Illumina_PE workflow to generate consensus genome assemblies and organism-specific characterization for non-SARS-CoV-2 viruses, including Monkeypox and HIV. These results help to glean critical insights in pathogen classification, drug resistance, and subsequent genomic epidemiological investigations, which establishes the tiled amplicon and modular bioinformatics workflow approach as a flexible solution for public health pathogen genomics.

---PAGEBREAK---

Control Number: 2022-A-195-NGS

Topic 1: Epidemiological Cues: NGS in Clinical and Public Health Microbiology

Topic 2:

Publishing Title: Comparing Genomic Methods for *Salmonella* Enteritidis

Author Block: **S. Holtsmark Nielsen**, P. Gymoese, E. Litrup;
Statens Serum Institut, Copenhagen, DENMARK.

Abstract Body: **Introduction** *Salmonella* Enteritidis is the most common serovar reported in human *Salmonella* infections in Europe. The most frequent sequence type is ST11 and the population of ST11 in Europe consists of two clades that are highly clonal and limited diversity is seen within these clades. The high clonality causes difficulties in surveillance and outbreak investigations and commonly used WGS analysis such as core-SNP analysis or cgMLST can give troubling results. We compared the use of cg/wgMLST analysis versus SNP analysis with different references with the aim of gaining more resolution in our ST11 population.**Material and Methods** The dataset consist of two Danish outbreaks. One caused by eggs (28 cases), the other is still under investigation (16 cases). 20 genomes of other closely related strains, not a part of the outbreaks, were also included in the analyses. cgMLST and wgMLST were calculated in BioNumerics8.1 (Applied Maths) using the Enterobase scheme and the BioNumerics wgMLST scheme. The SNP analysis was completed with two different reference genomes. The NCBI reference sequence AM933172 and an in-

silico hybrid reference created from the consensus of eight Oxford Nanopore sequenced ST11 strains. Bowtie2 (v. 2.4.2) local mapping allowing one mismatch and bcftools (v. 1.10.2) consensus caller was used for the analysis. SNPs with 10X depth and quality 20 was kept for further analyses. IQ-tree (v. 2.2.0.3) with substitution model GTR+G and 1000 bootstraps was used to build the maximum likelihood tree. Allele vs SNP differences was compared using R (v. 4.2.1) dist.dna function (ape package v. 5) and visualized with pheatmap (v. 1.0.12) and trees visualized with ggtree (Bioconductor 3.15). **Results** The hybrid reference were initially tested on a part of the Danish Enteritidis surveillance collection. These results indicated higher resolution than SNP analysis with AM933172. Therefore, we decided to investigate the use of the hybrid reference in two outbreaks.From these analyses the hybrid reference added more variation in many cases and therefore a better possibility to cluster the isolates correctly. Within the current outbreak we were able to differentiate and exclude three probable cases. From the other outbreak two samples was much more divergent (32 and 43 SNPs) using the hybrid reference, these were closely related from all other analyses. However, there are still uncertainties using a hybrid reference. **Conclusion** Due to the high clonality of Enteritidis more than the core genes are needed to attempt at clustering the isolates and conduct useful outbreak investigations. Using the hybrid reference slightly more resolution within clusters are detected, though clustering is still challenging.

---PAGEBREAK---