Achievement of rapid whole genome coverage of bacterial pathogens at 1 CFU/mL in blood

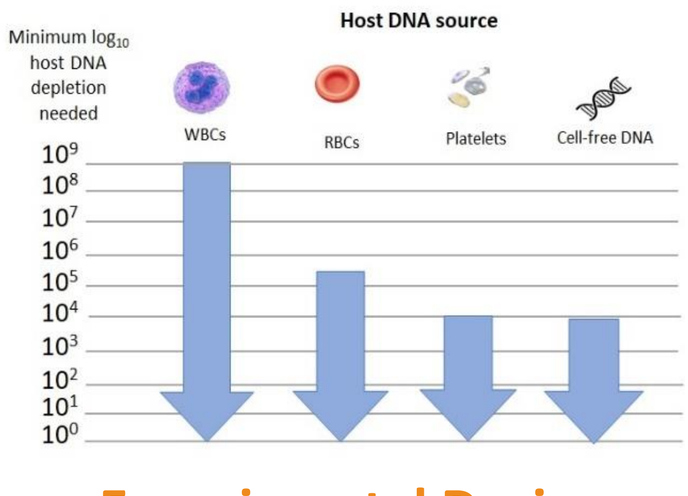
Nicole Billings*, Chiahao Tsui*, Ian Herriott*, Febriana Pangestu*, Michael Kumcu*, Elliott Bartsch*, Miriam Huntley*
* Day Zero Diagnostics, Inc.

Introduction

There has been limited success using whole genome sequencing for culture-free diagnosis of bacterial bloodstream infections in part due to the ~10^8 ratio of human to bacterial DNA. Here we report successful WGS of pathogen DNA from blood spiked with 1 CFU/mL of *Escherichia coli*, *Klebsiella pneumoniae, Enterobacter cloacae, Acinetobacter baumannii,* or *Staphylococcus aureus*, processed using our in-house bacterial enrichment pipeline and sequenced with the ONT MinION. Overall the method demonstrates a proof-of-concept use of rapid sequencing for diagnosing bacterial bloodstream infections directly from clinical samples.

Challenges to Culture-Free WGS from Blood

1. Absolute abundance: at clinically relevant loads of 1 CFU/mL, there are only 10s of femtograms of bacterial DNA in a clinical sample.
2. Relative abundance: human DNA outnumbers bacterial DNA by 8-9 orders of magnitude. Sources of human DNA include white blood cells, red blood cells, platelets, and cell-free DNA.
3. Blood itself and blood collection containers contain molecular amplification inhibitors, such as hemoglobin and SPS.

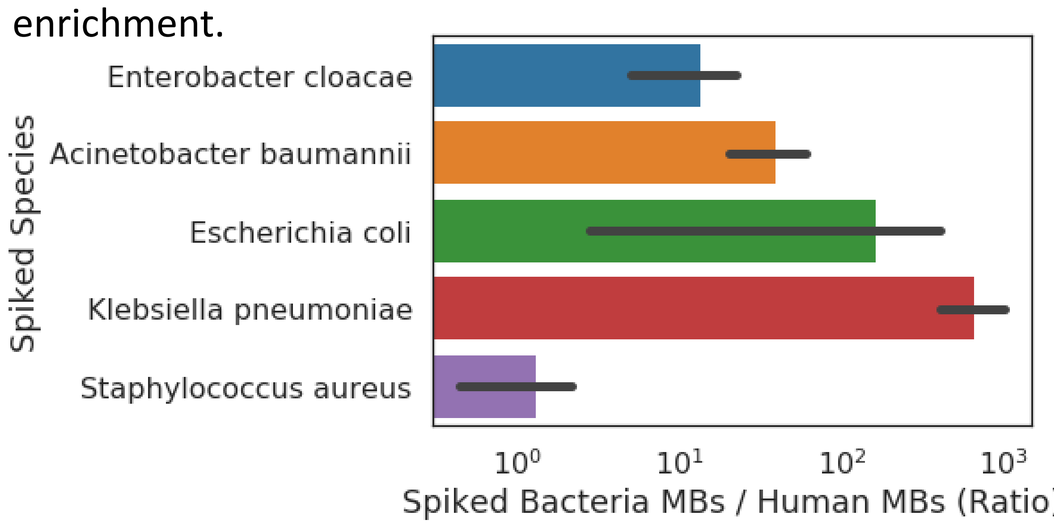


Experimental Design

Stocks were prepared from bloodstream or wound drainage derived clinical isolates of *Acinetobacter baumannii, Escherichia coli*, *Enterobacter cloacae, Klebsiella pneumoniae,* and *Staphylococcus aureus*. For each species, 4 samples of healthy whole blood (2 samples per donor) were spiked with bacterial stocks at 1 CFU/mL concentration. We processed the samples using our in-house bacterial enrichment pipeline *Blood2Bac* followed by isothermal whole genome amplification. Libraries were prepared using ONT SQK-RPB004 and sequenced on the MinION, yielding 7-820 Mbp per sample. Reads were classified using Kraken to determine relative abundance of species.

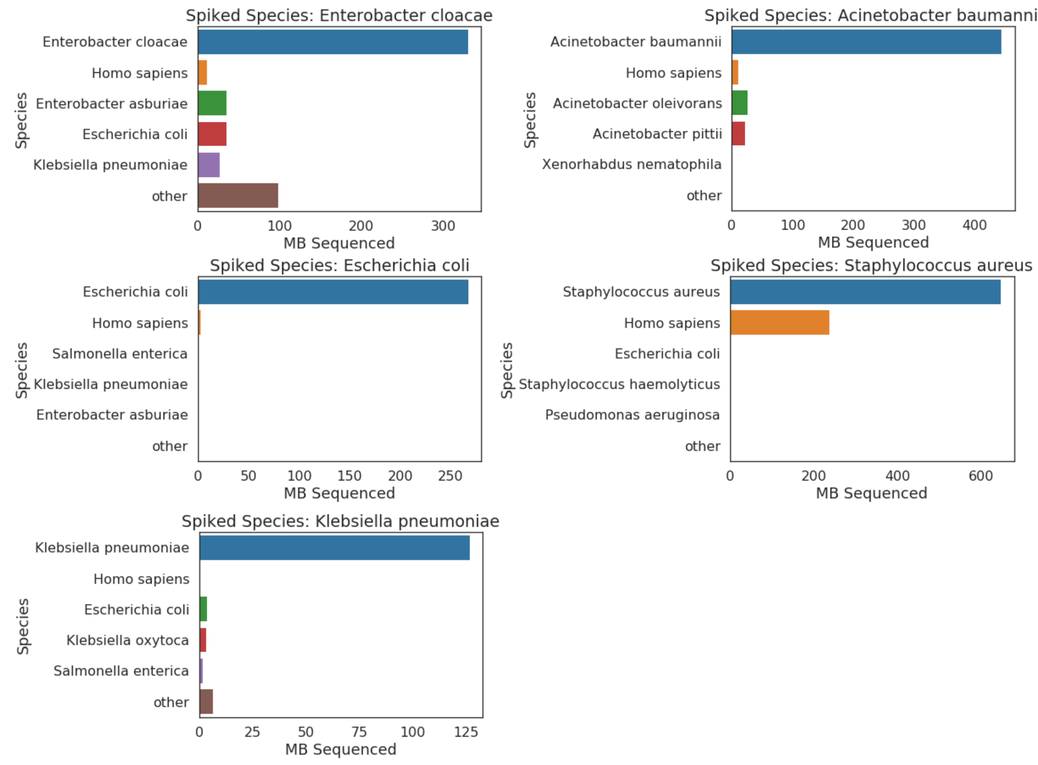
Achievement of 10^7-10^{11} Fold Enrichment of Bacterial DNA

Across all 20 experiments, we achieved a final ratio of the spiked pathogen DNA to human DNA of 10^{-1}-10^3, representing >10^7 fold enrichment.



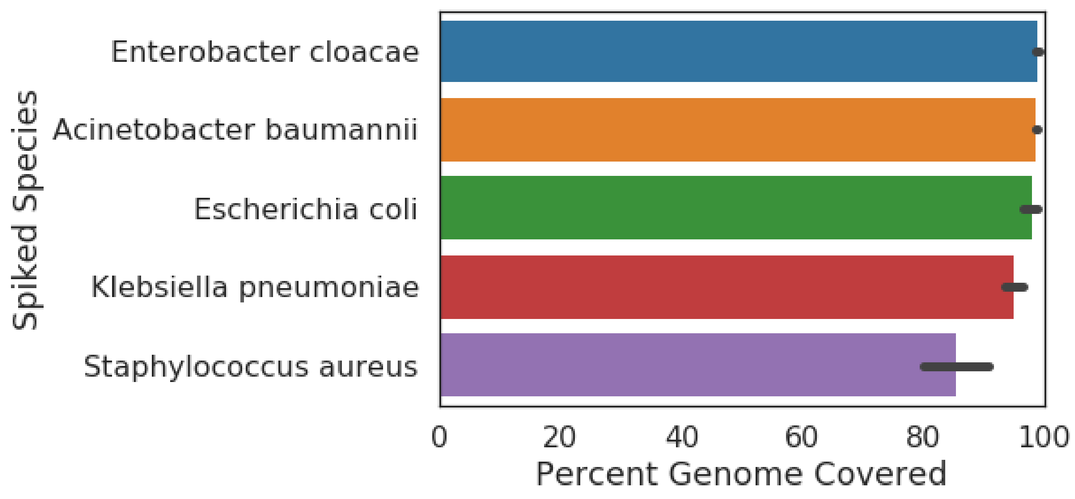
Few Contaminants in Sequenced Data

Representative examples of the breakdown of read classification from these experiments show few contaminants or misclassified reads. *E. cloacae* experiments had the most reads classified as extraneous microbial species, while *S. aureus* experiments had the most human contaminant DNA.



Whole Genome Coverage

We assessed what percent of the genome of the spiked isolate was represented in the data. For each experiment, we removed all human reads (as identified by Kraken) and mapped the remainder with minimap2 to matched short-read assemblies (created for each isolate with Illumina NextSeq 2x150 data and assembled with Spades 3.8.1). 19 samples are represented below (one *A. baumannii* experiment yielding <8MB total sequenced reads was excluded from this analysis).



Predicting Antimicrobial Resistance Phenotype

We had previously developed a custom machine learning algorithm to predict AMR from genomic sequences alone, trained with pathogen short-read WGS data. We used these models to predict binary AMR for the 4 *K. pneumoniae* (shown at right) and 4 *E. coli* experiments after removing human reads and correcting with canu. We achieved 88% prediction accuracy across 72 individual predictions.

K. pneumoniae Predictions

DAYZERO DIAGNOSTICS