illumina®

Accurate surveillance of healthcare-acquired infections with bacterial genome sequencing

Comprehensive isolate discrimination and characterization through microbial WGS paired with the user-friendly bioMérieux EPISEQ CS software is a reliable method for outbreak detection.

Introduction

Healthcare-acquired infections (HAIs) are a major healthcare concern, especially in critically ill and immunocompromised patients. The ability to prevent such infections could be facilitated by the development of standard infection control practices to identify and monitor pathogenic bacterial strains in the healthcare facility environment. Laboratory methods, such as qPCR and mass spectrometry, provide rapid identification of pathogens, allowing for timely treatment decisions. However, these methods are insufficient for tracing outbreaks or performing transmission investigations.

Next-generation sequencing (NGS) allows for the complete characterization of bacterial genomes, including information for subtyping, differentiating between isolates, or highlighting isolates in a cluster, as well as information about antimicrobial resistance (AMR) and virulence.[1] Analyzing and interpreting sequencing data quickly allows infection control personnel to respond rapidly to potential outbreaks and trace them back to the source to help prevent further transmission and infection. With a comprehensive whole-genome sequencing (WGS) workflow that can be completed in two days, the spread of pathogens responsible for HAIs can be efficiently monitored and properly managed.

Pseudomonas aeruginosa is a multidrug-resistant pathogen recognized for its ubiquity and association with serious illnesses. Drug-resistant strains have been found in ventilator-associated pneumonia and various sepsis syndromes, leading to increased mortality in hospitalized patients.[2] Through mutation and acquisition of resistance elements, these bacteria have developed their own local populations.

This application note highlights the use of the bioMérieux EPISEQ CS cloud service as part of a comprehensive WGS solution that includes Illumina library preparation and sequencing to characterize isolates of *P. aeruginosa*, one of the most widespread causes of HAIs across Europe. From different environmental sources, four strains were selected based on the similarity of their antimicrobial susceptibility patterns that makes subtyping difficult with traditional methods.

The components of this study were selected to demonstrate improvements in library preparation, sequencing, and analysis in terms of performance and ease of use, making the approach accessible to laboratories without prior NGS experience. Illumina DNA Prep* is an innovative method that enables a quick and easy workflow from extracted DNA or directly from bacterial colonies. The smallest of the Illumina sequencing instruments, the iSeq™ 100 System, can sequence up to six bacterial genomes in a single batch. This study also demonstrates how multiplexing on the iSeq 100 System enables analysis of bacteria with large genomes, like *P. aeruginosa* (genome size 5.5-7.7 Mb) with enough coverage for full characterization. bioMérieux EPISEQ CS provides automated analysis of WGS data within an hour of sequencing completion. The entire NGS workflow, from isolate to report, can be completed in under 48 hours.

Library preparation and sequencing

P. aeruginosa isolates exhibiting multiple drug resistance (MDR) were obtained from a university hospital in France between February 2004 and August 2005 from four different specimen sites (sputum, stool, mouth, or catheter). After DNA extraction, libraries were prepared using Illumina DNA Prep[3] and sequencing was performed on the iSeq 100 system using 2 × 150 bp reads (Figure 1).

* Formerly available as Nextera™ DNA Flex Library Preparation Kit

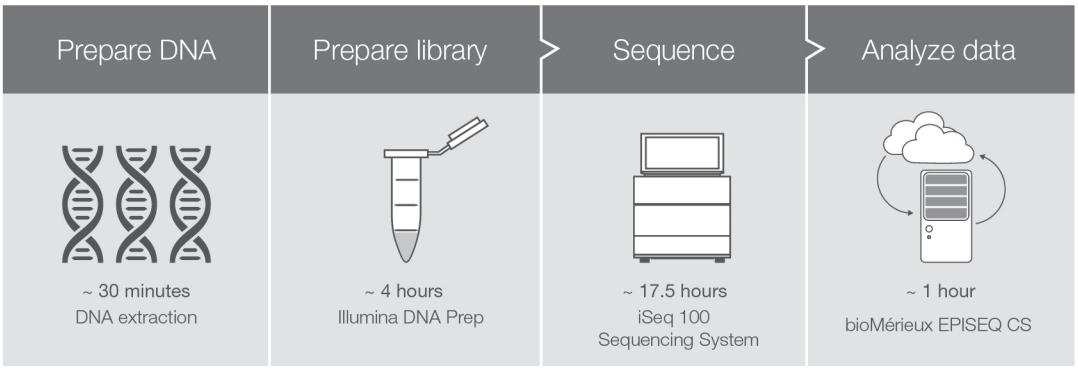

Figure 1: Bacterial whole-genome sequencing workflow—In a streamlined, comprehensive workflow, bacterial genomes can be sequenced and analyzed within two days.

Data analysis

WGS offers the highest possible resolution of closely related microbial genomes. From the raw sequence data, bioMérieux EPISEQ CS automatically performs QC analysis, *de novo* assembly, multilocus sequence typing (MLST), whole-genome MLST (wgMLST), resistome characterization, virulome characterization, and phylogenetic analysis. This whole-genome level of bacterial strain typing and characterization allows users to identify the infectious pathogen source and define transmission pathways quickly.

Accurate strain identification by MLST analysis

For determining bacterial relatedness, MLST is a common procedure that characterizes isolates of bacterial species using the sequences of internal fragments (450-500 bp) of 7-8 seven housekeeping genes. The bioMérieux EPISEQ CS software includes a curated collection of 30,000+ reference genomes belonging to a menu of 13 bacterial HAI-related species (Table 1). Using MLST data, it is observed that the allelic profile of the four analyzed samples is the same and corresponds to the *P. aeruginosa* ST235 strain from the collection in the database. At this level of resolution (MLST), it is not possible to tell if the four isolates form a cluster or not (Figure 2).

Table 1: Thirteen bacterial species that cover the vast majority of HAI-related organisms

Acinetobacter baumannii	*Klebsiella aerogenes*
Burkholderia cepacia complex	*Klebsiella oxytoca*
Clostridioides difficile	*Klebsiella pneumoniae*
Enterobacter cloacae complex	*Pseudomonas aeruginosa*
Enterococcus faecalis	*Serratia marcescens*
Enterococcus faecium	*Staphylococcus aureus*
Escherichia coli/Shigella	

Bacterial species identified by bioMérieux and analyzed using EPISEQ CS.

High resolution epidemiological analysis of isolates from wgMLST data

The bioMérieux EPISEQ CS phylogenetic analysis is based on wgMLST, a typing approach based on genome-wide gene-by-gene comparisons, including several thousand loci (~1500-4000) not limited to core genes. For hospital outbreak investigations, wgMLST can identify subtle differences often overlooked by other genotyping methods such as MLST or pulsed-field gel electrophoresis (PFGE). The wgMLST-based dendrogram of the four samples in this study demonstrates that one of the isolates is slightly more distant (99.74% similarity) from the other three (99.93% similar). The three isolates similar enough to be classified as a cluster are grouped together in the dendrogram and highlighted in red in the similarity table (Figure 3). WGS data reveals that the distinct isolate has a slightly smaller (6.66 Mb) genome than the other three samples (6.77 Mb); this 110 kb difference in genome size is enough to explain the separation in the clustering pattern (data not shown).

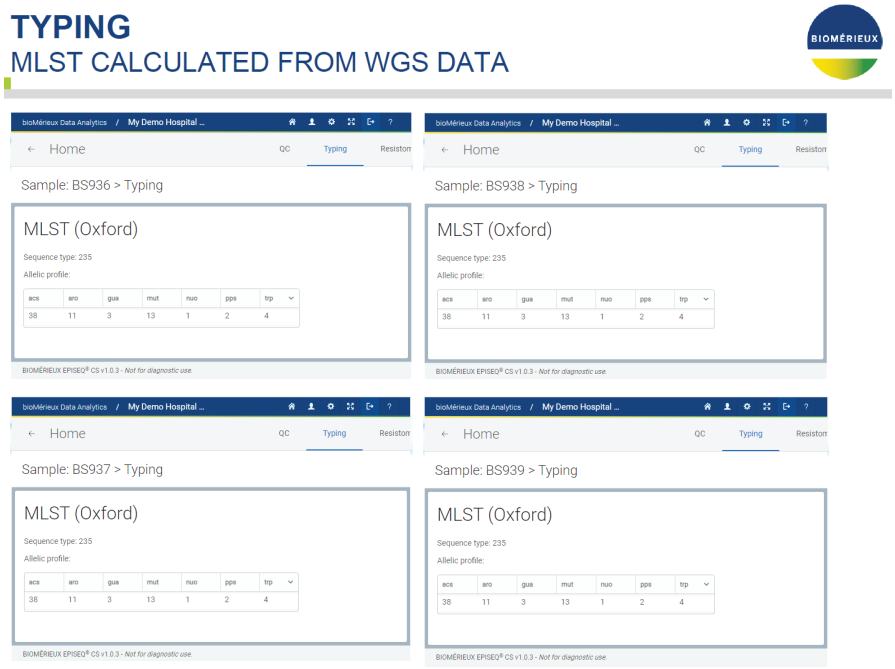

Figure 2: Report of MLST typing calculated from WGS—WGS data easily characterizes the four *P. aeruginosa* isolates as ST235 subtype.

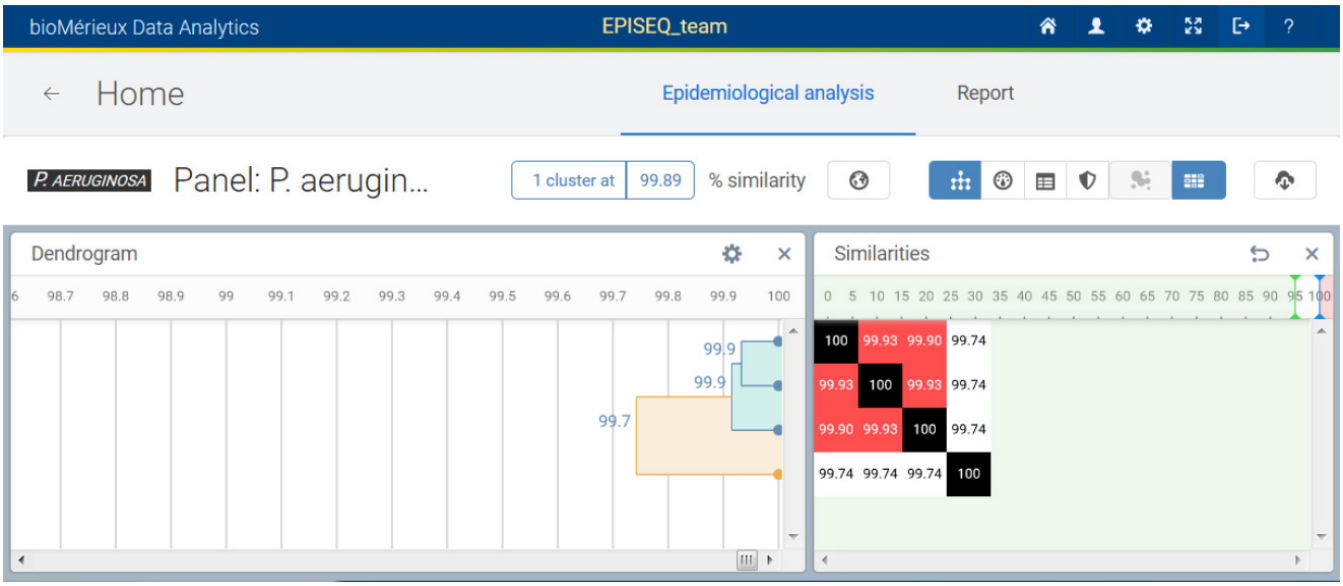

Figure 3: wgMLST dendrogram showcases sample differences—WGS data easily highlights subtle differences between isolates, providing visual clues of which isolates are similar (cyan) and which are distinct (orange). bioMérieux EPISEQ CS automatically sets thresholds for wgMLST allelic similarity percentage used to consider isolates as part of a cluster (red).

Resistome and virulome prediction

To further characterize isolates, bioMérieux EPISEQ CS provides resistome and virulome information. For genes with known mutations conferring AMR, the software showed no differences between the four strains, in agreement with their identical antibiotic resistance pattern determined by traditional phenotypic methods, such as antibiotic susceptibility testing (AST). WGS data also showed that the *SoxR* gene was present in three of the four strains, but was missing in the isolate with a smaller genome. In *P. aeruginosa*, *SoxR* can directly upregulate the expression of the MexGHI-OpmD (multidrug) efflux pump. This result suggests that the *SoxR*-deficient isolate may exhibit a different antimicrobial susceptibility under certain conditions. All other detected antibiotic resistance genetic markers found in the four strains confirmed that they are MDR organisms. For virulome, no particular virulence features were found in any of the isolates.

Conclusion

HAIs are a major healthcare concern attributing to high patient morbidity and mortality and increased healthcare costs. Using an NGS workflow with bioMérieux EPISEQ CS analysis, it is possible to identify and distinguish between closely related bacterial strains. With NGS innovations such as fast Illumina DNA Prep and the small-form iSeq 100 System, bacterial genomes can be fully sequenced and analyzed within two days.

bioMérieux EPISEQ CS expands upon traditional MLST-based methods to integrate significantly more input data. Using information available from WGS, this cutting-edge level of bacterial strain typing and characterization enables rapid identification of the source of infectious pathogens and elucidation of transmission pathways. The whole-genome data analysis process is completely automated and requires no bioinformatics expertise.

P. aeruginosa, a bacterial pathogen commonly found in hospital infections, was chosen to demonstrate the utility of an NGS workflow because of the abundance of closely related multidrug-resistant strains that are difficult to distinguish by traditional microbiological methods. The bioMérieux EPISEQ CS software rapidly confirmed previously known AST results, identified each of the four isolated strains as ST235, and further characterized one isolate as different from the three others that formed a cluster. Additional information was provided by automated features of bioMérieux EPISEQ CS, such as differences in genome size and a potential difference in the AMR profile harbored by a specific isolate.

References

1. Mirande C, Bizine I, Giannetti A, Picot N, van Belkum A. Epidemiological aspects of healthcare-associated infections and microbial genomics. *Eur J Clin Microbiol Infect Dis*. 2018;37(5):823-831.
2. Bassetti M, Vena A, Croxatto A, Righi E, Guery B. How to manage Pseudomonas aeruginosa infections. *Drugs Context*. 2018;7:212527.
3. Illumina (2018) Illumina DNA Prep Reference Guide. Accessed May 22, 2020.

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com
 Pub. no. 1270-2019-004-C QB7364

illumina®