Evaluation of the BIOMÉRIEUX EPISEQ® Software for whole-genome multi-locus sequence typing-based (wgMLST) Bacterial Strain Typing

Suzane Silbert[1], PhD; Amorce Lima[1], PhD
[1]Tampa General Hospital, Tampa, FL
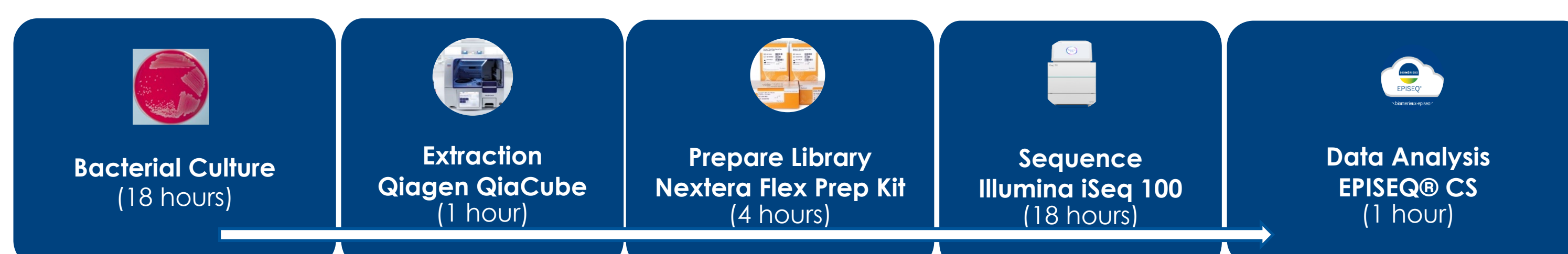Contact: ssilbert@tgh.org

BACKGROUND

- Whole-genome sequencing (WGS) technologies have revolutionized our ability to identify, track, and manage hospital-acquired infections (HAIs) outbreaks.
- bioMérieux EPISEQ® CS is a novel, fully automated bioinformatics tool designed for routine use in the clinical setting for WGS-based bacterial strain typing.

OBJECTIVE

- To compare the bioMérieux EPISEQ®CS and the DiversiLab System results to four groups of different species commonly isolated in hospital-acquired infections

METHODS

- A total of 72 characterized bacterial strains were included in this study: MRSA (n=30), *C. difficille* (n=15), *P. aeruginosa* (n=17), *A. baumannii* (n=10).
- Reference sequences for *E. coli* ATCC 25922 and *S. aureus* ATCC 25923 were downloaded from NCBI GenBank.
- Bacteria DNA was extracted using the QIAcube (Qiagen, Hilden, Germany).
- DNA libraries were prepared using the Illumina Nextera Flex kit and sequenced on the Illumina iSeq 100.
- FASTQ files containing sequencing data were uploaded into EPISEQ® CS for sequencing assembly, QC metrics analysis, generation of dendrogram, and similarity matrix. The accuracy of the software was assessed based on comparison of previously sequenced and analyzed ATCC strains.
- Reproducibility was established based on wgMLST similarity scores, MLST alleles and resistance markers using *A. baumannii* (n=3), *C. difficile* (n=4), *P. aeruginosa* (n=-5) and *S. aureus* (n=4) isolates in duplicate.
- The analytical specificity was assessed by analyzing in silico mixed and contaminated clinical samples.
- Bacterial clustering resulted from wgMLST using EPISEQ® CS was compared with those from rep-PCR DiversiLab.

Bacterial Culture (18 hours) → Extraction Qiagen QiaCube (1 hour) → Prepare Library Nextera Flex Prep Kit (4 hours) → Sequence Illumina iSeq 100 (18 hours) → Data Analysis EPISEQ® CS (1 hour)

RESULTS

Figure 1: Cluster analysis comparison between EPISEQ® CS and DiversiLab using Sankey diagram

A. baumannii (DiversiLab v. EPISEQ® CS)   C. difficile (DiversiLab v. EPISEQ® CS)   P. aeruginosa (DiversiLab v. EPISEQ® CS)   S. aureus (DiversiLab v. EPISEQ® CS)
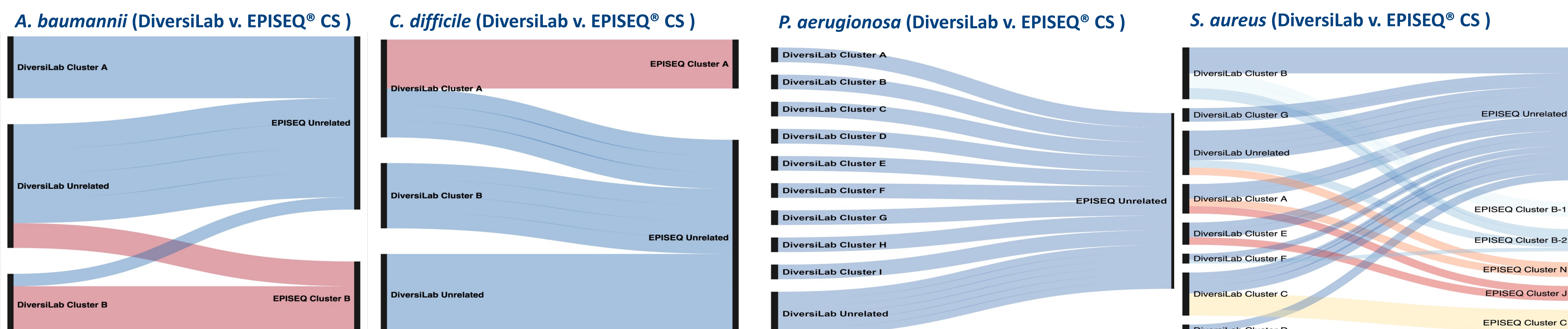
Table 1.

Organisms	Similarity Score Cutoff		Isolates	EPISEQ® CS Clusters (n)		*DiversiLab Clusters (n)		Simpson's Diversity Index	
	EPISEQ® CS	DiversiLab		Unrelated	Related	Unrelated	Related	EPISEQ® CS	DiversiLab
A. baumannii	≥98.70%	94.50%	10	7	B (n=3)	5	A (n=2); B (n=3)	0.933	0.911
C. difficile	≥99.10%	94.50%	15	12	A (n=3)	5	A (n=5); B (n=4)	0.971	0.800
P. aeruginosa	≥99.89%	94.50%	17	17	0	6	A (n = 2); B (n = 2) C (n = 2); D (n = 2) E (n = 2); F (n = 2) G (n = 2); H (n = 2); I (n = 2)	1	0.978
S. aureus	≥98.68%	94.50%	30	18	B-1 (n=2); B-2 (n=3) C (n=3); J (n=2); N (n=2)	6	A (n = 4); B (n = 9) C (n = 8); D (n = 2) E (n = 3); F (n = 3); G (n = 3)	0.963	0.886

Table 2. Reproducibility

Organism	*A. baumannii*	*C. difficile*	*P. Aeruginosa*	*S. aureus*	Overall
Isolates	3	4	5	5	17
Similarity Score(Species Specific)	99.96 ± 0.035	99.99 ± 0.015	97.36 ± 1.262	99.99 ± 0.020	99.16 ± 1.416
MLST Alleles	19/21	28/28	29/35	35/35	93.28% (111/119)
Resistant Marker Alleles	57 Alleles (342/342)	12 Alleles (96/96)	58 Alleles (422/464)	48 Alleles (480/480)	96.96% (1340/1382)
Number of Core Loci	1393	1999	1480	2117	N/A
Total Number of Loci	5633	8745	15143	3904	N/A

Table 3. Accuracy

Organism	ATCC Strain	Similarity to Reference	Assembly Length (Mb)	Core Loci Present
E. coli (ASM74325v1)	ATCC 25922	N/A	5.2	99.50%
E. coli A	ATCC 25922	99.85%	5.14	99.50%
E. coli B	ATCC 25922	99.78%	5.14	99.50%
S. aureus (ASM75620v1)	ATCC 25923	N/A	2.81	98.30%
S. aureus A	ATCC 25923	99.85%	2.77	98.30%
S. aureus B	ATCC 25923	99.85%	2.77	98.30%

CONCLUSIONS

- Differences in bacterial clustering was observed between EPISEQ® CS and DiversiLab.
- The differences were expected as one method is based on the whole bacterial genome and the other on PCR analysis of a few genomic regions of the bacteria.
- Built-in genome assembly, data analysis functions, and quality control metrics highlight some of the important features of EPISEQ® CS and decrease its barriers to implementation.
- EPISEQ® CS enables a comprehensive wgMLST analysis and provides a more reliable method for bacterial strain typing.

TGH Tampa General Hospital   BIOMÉRIEUX ANTIMICROBIAL STEWARDSHIP CENTER OF EXCELLENCE