Development and Validation of Metagenomics Sequencing Pipelines for Biosurveillance and Diagnostics

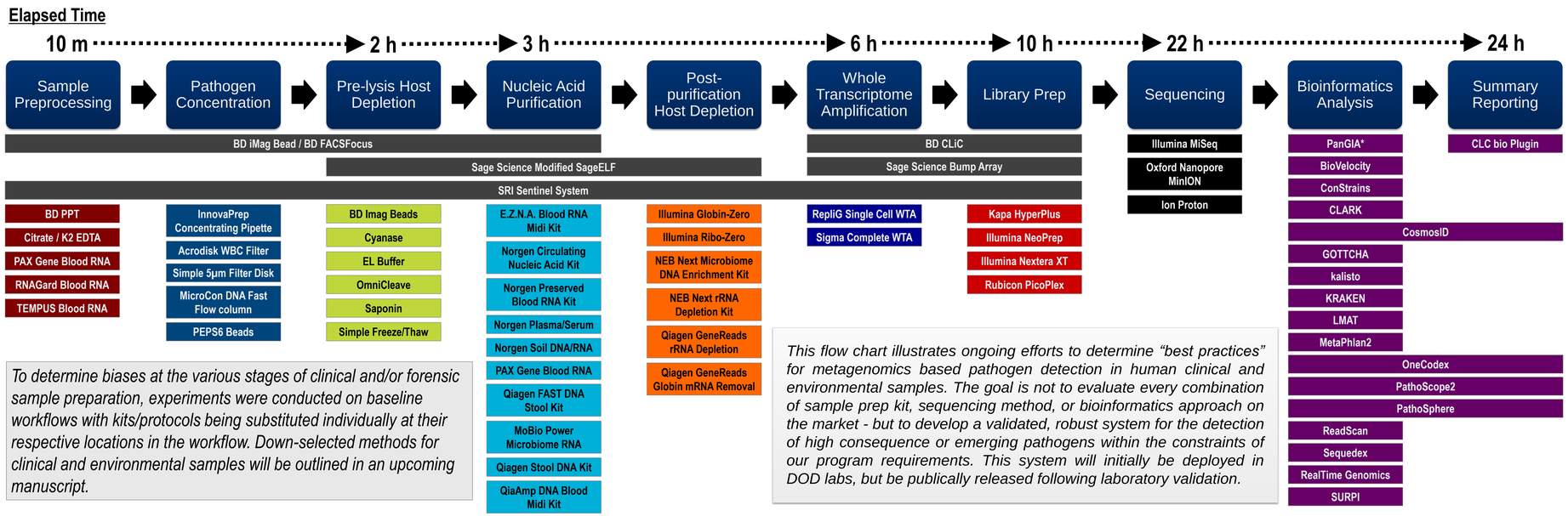
JA Russell[1], JL Jacobs[1], R Winegar[2], C Zimmerman[2], JR Aspinwall[2], JS Bagnoli[1], K Parker[2], J Stone[2], B Campos[2], T Slezak[3], PS Chain[4], KW Davenport[4], P Li[4], JJ Anderson[5], KA Bishop-Lilly[6], KG Frey[6], T Postlethwaite[7], T Spain[7], J Jakobsen[8], C Boysen[8], K Werking[2], M Cassler[2]

MRIGlobal
National Solutions Worldwide Impact

ABSTRACT

Next generation sequencing (NGS) has the potential to allow unbiased detection and characterization of biothreat agents and emerging pathogens from a variety of clinical and environmental samples. This capability would greatly benefit multiple applications, including microbial forensics, biosurveillance, clinical detection and clinical diagnostics. However, current sample to sequence pipelines are complex, and there is a growing need for them to be simplified, standardized, and validated before results can be made comparable across multiple laboratories. Specific needs include standard reference materials, simplified sample and library preparation, trusted reference databases, robust bioinformatics pipelines, and clear regulatory pathway. MRIGlobal is leading a large team of vested organizations in developing and validating methods for accelerating the use of NGS as a powerful tool for the detection of infectious disease agents. Here we present the overall the scope of the project and the trajectory of its development.

Evaluation All Stepwise Components Integral to a Metagenomics Workflow for Pathogen Detection



To determine biases at the various stages of clinical and/or forensic sample preparation, experiments were conducted on baseline workflows with kits/protocols being substituted individually at their respective locations in the workflow. Down-selected methods for clinical and environmental samples will be outlined in an upcoming manuscript.

This flow chart illustrates ongoing efforts to determine "best practices" for metagenomics based pathogen detection in human clinical and environmental samples. The goal is not to evaluate every combination of sample prep kit, sequencing method, or bioinformatics approach on the market - but to develop a validated, robust system for the detection of high consequence or emerging pathogens within the constraints of our program requirements. This system will initially be deployed in DOD labs, but be publically released following laboratory validation.

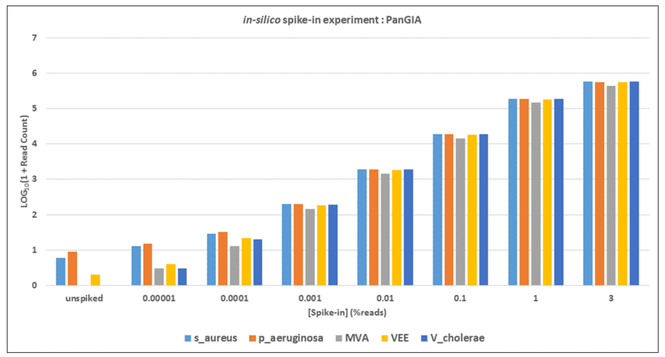
Program Requirements

This project aims to develop a robust, metagenomics based pathogen detection system that meets the following operational requirements:

- Capable of handling both clinical and environmental sample types.
- Deployable to fixed laboratories in OCONUS operation settings.
- Sample to answer within actionable time frames.
- Capable of detecting all known pathogens at clinically relevant levels.
- A single workflow that includes gram(-) and gram(+) bacteria, spore forming bacteria, and both RNA and DNA viruses.
- Reliable, "gold-standard" database of reference genomes for bioinformatics analysis.

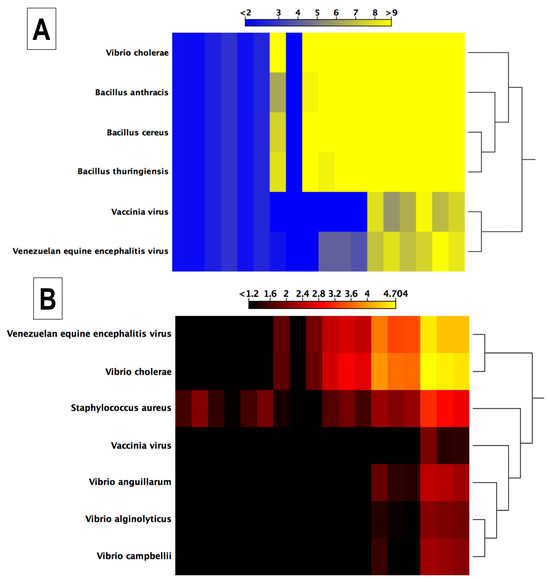
***AFFILIATIONS**: 1) MRIGlobal, Rockville, MD. 2) MRIGlobal, Palm Bay, FL. 3) Lawrence Livermore National Laboratory, Livermore, CA. 4) Los Alamos National Laboratory, Los Alamos, NM. 5) Defense Threat Reduction Agency, Fort Detrick, MD. 6) Naval Medical Research Center, Frederick, MD. 7) Draper Laboratory, St. Petersburg, FL. 8) Qiagen, Redwood City, CA.* ***FUNDING**: Financial support from the Defense Threat Reduction Agency (HDTRA1-15-C-0013):* Distribution Statement A: Approved for Public Release; distribution is unlimited.

In-Silico Datasets



*The PanGIA taxonomic classifier for metagenomic reads has been developed in collaboration with wet-lab efforts to determine best practices in detection of a variety of pathogens from a variety of operational sample matrices. This model has helped the development and integration of features for specific pathogen detection. In-silico performance testing of the PanGIA pipeline **(above)** has shown consistent detection of viral and bacterial pathogens above background signal down to the 0.0001% spike-in level. In this analysis, 150bp PE fastq reads were generated from high-quality genomes, using an Illumina error-model, and spiked into a 20 million-read sequencing dataset from human blood. In simulations of 'real-world' samples **(right)**, where actual cells/virions were spiked into soil **(A)** and blood **(B)**, targets are detected down to 1e3 dilutions.*

Real-world '*spike-in*' Datasets



FUTURE DEVELOPMENTS

This multi-year effort will go beyond just pathogen detection, and include

- Multi-lab validation and training OCONUS partner laboratories;
- Provide continued long-term sustainment of OCONUS laboratory activies;
- Sequence based pathogen characterization for antibiotic resistance genes and virulence factors;
- Strain level identification (where possible);
- Automated data compression and report delivery.

In addition, the results of the laboratory and bioinformatics comparative studies will be released publically. The source code for the PanGIA bioinformatics pipeline will be publically released as well.

Contact information

Dr. Mike Cassler

MRIGlobal Dx
1470 Treeland Boulevard, SE
Palm Bay, FL 32909-2211

T: 321 - 308 - 6900
E: mcassler@mriglobal.org