Validation of epiXact®: Robust bacterial relatedness and outbreak detection pipeline for WGS data

Mohamad Sater[1], Nicole Billings[1], Timothy Farrell[1], Defne Surujon[1], Ian C. Herriott[1], Grace Cox[1], Alfred Wong[1], Mihae Platt[1], Miriam Huntley[1]

[1] Day Zero Diagnostics, Inc. (Boston, MA)

Introduction

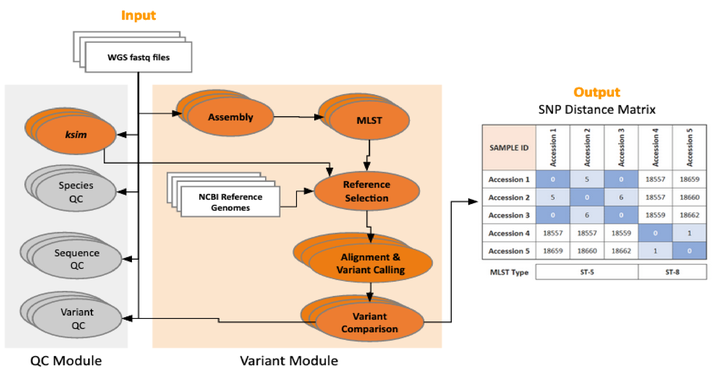
Although hospital acquired infection (HAI) prevention and management has improved over the past decade, an estimated 35-70% of HAIs remain preventable. Whole genome sequencing (WGS), in combination with single nucleotide polymorphism (SNP) analysis to measure clonality and genomic relatedness, has emerged as the most accurate tool for definitive HAI tracking. However, the speed and accessibility of WGS analyses has limited the translation of this technology from the research laboratory to reliable use in infection control settings. To this end we developed epiXact®, an automated computational pipeline that can rapidly and robustly detect pathogen relatedness from WGS data and therefore enable transmission detection. In this work we performed a large-scale validation of the epiXact pipeline to measure its accuracy and reproducibility for clonality detection.

epiXact®

DZD's Illumina-based commercial HAI sequencing and analysis service, is used by multiple partnering hospitals to investigate a wide variety of suspected outbreaks. When infection control (IC) suspects transmission, cultured bacterial samples are sent to DZD. The epiXact test is a "rule-in/rule-out" test of an infection transmission event based on clonality of 2 or more bacterial isolates. This test is designed as a stand-alone test to provide clinicians with a definitive measure of the relatedness of bacterial isolates. Results are reported back in 1-2 days allowing Infection Control to use the definitive genomic evidence to inform decisions regarding ward cleaning, staff screening, and equipment contamination.

We have investigated a wide variety of suspected outbreaks (in both clinical and laboratory settings) at the request of partnering institutions using the epiXact pipeline to determine sample genomic relatedness. In 36 recently examined suspected outbreaks, bacterial samples were sent to Day Zero Diagnostics for clonality assessment. Across these cases we sequenced and analyzed a total of 116 bacterial samples encompassing 15 species/pathogen types (e.g. MRSA, ESBL, CRE). 25 (70%) of the 36 cases were found to contain one or more clonal clusters, providing evidence for an outbreak, while the remaining 11 cases did not. In all cases we reported back results within 24-48 hours from sample receipt. Infection control specialists were able to use the clonality determination results to inform decisions regarding ward cleaning, staff screening, and equipment contamination.

epiXact® bioinformatic analysis pipeline



Validation Design

We designed and performed a validation to test the fidelity of the epiXact pipeline in accordance with CLIA regulations. We tested 5 representative species frequently implicated in HAIs, with varied gram stain, genomic DNA GC content, and genome sizes. Clonal and non-clonal bacterial samples from the 5 species were sequenced with an Illumina iSeq100 to generate a blinded test dataset of 76 samples. An *in silico* dataset was also generated for testing, with 40 simulated sequenced libraries across 5 species. For 5 different species (*S. aureus, E. faecium, A. baumannii, P. aeruginosa and K. pneumoniae*), we simulate 8 sequenced libraries in silico from a representative NCBI reference genome, where one sequenced library represents the reference strain and the other 7 sequenced libraries have between 5 and 200 synthetic SNPs introduced to their sequences (relative to the reference). The data was input into the automated epiXact pipeline, which computed the number of SNP differences between pairs of samples to quantify relatedness. Pairs of samples were classified as clonally related, closely related, or not closely related depending on the pairwise SNP distances.

Table 1: Select validation species

Species	Description/ Characteristics
Enterococcus faecium	Gram positive, low GC content, 2.9MB genome
Escherichia coli	Gram negative, medium GC content, 4.6MB genome
Pseudomonas aeruginosa	Gram negative, high GC content, 6.5MB genome
Staphylococcus aureus	Gram positive, low GC content, 2.8MB genome
Clostridioides (Clostridium) difficile	Gram positive, low GC content, 2.8MB genome

Table 2: Species NCBI reference genome list

Species	NCBI Reference genome
Enterococcus faecium	GCF_001721005
Klebsiella pneumoniae	GCF_002258055
Pseudomonas aeruginosa	GCF_003028335
Staphylococcus aureus	GCF_001278745
Acinetobacter baumannii	GCF_001908295

Table 3: Validation strains selected from known bacterial biobanks

Species	Source	DZD sample ID	MLST
Enterococcus faecium	IHMA	*dzd-ihma-69*	ST-1451
Enterococcus faecium	IHMA	*dzd-ihma-80*	ST-1451
Enterococcus faecium	BEI	*dzd-bei-250*	ST-78
Enterococcus faecium	BEI	*dzd-bei-252*	ST-78
Escherichia coli	CDC	*dzd-cdc-78*	ST-101
Escherichia coli	CDC	*dzd-cdc-88*	ST-101
Escherichia coli	CDC	*dzd-cdc-152*	ST-131
Escherichia coli	CDC	*dzd-cdc-175*	ST-131
Pseudomonas aeruginosa	CDC	*dzd-cdc-204*	ST-233
Pseudomonas aeruginosa	CDC	*dzd-cdc-221*	ST-233
Pseudomonas aeruginosa	CDC	*dzd-cdc-232*	ST-17*
Pseudomonas aeruginosa	CDC	*dzd-cdc-277*	ST-17*
Staphylococcus aureus	CDC	*dzd-cdc-474*	ST-5
Staphylococcus aureus	CDC	*dzd-cdc-480*	ST-5
Staphylococcus aureus	CDC	*dzd-cdc-476*	ST-45
Staphylococcus aureus	CDC	*dzd-cdc-478*	ST-45
Clostridioides (Clostridium) difficile	ATCC	*dzd-atcc-50*	ST-46
Clostridioides (Clostridium) difficile	ATCC	*dzd-atcc-51*	ST-46
*Clostridioides** (Clostridium) difficile*	ATCC	*dzd-atcc-52*	ST-11

*MLST identified as ST-111 (confirmed by CDC and epiXact bioinformatic pipeline)
**Limited by sourcing paired MLST *C. difficile* isolates.

epiXact® Performance

Clonality calls were made for a total of 472 pairs of samples. Across the 5 species, **epiXact achieved 100% analytical sensitivity and 98.5% analytical specificity in determining clonality** (Table 1), and 100% repeatability. High accuracy was also achieved on species identification (100%) and MLST determination (98.2%), which are also automatically computed by the epiXact pipeline. There were no false negatives (clonal samples predicted as non-clonal by epiXact), and 6 false positives (non-clonal samples predicted as clonal by epiXact). All false positive calls were between libraries originating from two isolates in the *E. coli* dataset (dzd-cdc-78 and dzd-cdc-88 from the CDC AR Isolate Bank), having the same MLST. We retrieved from the NCBI SNP tree viewer the pre-computed SNP distance on the publicly available assembled genomes of these two samples and found that the two assemblies only differ by 12 SNPs, indicating these samples to in fact be CLONAL. This suggests epiXact was able to successfully identify an unmarked case of clonal pairs in the CDC AR Isolate Bank dataset (communicated to CDC). For the sake of concordance with the original validation plan, we consider these isolates non-clonal when computing performance statistics. The *in silico* validation similarly yielded highly accurate results with a mean absolute error in SNP distance of 2.6 SNPs.

Table 4: Summary of epiXact performance across large scale validation including *in silico* simulated dataset.

Test Output	Metric	Value
Clonality	**Analytical Sensitivity**	100.00%
Clonality	**Analytical Specificity**	98.53%
Clonality	**Accuracy**	98.73%
Clonality	**Precision (Repeatability/Variability)**	100.00%
Species Identification	**Accuracy**	100.00%
MLST Assignment	**Accuracy**	98.2%
***in silico* Clonality**	**Average Positive Predictive Value**	100.00%
***in silico* Clonality**	**Average Recall**	100.00%
***in silico* SNP Distance**	**Mean Absolute Error**	2.6 SNPs

Conclusions

Overall, epiXact demonstrated high accuracy for determining clonality between bacterial isolates. This CLIA validated pipeline may therefore be a highly relevant tool for infection control, enabling automated WGS analysis of pathogens suspected of hospital transmission.

References

Rhee et al. Clinical Infectious Diseases 2022 March.
Zellmer et al. Clinical Infectious Diseases 2021 June, Vol 72, Issue 11.
Shenoy et al. Infection Control & Hospital Epidemiology 2020 May;41(5):531-538
DeFilipp et al. The New England Journal of Medicine 2019 October.

Contact information:

Mohamad Sater, PhD mohamad@dayzerodiagnostics.com
epiXact Team epiXact@dayzerodiagnostics.com
URL https://dayzerolab.com/