# THE CLINICAL UTILITY OF NANOPORE SEQUENCING FOR RAPID DIAGNOSIS OF ACUTE INVASIVE INFECTION IN NORMALLY STERILE BODY SITES

## LAO HIU YIN

## PhD

## The Hong Kong Polytechnic University

## 2025

The Hong Kong Polytechnic University

Department of Health Technology and Informatics

# The Clinical Utility of Nanopore Sequencing for Rapid Diagnosis of Acute Invasive Infection in Normally Sterile Body Sites

Lao Hiu Yin

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

August 2024

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.


_____ (Signed)


\_\_\_\_Lao Hiu Yin_____ (Name of student)

# Abstract

For a considerable period, culture has been the dominant method in clinical laboratories for pathogen identification. Nonetheless, the prolonged incubation period associated with culture significantly extends the sample-to-report time. In critical medical situations, like acute invasive infections, mortality rates escalate with delays. Empirical treatment prescriptions may result in suboptimal or ineffective treatment, while the utilization of broad-spectrum antimicrobials can promote selective pressure for antimicrobial-resistant strains.

The advent of culture-independent nanopore sequencing, enabling long-read and real-time sequencing, presents a promising tool for rapid diagnosis in clinical laboratories. This study aimed to evaluate the clinical utility of nanopore sequencing for rapid diagnosis of acute invasive infections, with reference to traditional culture. Given the absence of a standardized protocol for interpreting sequencing data in clinical contexts, this study also aimed to establish a comprehensive nanopore-based sequencing workflow—from DNA extraction to data analysis. The detection threshold for distinguishing pathogens and background contaminants was determined. Furthermore, the clinical utility of nanopore targeted sequencing and unbiased metagenomic sequencing was evaluated and compared.

The performance of nanopore 16S rRNA gene sequencing (Nanopore 16S) was first evaluated by comparing with the traditional Sanger (Sanger 16S) and short-read Illumina 16S rRNA (Illumina 16S) gene sequencing in taxonomic assignment of 172 MALDI-TOF MS-unidentifiable clinical isolates. It was found that the diagnostic accuracy of Nanopore 16S was 96.36%, identical to that of Sanger 16S, and much higher than that of Illumina 16S (69.07%). Despite the lower read accuracy, sequencing the full-length 16S rRNA gene using Nanopore 16S provided better taxonomic resolution compared to the short-read Illumina 16S, which only sequenced the V3-V4

region in the commercially available kit. Additionally, the study demonstrated the potential of sequencing technologies to uncover novel species, leading to the confirmation of a novel species, *Scrofimicrobium appendicitidis*, through whole-genome sequencing and phylogenomic analysis.

Subsequently, the performance of Nanopore 16S for direct pathogen identification in normally sterile body fluids was evaluated with reference to culture results. Additionally, the performance of three analysis pipelines, including Epi2me, Emu, and NanoCLUST, was compared. Results showed that Nanopore 16S coupled with Emu demonstrated the highest concordance with the culture results. The concordance between culture and Emu was 97.7% among the 128 monomicrobial samples, compared to 85.2% for Epi2me and 79.7% for NanoCLUST. For the 230 cultured species in the 65 polymicrobial samples, Emu correctly identified 81.7% of cultured species, compared to 75.7% for Epi2me and 54.3% for NanoCLUST. To differentiate potential pathogens from background in Nanopore 16S, a threshold of relative abundance ($T_{RA}$) at 0.058 was established through ROC analysis of the monomicrobial samples. However, a threshold could not be determined for the polymicrobial samples, it was presented as a random classifier in ROC analysis. The limit of detection of Nanopore 16S was found to be 90 CFU/ml.

Nanopore targeted sequencing (NTS) and nanopore metagenomic sequencing (NMgS) workflows were developed and their performance for pathogen identification and AMR detection were compared with culture results. Of the 229 species cultured from 138 body fluids, NTS successfully identified 80.35% of the species, with 79.48% meeting the threshold of 0.058 TRA and 74.24% having a minimum of 10 classified reads. In contrast, NMgS identified 60.70% of the cultured species in the 138 body fluids, with only 41.48% of samples surpassing the threshold of 10 in Bracken. Among the 20 samples containing AMR ESKAPE pathogens, NTS detected associated AMR genes in 14 samples (70.0%). Out of the 24 AMR ESKAPE pathogens within these 20

samples, NMgS successfully identified AMR genes in association with 6 of the ESKAPE pathogens (25.0%).

In conclusion, this study demonstrated the clinical utility of Nanopore sequencing for rapid diagnosis in clinical microbiology. The diagnostic accuracy of Nanopore 16S was comparable to Sanger 16S and outperformed Illumina 16S. Moreover, the high-throughput nature of Nanopore sequencing enables direct bacteria detection from samples, bypassing the time-consuming culture process. The heightened sensitivity of Nanopore targeted sequencing renders it ideal for routine clinical microbiology diagnoses, whereas unbiased Nanopore metagenomic sequencing is advantageous in identifying infections of unknown etiology.

# Publications arising from the thesis

**Journal articles:**

**Lao, H.-Y.**, Ng, T. T.-L., Wong, R. Y.-L., Wong, C. S.-T., Lee, L.-K., Wong, D. S.-H., Chan, C. T.-M., Jim, S. H.-C., Leung, J. S.-L., Lo, H. W.-H., Wong, I. T.-F., Yau, M. C.-Y., Lam, J. Y.-W., Wu, A. K.-L., & Siu, G. K.-H. (2022). The clinical utility of two high-throughput 16S rrna gene sequencing workflows for taxonomic assignment of unidentifiable bacterial pathogens in matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Journal of Clinical Microbiology*, *60*(1). https://doi.org/10.1128/jcm.01769-21

**Lao, H.-Y.**, Wong, L. L.-Y., Hui, Y., Ng, T. T.-L., Chan, C. T.-M., Lo, H. W.-H., Yau, M. C.-Y., Leung, E. C.-M., Wong, R. C.-W., Ho, A. Y.-M., Yip, K.-T., Lam, J. Y.-W., Chow, V. C.-Y., Luk, K. S., Que, T.-L., Chow, F. W., & Siu, G. K.-H. (2024). The clinical utility of nanopore 16S rrna gene sequencing for direct bacterial identification in normally sterile body fluids. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1324494

**Lao, H.-Y.**, Wong, A., Ng, T. T.-L., Wong, R. Y.-L., Yau, M. C.-Y., Lam, J. Y.-W., & Siu, G. K.-H. (2025). *Scrofimicrobium appendicitidis* sp. nov., isolated from a patient with ruptured appendicitis. *International Journal of Systematic and Evolutionary Microbiology, 75*(1). https://doi.org/10.1099/ijsem.0.006633

**Conference presentations:**

**Lao, H.-Y.**, & Siu, G. K.-H. The clinical utility of nanopore 16S rRNA gene sequencing on identifying pathogens in direct specimens. The 33rd European Congress of Clinical Microbiology & Infectious Diseases. 15-18 April 2023. Copenhagen, Denmark.

**Lao, H.-Y.**, & Siu, G. K.-H. Optimising a nanopore-based metagenomic workflow for low microbial biomass clinical samples. The 34th European Congress of Clinical Microbiology & Infectious Diseases. 27–30 April 2024. Barcelona, Spain.

## Other publications

Tafess, K., Ng, T. T., **Lao, H. Y.**, Leung, K. S., Tam, K. K., Rajwani, R., Tam, S. T., Ho, L. P., Chu, C. M., Gonzalez, D., Sayada, C., Ma, O. C., Nega, B. H., Ameni, G., Yam, W. C., & Siu, G. K. (2020). Targeted-sequencing workflows for comprehensive drug resistance profiling of mycobacterium tuberculosis cultures using two commercial sequencing platforms: Comparison of analytical and diagnostic performance, turnaround time, and cost. *Clinical Chemistry*, *66*(6), 809–820. https://doi.org/10.1093/clinchem/hvaa092

Siu, G. K.-H., Lee, L.-K., Leung, K. S.-S., Leung, J. S.-L., Ng, T. T.-L., Chan, C. T.-M., Tam, K. K.-G., **Lao, H.-Y.**, Wu, A. K.-L., Yau, M. C.-Y., Lai, Y. W.-M., Fung, K. S.-C., Chau, S. K.-Y., Wong, B. K.-C., To, W.-K., Luk, K., Ho, A. Y.-M., Que, T.-L., Yip, K.-T., … Yip, S. P. (2020). Will a new clade of SARS-COV-2 imported into the community spark a fourth wave of the COVID-19 outbreak in Hong Kong? *Emerging Microbes & Infections*, *9*(1), 2497–2500. https://doi.org/10.1080/22221751.2020.1851146

Leung, K. S.-S., Ng, T. T.-L., Wu, A. K.-L., Yau, M. C.-Y., **Lao, H.-Y.**, Choi, M.-P., Tam, K. K.-G., Lee, L.-K., Wong, B. K.-C., Man Ho, A. Y., Yip, K.-T., Lung, K.-C., Liu, R. W.-T., Tso, E. Y.-K., Leung, W.-S., Chan, M.-C., Ng, Y.-Y., Sin, K.-M., Fung, K. S.-C., … Siu, G. K.-H. (2021). Territory wide study of early coronavirus disease outbreak, Hong Kong, China. *Emerging Infectious Diseases*, *27*(1), 196–204. https://doi.org/10.3201/eid2701.201543

Chan, C. T.-M., Leung, J. S.-L., Lee, L.-K., Lo, H. W.-H., Wong, E. Y.-K., Wong, D. S.-H., Ng, T. T.-L., **Lao, H.-Y.**, Lu, K. K., Jim, S. H.-C., Yau, M. C.-Y., Lam, J. Y.-W., Ho, A. Y.-M., Luk, K. S., Yip, K.-T., Que, T.-L., To, K. K.-W., & Siu, G. K.-H. (2022). A low-cost Taqman minor groove binder probe-based one-step RT-qPCR assay for rapid identification of N501Y variants of SARS-COV-2. *Journal of Virological Methods*, *299*, 114333. https://doi.org/10.1016/j.jviromet.2021.114333

Zhu, L., Lee, A. W.-T., Wu, K. K.-L., Gao, P., Tam, K. K.-G., Rajwani, R., Chaburte, G. C., Ng, T. T.-L., Chan, C. T.-M., **Lao, H. Y.**, Yam, W. C., Kao, R. Y.-T., & Siu, G. K. (2022). Screening repurposed antiviral small molecules as antimycobacterial compounds by a lux-based phop promoter-reporter platform. *Antibiotics*, *11*(3), 369. https://doi.org/10.3390/antibiotics11030369

Leung, K. S.-S., Tam, K. K.-G., Ng, T. T.-L., **Lao, H.-Y.**, Shek, R. C.-M., Ma, O. C., Yu, S.-H., Chen, J.-X., Han, Q., Siu, G. K.-H., & Yam, W.-C. (2022). Clinical utility of target amplicon sequencing test for rapid diagnosis of drug-resistant mycobacterium tuberculosis from respiratory specimens. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.974428

Rajwani, R., Galata, C., Lee, A. W., So, P.-K., Leung, K. S., Tam, K. K., Shehzad, S., Ng, T. T., Zhu, L., **Lao, H. Y.**, Chan, C. T.-M., Leung, J. S.-L., Lee, L.-K., Wong, K. C., Yam, W. C., & Siu, G. K.-H. (2022). A multi-omics investigation into the mechanisms of hyper-virulence in mycobacterium tuberculosis. *Virulence*, *13*(1), 1088–1100. https://doi.org/10.1080/21505594.2022.2087304

Su, J., Lui, W. W., Lee, Y., Zheng, Z., Siu, G. K.-H., Ng, T. T.-L., Zhang, T., Lam, T. T.-Y., **Lao, H.-Y.**, Yam, W.-C., Tam, K. K.-G., Leung, K. S.-S., Lam, T.-W., Leung, A. W.-S., & Luo, R. (2023). Evaluation of mycobacterium tuberculosis enrichment in metagenomic samples using ont adaptive sequencing and amplicon sequencing for identification and variant calling. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-32378-x

Lee, A. W.-T., Chan, C. T.-M., Wong, L. L.-Y., Yip, C.-Y., Lui, W.-T., Cheng, K.-C., Leung, J. S.-L., Lee, L.-K., Wong, I. T.-F., Ng, T. T.-L., **Lao, H.-Y.**, & Siu, G. K.-H. (2023). Identification of microbial community in the urban environment: The concordance between conventional culture and nanopore 16S rrna sequencing. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1164632

Ng, T. T.-L., Su, J., **Lao, H.-Y.**, Lui, W.-W., Chan, C. T.-M., Leung, A. W.-S., Jim, S. H.-C., Lee, L.-K., Shehzad, S., Tam, K. K.-G., Leung, K. S.-S., Tang, F., Yam, W.-C., Luo, R., & Siu, G. K.-H. (2023). Long-read sequencing with hierarchical clustering for antiretroviral resistance profiling of mixed human immunodeficiency virus quasispecies. *Clinical Chemistry*, *69*(10), 1174–1185. https://doi.org/10.1093/clinchem/hvad108

Wu, W.-G., Shum, M. H.-H., Wong, I. T.-F., Lu, K. K., Lee, L.-K., Leung, J. S.-L., **Lao, H.-Y**., Lee, A. W.-T., Hau, P.-T., Chan, C. T.-M., Wong, H. F.-T., Fung, S. K.-Y., Wong, S. C.-Y., Ng, I. C.-F., Ng, T. T.-L., Chow, N., Ho, A. Y.-M., Hung, M. F., Chow, F. W.-N., … Siu, G. K.-H. (2023). Probable airborne transmission of Burkholderia pseudomallei causing an urban outbreak of melioidosis during typhoon season in Hong Kong, China. *Emerging Microbes & Infections, 12*(1). https://doi.org/10.1080/22221751.2023.2204155

Tafess, K., Ng, T. T.-L., Tam, K. K.-G., Leung, K. S.-S., Leung, J. S.-L., Lee, L.-K., **Lao, H. Y.**, Chan, C. T.-M., Yam, W.-C., Wong, S. S., Lau, T. C.-K., & Siu, G. K.-H. (2024). Genetic mechanisms of co-emergence of inh-resistant mycobacterium tuberculosis strains during the standard course of antituberculosis therapy. *Microbiology Spectrum*, *12*(4). https://doi.org/10.1128/spectrum.02133-23

Lee, A. W.-T., Ng, I. C.-F., Wong, E. Y.-K., Wong, I. T.-F., Sze, R. P.-P., Chan, K.-Y., So, T.-Y., Zhang, Z., Ka-Yee Fung, S., Choi-Ying Wong, S., Tam, W.-Y., **Lao, H.-Y.**, Lee, L.-K., Leung, J. S.-L., Chan, C. T.-M., Ng, T. T.-L., Zhang, J., Chow, F. W.-N., Leung, P. H.-M., & Siu, G. K.-H. (2024). Comprehensive identification of pathogenic microbes and antimicrobial resistance genes in food products using nanopore sequencing-based metagenomics. *Food Microbiology*, *121*, 104493. https://doi.org/10.1016/j.fm.2024.104493

## Acknowledgements

I am profoundly grateful to Prof. Gilman Siu for offering me the opportunity to pursue my Ph.D. studies at The Hong Kong Polytechnic University. Throughout the four-year journey, his patient guidance and unwavering support were crucial in enabling me to complete my research. His mentorship and expertise have been instrumental in shaping my research, academic growth, and future career path.

I extend my heartfelt thanks to my teammates for their technical assistance and insightful suggestions that helped me overcome the challenges encountered during my project. I am especially grateful to Dr. Timothy Ng for his expertise in bioinformatics and for drafting the shell script for automated analysis, and to Dr. Annette Wong for constructing the complete circular chromosome and phylogenomic tree for the novel species. I also extend my thanks to Mr. Ryan Wong, Ms. Celia Wong, Ms. Lily Wong, and Mr. Thomas Hung for their assistance with DNA extraction and library preparation.

Thanks should also go to the technical staff and colleagues in the general office of the Department of Health, Technology, and Informatics. They were helpful in providing resources and assistance whenever needed, making my study journey much smoother.

I would like to extend my heartfelt gratitude to all the collaborators from Pamela Youde Nethersole Eastern Hospital, Princess Margaret Hospital, United Christian Hospital, Tuen Mun Hospital, Prince of Wales Hospital, and Kwong Wah Hospital for their invaluable contribution of samples and routine culture results to my study. I am especially grateful to the clinical microbiology laboratory at Pamela Youde Nethersole Eastern Hospital for their assistance in performing

biochemical tests for the novel species, as well as the chemical pathology laboratory at Princess Margaret Hospital for their efforts in capturing electron-microscopic images of the novel species.

Lastly, I would like to thank my family for their love, understanding, and encouragement throughout this lengthy journey. Their support allowed me to focus on my studies with peace of mind.

Here comes the end of my Ph.D. journey. This thesis represents not only my efforts but the collective support of all those who have aided me along the way.

# Table of Contents

# Abbreviations

Amplicon sequence variant (ASV)

AMPure XP beads (AXP)

Antimicrobial resistance (AMR)

Antimicrobial susceptibility test (AST)

Average nucleotide identity (ANI)

Benchmarking Universal Single-Copy Orthologs (BUSCO)

Bloodstream infections (BSIs)

Bronchoalveolar lavage fluid (BALF)

Carbapenem-resistant *Acinetobacter baumannii* (CRAB)

Carbapenem-resistant *Enterobacterales* (CRE)

Carbapenem-resistant *Klebsiella pneumoniae* (CRKP)

Carbapenem-resistant *Pseudomonas aeruginosa* (CRPA)

Cell-free DNA (cfDNA)

Cerebrospinal fluid (CSF)

Central nervous system (CNS)

Centre for Health Protection (CHP)

Clinical and Laboratory Standards Institute (CLSI)

Coding sequences (CDSs)

Computed tomography (CT)

Extend Spectrum Beta-Lactamase-producing *Enterobacterales* (ESBL-PE)

Genome Taxonomy Database (GTDB)

High throughput sequencing (HTS)

Internal transcribed spacer (ITS)

Invasive bacterial infections (IBIs)

Latex agglutination test (LAT)

Melting temperatures (Tm)

Metagenomic next-generation sequencing (mNGS)

Methicillin-resistant *Staphylococcus aureus* (MRSA)

Mobile genetic elements (MGEs)

Minimal inhibitory concentration (MIC)

Nanopore metagenomic sequencing (NMgS)

Nanopore targeted sequencing (NTS)

National Center for Biotechnology Information (NCBI)

Next-generation sequencing (NGS)

Nucleic acid amplification test (NAAT)

Orthologous Average Nucleotide Identity Tool (OAT)

Oxford Nanopore Technology (ONT)

Penicillin-binding proteins (PBPs)

Polymerase Chain Reaction (PCR)

Positive predictive value (PPV)

Real-time PCR (qPCR)

Receiver operating characteristic (ROC)

Single molecule real-time (SMRT)

Standard deviation (SD)

Staphylococcal cassette chromosome mec (SCCmec)

Super accurate (SUP)

The Global Burden of Disease (GBD)

Third-generation sequencing (TGS)

Vancomycin-resistant *Enterococci* (VRE)

Vancomycin-resistant *Staphylococcus aureus* (VRSA)

Whole-genome sequencing (WGS)

World Health Organization (WHO)

# Chapter 1: General introduction

## 1.1 Background and significance of rapid pathogen identification in acute invasive infections

A pathogen is commonly defined as a microorganism that causes disease to its host, which can originate from diverse groups, including bacteria, viruses, fungi, or parasites (1). According to the World Health Organization (WHO), infection is one of the top ten leading causes of death in 2019 (2). Bacteria are the most common infectious agents and have raised most concern globally. According to a systematic analysis for the global burden of disease (GBD) study conducted by the Lancet, there were 13.7 million infection-related deaths globally in 2019 and 7.7 million of the deaths were associated with 33 common bacterial pathogens (3). Infections associated with these bacteria represented the second leading cause of death in 2019, which comprised 13.6% of all global deaths (3). In addition, the emergence of drug-resistant bacteria poses a significant threat to public health. There were 4.95 million deaths associated with bacterial antimicrobial resistance (AMR), and 1.27 million deaths were specifically attributed to bacterial AMR (4).

Rapid pathogen and AMR identification in clinical labs is crucial for optimizing treatment and reducing the mortality rate, especially in cases of acute invasive infections. Invasive bacterial infections (IBIs) refer to the isolation of microorganisms from a normally sterile body fluid such as blood, cerebrospinal fluid (CSF), pleural fluid, pericardial fluid, joint fluid, peritoneal fluid, or deep tissue abscess (5). In the case of acute IBIs, the pathogens invade rapidly and cause damage to the normally sterile body sites. More importantly, acute IBIs have the capacity to spread beyond the initial site of infection and affect multiple organs. In addition, severe microbial infections can lead to sepsis, a life-threatening condition characterized by a dysregulated host response to infection (6), with a mortality rate of 20-50% (7) and the potential for multiple organ dysfunctions.

The emergence of antibiotic resistance has become a significant global concern in recent years (8). One group of particular concern is the ESKAPE pathogens, which include *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species (8, 9). Some studies include *Enterobacterales* instead of *Enterobacter* species in the acronym to encompass other gram-negative enteric pathogens, such as *Escherichia coli*, that are commonly associated with drug resistance (10, 11). These pathogens are notorious for their ability to "escape" the effects of multiple antibiotics. The emergence and spread of antibiotic resistance among ESKAPE pathogens pose a significant challenge in clinical settings, limiting treatment options and contributing to increased morbidity, mortality, and healthcare costs (11, 12).

Commonly, patients with acute IBIs are prescribed with empirical treatment since it takes at least 24 to 48 hours to obtain culture results in clinical laboratories (13). However, the unnecessary use of broad-spectrum antibiotics due to the lack of rapid diagnostic methods for bacterial identification and AMR detection can cause selective pressure on commensal bacteria, increasing the risk of secondary infections (14). Also, patients infected with drug-resistant bacteria are more susceptible to inadequate empirical treatment (15), which is associated with higher mortality rate (15-17). A study by Kumar et al. showed that every hour of delay in effective antimicrobial treatment increases the mortality rate by 7.6% in patients with sepsis (18). In the case of bloodstream infections, the mortality rate due to the delay of effective treatment increased with time, from 15.9% after a 2-day delay to 57% after a 5-day delay (19). Another systematic review study also showed that a delay of >6h in administration of appropriate antibiotics conferred an 8.4-fold increase in mortality from meningitis (20). Therefore, early and accurate identification of

pathogens and their AMR profile in acute IBIs is essential for providing appropriate treatment and improving patient outcomes.

## 1.2 Antimicrobial resistance and AMR genes commonly associated with ESKAPE pathogens

ESKAPE pathogens, prioritized by the WHO for multidrug resistance, are frequently associated with nosocomial infections, presenting a substantial threat to public health due to the limited choice of treatment (21, 22). The pathogens develop antimicrobial resistance through genetic mutation and the acquisition of mobile genetic elements (MGEs) (9). MGEs allow transfer of AMR genes between different species of bacteria, further exacerbating the issue. The most prevalent superbugs worldwide include carbapenem-resistant *Enterobacterales* (CRE), carbapenem-resistant *Klebsiella pneumoniae* (CRKP), methicillin-resistant *Staphylococcus aureus* (MRSA), Extend Spectrum Beta-Lactamase-producing *Enterobacterales* (ESBL-PE), vancomycin-resistant *Enterococcus* (VRE), carbapenem-resistant *Pseudomonas aeruginosa* (CRPA), and carbapenem-resistant *Acinetobacter baumannii* (CRAB) (23, 24).

### 1.2.1 Resistance to β-lactams

β-lactam is the most frequently used class of antimicrobial agent (24-26). The first identified β-lactam and antibiotic, penicillin, was discovered in the *Penicillium* mold in 1928 and successfully isolated in 1939 (27). The β-lactams are signature for carrying a β-lactam ring in their structures. They act on the bacteria by binding to the penicillin-binding proteins (PBPs), which is a transpeptidase that is responsible for the crosslinking of peptidoglycan in bacterial cell wall

synthesis (25). The weakened cell wall is more susceptible to the osmotic stress and eventually leads to bacterial cell death (28). However, resistance to β-lactams emerged rapidly after the release of penicillin for clinical use, the plasmid-mediated β-lactamases facilitate the spread of β-lactamases among various bacterial species (29).

### 1.2.1.1 The mecA gene

The mecA gene encodes an altered penicillin-binding protein 2a (PBP2a), which has low binding affinity to most β-lactam antibiotics. It is commonly found in the methicillin-resistant *Staphylococcus aureus* (MRSA). The prevalence of mecA in clinically isolated *Staphylococcus aureus* ranges from 38% -70.2% (30-33), while the prevalence of mecA in clinically isolated MRSA ranges from 90.2% to 97.3% (31, 34, 35). In the case of Hong Kong, there is an increasing trend of MRSA. The report of Centre for Health Protection (CHP) showed that the prevalence of MRSA in clinically isolated *S. aureus* increased from 37.0% in 2008 to 40.9% in 2023 (36).

### 1.2.1.2 β-lactamases

β-lactamase is a class of enzyme produced by bacteria that hydrolyzes the β-lactam ring in β-lactam antibiotics. β-lactamases can be divided into classes A, B, C, and D based on Ambler molecular classification. Classes A, C, and D are serine β-lactamases while class B are metallo-β-lactamases which contain serine and zinc in the active site respectively (37). Additionally, β-lactamases can be divided into three main groups by the Bush-Jacobi-Medeiros functional classification in which the β-lactamases are classified based on the similarity in the substrates and

inhibitors profiles (38). The classification of some common β-lactamases was shown in Table 1-1.

**Table 1-1: The classification of some common β-lactamases**

| Ambler Classification | active site | Enzymes | Substrates | Inhibitors | Bush-Jacobi-Medeiros classification |
|---|---|---|---|---|---|
| Class A | serine | TEM-1, TEM-2, SHV-1 | Penicillins, early cephalosporins | β-lactamase inhibitors | Group 2b |
| | | ESBLs: TEM, SHV, CTX-M families | Penicillins, extended spectrum cephalosporins and monobactams | β-lactamase inhibitors | Group 2be |
| | | Carbapenemases: KPC | Carbapenems | β-lactamase inhibitors | Group 2f |
| Class B | zinc | subclass B1: IMP, VIM, NDM | Carbapenems | EDTA | Group 3a |
| Class C | serine | AmpC family | Penicillins, extended spectrum cephalosporins and monobactams | none | Group 1 |
| Class D | serine | ESBLs: OXA family | Penicillins, extended spectrum cephalosporins and monobactams | β-lactamase inhibitors | Group 2 |
| | | Carbapenemases: OXA-48 | Carbapenems | β-lactamase inhibitors | Group 2df |

## 1.2.1.3 Extended-spectrum β-lactamases (ESBLs)

ESBLs are commonly defined as the β-lactamases that confer resistance to the penicillins, first to third generations of cephalosporins, and monobactams, but their activities can be inhibited by the β-lactamase inhibitors (38, 39). ESBLs were first identified as the derivatives of the plasmid-encoded narrow-spectrum beta-lactamases (TEM-1, TEM-2, or SHV-1), the slight alteration of

amino acid sequences in the derivatives significantly enhances the enzymatic activity (40). Other well-known ESBLs include Cefotaximase-Munich (CTX-M)- and oxacillinase (OXA)-type ESBLs. Currently, the CTX-M-type ESBLs, in particular CTX-M-15, are the most frequently identified ESBL type worldwide (39, 41, 42). The *Enterobacteriaceae* family is the main group of ESBL-producing organisms, in particular *Escherichia coli* and *Klebsiella pneumoniae* (39, 40). A recent systematic review showed that pooled prevalence of ESBL-producing *E. coli* and *K. pneumoniae* in hospitals worldwide were 49% and 45% respectively (43). The global prevalence of blaCTX−M, blaTEM, blaSHV and blaOXA genes in human-related ESBL-producing *E. coli* and *K. pneumoniae* was 35.9%, 32.3%, 31.7%, and 2.6%, respectively (43). In Hong Kong, there is an increasing trend of ESBL-PE (44). According to CHP, the prevalence of ESBL-producing *E. coli* increased from 20.6% in 2008 to 26.9% in 2023 (45), while the prevalence of ESBL-producing *Klebsiella* strains increased from 15.6% in 2008 to 21.4% in 2023 (46). The predominant genotype among ESBL-producing *E. coli* in Hong Kong is blaCTX-M, accounting for over 90% of the ESBL-producing *E. coli* strains (44, 47).

### 1.2.1.4 AmpC β-lactamases

The AmpC β-lactamases belong to class C or group 1 in Ambler and Bush–Jacoby–Medeiros classification systems, respectively. Similar to the ESBLs, AmpC β-lactamases can hydrolyze the penicillins, first to third generations of cephalosporins, and monobactams. However, the activity of AmpC β-lactamases is not affected by the ESBLs inhibitors (48, 49). Although AmpC β-lactamases were initially identified in the chromosomes of certain gram-negative bacteria such as *Enterobacter cloacae* and *Citrobacter freundii*, plasmid-mediated AmpC (pAmpC) β-lactamases were later reported in many studies (50). The hyperexpression of chromosomally encoded AmpC

(cAmpC) β-lactamases can be either induced by the presence of certain β-lactams or resulted from the mutations in the regulatory genes (51). In contrast, the pAmpC β-lactamases are constitutively expressed and the blaAmpC genes can be transferred to bacteria lacking cAmpC genes (51). While the most frequently identified pAmpC gene worldwide is blaCMY-2 (50), some other common pAmpC β-lactamases include FOX, MOX, DHA, MIR, and ACT (38, 50, 52).

## 1.2.1.5 Carbapenemases

Carbapenemases are the β-lactamases that confer resistance to most β-lactam antibiotics, including penicillins, cephalosporins, monobactams, and carbapenems (53). The emergence of carbapenemases raises great concern worldwide since the broad spectrum carbapenems possess great potency against both gram-positive and gram-negative bacteria and are considered last-line antibiotics (54). The most common and widespread carbapenemases include *Klebsiella pneumoniae* carbapenemase (KPC), OXA-type carbapenemases OXA-48, Verona Integron-encoded Metallo-β-lactamase (VIM), Imipenemase (IMP), and New Delhi metallo-β-lactamase (NDM), which are mainly disseminated in *K. pneumoniae* and *E. coli* (55, 56). In Hong Kong, the prevalence of carbapenem-resistant *E. coli* and *Klebsiella* strains in 2023 was 0.6% and 1.7%, respectively (57, 58). In addition to carbapenem-resistant *Enterobacteriaceae* (CRE), carbapenem-resistant *Pseudomonas aeruginosa* (CRPA), and carbapenem-resistant *Acinetobacter baumannii* (CRAB) are also categorized as critical priority bacteria by WHO (59). The prevalence of CRAB was much higher than that of CRE in Hong Kong, 47.8% of clinically isolated *Acinetobacter* was resistant or intermediate to carbapenem, according to the statistics from CHP in 2023 (60). In the study by Cheng et al., the prevalence of CRPA in Hong Kong was approximately 7% (44).

## 1.2.2 Resistance to vancomycin

Vancomycin, a glycopeptide antibiotic, was initially isolated in 1953 from a soil bacterium, *Amycolatopsis orientalis*, collected in Borneo by Dr. Edmund Kornfeld (61). Vancomycin is effective against most gram-positive bacteria and is commonly used to treat multidrug-resistant bacteria, including MRSA (61-63). It exerts its bactericidal effect through the inhibition of cell wall synthesis in bacteria. By binding to the D-Ala-D-Ala terminus of the bacterial cell wall, vancomycin blocks the cross-linking of the peptidoglycan strands, which in turn weakens the bacterial cell wall, leading to cell lysis and death (61). The emergence of vancomycin-resistant bacteria, particularly vancomycin-resistant *Enterococci* (VRE), poses a significant challenge in the treatment of gram-positive bacterial infections.

Vancomycin resistance is most commonly observed in *Enterococcus faecium*, followed by *Enterococcus faecalis*. The prevalence of VRE varies across regions. In Asia, the pooled prevalence of VRE was 8.10%, with resistance rates of 22.40% in *E. faecium* strains and 3.70% in *E. faecalis* strains (64). In the United States, 82.2% of *E. faecium* strains recovered from bloodstream infections were vancomycin-resistant, in contrast to only 9.8% of vancomycin-resistant *E. faecalis* strains (65). In Europe, the pooled prevalence of VRE among all enterococcal isolates from patients with nosocomial infections was 7.3% (66), and the rate of vancomycin resistance among *E. faecium* isolates from blood cultures is 17.3% (67). According to the report from CHP, the prevalence of VRE among clinically isolated *Enterococci* in Hong Kong was 0.9% in 2023 (68). Addressing enterococcal infections presents a challenge in treatment due to their intrinsic resistance to several critical antimicrobial agents such as cephalosporin, lincomycin, and cotrimoxazole, alongside their restricted susceptibility to penicillin and aminoglycosides (69). The

development of additional antimicrobial resistance in *Enterococci* through acquisition of plasmids or transposons, particularly vancomycin resistance, greatly constrains therapeutic strategies.

Vancomycin resistance in bacteria primarily occurs through modifications in the D-Ala-D-Ala binding site on peptidoglycan precursors. This resistance can arise from eight distinct van gene clusters (vanA, vanB, vanD, vanE, vanG, vanL, vanM, and vanN), with vanA and vanB being the most prevalent worldwide (62, 70). Both vanA and vanB gene clusters synthesize peptidoglycan precursors with a D-Ala-D-lactate terminus, significantly decreasing their affinity for vancomycin (71). VanA-resistant strains exhibit high levels of resistance to vancomycin and teicoplanin, whereas vanB-resistant strains show varying degrees of resistance to vancomycin but remain susceptible to teicoplanin (71). Furthermore, successful transmission of vancomycin resistance genes from vancomycin-resistant *Enterococci* (VRE) to *Staphylococcus aureus* have been reported, leading to the emergence of vancomycin-resistant *S. aureus* (VRSA) (72). This transmission underscores the potential for the spread of resistance mechanisms between VRE and MRSA, complicating treatment strategies and necessitating surveillance and control measures in healthcare settings.

## 1.3 Overview of traditional diagnostic methods for IBIs and their limitations

For suspected IBI cases, the body fluids from the suspected body sites will be collected and transferred to clinical microbiology laboratories for diagnosis. Conventionally, pathogen identification in microbiology laboratories includes four main techniques: microscopy, antigen detection, serology, and culture (73, 74).

### 1.3.1 Microscopic examination

Microscopic examination of direct clinical specimens is an inexpensive method for rapid identification of pathogens such as bacteria, fungi, and parasites. It is usually the first step in pathogen identification (75), presumptive diagnosis can be made by observing the cellular morphologies and physical characteristics of the microorganisms. Gram staining is commonly used in microscopic examination of body fluids (76), which detects and classifies bacterial species into gram-positive and gram-negative bacteria based on their differences in cell wall composition (77). While gram-positive bacteria have a thick peptidoglycan layer within the cell wall, gram-negative bacteria are characterized by a thin peptidoglycan layer in cell wall and are surrounded by an outer membrane containing lipopolysaccharide (78). The thick peptidoglycan layer allows gram-positive bacteria to retain the primary stain crystal violet, resulting in a purple color in cell wall. In contrast, gram-negative bacteria lose the primary stain during decolorization, resulting in a pink color from the counterstain safranin (79). Body fluids received in large volume are concentrated through centrifugation before Gram staining and culture to increase the sensitivity (80). The Gram staining result gives preliminary clue about the pathogens and allows initial evaluation of the choice of empirical treatment (76). However, the sensitivity of microscopic examination for microbial detection is low, and further confirmation is usually required for taxonomic identification.

### 1.3.2 Antigen detection and serology

Antigen detection and serology are two techniques used in microbiology to diagnose specific pathogens in patient samples. While antigen detection tests directly detect specific microbial antigens in clinical samples, serology detects antibodies produced by the host's immune system in

response to the pathogen (73). Antigen detection tests provide a relatively narrow diagnostic window, as they are most effective during the active phase of infection when the antigen is present in detectable concentrations. In contrast, serological tests offer a broader diagnostic window due to the longer persistence of antibodies in the bloodstream, which facilitates the diagnosis of both acute and past infections, disease surveillance, and vaccination monitoring (81).

Immunoassay is a technique commonly used in antigen and antibody detection and quantification. It is based on the recognition and binding of target molecules by specific antibodies or antigens, offering rapid detection of specific infections with high specificity. Latex agglutination test (LAT) is commonly used for detection of the presence of antigens of specific pathogens, including *Haemophilus influenzae* type b, *Streptococcus* group B, *Streptococcus pneumoniae*, *Neisseria meningitidis* (groups A, B, C, Y or W135) and *Escherichia coli* K1, in CSF for diagnosis of acute bacterial meningitis (82, 83). The latex particles are coated with antigen-specific antibodies, agglutination will be observed if the target antigens are present (83).

Although the culture-independent immunoassays offer rapid detection of microorganisms, there are still some limitations. Immunoassays may not be effective during the early stages of infection when the antibodies or the pathogen-specific antigens have not yet reached detectable levels, which may result in false-negative results. Interferences in immunoassays, such as cross-reactivity with non-targeted molecules, may lead to a misinterpretation of the result (84). The sensitivity and specificity of LAT in bacterial meningitis detection reported in various studies ranged from 60% to 93% (85). Also, study showed that the sensitivity of LAT to detect bacterial antigens in culture-negative CSF was only 7% (86), other high sensitivity diagnostic tests such as molecular tests are required for the diagnosis of culture-negative bacterial meningitis.

### 1.3.3 Culture-based methods

Currently, culture is still the most widely used method and considered the gold standard for pathogen identification in clinical microbiology (87, 88). Normally sterile body fluids other than blood are inoculated onto multiple solid media, mainly aerobic and anaerobic blood agar, chocolate agar, and MacConkey agar plates, accompanied with an enrichment broth like thioglycolate broth for cultivation and isolation of pathogens (89, 90). Once isolated colonies are obtained on solid media, bacterial identification is performed using matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS), which identifies bacterial by comparing the peptide mass fingerprint (PMF) generated from the sample with the reference database (91). MALDI-TOF MS relies on pure culture for accurate identification since polymicrobial samples will cause mixed PMF. After bacterial identification, an antimicrobial susceptibility test (AST) is performed to obtain the drug resistance profile of the isolated pathogens.

For bloodstream infections (BSIs), blood culture is the gold standard for detection of bacteremia (92). Due to the low bacterial load in blood samples, which mainly falls within 1-100 CFU/ml, blood culture is required to enrich the bacteria before inoculated on solid media for isolation (93). Approximately 20ml of blood is inoculated in aerobic and anaerobic blood culture bottles respectively and incubated in an automated blood culture system such as BD BACTEC™ FX Blood Culture System (94). The automated blood culture system continuously detects the increase in $CO_2$ generated by growing bacteria, culture-positive samples are flagged in the system. Gram staining is then performed to confirm the presence of bacteria in positive blood cultures and provide preliminary diagnosis of the etiology of BSIs (94). After that, positive blood culture samples are inoculated on solid media to obtain isolated colonies for the MALDI-TOF MS-based

identification and AST. Apart from blood samples, the automated blood culture system has been used to culture other types of normally sterile body fluids (89).

One of the main limitations of culture is the long incubation time, resulting in a long sample-to-report time, which may affect the patient outcomes. In general, it takes 24 to 48 hours to obtain isolated colonies on solid media. The incubation time is even longer in blood culture, it may take up to 5 days to obtain positive results (93), and another 48 hours for isolating the bacteria from solid media and AST. In addition, blood culture is susceptible to contamination, normal flora on skin may be introduced in blood samples during blood taking (95). Besides, false negative results may be obtained due to the presence of fastidious or uncultivable bacteria, or prescription of empirical antibiotics (96, 97). Also, culture tends to overlook anaerobic bacteria, especially in the presence of mixed culture (98).


## 1.4 Overview of traditional antimicrobial susceptibility test (AST) and their limitations

One of the earliest AST methods is the broth dilution test, which involves the cultivation of bacterial isolates in a series of twofold dilutions of antimicrobial agents in liquid growth media in order to determine the minimal inhibitory concentration (MIC) (99). While the miniaturization of broth dilution test through the use of microdilution panels has improved its cost-effectiveness and feasibility, the preparation process can be laborious and time-consuming. This is due to the requirement of dispensing bacterial suspensions into multiple dilutions of a single antimicrobial agent. Additionally, it is important to note that each panel allows testing only one sample at a time. Similar to the broth dilution test, agar dilution test also determines the MIC of antimicrobial agents. However, agar plates with different concentrations of the antimicrobial agent are used instead of liquid culture media. An advantage of agar dilution test over broth dilution test is that multiple

bacterial suspension can be inoculated on the same set of agar plates simultaneously for MIC determination (100). Nevertheless, these labor-intensive ASTs are not frequently performed in clinical laboratories.

Disc diffusion test is the most widely used AST in routine clinical laboratories. It involves placing paper discs with desired concentration of antimicrobial agents onto an agar plate inoculated with the test microorganism. The antimicrobial agent diffuses into the agar gradually during incubation, creating a concentration gradient. The diameters of inhibition growth zones are measured and compared with reference breakpoints for the interpretation of susceptibility (101). However, definite MIC of antimicrobial agents could not be determined by this qualitative method. A combination of dilution method and diffusion method is the antimicrobial gradient method (Etest) (101). The Etest strips, which contain a concentration gradient of the antimicrobial agent, are placed on an agar plate inoculated with the test microorganism. MIC is determined by observing the intersection of the elliptical zone of inhibition with the Etest strip after incubation. Nonetheless, this approach will be costly if numerous antimicrobial agents are tested (99).

The advent of automated AST systems significantly reduces the hands-on time and incubation time for the testing. The most commonly used automated AST systems in clinical laboratories include VITEK®2 from bioMérieux, MicroScan WalkAway from Beckman Coulter, and Phoenix from BD Diagnostics (102, 103). These automated systems allow both pathogen identification and AST for clinical isolates. While modified biochemical tests with colorimetric or fluorometric detection methods are employed for pathogen identification, broth microdilution test is used for the MIC determination in these systems (104). However, prior preparation of bacterial suspension in desired concentration is required for performing tests in these automated systems. Also, different panels are required for gram-positive and gram-negative bacterial identification, as well

as the AST. Therefore, bacterial identification and drug resistance detection in these automated systems are limited by the database and panels of the systems. In addition, discrepancies of bacterial identification and antimicrobial resistance by the automated systems have been reported (105, 106).

Conventional phenotypic AST methods require a defined inoculum of test microorganism and specific culture medium to evaluate the growth inhibition of the microorganism (100). Also, an incubation time of 16 to 20 hours is necessary to obtain the result. The incubation time could be even longer, depending on the testing methods and species (101). Since delays in appropriate antimicrobial treatment are associated with prolonged hospitalization and higher mortality rate (102), rapid ASTs are necessary to improve the patient outcomes.

## 1.5 Overview of current molecular diagnostic methods and their limitations

Molecular diagnostics in microbiology refers to the group of techniques that detect, identify, and characterize microorganisms at the genetic level. Nucleic acids-based molecular diagnostics involves the extraction and analysis of microbial DNA or RNA to detect specific genetic markers or sequences, such as 16S rRNA gene and AMR genes, of microorganisms in clinical specimens. Since molecular diagnostics is culture-independent, it allows rapid diagnosis of IBIs and minimizes the occupational exposure to high concentration of infectious agents (107). Also, molecular diagnostics can serve as a complementary method for bacterial identification in culture-negative samples suspected with IBIs (97, 108, 109). The most used molecular diagnostic techniques currently in clinical microbiology laboratories include PCR, DNA microarray and sequencing (110).

### 1.5.1 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) is a pioneering molecular technique developed by Kary Mullis in 1980s (111). It is a kind of nucleic acid amplification test (NAAT) that allows specific amplification of targeted regions in genetic materials with the help of DNA polymerases and primers. PCR technology relies on prior knowledge of the targeted DNA sequence in order to design specific primers for amplification. It can be used for detection of specific pathogens in clinical microbiology by amplification of the species-specific genes. For instance, PCR of 16S rRNA gene detects the presence of bacteria, uidA gene shows the presence of *Escherichia coli*, pbp5 gene shows the presence of *Enterococcus faecalis* (112). Commonly, multiplex PCR is used to detect the presence of a group of targeted pathogens in clinical specimens (112-114).

Since the conventional PCR only amplifies the targeted sequences, the visualization or detection of amplicons relies on agarose gel electrophoresis in early diagnostic PCR (110). The advent of real-time PCR (qPCR), which was developed by Russ Higuchi in 1992 (115), facilitates the use of PCR-based microbial identification in clinical microbiology laboratories. Real-time PCR allows both detection and quantification of the target sequences (116). Similar to conventional PCR, primers with specific sequences are used to amplify target sequences in qPCR. In addition to the primers, fluorescent probes or intercalating dyes are employed in qPCR for the detection and quantification of target sequences. Upon binding to the target DNA during the amplification, fluorescence signals are emitted from the probes or dyes and captured by a detector in real-time. The cycle number that fluorescence signal reaches the detectable threshold, which is known as $C_T$ value, can be used to quantify the target sequence by linear regression of standard curve (110). The development of automated qPCR platforms, such as Cepheid GeneXpert® Systems and Luminex Aries® systems (117), allows rapid microbial and AMR detection in clinical laboratories.

Although qPCR allows rapid diagnosis in clinical laboratories, the number of targets in a qPCR run is highly limited by the availability of different fluorescent dyes with distinct emission spectra (107, 118). The overlap between emission spectra of different fluorophores will lead to crosstalk of fluorescent signals, resulting in false-positive result. In general, three to four fluorescent dyes are used in a qPCR run to prevent crosstalk of fluorescent signals (119). Besides, highly multiplex PCR increases the chance of interactions of primers, the formation of primer-dimers decreases the sensitivity of the assay (120). Therefore, the assay kits available for automated qPCR platforms detect only specific groups of clinically important pathogens or AMR genes, such as methicillin-resistant Staphylococcus aureus (MRSA) and AMR genes associated with carbapenem-resistance. Other diagnostic tests are required for detecting pathogens not included in the assay kits.

Nested PCR is a technique that can increase both sensitivity and specificity of PCR by performing two successive amplification processes. In the first round of PCR, a set of "outer" primers amplifies the target sequence with extended flanking regions. In the second round of PCR, a set of "inner" primers amplifies the target sequence using the PCR product from the initial PCR as template (121). This technology was employed in multiplex PCR of an automated molecular diagnostic platform, namely BioFire® FilmArray® System, for rapid pathogen identification and AMR genes detection (121). DNA melting curve analysis in the system enables the detection of multiple targets within a single test by utilizing the distinct melting temperatures (Tm) of each target (121). Therefore, this system provides comprehensive panels that detect a variety of pathogens and resistance markers (107). For instances, the BioFire® FilmArray® Meningitis Encephalitis panel detects 14 common pathogens in CSF that cause infections in central nervous system (CNS), the BioFire® Blood Culture Identification 2 panel detects 43 pathogens and 10 antimicrobial resistance genes associated with BSI, and the BioFire® Joint Infection panel detects

31 pathogens and 8 AMR genes in joint infections. Despite the ability to detect comprehensive targets in FilmArray® system, additional tests are required to detect pathogens and their associated resistance that are not included in the panels. Besides, the FilmArray® system has a relatively low throughput, enabling the simultaneous detection of up to 12 samples within a single machine (122). Also, false-positive detection of mecA gene in FilmArray® system was reported. A study by Bhatti et al. showed that false-positive mecA gene reported by FilmArray® system could be due to the presence of an altered staphylococcal cassette chromosome mec (SCCmec) element which lacks a functional mecA gene (123).

## 1.5.2 DNA microarray

DNA microarray is a technology that enables simultaneous detection of thousands of targets based on the hybridization between specific DNA probes immobilized on the microarray surface and the complementary targeted nucleic acid sequences from the samples (124). The main steps in DNA microarray test include extraction of nucleic acids, amplification of targets by broad-range or multiplex PCR, labelling of PCR products with fluorescent dye, and hybridization of labelled nucleic acids with probes on DNA microarray (125, 126). The DNA microarray is then scanned to detect specific interactions between probes and targets by fluorescence (125, 127). This technology has been employed for rapid identification of pathogens and AMR genes in clinical microbiology. Several studies have developed and utilized DNA microarray for rapid pathogen identification in invasive infections. A DNA microarray assay was developed by Hou et al. to detect 7 common bacterial pathogens in CSF, results showed that DNA microarray had a higher sensitivity (87.5%) than the conventional culture (58.3%) for pathogen identification in 24 CSF samples since culture failed to detect pathogens from patients with prior treatment (128). Järvinen

et al. developed a DNA microarray to detect 12 pathogens and the mecA gene in 146 blood culture-positive and 40 blood-culture negative samples, the assay showed a sensitivity of 96% and a specificity of 98% (129). Another study by Spiess et al. developed a microarray to detect 14 fungal pathogens in blood, bronchoalveolar lavage fluid (BALF), and tissue samples from immunocompromised neutropenic patients (130). The microarray was able to detect fungal pathogens at low detection threshold, ranging from 1pg to 500pg of DNA, depending on the fungal species (130). The design of DNA microarray is critical to obtain reliable results, cross-hybridization between probes and non-target sequences can result in false positive results (131).

Apart from the manually performed in-house DNA microarray, there are automated DNA array systems for pathogen and AMR detection available in the market, which may promote the adaptation of DNA microarray for rapid diagnosis in clinical laboratories. The Prove-it™ Sepsis StripArray system from Mobidiag can detect 60 bacteria and 3 AMR genes from blood cultures, and the Verigene® System from Nanosphere can detect 12 gram-positive bacteria and 3 AMR genes in the Gram-Positive Blood Culture assay, and 8 gram-negative bacteria and 6 AMR genes in the Gram-Negative Blood Culture assay (132, 133). However, the number of targets and sample types are restricted by the detection panels provided by the manufacturers of the automated systems. In addition to the species-specific genes, 16S rRNA gene is commonly used in microarray to detect the presence of bacteria. For samples that are positive in 16S rRNA gene but failed to be identified by the panel in microarray, sequencing is required to further confirm the identity of the bacteria.

### 1.5.3 Sequencing

The advent of sequencing technologies has revolutionized various fields, including clinical microbiology, especially after the emergence of NGS technologies in the 2000s (134). Sequencing allows rapid genotypic identification of pathogens in infectious diseases, improving patient outcomes. There are different approaches to sequencing, such as whole-genome sequencing (WGS), targeted sequencing, and metagenomic sequencing. Each approach has its own advantages and limitations, with WGS providing the most comprehensive genetic information of an organism. Metagenomic sequencing, on the other hand, allows for the identification of all genetic material present in a sample, including unknown pathogens. Targeted sequencing focuses on specific regions of interest in the genome, making it a cost-effective option for identifying known pathogens in clinical settings. While WGS is more applicable to outbreak investigation, targeted sequencing and metagenomic sequencing are more commonly used in clinical diagnostics.

### 1.5.3.1 DNA barcoding – targeted sequencing

The concept of DNA barcoding was first introduced by Hebert et al. in 2003 (135), which refers to the molecular tool for rapid species identification based on short and standardized gene regions (136, 137). The barcode sequence of unidentified species is compared to the reference barcode sequences in the database for identification (138). An ideal DNA barcode should exhibit high genetic divergence at species level, contain conserved flanking regions across species for universal amplification, and have a relative short sequence length (500-800bp) for efficient amplification (136, 137). The 16S rRNA gene is commonly used for bacterial identification, while the internal transcribed spacer (ITS) region is the standard DNA barcode for fungal identification (138, 139). Since genotypic identification offers greater reliability in species identification compared to

phenotypic identification, 16S rRNA gene sequencing has been the "gold standard" for definitive

bacterial identification, particularly in cases where phenotypic-based method or MALDI-TOF MS

yields ambiguous profiles (140-142).

Conventionally, Sanger sequencing is used for 16S rRNA gene sequencing in clinical

microbiology laboratories (143). However, it is not routinely performed in clinical laboratories for

bacterial identification since Sanger sequencing is laborious, time consuming, and has limited

throughput (144). Furthermore, Sanger 16S rRNA gene sequencing could not identify mixed

species in polymicrobial samples, multiple copies of 16S rRNA genes would lead to overlapping

in the chromatogram (145). The emergence of next-generation sequencing (NGS) in the 2000s

allows high throughput massively parallel sequencing with lower cost and higher speed (134),

which could be an alternative to Sanger sequencing for bacterial identification in clinical

laboratories. Unlike Sanger sequencing, NGS can identify mixed species in polymicrobial samples,

additional incubation time to obtain pure cultures for identification is not necessary. The most

well-known and widely used NGS platform is Illumina sequencing. Despite the high read accuracy,

taxonomic resolution of Illumina 16S rRNA gene sequencing is limited by the relatively short read

length of the technology (up to 2x300bp), only partial 16S rRNA gene region can be sequenced

(146). The nine hypervariable regions of 16S rRNA gene exhibit different discriminatory power

for bacterial species identification due to different degrees of sequence divergence, and no single

region can differentiate all bacteria (147). The choice of variable regions for Illumina 16S rRNA

gene sequencing depends on the sample type, but V3-V4 regions of 16S rRNA gene is commonly

used due to its good balance of length and variability (148).

Undoubtedly, full-length 16S rRNA gene has better discriminatory power than partial 16S rRNA

gene for bacterial identification at species level. High throughput sequencing of the entire 16S

rRNA gene became possible when the third-generation sequencing (TGS), also known as long-read sequencing, was launched in 2010s. Pacific Biosciences presented the first single molecule real-time (SMRT) sequencer in 2011 (149), while Oxford Nanopore Technology (ONT) introduced the nanopore sequencing in 2014 (150). Although TGS technologies can produce long reads with an average length > 10 kbp, they have the common limitation: relatively high sequencing error rate (ranging from 10-20%) (151, 152) compared to NGS technologies (>0.1%) (153) and Sanger sequencing (0.001%) (154). Nevertheless, several studies have indicated a preference for long-read sequencing technologies over Illumina sequencing for species-level taxonomic classification based on 16S rRNA gene sequencing, despite the higher sequencing error rate (141, 155, 156).

### 1.5.3.2 Unbiased metagenomic sequencing

Although 16S rRNA gene sequencing can rapidly identify the bacterial species, it provides no information about the AMR profile of the bacterial pathogens. In addition, studies showed that some species could be preferentially amplified during PCR amplification of 16S rRNA gene, affecting the resulting microbial composition (157, 158). In this case, unbiased shotgun metagenomic sequencing could be an alternative sequencing approach to characterize the pathogens. Metagenomics refers to the collective study of genetic content from a microbial community (159), allowing the detection of all kinds of pathogens and their respective AMR and virulence genes in a sample (160). In combination with rapid data analysis, and comprehensive reference databases, clinical metagenomics holds significant potential in improving the diagnostic yield for syndromic testing of invasive infections, as it eliminates the need for a priori hypotheses and enables the identification of a wide range of pathogens (161).

Many studies have shown the potential of clinical metagenomics in diagnostic microbiology. In a case report of Wilson et al., unbiased metagenomic next-generation sequencing (mNGS) successfully detected an uncommon pathogen, *Leptospira*, in the CSF of a patient with meningoencephalitis within a rapid timeframe of 48 hours, greatly improved the patient outcomes (162). Given that the conventional testing failed to identify the etiology of meningoencephalitis for months, this finding underscores the diagnostic potential of unbiased mNGS in detecting elusive pathogens that may evade traditional diagnostic approaches. Unbiased mNGS has been applied across a range of clinical specimens to identify the causative agents of invasive infections, including those affecting the CNS (162-166), bloodstream (167-171), respiratory system (172-175), and joints (176-179). Due to its capability for real-time sequencing and shorter turnaround time, nanopore sequencing has applied for unbiased metagenomic sequencing of various clinical specimens (180-183), despite the relatively low read accuracy. Studies reported that the sample-to-result time of nanopore-based metagenomic sequencing workflows could be as short as 6-9 hours (180, 184, 185), compared to 48-72 hours in Illumina sequencing (186).

However, there are limitations in clinical metagenomics. One of the major challenges is the presence of large amount of host DNA in the sample significantly lowers the sequencing depth of microbial DNA and therefore lowers the sensitivity in microbial detection (187). Since the human genome (around 3.2 Gb) is approximately a thousand times larger than the bacterial genome (3.6 Mb on average), the mere presence of a small number of human cells can overwhelm the DNA components of microorganisms (188). Therefore, host DNA depletion is necessary to increase the sequencing depth of microbial DNA, especially in clinical samples with high concentrations of human DNA. Another challenge is that there is no consensus in sequencing depth required for detecting and characterizing the genome of interest (160). In addition, there is a lack of

standardized methods, bioinformatic pipelines, and databases for clinical metagenomics. The interpretation of findings is also challenged by contaminants from the environment and normal flora. Besides, the cost of clinical metagenomics is much higher than that of the conventional culture-based method for microbial identification.

## 1.6 Introduction to nanopore sequencing as a promising rapid diagnostic tool

Nanopore sequencing has emerged as a promising rapid diagnostic tool with transformative potential in various fields, including clinical diagnostics. Although the concept of nanopore sequencing was first proposed in the 1980s, it takes three decades to develop the technology and finally launched the first nanopore sequencer – MinION in 2014 (189). The scaled-up sequencers, GridION and PromethION, were also released by ONT later in 2017 and 2018, respectively. Nanopore sequencing relies on nanopores, which are proteins embedded into an electrically resistant membrane, for real-time sequencing. As the DNA or RNA molecule passes through a nanopore, the change in electric currents due to presence of different nucleotides is measured and translated to DNA sequences (150).

Nanopore sequencing offers several advantages over other sequencing technologies, including real-time sequencing, portability, low set up cost, ability to generate long reads, and the availability of cloud-based analysis pipelines. These features make it well-suited for rapid and on-site diagnostic applications. The real-time sequencing coupled with cloud-based analysis pipelines allows timely and actionable results, facilitating prompt clinical decision-making and patient management. The sequencer, MinION, is a pocket-sized device that can connect to a laptop for sequencing. The starter pack of nanopore sequencing costs only $1999, compared to $128 K for Illumina MiSeq and $695 K for PacBio RS (190). The portability and low instrumental cost of

nanopore sequencing make it particularly useful in resource-limited settings. With ongoing technological advancements and the growing availability of dedicated bioinformatics tools, it is believed that nanopore sequencing holds great promise as a rapid diagnostic tool that can revolutionize clinical practice and improve patient outcomes.

## 1.6.1 Review of existing literature on nanopore sequencing for clinical diagnosis

There is a growing interest in the adoption of nanopore sequencing for rapid diagnosis of infectious disease in clinical microbiology. Several pilot studies have evaluated the utility of nanopore sequencing for pathogen identification and AMR detection in various clinical specimens. These studies were mainly based on two sequencing approaches: targeted sequencing and unbiased metagenomic sequencing. Both approaches have their own advantages and limitations. While targeted sequencing is more sensitive and cost-effective, unbiased metagenomic sequencing allows detection of all kinds of pathogens, including novel species, as well as the AMR genes. However, further research is needed to optimize the use of nanopore sequencing in clinical settings and to establish standardized protocols for data analysis and interpretation.

### 1.6.1.1 Nanopore targeted sequencing (NTS)

NTS has been used for rapid diagnosis of bloodstream infections, which has shown promising results in terms of speed and sensitivity. Hong et al. evaluated the performance of NTS on 202 blood samples with reference to traditional blood cultures and Sanger sequencing (191). In this study, the 16S rRNA gene, the ITS gene, and a group of viral specific genes were amplified and sequenced for bacterial, fungal, and viral identification, respectively. NTS was positive in 63.36%

of samples, compared to 14.85% in blood cultures. The sensitivity and specificity of the NTS assay were 92.11% and 78.41%, respectively. The sample-to-report time could be reduced to 6 − 24 hours in NTS assay, and sufficient data could be generated for subsequent analysis within 1 hour of sequencing. Another study used NTS to identify pathogens in 387 blood samples using pathogen specific primers instead of the universal primers for the 16S rRNA gene and the ITS gene (192). In addition to the pathogen specific primers that targeted 114 BSI associated pathogens (57 bacteria, 21 viruses, 17 fungi, 12 special pathogens, and 7 parasites), 27 AMR genes were also included for AMR detection in this NTS assay. The NTS result was validated by routine blood cultures and plasma mNGS. As shown in the study of Hong, NTS also had a higher positive rate (69.5%) than the blood cultures (33.9 %) in this study, though it was slightly lower than that of mNGS (74.7 %). Also, a high diagnostic consistency was observed between NTS and mNGS (90.2 %), compared to 60.2% between NTS and blood cultures. The utilization of pathogen specific primers in NTS enhanced the specificity of the assay, a specificity of 90.1% was achieved, while the sensitivity of the NTS assay was 84.0 %. The concordance between the genotypic and phenotypic AMR profiles was 80.6%, suggesting that genotypic AMR profiles could be used to predict potential antibiotic resistance in case isolated colonies could not be obtained from blood cultures. The majority of samples could generate results with 3.5 hours of sequencing and the turnaround time of NTS was about 7 hours.

Apart from direct blood samples, the utility of NTS for rapid diagnosis of various specimen types was also evaluated. In a retrospective study, Fu and colleagues evaluated the performance of NTS with reference to Sanger sequencing and traditional culture for pathogen identification in various clinical samples, including blood, pleural fluid, ascitic fluid, CSF, wound drainage, BALF, and urine (193). The 16S rRNA gene, ITS gene, and rpoB gene were amplified and sequenced using

nanopore and Sanger sequencing methods for the identification of bacteria, fungi, and Mycobacterium species, respectively. The concordance between NTS and Sanger sequencing was 85.2% on average but varied among different sample types. The sensitivity of NTS (94.5%) was much higher than that of culture (37.8%). The reported sample-to-report time of NTS was about 8 to 14 hours. A prospective study of NTS for rapid diagnosis of CNS infections in 50 CSF samples was also performed by Fu et al. recently (194). NTS, mNGS, and culture were performed in parallel for pathogen identification. 16S rRNA gene and ITS gene were amplified and sequenced in NTS. Discrepancies between NTS and mNGS were further validated using Sanger sequencing. Similar to the previous BSI study, NTS showed a higher positive rate (76.0%) than culture (16.0%), but a lower positive rate than mNGS (94.0%). There are 70.0% of NTS-positive strains were verified by mNGS, and 86.7% of NTS-positive strains were verified by further confirmation using Sanger sequencing. However, the problem of relatively high false positive rate due to environmental contaminations in NTS was highlighted in the study. Positive NTS results were obtained in 16 out of 19 cases diagnosed with non-CNS infections, resulting in a low specificity (15.8%) and positive predictive value (PPV) (57.9%) in NTS. The inclusion of extraction control and no-template control (191) and application of detection threshold (192) in NTS assay might help to discriminate the contaminants from true pathogens. Also, NTS positive results must be interpreted with careful consideration for the clinical manifestations.

### 1.6.1.2 Nanopore metagenomic sequencing (NMgS)

Similar to NTS, NMgS has been used for the rapid diagnosis of BSI. However, due to the low microbial load in infected blood samples and challenges in discriminating background contaminants, positive blood cultures have commonly been enrolled for NMgS instead of direct

blood samples (195). The enrichment of microbes in blood cultures enhances the sensitivity of NMgS and reduces the sequencing time for pathogen identification. A NMgS study by Taxt et al. demonstrated the detection of spiked pathogens from positive blood cultures after 10 minutes of sequencing, and the AMR genes in monomicrobial blood cultures within one hour of sequencing (196). The bacterial concentration in monomicrobial positive blood cultures ranging from $2.6 \times 10^7$ to $1.6 \times 10^9$ CFU/ml. Besides, 95% genome coverage could be achieved in monocultures within 8 hours, and a steep increase in genome coverage was observed within the first two to four hours of sequencing. Although this study showed that NMgS of positive blood cultures allowed rapid pathogen identification and AMR genes detection within four hours since the positivity of blood cultures, it was based on positive blood cultures of spiked microorganisms. Another study by Harris and colleagues performed NMgS on 52 positive blood cultures from intensive care patients with sepsis (195). Adequate reads could be generated for downstream species identification and AMR gene detection analysis within four hours of sequencing. While the concordance between NMgS and MALDI-TOF MS for pathogen identification was 94.2%, the concordance between genotypic and phenotypic AST was 89.3%. The reporting time, starting from blood culture positivity, ranges from 8 to 16 hours. Although sample to report time could be further reduced by shortening blood culture incubation time (185), sequencing of all blood cultures before positivity could be impractical and costly (195).

The challenges of pathogen detection from direct blood samples by NMgS were also illustrated by the study of Gu et al. (186). This study compared the performance of nanopore- and Illumina-based metagenomic sequencing of cell-free DNA (cfDNA) in 182 body fluid samples with reference to culture and PCR testing. Multiple body fluid types were enrolled for the study, including abscess aspirate, synovial fluid, pleural fluid, ascitic fluid, CSF, BAL, and others. It was

found that the sensitivity and specificity of nanopore and Illumina sequencing were comparable. For bacterial identification, the sensitivity and specificity of nanopore sequencing were 75.0% and 81.4%, respectively, compared to 79.2% and 90.6% in Illumina sequencing, respectively. For fungal identification, the sensitivity and specificity of nanopore sequencing was 90.9% and 100%, respectively, compared to 90.6% and 89.0% in Illumina sequencing, respectively. Moreover, the performance of metagenomic sequencing was comparable across all body fluid types, except for blood plasma. The pathogen cfDNA burden in local body fluid was found to be 160-fold higher than in plasma from the same patient. A higher sequencing depth is necessary for metagenomic sequencing of direct blood samples for pathogen detection and thus increases the cost of testing. The sample-to-report time of NMgS was 6 hours, and pathogen identification could be achieved within one hour of sequencing. It is worth noting that one sample was sequenced per flow cell in NMgS to attain the short turnaround time in this study. To simplify the extraction workflow and shorten the turnaround time, cfDNA was extracted and sequenced in this study to eliminate the host DNA depletion and bead beating steps. However, metagenomic sequencing of cfDNA provides no information about the pathogen's AMR profile (197).

The NMgS of genomic DNA in various body fluid types was evaluated in the study of Zhao and colleagues (198). A total of 297 suspected infectious fluid samples were enrolled in the study, including 109 plasma, 41 BALFs, 40 CSFs, 36 sputum, 24 serous cavity effusions, 24 urine, and 23 abscess aspirates. Saponin-based host depletion was performed before DNA extraction, except for plasma samples. Cell-free DNA was extracted from plasma for NMgS. Similarly, a higher positive rate was observed in NMgS (67.34%) than in culture (33.95%). With reference to the composite standard, the overall positive percentage agreement (PPA) and negative percentage agreement (NPA) of NMgS were 89.7% and 69.0%, respectively. Despite a high PPA (100%) in

microbial-colonized sites, the NPA (14.3%−50.0%) was relatively lower than the overall NPA due to the presence of normal flora. In contrast, high PPA and NPA were obtained at the sterile body sites. The PPA of the abscess aspirate and CSF were 95.7% and 88.2%, respectively. The NPA of the abscess aspirate and CSF were 100% and 95.7%, respectively. A relatively lower PPA and NPA were also observed in plasma samples, which were 77.6% and 66.7%, respectively, indicating the challenge of performing NMgS in direct blood samples. The top three frequently detected pathogens from sterile body sites in this study were *K. pneumoniae*, *E. coli*, and *S. aureus*, which were members of the "ESKAPE" pathogens, illustrating the need for rapid AMR detection in invasive infections. This study also tried to evaluate the performance of NMgS for AMR detection. AMR genes were detected in 56 microbial stains by NMgS; however, only 23 of the strains could be cultured and validated using phenotypic AST. Due to the low positive rate of culture, genotypic AMR results failed to be validated in this study; further study with a larger sample size is required for evaluating the clinical utility of NMgS.

## 1.6.2 Challenges of implementation of nanopore sequencing in clinical settings

Nanopore sequencing offers several advantages over conventional diagnostic tests, including a short turnaround time, long-read sequencing, a hypothesis-free approach, syndromic testing, real-time analysis, high sensitivity, and low setup costs, but its adoption in clinical settings faces several significant challenges. One of the primary hurdles is the requirement of standardized protocols and quality control measures to ensure consistent and reliable results across different laboratories and healthcare facilities. Setting guidelines for data analysis and interpretation is critical yet challenging to distinguish pathogen microbes from commensal or environmental contaminants in sequencing-based diagnostics, especially for samples from non-sterile body sites (199). The inclusion of extraction control and no-template control, applying a detection threshold, and

integrated consideration of clinical manifestations are recommended to distinguish the pathogens from background contaminants (191, 192). However, there is no consensus on the setting of detection threshold; the optimal detection threshold might vary among different sequencing platforms, workflows, and sample types. Also, the use of different classifiers and databases could lead to different taxonomic classification, especially for closely related species with high average nucleotide identity (ANI) (200). Additionally, the integration of nanopore sequencing for routine diagnosis demands substantial investment in initial setup costs, reagent expenses, data storage requirements and training of personnel, which can be prohibitive for many healthcare institutions. Moreover, the interpretation of complex genomic data generated by nanopore sequencing requires advanced bioinformatics tools and expertise, posing another barrier to widespread clinical implementation. Despite these challenges, the potential of nanopore sequencing to revolutionize clinical diagnostics by enabling rapid and accurate identification of pathogens and antimicrobial resistance profiles makes it a promising technology worth further exploration and refinement.

## 1.7 Bioinformatic tools

Bioinformatic analysis is a crucial process in converting sequencing data into a clinical report. The bioinformatic analysis varies with sequencing platforms, sequencing approaches, and nature of samples. Those developed for short-read data might not be applicable for long-read data. Typically, the first step of bioinformatic analysis is quality control of sequencing reads, encompassing tasks such as barcode and adapter trimming, base quality filtering, and read length assessment. In unbiased metagenomics sequencing, an additional step commonly involves the removal of host DNA during quality control. To classify reads, sequences are aligned to reference sequences within a database, after which their abundance is estimated. Sequence assembly, frequently

employed in WGS and unbiased metagenomics studies, is used to reconstruct genomes. Variant calling is used to identify genetic variations within specific genes, such as single nucleotide polymorphisms linked to antimicrobial resistance.

The choice of bioinformatic tool can substantially affect the interpretation of sequencing data since different tools employ different algorithms. An ideal bioinformatic tool for clinical diagnosis should be accurate in pathogen identification and antimicrobial resistance detection, short computation time, and low computer resource requirement. Lengthy computation times can delay sample-to-report timelines, while high computational demands may hinder concurrent analysis of multiple sequencing batches. BLAST, a renowned classifier and the "gold standard" for sequence comparison, is characterized by its comparatively lower computation time (201). Kraken 2 (202) is known for its superior combination of speed, memory efficiency, and accuracy compared to other tools and is frequently employed for metagenomic analysis (203). Given the higher error rates associated with nanopore sequencing, specialized tools like NanoCLUST (204) and Emu (205) have been created to classify full-length 16S rRNA genes produced through nanopore sequencing. ONT also developed a bioinformatics platform, Epi2me, which provides a suite of analysis tools and workflows tailored for the analysis of data generated by nanopore sequencing.

### 1.7.1 Epi2me

Epi2me is a cloud-based analysis platform that provides real-time processing of sequencing data, facilitating quick insights and decision-making. The software is designed to be user-friendly, making bioinformatic analysis accessible to users without prior expertise. Users can easily upload FASTQ files to Epi2me and select appropriate workflows for analysis. Since all computations are performed in the cloud, high-performance computing resources are not necessary. Data can be

conveniently accessed through the Epi2me web interface. The platform offers specific workflows such as the 16S workflow, which uses BLAST and the NCBI 16S ribosomal RNA database for taxonomic classification, and the WIMP workflow, which utilizes Centrifuge and NCBI databases for metagenomic taxonomic classification.

Unfortunately, the web-based Epi2me service was discontinued in 2024. Epi2me Labs, a new software set to replace the web-based version, will provide both local and cloud-based analysis capabilities. In Epi2me Labs, taxonomic classifiers in the 16S and metagenomic workflows have been changed to Kraken 2 or Minimap2. While Epi2me offers straightforward and rapid analysis, it may have limitations in flexibility, such as the inability to perform additional steps like host DNA removal from datasets. Other bioinformatic tools may be required for more advanced analysis.

## 1.7.2 Kraken 2

Kraken 2 (202) is a highly efficient and widely used bioinformatics tool for taxonomic classification of metagenomic sequencing data. It matches k-mers from query sequences to the lowest common ancestors (LCA) in a taxonomy tree to determine the classification. By utilizing a minimizer strategy and spaced seeds approach, Kraken 2 enhances matching sensitivity while reducing memory requirements. This tool also employs a compact hash table (CHT) to store taxonomic information efficiently, using 32 bits per taxonomy record compared to the 96 bits in its predecessor, Kraken 1. These optimizations in Kraken 2 significantly improve memory efficiency and speed up the taxonomy searching process without compromising classification accuracy. With its high accuracy and reduced computation time in metagenomic analysis, a few

proof-of-principle studies adopted this bioinformatic tool for nanopore sequencing based rapid pathogen identification in clinical samples from urgent life-threatening cases (185, 195).

### 1.7.3 NanoCLUST

NanoCLUST (204) is a specialized bioinformatics tool designed for the classification of full-length 16S rRNA gene sequences generated through nanopore sequencing. This pipeline is tailored to address the higher error rates typically associated with nanopore sequencing data. Sequencing reads are converted to normalized 5-mer frequency vector, proceeds with UMAP projection and HDBSCAN clustering. The read with highest average intracluster ANI after pairwise alignment between reads within a cluster is used for draft assembly, while the remaining reads are used for polishing. The resulting consensus sequence is used for bacteria identification using BLASTn and NCBI RefSeq 16S database. The abundance estimation is based on the UMAP projection and HDBSCAN clustering.

NanoCLUST outperformed other classifiers such as Kraken 2, Bracken, Centrifuge, and Epi2me by accurately estimating bacteria species in mock sample (204). While Kraken 2, Bracken, Centrifuge, and Epi2me tended to identify a significantly higher number of species, NanoCLUST provides superior accuracy in species richness, taxonomic identification and abundance estimation. A minimum of 32 GB of RAM is recommended for conducting standard analyses with NanoCLUST.

## 1.7.4 Emu

Emu (205) is another bioinformatics tool tailored for taxonomic assignment of nanopore 16S sequencing data. The process begins by calculating the maximum alignment likelihood for each sequencing read against a reference sequence, using the lowest normalized occurrence of mismatches, insertions, deletions, and soft clipping. Emu then initializes a composition vector with an even distribution of taxonomic identifications. Subsequently, the tool calculates the likelihood of taxonomic identification for each sequencing read using Bayes' theorem. Through iterative refinement of taxonomic likelihoods and redistribution of sample composition, Emu improves its accuracy until no further enhancements can be achieved. This iterative approach minimizes false positives and reduces the misidentification of closely related species. Emu has demonstrated its effectiveness in accurately distinguishing and estimating the abundance of closely related species with high ANI, such as *Bacillus*, *Salmonella*, and *Desulfosporosinus*.

## 1.8 Research aims and objectives

This study aimed to evaluate the clinical utility of nanopore sequencing for rapid diagnosis of acute invasive infections. Initially, the performance of nanopore 16S rRNA gene sequencing in clinical isolates was evaluated by comparing it with Sanger and Illumina sequencing methods. As 16S rRNA gene sequencing is commonly employed for the definitive identification of samples with uncertain identities in culture-based identification, clinical isolates that could not be identified by MALDI-TOF MS were included in the comparison. Subsequently, the study evaluated the utility of nanopore 16S rRNA gene sequencing in direct clinical specimens in comparison to traditional culture. The final goal was to develop an optimized nanopore sequencing workflow for pathogen identification and AMR detection, while offering insights for the standardization of sequencing-based protocols. Two workflows were developed - nanopore targeted sequencing (NTS) and nanopore metagenomic sequencing (NMgS) - and their performance was evaluated.

Objectives:

(1) To compare the performance of nanopore sequencing with Illumina sequencing and traditional Sanger sequencing in bacterial identification using 16S rRNA gene sequencing.

(2) To illustrate the utility of sequencing technologies for identifying samples that failed to be identified by traditional culture-based methods.

(3) To evaluate the clinical utility of nanopore 16S rRNA gene sequencing for bacterial identification in direct body fluids.

(4) To establish the optimal detection threshold of nanopore 16S rRNA gene sequencing for pathogen identification, using traditional culture as a reference.

(5) To develop optimized NTS and NMgS workflows, from sample extraction to data analysis, for bacterial and fungal identification and AMR gene detection from direct body fluids.

(6) To compare and evaluate the clinical utility of NTS and NMgS for pathogen identification and AMR prediction from direct body fluids.

# Chapter 2: The clinical utility of Nanopore 16S for taxonomic assignment of unidentifiable bacterial pathogens in MALDI-TOF MS

## 2.1 Introduction

Nowadays, clinical laboratories mainly rely on MALDI-TOF MS for bacterial identification (206). However, failure in MALDI-TOF MS-based identification can result due to the high similarity of mass spectra among closely related species, a lack of reference spectra, or poor sample quality (207). Lau and colleagues reported that MALDI-TOF MS failed to identify 37 out of 67 (55%) phenotypically "difficult-to-identify" clinically important bacteria (208). Studies reported that anaerobes showed a higher failure rate in MALDI-TOF MS-based identification than aerobes in general, particularly *Actinomyces* spp., *Peptostreptococcus* spp., *Prevotella* spp., and *Fusobacterium* spp. (209-212). However, certain aerobes were also reported to be poorly identified by MALDI-TOF MS, including *Streptomyces* spp., *Achromobacter* spp., *Acinetobacter* spp., *Chryseobacterium* spp., *Moraxella* spp., and the partially acid-fast *Nocardia* spp. (209, 213, 214). When MALDI-TOF MS fails to identify the bacterial isolates from clinical samples, 16S rRNA gene sequencing is commonly performed for definite species-level identification.

In clinical laboratories, Sanger sequencing has conventionally been employed for 16S rRNA gene sequencing, offering high accuracy but a lengthy workflow and limited throughput. High throughput sequencing (HTS) technologies could be a potential alternative for 16S rRNA gene sequencing in clinical settings. While short-read NGS such as Illumina sequencing generates highly accurate reads, long-read TGS like nanopore sequencing allow sequencing of the entire 16S rRNA gene. Nonetheless, taxonomic resolution may be limited by targeted variable regions of 16S rRNA gene in Illumina sequencing or the relatively higher error rate ($8 - 15\%$) in nanopore

sequencing (215). This study aimed to compare the performance of long-read nanopore sequencing with traditional Sanger sequencing and short-read Illumina sequencing in identifying clinical isolates that could not achieve definite species-level identification via MALDI-TOF MS, using 16S rRNA gene sequencing (216).

Both Illumina and nanopore sequencing platforms produce commercially available 16S rRNA gene sequencing kits and analysis pipelines, which are 16S Metagenomic Sequencing Library Preparation workflow (Nextera XT Index kit v2) and MiSeq Reporter Software (MSR) from Illumina, and 16S Barcoding Kit 1-24 (SQK-16S024) and Epi2me from ONT, respectively. These two commercial workflows were used to identify the bacterial isolates with no reliable identification from MALDI-TOF MS, with reference to the traditional Sanger sequencing. Additionally, an in-house BLAST+ (v2.11.0) analysis was performed in order to evaluate the performance of the two built-in analysis pipelines from Illumina (MSR) and ONT (Epi2me). Considering the challenges in evaluating diagnostic accuracy without a perfect gold standard, a composite reference standard was employed in this study. The composite reference standard was derived from the combined results obtained through Sanger sequencing and the two HTS platforms. In cases of discordance among the three platforms, WGS was conducted for definite bacterial identity confirmation.

## 2.2 Materials and methods

### 2.2.1 Sample collection and DNA extraction

Clinical isolates that were unable to be classified at the species level (score < 2.00) by the IVD MALDI Biotyper (Bruker Daltonics, Bremen, Germany) were included in this study. To minimize random errors, the MALDI-TOF MS procedures were repeated twice. A total of 172 clinical isolates from different specimen sources (Supplementary 1) were collected from the clinical microbiology laboratory of Pamela Youde Nethersole Eastern Hospital in Hong Kong. With reference to the 16S rRNA gene sequencing results, these clinical isolates were failed to be identified by MALDI-TOF MS due to (1) the absence of a reference spectrum in the database for 81 samples, (2) the inclusion of certain highly pathogenic species in the Brucker Security-Relevant library (Security Library 1.0) instead of the standard database (BD-6763) for two samples, and (3) poor-quality protein spectra for 89 samples. Upon receival of the clinical isolates, DNA extraction was performed using AMPLICOR® Respiratory Specimen Preparation Kit (Roche, Basel, Switzerland), followed by a bead-based purification using 1.8x AMPure XP beads (Beckman Coulter, California, USA).

### 2.2.2 Sanger 16S rRNA sequencing (Sanger 16S)

In Sanger 16S, the entire 16S rRNA gene was amplified using the forward primer 27F (5´-AGAGTTTGATCMTGGC-3´) and the reverse primer 1492R (5´-TACCTTGTTACGACTT-3´) (217). The PCR reaction mixture was prepared by mixing 36.7 µl of nuclease-free water, 5 µl of 10x PCR buffer, 1 µl of 10 mM dNTP mix (NEB, Ipswich, Massachusetts, USA), 1 µl of each 25 µM primer, 0.3 µl of HotStarTaq Plus DNA Polymerase (Qiagen, Hilden, Germany), and 5 µl of DNA template. PCR conditions included an initial denaturation at 96 °C for 8 min, 37 cycles at

94 °C for 1 min, 37 °C for 2 min, and 72 °C for 2 min 30 s, followed by a final extension at 72 °C for 10 min, and a hold at 4 °C. Amplicons were purified using ExoSAP-IT reagent (Thermo Fisher Scientific, Waltham, MA, USA) and proceeded to subsequent cycle sequencing.

In cycle sequencing, a reaction mixture was prepared by mixing 13 µl of nuclease-free water, 1 µl of BigDye® Terminator v3.1 Ready Reaction Mix (Thermo Fisher Scientific), 3.5 µl of 5× sequencing buffer, 1 µl of 3.2 µM primer, and 1.5 µl of purified PCR product. One primer was added to one reaction mixture, for a total of 8 primers (Table 2-1), and therefore 8 reactions were prepared for each sample. The PCR conditions were 96 °C for 1 min, 25 cycles at 96 °C for 10 s, 37 °C for 30 s, and 60 °C for 4 min, and finally holding at 4 °C. The sequencing products were purified with 75% isopropanol, resuspended in 12 µl of Hi-Di™ Formamide (Thermo Fisher Scientific), and proceeded to capillary electrophoresis analysis using the Applied Biosystems® 3130 Genetic Analyzer (Thermo Fisher Scientific).

The raw trace files of the 8 reactions in each sample were aligned to generate a consensus 16S rRNA gene sequence using the Staden Package (v2.0.0b11). To determine the identity of each sample, the consensus sequence was analyzed using the Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The 16S ribosomal RNA sequence database and default parameters were used for the BLAST query. The classified species with the lowest E value and the highest percentage identity were determined as the identified species for each sample.

**Table 2-1: The cycle sequencing primers used in Sanger 16S**

| Primer | Primer sequence (5' - 3') | References |
|--------|---------------------------|------------|
| 27F | AGAGTTTGATCMTGGC | (217) |
| 343F | TACGGRAGGCAGCAG | (218) |
| 357R | CTGCTGCCTYCCGTA | (218) |
| 784F | AGGATTAGATACCCTGGTA | (219) |
| 803R | CTACCRGGGTATCTAATCC | (220) |
| 1099F | GYAACGAGCGCAACCC | (218) |
| 1114R | GGGTTGCGCTCGTTRC | (218) |
| 1492R | TACCTTGTTACGACTT | (217) |

2.2.3 Illumina 16S rRNA sequencing (Illumina 16S)

For Illumina 16S, library preparation was performed using the 16S Metagenomic Sequencing Library Preparation workflow (Nextera XT Index kit v2) according to the manufacturer's protocol. In the first round of PCR, V3 and V4 variable regions of the 16S rRNA gene were amplified using the 16S Amplicon PCR Forward Primer (5´-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3´) and 16S Amplicon PCR Reverse Primer (5´-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3´). The underlined bases are overhang adapter sequences for attachment of the indexed adapters in the subsequent barcode PCR. The amplicon was about 460 bp in length. After purification, a unique indexed sequencing adapter from Nextera XT Index Kit v2 (Illumina, San Diego, California, USA) was added to each sample in the second round of PCR. After that, a second bead-based purification was performed.

The purified libraries were quantified through qPCR using the LightCycler® 480 Instrument II (Roche) and QIAseq™ Library Quant Assay Kit (Qiagen). Also, the size of the libraries was determined using the 2100 Bioanalyzer system (Agilent, Santa Clara, CA, USA) coupled with the

High Sensitivity DNA Kit (Agilent). Subsequently, the libraries were diluted to a concentration of 4 nM and pooled into a single tube. The pooled library was denatured using 0.2 N NaOH, followed by further dilution to a final concentration of 9 pM. To enhance sequencing diversity, 15% of a 9 pM PhiX control prepared from the PhiX Control Kit v3 (Illumina) was added as a spike-in. The prepared library was loaded onto the MiSeq sequencer (Illumina) and subjected to sequencing using MiSeq Reagent Kits v3 (Illumina). The total sequencing time for Illumina 16S was 56 hours.

On-instrument data analysis was performed using MSR (v2.6.2.3) in the MiSeq system. The sequencing reads were classified with reference to the Greengenes database (v13.5, May 2013) (http://greengenes.lbl.gov/) by MSR. Additionally, the sequencing reads were classified using BLAST+ (v2.11.0) with reference to the NCBI 16S ribosomal RNA database. Briefly, paired-end reads were merged using the "make.contigs" command in Mothur (v1.44.3) (221). The reads were then filtered using the "screen.seqs" command to remove sequences that were smaller than 400 bp, larger than 500 bp, or contained any ambiguous bases. The resulting reads were classified by BLAST+ using an in-house Python script (https://github.com/siupenyau/Pocket_16S/tree/7d3fa9d73a6a35afb47e40e7850cef72b4b91a22). The percentage identity and percentage query coverage were set at 90% in the BLAST+ analysis. Any disagreements between the results of MSR and BLAST+ were resolved using another pipeline, nf-core/ampliseq (https://github.com/nf-core/ampliseq) developed by Straub et al. (222).

### 2.2.4 Nanopore 16S rRNA gene sequencing (Nanopore 16S)

For Nanopore 16S, libraries were constructed using the 16S Barcoding Kit 1-24 (SQK-16S024) following the manufacturer's protocol. The full-length 16S rRNA gene of the samples was amplified using the barcoded 16S primers, followed by a post-PCR clean up. The purified amplicons were quantified using the Qubit 2.0 Fluorometer (Thermo Fisher Scientific) with the Qubit™ 1X dsDNA HS Assay Kit (Thermo Fisher Scientific). A total of 24 barcodes were provided in the kit and therefore up to 24 samples were pooled and sequenced per run. The pooled library was ligated with the rapid adapter, loaded on flow cell FLO-MIN106 R9.4.1, and sequenced on the MinION sequencer for 4 hours.

During the sequencing, real-time analysis was performed using the cloud-based analysis platform Epi2me provided by ONT. The passed fastq files (quality score ≥7) were uploaded on Epi2me and analyzed using the FASTQ 16S workflow (v2020. 04. 06). The sequencing reads were compared with reference sequences in the NCBI 16S ribosomal RNA database for taxonomic classification. Both minimum percentage coverage and minimum percentage identity were set at 90% in the Epi2me analysis. Similar to Illumina 16S, the sequencing data was also further analyzed using BLAST+ in Nanopore 16S. The disagreements between Epi2me and BLAST+ were further analyzed using another pipeline, NanoCLUST (https://github.com/genomicsITER/NanoCLUST) (204).

### 2.2.5 Whole genome sequencing (WGS)

Samples that exhibited inconsistent taxa identification based on Sanger 16S, Illumina 16S, and Nanopore 16S tests underwent WGS on the ONT platform to confirm their definitive identities. Library preparation was conducted using the transposase-based rapid barcoding kit (SQK-

RBK110.96) from ONT. The pooled library was loaded onto the flow cell FLO-MIN106 R9.4.1 and sequenced using the GridION for 48 hours in high-accuracy base calling mode. The resulting fastq files were uploaded to Epi2me and analyzed using the WIMP workflow (v2021.03.05) for taxonomic classification.

Apart from read-by-read classification in the WIMP analysis, *de novo* assembly was performed for further analysis. The sequencing reads of each sample were assembled using Shasta (v0.7.0) (https://github.com/chanzuckerberg/shasta), followed by re-alignment of sequencing reads to the assembled consensus sequences using minimap2 (v2.17-r941) and samtools (v1.10). Consensus sequences were first polished using MarginPolish (v1.3.dev-5492204) (https://github.com/UCSC-nanopore-cgl/MarginPolish) and then further polished using homopolish (v0.2.1) (https://github.com/ythuang0522/homopolish) (223). Additionally, the sequencing reads underwent another *de novo* assembly pipeline to avoid bioinformatic bias. The sequencing reads were first assembled using miniasm (v0.3-r179) (https://github.com/lh3/miniasm/releases/tag/v0.3), followed by an all-vs-all read self-mapping using minimap2. Then, the sequencing reads were re-aligned to the consensus sequences generated by miniasm using minimap2. Finally, the consensus sequences were subjected to two rounds of polishing using racon (v1.4.3) (https://github.com/isovic/racon). For taxonomic classification, the longest polished consensus sequence of each sample was analyzed using BLAST+ with reference to the NCBI Prokaryotic RefSeq Genomes database. Furthermore, the ANI between the consensus sequences and the best-matched reference genomes was calculated using an ANI calculator (https://www.ezbiocloud.net/tools/ani) (224). An ANI value greater than 94% indicated that the samples belonged to the same species as the best-matched genomes.

## 2.2.6 Data and statistical analysis

The concordance between Sanger 16S and the two HTS sequencing platforms were calculated by comparing the taxonomic classifications obtained from the Nanopore 16S and Illumina 16S with that of Sanger 16S. To evaluate the diagnostic accuracy, a composite 16S rRNA sequencing result obtained from the three sequencing platforms was considered as the reference standard (Figure 2-1). Taxa that were identical across at least two sequencing platforms were considered as reference taxa. For samples where there were complete discrepancies in the species identified by the three sequencing platforms, WGS was performed to confirm the reference taxa.

**Figure 2-1. The workflow for constructing composite reference standards**
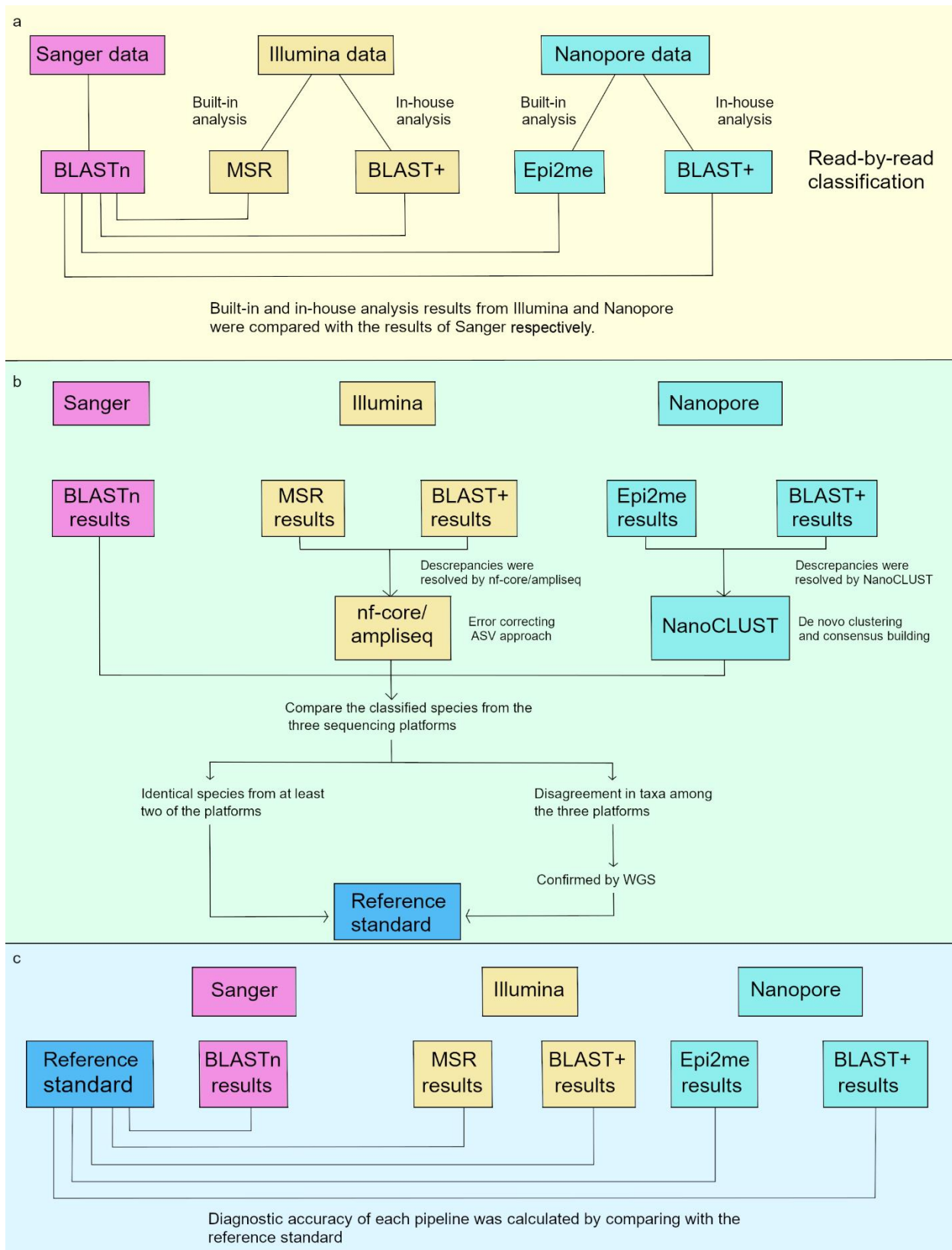
## 2.3 Results

### 2.3.1 Statistics of sequencing reads generated from the two HTS workflows

For Illumina 16S, an average of 113,381 reads were generated per sample and subjected to the MSR analysis. The paired-end reads were then merged and filtered using Mothur before proceeding to BLAST+ analysis. As a result, an average of 68,652 reads per sample was retained for BLAST+ analysis.

For Nanopore 16S, an average of 51,796 passed reads (QSCORE ≥ 7) were generated per sample by the sequencing platform MinKNOW. However, after uploading the passed reads on Epi2me, an average of 51,419 reads per sample were obtained after quality control and proceeded to classification in the FASTQ 16S workflow. This slight difference in the number of average reads per sample was attributed to the use of different algorithms in the demultiplexing step between Epi2me and Guppy of MinKNOW.

### 2.3.2 The percentage of classified reads from the two HTS workflows

The distribution of the percentage of classified reads by the built-in analysis and in-house BLAST+ analysis of the two HTS sequencing platforms is presented in Figure 2-2. For Illumina 16S, the percentage of classified reads per sample varied greatly among the samples in the MSR analysis. Also, the average percentage of classified reads per sample from MSR was much lower than that from BLAST+, which was 45.74% and 94.02%, respectively. For Nanopore 16S, an average of 76.03% of the total reads was classified by Epi2me at the species level, compared to 53.56% in BLAST+.

A lower percentage of classified reads in BLAST+ results was observed in Nanopore 16S compared to Illumina 16S, it is likely due to the lower read accuracy of Nanopore sequencing. In the in-house BLAST+ analysis, 90% query coverage and 90% percent identity thresholds were used. Given that Nanopore sequencing read accuracy ranges from approximately 85% to 95%, compared to over 99.9% for Illumina sequencing, more Nanopore reads might fall below the 90% identity threshold when aligned to the reference sequence. This discrepancy might lead to a higher percentage of unclassified reads in Nanopore 16S.

**Figure 2-2. The boxplots showing the distribution of the percentage of classified reads of all samples by in (a) Illumina 16S and (b) Nanopore 16S.**
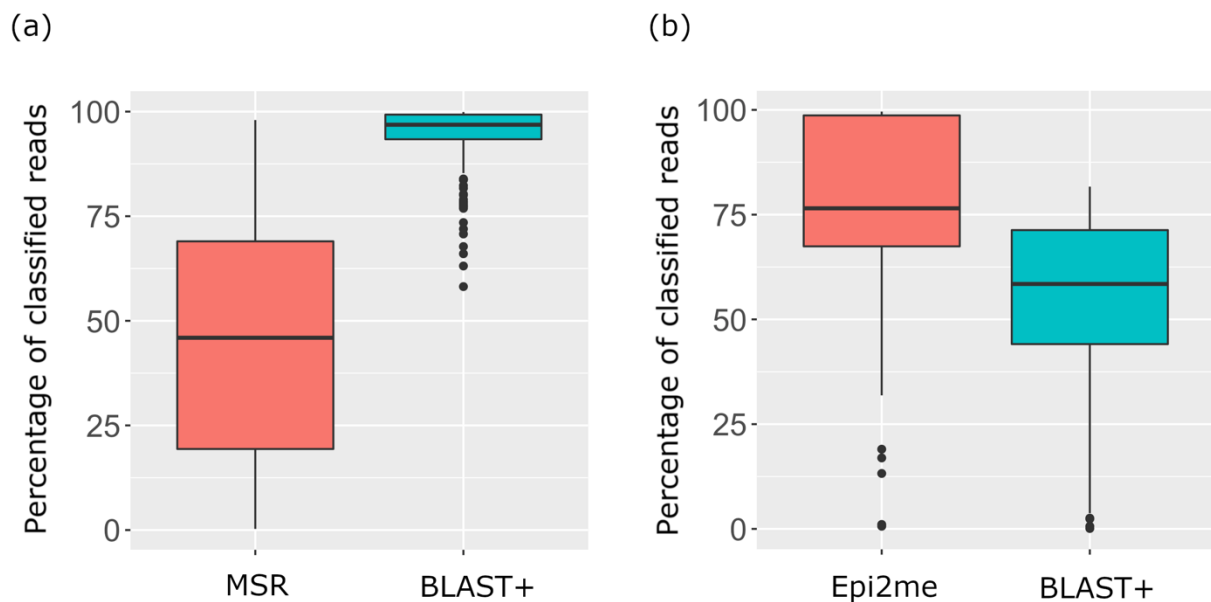
### 2.3.3 The concordance between the two HTS workflows and Sanger 16S

The best-matched species obtained from each analysis pipeline in each sequencing workflow are listed in Supplementary 2. Figure 2-3 shows the percentage of samples from the two HTS workflows that matched Sanger 16S results at each of the species, genus, and family levels. The concordance in species-level identification among the three sequencing platforms is shown in Figure 2-4.

In the Illumina 16S workflow, MSR and BLAST+ achieved concordance rates of 33.14% (57/172) and 65.70% (113/172), respectively, at the species level compared to Sanger 16S. In MSR, 9.30% of samples (16/172) were unmatched even at the family level, while all samples matched at the family level or below in BLAST+. Notably, there was low concordance between MSR and BLAST+ results, with only 32.56% (56/172) agreement among classified species. Furthermore, only 28.49% of samples (49/172) exhibited complete agreement in the classified species among the MSR, BLAST+, and Sanger datasets. The 116 samples with discrepant taxa inferred by MSR and BLAST+ were further analyzed using nf-core/ampliseq. Of which, only 41 samples were classified at the species level by nf-core/ampliseq, with 28 (24.14%) of them matching the results of BLAST+ and 4 (3.45%) matching the results of MSR. For the nine out of forty-one samples that failed to reach agreement at the species level, all of them matched the results of BLAST+ at the genus level. A total of 75 samples were only classified at the genus level or above by nf-core/ampliseq, and all of them matched the genus or family inferred by BLAST+. The concordance between the resolved Illumina 16S and Sanger 16S was 63.95% (110/172).

In the Nanopore 16S workflow, the concordance of species-level identification with Sanger 16S was 87.79% (151/172) for Epi2me and 83.14% (143/172) for BLAST+. Two unmatched samples (1.16%) were reported by Epi2me and BLAST+. The concordance between Epi2me and BLAST+

results was 80.23% (138/172). Additionally, 76.74% of samples (132/172) exhibited agreement in the classified species among Epi2me, BLAST+, and Sanger datasets. The 34 discrepancies among Epi2me and BLAST+ were resolved using NanoCLUST, with 38.24% (13/34) of them matching the results of Epi2me and 50.00% (17/34) of them matching the results of BLAST+. Four samples remained unresolved at the species level, with three matching at the genus level and one completely unmatched (R131). The concordance between the resolved Nanopore 16S and Sanger 16S was 89.53% (154/172).

**Figure 2-3. The concordance between bacterial taxa inferred by the two HTS workflows and the Sanger sequencing.**



The concordance between bacterial taxa generated from HTS technologies and the reference Sanger sequencing

| | MSR (Illumina) | BLAST+ (Illumina) | Epi2me (Nanopore) | BLAST+ (Nanopore) |
|---|---|---|---|---|
| ■ Not match | 9.30% | 0.00% | 1.16% | 1.16% |
| ■ Matched at family level | 18.02% | 1.16% | 0.58% | 1.16% |
| ■ Matched at genus level | 39.54% | 33.14% | 10.47% | 14.54% |
| ■ Matched at species level | 33.14% | 65.70% | 87.79% | 83.14% |
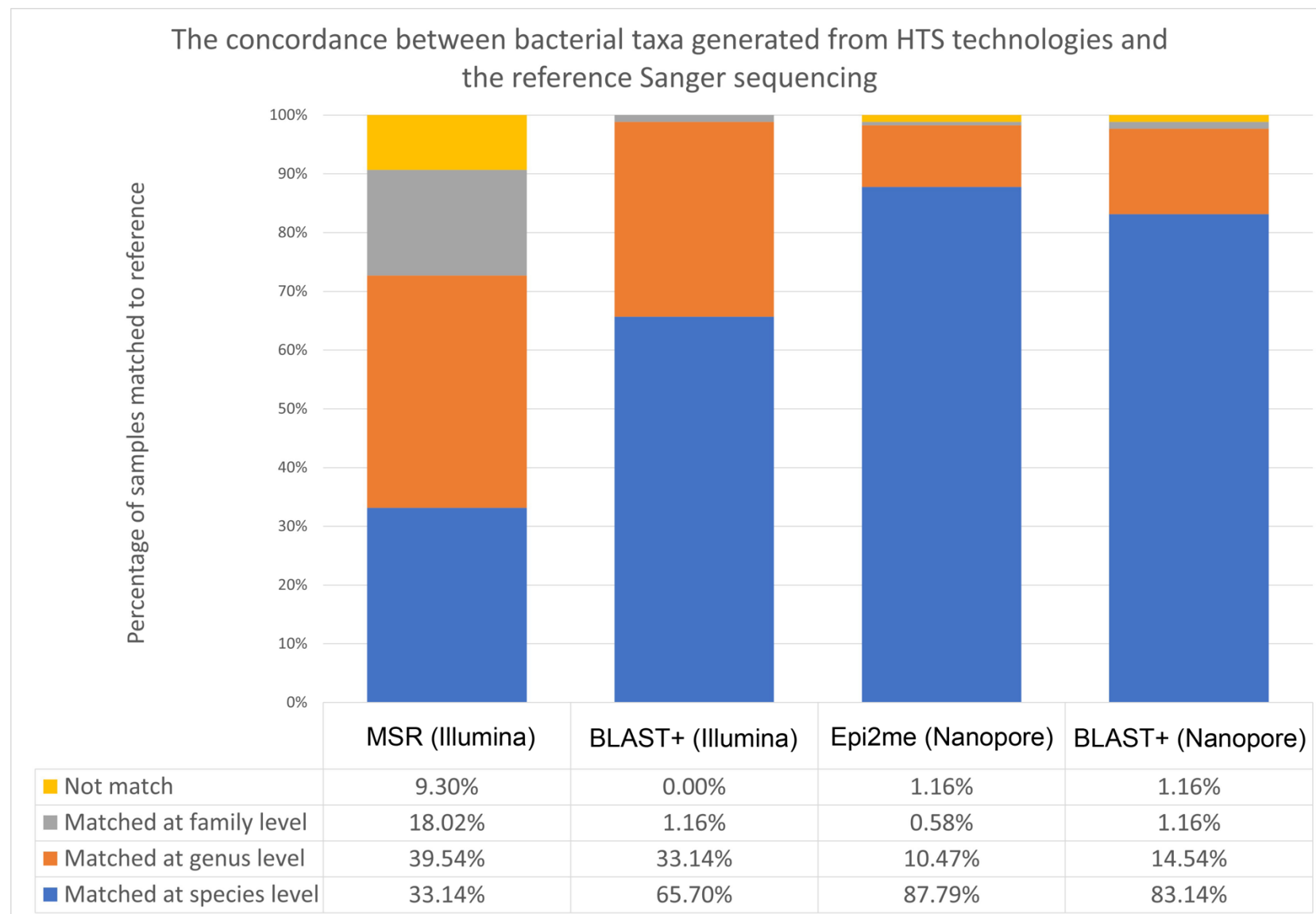
**Figure 2-4. The Venn Diagram showing the concordance of bacterial taxa inferred by different 16S rRNA sequencing platforms. (a) The concordance of best-matched species between Illumina 16S and Sanger 16S. (b) The concordance of best-matched species between Nanopore 16S and Sanger 16S. (c) The concordance of best-matched species among Sanger 16S, resolved Illumina 16S, resolved Nanopore 16S, and the reference standards.**
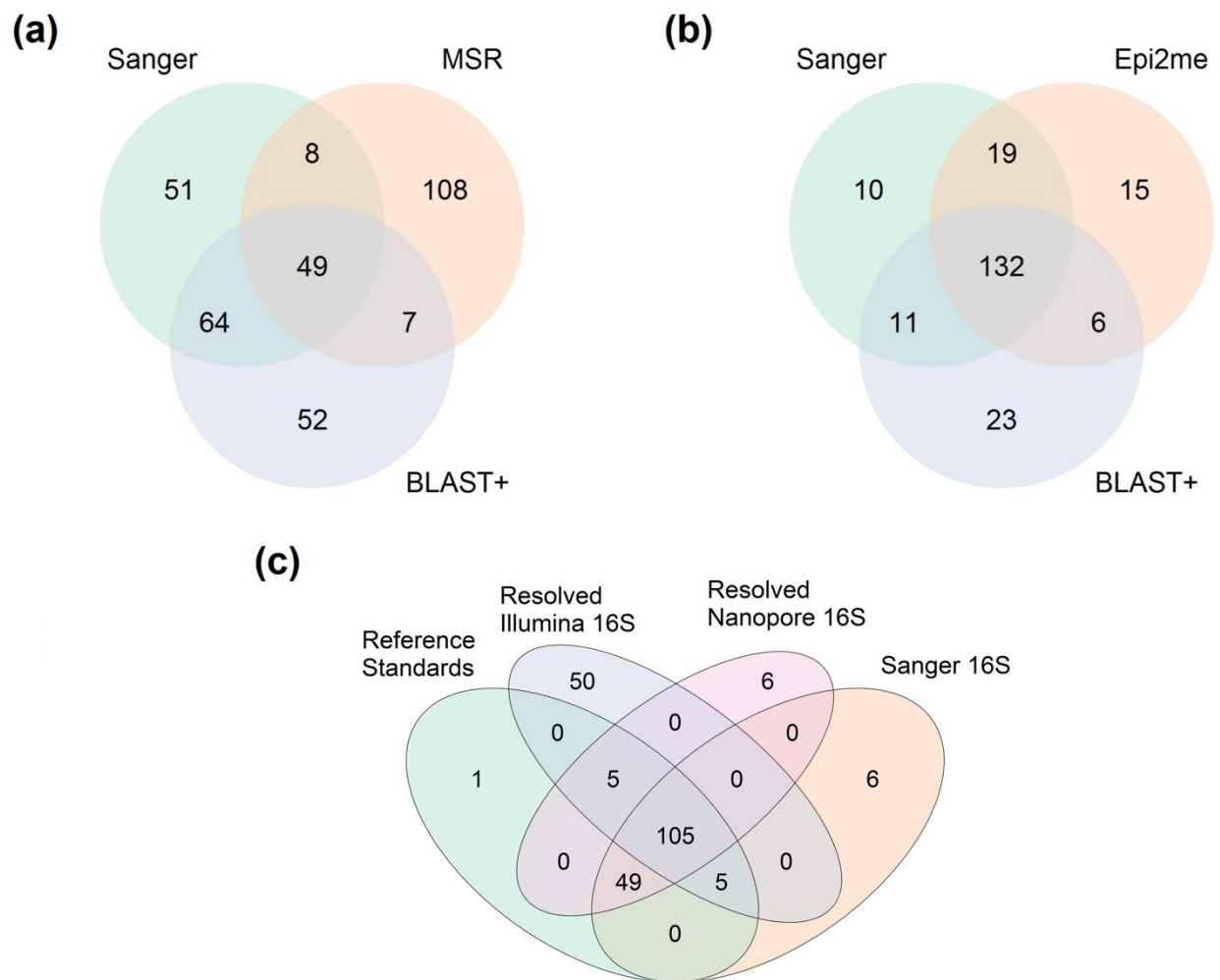
### 2.3.4 WGS results for bacterial isolates with completely discordant taxa

A total of eight samples exhibited complete discordance in bacterial species identification by the three 16S rRNA gene sequencing workflows. To determine the definite taxa, WGS was performed. Additionally, two ATCC reference strains (*Klebsiella pneumoniae* BAA3079 and *Staphylococcus aureus* BAA3114) were sequenced and analyzed alongside the eight discordant samples to validate the WGS workflow and genome assembly pipelines. Both reference strains yielded consensus sequences of >3Mb, covering 94% of reference genomes with 99% identity, indicating the reliability of the WGS protocol for constructing consensus prokaryotic genomes (Table 2-2).

Surprisingly, seven out of eight samples still failed to be identified using WGS, with query coverage below 70% for the longest consensus sequences and ANI below 94% (the threshold for the same species should be >94%) (Table 2-2). These samples did not match any of the published bacterial genomes, indicating that they are possibly novel bacterial species. For samples R121 and R131, the query coverage was even as low as 6.04–6.29%. Interestingly, these two samples also exhibited a remarkably low percentage of classified reads exclusively on the Nanopore 16S workflow compared to other discordant samples (Table 2-3). This suggests that the full-length 16S rRNA genes of R121 and R131 have great variances with the reference sequences available in the database. Therefore, R121 and R131 might belong to a new genus. The seven possibly novel species samples were excluded from the subsequent evaluation of diagnostic accuracy due to the lack of a definite identity.

The remaining one sample, R062, was identified as *Klebsiella michiganensis* using WGS. The consensus sequence of R062 generated using Shasta exhibited a query coverage of 92.17%, percentage identity of 99.17%, and ANI of 98.71% to *Klebsiella michiganensis* (NZ_CP060111.1).

**Table 2-2. WGS results for the samples with complete discordant taxonomic assignment by Sanger, Illumina, and Nanopore 16S**

| Sample ID | Species inferred by Sanger 16S | Species inferred by resolved Illumina 16S [b] | Species inferred by resolved Nanopore 16S [c] | Best-matched Species by WGS (reference genome) | WGS | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Genome assembly method | | | | | |
| | | | | | Shasta | | | Miniasm | | |
| | | | | | Query coverage (%) | Identity (%) | ANI (%)[d] | Query coverage (%) | Identity (%) | ANI (%)[d] |
| BAA 3079[a] | Not applicable | Not applicable | Not applicable | *Klebsiella pneumoniae* (NC_016845.1) | 99.00 | 97.00 | 98.92 | 92.13 | 99.40 | 99.14 |
| BAA 3114[a] | Not applicable | Not applicable | Not applicable | *Staphylococcus aureus* (NC_007795.1) | 94.06 | 99.95 | 99.30 | 88.39 | 99.92 | 99.23 |
| R001 | *Kocuria koreensis* | *Kocuria massiliensis* | *Kocuria spp.* | *Kocuria massiliensis* (NZ_LT835161.1) | 42.21 | 87.44 | 78.29 | 42.42 | 87.41 | 78.55 |
| R006 | *Kocuria koreensis* | *Kocuria massiliensis* | *Kocuria spp.* | *Kocuria massiliensis* (NZ_LT835161.1) | 43.04 | 79.12 | 78.49 | 42.04 | 87.49 | 78.44 |
| R062 | *Klebsiella grimontii* | *Enterobacter cloacae* | *Yokenella regensburgei* | *Klebsiella michiganensis* (NZ_CP060111.1) | 92.17 | 99.17 | 98.71 | 86.30 | 98.99 | 98.69 |
| R120 | *Brachybacterium conglomeratum* | *Brachybacterium faecium* | *Brachybacterium paraconglomeratum* | *Brachybacterium saurashtrense* (NZ_CP031356.1) | 62.15 | 85.18 | 82.30 | 62.30 | 85.12 | 82.39 |
| R121 | *Schaalia odontolytica* | *Schaalia vaccimaxillae* | *Sphingomonas paucimobilis* | *Schaalia odontolytica* (NZ_CP046315.1) | 6.07 | 78.55 | 70.34 | 6.04 | 78.24 | 70.86 |
| R131 | *Schaalia odontolytica* | *Schaalia vaccimaxillae* | No reliable ID | *Schaalia odontolytica* (NZ_CP046315.1) | 6.19 | 82.12 | 71.21 | 6.29 | 78.25 | 71.26 |
| R158 | *Microbacterium ginsengiterrae* | *Microbacterium assamensis* | *Microbacterium foliorum* | *Microbacterium foliorum* (NZ_CP041040.1) | 65.41 | 84.52 | 82.24 | 65.21 | 84.51 | 82.15 |
| R181 | *Sphingomonas yabuuchiae* | *Sphingomonas paucimobilis* | *Sphingomonas sanguinis* | *Sphingomonas hominis* (NZ_JABULH010000007.1) | 31.48 | 89.67 | 82.09 | 30.68 | 89.59 | 81.95 |

[a] BAA3079 and BAA3114 were QC sample, which were sequenced and analyzed in parallel with the discordant samples for WGS and bioinformatics analysis.

[b] Discordant samples between MSR and Illumina_BLAST+ were resolved by nf-core/ampliseq.

[c] Discordant samples between Epi2me and NanoBLAST+ were resolved by NanoCLUST.

[d] Average Nucleotide Identity (ANI) > 94% indicated that the samples belong to the same species as the best-matched genomes.

**Table 2-3. The percentage of classified reads for the 8 completely discordant samples in Illumina and Nanopore 16S.**

| Sample ID | No. of classified reads/ Total no. of reads (%) | | | |
| --- | --- | --- | --- | --- |
| | Illumina 16S | | Nanopore 16S | |
| | MSR | BLAST+ | Epi2me | BLAST+ |
| R001 | 18829/127460 (14.77) | 71270/74207 (96.04) | 34806/56074 (62.07) | 38772/56459 (68.67) |
| R006 | 15811/97292 (16.25) | 58201/59409 (97.97) | 36688/58626 (62.58) | 41190/59424 (69.32) |
| R062 | 42762/156000 (27.41) | 45512/57624 (78.98) | 27982/44518 (62.86) | 25580/44499 (57.48) |
| R120 | 113559/133543 (85.04) | 90700/91045 (99.62) | 39733/52248 (76.05) | 37655/52302 (72.00) |
| R121 | 91842/139327 (65.92) | 73990/76698 (96.47) | 313/49456 (0.63) | 49/49545 (0.10) |
| R131 | 79486/117795 (67.48) | 68220/71951 (94.81) | 994/94124 (1.06) | 586/91586 (0.64) |
| R158 | 23346/86042 (27.13) | 36253/37774 (95.97) | 41530/41945 (99.01) | 19952/44686 (44.65) |
| R181 | 72948/91201 (79.99) | 46722/47516 (98.33) | 38118/51271 (74.35) | 35013/51291 (68.26) |

## 2.3.5 Diagnostic accuracy of the three 16S rRNA sequencing workflows

To determine the diagnostic accuracy of each workflow, composite reference standards were constructed by combining the results of Sanger 16S, Illumina 16S, Nanopore 16S, as well as the WGS. Table 2-4 presents a list of the samples that showed disagreement between each sequencing workflow and the reference standards. Table 2-5 summarizes the diagnostic performance of each sequencing workflow.

For Illumina 16S, MSR and BLAST+ achieved diagnostic accuracies of 35.76% and 71.52%, respectively. It is worth noting that the diagnostic accuracy of resolved Illumina 16S using nf-core/ampliseq was even lower than that of using BLAST+ alone (69.07% vs. 71.52%). This suggests that BLAST+ was the most optimized analysis pipeline for Illumina 16S in this study.

Nanopore 16S demonstrated a higher diagnostic accuracy than Illumina 16S, regardless of the analysis pipelines. The diagnostic accuracy of Epi2me and BLAST+ was comparable, which was

89.09% and 89.70%, respectively. Moreover, the resolved Nanopore 16S achieved diagnostic accuracy of 96.36%, which was equivalent to that of Sanger 16S.

**Table 2-4. The samples with mismatched taxa inferred by at least one sequencing platform**

| Sample ID | Species-level ID (Reference Standard) | Sanger 16S | | Illumina 16S | | Nanopore 16S | |
|---|---|---|---|---|---|---|---|
| | | Classified species from Sanger 16S [a] | Identity against the reference (%) | Classified species from resolved Illumina 16S [a] | Identity against the reference (%) | Classified species from resolved Nanopore 16S [a] | Identity against the reference (%) |
| R003 | *Pseudoglutamicibacter albus* | *Pseudoglutamicibacter cumminsii* | 99.26% | *Pseudoglutamicibacter albus* | matched | *Pseudoglutamicibacter albus* | matched |
| R013 | *Microbacterium hominis* | *Microbacterium hominis* | matched | *Microbacterium aerolatum* | 97.47% | *Microbacterium hominis* | matched |
| R017 | *Microbacterium hominis* | *Microbacterium hominis* | matched | *Microbacterium aerolatum* | 97.47% | *Microbacterium hominis* | matched |
| R021 | *Microbacterium hominis* | *Microbacterium hominis* | matched | *Microbacterium aerolatum* | 97.47% | *Microbacterium hominis* | matched |
| R024 | *Bacillus idriensis* | *Bacillus idriensis* | matched | *Bacillus idriensis* | matched | *Bacillus indicus* | 97.62% |
| R025 | *Varibaculum cambriense* | *Varibaculum cambriense* | matched | *Varibaculum anthropi* | 98.50% | *Varibaculum cambriense* | matched |
| R026 | *Varibaculum cambriense* | *Varibaculum cambriense* | matched | *Varibaculum anthropi* | 98.50% | *Varibaculum cambriense* | matched |
| R036 | *Corynebacterium lowii* | *Corynebacterium lowii* | matched | *Corynebacterium bovis* | 93.29% | *Corynebacterium lowii* | matched |
| R040 | *Weissella cibaria* | *Weissella cibaria* | matched | *Weissella confusa* | 99.26% | *Weissella cibaria* | matched |
| R043 | *Proteus vulgaris* | *Proteus vulgaris* | matched | *Proteus alimentorum* | 99.64% | *Proteus vulgaris* | matched |
| R045 | *Brucella microti* | *Brucella microti* | matched | *Brucella papionis* | 99.86% | *Brucella microti* | matched |
| R047 | *Proteus cibarius* | *Proteus cibarius* | matched | *Proteus terrae* | 99.65% | *Proteus cibarius* | matched |
| R049 | *Dermacoccus barathri* | *Dermacoccus barathri* | matched | *Dermacoccus profundi* | 99.86% | *Dermacoccus barathri* | matched |
| R052 | *Arcanobacterium wilhelmae* | *Arcanobacterium wilhelmae* | matched | *Arcanobacterium pinnipediorum* | 96.60% | *Arcanobacterium wilhelmae* | matched |
| R053 | *Dermacoccus barathri* | *Dermacoccus barathri* | matched | *Dermacoccus profundi* | 99.86% | *Dermacoccus barathri* | matched |
| R056 | *Corynebacterium simulans* | *Corynebacterium simulans* | matched | *Corynebacterium glutamicum* | 93.74% | *Corynebacterium simulans* | matched |
| R058 | *Corynebacterium mastitidis* | *Corynebacterium mastitidis* | matched | *Corynebacterium tuberculostearicum* | 94.67% | *Corynebacterium mastitidis* | matched |
| R062 | *Klebsiella michiganensis* | *Klebsiella grimontii* | 99.20% | *Enterobacter cloacae* | 97.07% | *Yokenella regensburgei* | 98.56% |
| R063 | *Corynebacterium pilbarense* | *Corynebacterium pilbarense* | matched | *Corynebacterium coyleae* | 98.04% | *Corynebacterium pilbarense* | matched |
| R069 | *Eikenella corrodens* | *Eikenella corrodens* | matched | *Eikenella halliae* | 98.69% | *Eikenella corrodens* | matched |
| R071 | *Corynebacterium xerosis* | *Corynebacterium hansenii* | 99.07% | *Corynebacterium xerosis* | matched | *Corynebacterium xerosis* | matched |
| R072 | *Mycolicibacterium fortuitum* | *Mycolicibacterium fortuitum* | matched | *Mycolicibacterium arcueilense* | 98.96% | *Mycolicibacterium fortuitum* | matched |
| R073 | *Tessaracoccus oleiagri* | *Tessaracoccus oleiagri* | matched | *Tessaracoccus flavescens* | 95.95% | *Tessaracoccus oleiagri* | matched |
| R078 | *Vagococcus teuberi* | *Vagococcus teuberi* | matched | *Vagococcus martis* | 99.22% | *Vagococcus teuberi* | matched |

| R079 | *Corynebacterium xerosis* | *Corynebacterium hansenii* | 99.07% | *Corynebacterium xerosis* | matched | *Corynebacterium xerosis* | matched |
| R083 | *Tessaracoccus oleiagri* | *Tessaracoccus oleiagri* | matched | *Tessaracoccus flavescens* | 95.95% | *Tessaracoccus oleiagri* | matched |
| R086 | *Raoultella planticola* | *Raoultella planticola* | matched | *Raoultella planticola* | matched | *Klebsiella aerogenes* | 99.06% |
| R094 | *Corynebacterium xerosis* | *Corynebacterium hansenii* | 99.07% | *Corynebacterium xerosis* | matched | *Corynebacterium xerosis* | matched |
| R096 | *Streptomyces thermodiastaticus* | *Streptomyces thermodiastaticus* | matched | *Streptomyces thermoviolaceus* | 98.86% | Streptomyces thermodiastaticus | matched |
| R097 | *Pseudoxanthomonas helianthi* | *Pseudoxanthomonas helianthi* | matched | *Pseudoxanthomonas spadix* | 97.04% | *Pseudoxanthomonas helianthi* | matched |
| R098 | *Brachybacterium huguangmaarense* | *Brachybacterium huguangmaarense* | matched | *Brachybacterium huguangmaarense* | matched | *Brachybacterium nesterenkovii* | 97.84% |
| R104 | *Gordonia sputi* | *Gordonia sputi* | matched | *Gordonia otitidis* | 99.07% | *Gordonia sputi* | matched |
| R105 | *Gordonia sputi* | *Gordonia sputi* | matched | *Gordonia otitidis* | 99.07% | *Gordonia sputi* | matched |
| R107 | *Moraxella osloensis* | *Moraxella osloensis* | matched | *Enhydrobacter aerosaccus* | 99.19% | *Moraxella osloensis* | matched |
| R108 | *Staphylococcus saccharolyticus* | *Staphylococcus saccharolyticus* | matched | *Staphylococcus epidermidis* | 99.19% | *Staphylococcus saccharolyticus* | matched |
| R112 | *Citrobacter sedlakii* | *Citrobacter sedlakii* | matched | *Citrobacter youngae* | 98.32% | *Citrobacter sedlakii* | matched |
| R116 | *Tsukamurella tyrosinosolvens* | *Tsukamurella tyrosinosolvens* | matched | *Tsukamurella ocularis* | 99.86% | *Tsukamurella tyrosinosolvens* | matched |
| R123 | *Pseudoglutamicibacter albus* | *Pseudoglutamicibacter cumminsii* | 99.26% | *Pseudoglutamicibacter albus* | matched | *Pseudoglutamicibacter albus* | matched |
| R133 | *Nocardia brasiliensis* | *Nocardia brasiliensis* | matched | *Nocardia vulneris* | 99.31% | *Nocardia brasiliensis* | matched |
| R140 | *Moraxella lacunata* | *Moraxella lacunata* | matched | *Moraxella equi* | 99.38% | *Moraxella lacunata* | matched |
| R141 | *Ottowia beijingensis* | *Ottowia beijingensis* | matched | *Brachymonas denitrificans* | 93.33% | *Ottowia beijingensis* | matched |
| R148 | *Moraxella osloensis* | *Moraxella osloensis* | matched | *Enhydrobacter aerosaccus* | 99.19% | *Moraxella osloensis* | matched |
| R149 | *Ornithinibacillus californiensis* | *Ornithinibacillus californiensis* | matched | *Ornithinibacillus scapharcae* | 98.48% | *Ornithinibacillus californiensis* | matched |
| R151 | *Dermacoccus barathri* | *Dermacoccus barathri* | matched | *Dermacoccus profundi* | 99.86% | *Dermacoccus barathri* | matched |
| R153 | *Corynebacterium mastitidis* | *Corynebacterium mastitidis* | matched | *Corynebacterium tuberculostearicum* | 94.67% | *Corynebacterium mastitidis* | matched |
| R167 | *Moraxella osloensis* | *Moraxella osloensis* | matched | *Enhydrobacter aerosaccus* | 99.19% | *Moraxella osloensis* | matched |
| R175 | *Corynebacterium pollutisoli* | *Corynebacterium pollutisoli* | matched | *Corynebacterium humireducens* | 98.07% | *Corynebacterium pollutisoli* | matched |
| R176 | *Tsukamurella ocularis* | *Tsukamurella ocularis* | matched | *Tsukamurella ocularis* | matched | *Tsukamurella hominis* | 100.00% |
| R178 | *Acinetobacter soli* | *Acinetobacter soli* | matched | *Acinetobacter soli* | matched | *Acinetobacter lactucae* | 97.82% |
| R179 | *Corynebacterium lipophiloflavum* | *Corynebacterium lipophiloflavum* | matched | *Corynebacterium mycetoides* | 97.16% | *Corynebacterium lipophiloflavum* | matched |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| R180 | *Corynebacterium mastitidis* | *Corynebacterium mastitidis* | matched | *Corynebacterium tuberculostearicum* | 94.67% | *Corynebacterium mastitidis* | matched |
| R182 | *Fusobacterium nucleatum* | *Fusobacterium nucleatum* | matched | *Fusobacterium canifelinum* | 98.34% | *Fusobacterium nucleatum* | matched |
| R183 | *Parabacteroides faecis* | *Parabacteroides faecis* | matched | *Parabacteroides chongii* | 97.15% | *Parabacteroides faecis* | matched |
| R190 | *Bacillus xiamenensis* | *Bacillus xiamenensis* | matched | *Bacillus aerius* | 97.16% | *Bacillus xiamenensis* | matched |
| R192 | *Corynebacterium pilbarense* | *Corynebacterium pilbarense* | matched | *Corynebacterium ureicelerivorans* | 98.85% | *Corynebacterium pilbarense* | matched |
| R204 | *Prevotella scopos* | *Prevotella scopos* | matched | *Prevotella melaninogenica* | 98.10% | *Prevotella scopos* | matched |
| R205 | *Pasteurella multocida* | *Pasteurella multocida* | matched | *Pasteurella stomatis* | 93.74% | *Pasteurella multocida* | matched |
| R206 | *Staphylococcus cohnii* | *Staphylococcus cohnii* | matched | *Staphylococcus auricularis* | 98.16% | *Staphylococcus cohnii* | matched |
| R208 | *Achromobacter denitrificans* | *Achromobacter denitrificans* | matched | *Achromobacter xylosoxidans* | 99.15% | *Achromobacter denitrificans* | matched |
| R210 | *Bacillus licheniformis* | *Bacillus licheniformis* | matched | *Bacillus piscis* | 97.37% | *Bacillus licheniformis* | matched |

[a] The mismatched taxa were underlined

**Table 2-5: Diagnostic accuracies of the Sanger, Illumina and Nanopore 16S**

| Sequencing method | No. of sample analyzed | No. of samples with matched taxa | Diagnostic Accuracy (%) | 95% Confidence Interval |
|---|---|---|---|---|
| **Sanger 16S** | **165** | **159** | **96.36** | **92.25 - 98.65** |
| **Resolved Illumina 16S [a]** | **165** | **115** | **69.70** | **62.07 – 76.60** |
| Analyzed by MSR | 165 | 59 | 35.76 | 28.46 - 43.58 |
| Analyzed by BLAST+ | 165 | 118 | 71.52 | 63.98 - 78.26 |
| **Resolved Nanopore 16S [b]** | **165** | **159** | **96.36** | **92.25 - 98.65** |
| Analyzed by Epi2ME | 165 | 147 | 89.09 | 83.31 - 93.41 |
| Analyzed by BLAST+ | 165 | 148 | 89.70 | 84.02 - 93.88 |

[a] Discordant samples between MSR and BLAST+ were analyzed by nf-core/ampliseq, classified species in nf-core/ampliseq were considered as resolved identities in Illumina workflow.

[b] Discordant samples between Epi2me and BLAST+ were analyzed by NanoCLUST, classified species in NanoCLUST were considered as resolved identities in Nanopore workflow.

## 2.3.6 Comparison of turnaround time and running cost of the three workflows

The workflow and turnaround time for each platform are illustrated in Figure 2-5. For full-length 16S rRNA gene sequencing using the Sanger 16S platform, only 12 samples can be sequenced per run. In contrast, the ultra-high-throughput Illumina platform supports up to 384 samples per run, whereas the Nanopore platform is limited to 24 samples per batch due to the availability of sequencing barcodes. Excluding the time required for DNA extraction, the turnaround times for the Sanger 16S, Illumina 16S, and Nanopore 16S workflows are approximately 10.2 hours, 78 hours, and 8.25 hours per run, respectively. To sequence all 172 samples in this study, Sanger 16S required the longest time—153 hours—compared to 78 hours for Illumina 16S and 66 hours for Nanopore 16S. Although the sequencing time for Nanopore 16S extended up to 4 hours in this study, the sequencing can be stopped as soon as sufficient reads are obtained for downstream analysis.

In addition to its lower set-up cost, the running cost of the Nanopore 16S workflow is also significantly lower compared to the Sanger 16S and Illumina 16S workflows. For 172 samples, the cost of the Illumina 16S workflow per sequencing run is approximately US $4,931, translating to a cost of around US $28.7 per sample. When the sample size increases to 384, the cost per sequencing run for the Illumina 16S workflow rises to US $8,279, reducing the cost per sample to approximately US $21.6. Conversely, the cost per sequencing run for the Nanopore workflow, accommodating 24 samples, is US $424, yielding a cost per sample of approximately US $17.7. Nevertheless, both HTS workflows demonstrate lower running costs than the Sanger 16S workflow, which costs approximately US $25 per sample for sequencing the full-length 16S rRNA gene.

**Figure 2-5. The workflow and turnaround time of Sanger, Illumina and Nanopore 16S**
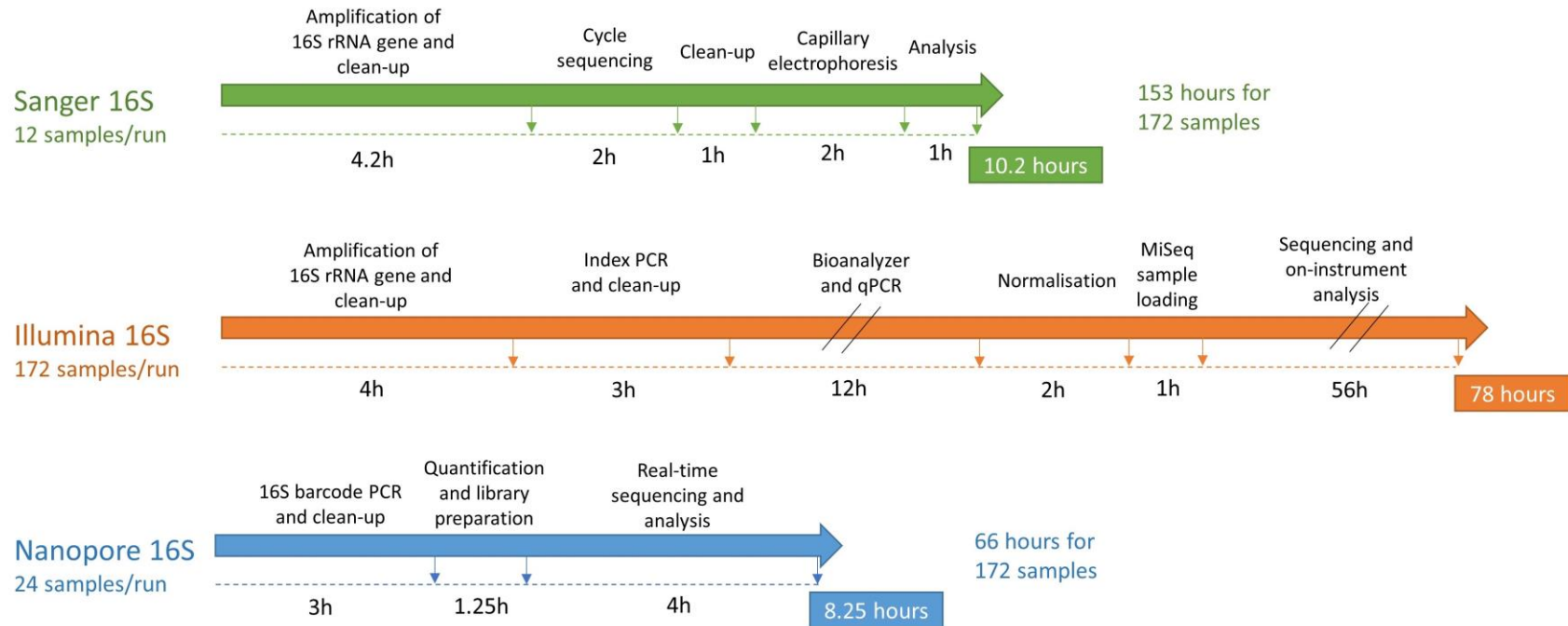
## 2.4 Discussion

While MALDI-TOF MS is effective for identifying most bacterial pathogens, clinical microbiology laboratories still rely on 16S rRNA gene sequencing to confirm the identities of "difficult-to-identify" isolates. With the advent of reduced costs, simplified protocols, and automated bioinformatics pipelines, HTS has emerged as a viable alternative to traditional Sanger sequencing for bacterial identification. This study is the first to compare and evaluate the clinical utility of two generations of HTS platforms: short-read NGS (Illumina 16S) and long-read TGS (Nanopore 16S) workflows, for the taxonomic assignment of bacterial pathogens unidentifiable by MALDI-TOF MS. To assess the performance of the built-in analysis pipelines from the Illumina (MSR) and Nanopore (Epi2me) platforms, the sequencing data from both platforms were subjected to additional analysis using BLAST+. By employing the same read-by-read classification approach as MSR and Epi2me, and being applicable to both Illumina and Nanopore data, BLAST+ serves as a suitable analysis tool for conducting intra- and inter-platform comparisons.

Results from the Illumina 16S and Nanopore 16S were compared against the results of the traditional Sanger 16S, respectively. For Illumina 16S, the concordance between MSR and Sanger 16S was notably low, with only 33.14% of samples matching the Sanger result at the species level. However, the concordance improved to 65.70% when BLAST+ was used. Previous studies have highlighted that different bioinformatic tools and 16S rRNA sequence databases can lead to varying taxonomic assignments, particularly at lower taxonomic levels (225, 226). The Greengenes database used by MSR was last updated in 2013 and lacks certain new bacterial taxa, which likely contributes to the lower agreement observed (226). Nevertheless, even when using the same aligner (BLAST) and database (NCBI 16S ribosomal RNA database), mismatches of

inferred taxa between Illumina 16S and Sanger 16S were present in 34.30% of samples. In contrast, the results of Nanopore 16S showed a higher concordance with Sanger 16S compared to Illumina 16S, irrespective of the analysis pipeline used. The concordance between Nanopore 16S and Sanger 16S was 87.79% when Epi2me was used for analysis, compared to 83.14% when BLAST+ was used. Despite this, species-level disagreement between Epi2me and BLAST+ was observed in 34 samples (19.77%).

In Illumina 16S workflow, discrepant samples between MSR and BLAST+ were further analyzed using nf-core/ampliseq. This pipeline employs an error-correcting amplicon sequence variant (ASV) approach, which has demonstrated better taxonomic classification performance than the operational taxonomic units (OTUs) clustering approach (222). However, when comparing the resolved Illumina 16S results with the reference standards, no improvement in diagnostic accuracy was observed.  This suggests that the diagnostic accuracy of the Illumina workflow was mainly constrained by the length and position of the variable regions of the 16S gene fragment being sequenced. In fact, sub-regions of the 16S rRNA gene do not capture sufficient sequence variation to discriminate closely related taxa, and different sub-regions have identification bias towards different bacterial taxa (227). Therefore, the V3–V4 regions used in the Illumina 16S workflow may have performed poorly in classifying the genera of mismatched samples (as shown in Table 2-4) down to the species level. It is worth noting that BLAST+ achieved high concordance with the reference at the genus level (98.79%), indicating that genus-level identification on the Illumina 16S workflow is reliable.

The relatively high read error rate of Nanopore sequencing could hinder it from resolving highly similar sequences in the Nanopore 16S workflow. Similarly, discrepancies between Epi2me and BLAST+ in Nanopore 16S were resolved using an additional pipeline with a different algorithm

for taxonomic classification. NanoCLUST, which generates clusters based on uniform manifold approximation and projection and classifies the representative consensus read in each cluster using BLAST, mitigates the effect of sequencing errors in individual reads. The diagnostic accuracy of Nanopore 16S increased from 89.09% (Epi2me) and 89.70% (BLAST+) to 96.36% when discrepancies were resolved by NanoCLUST. There were six samples that did not match the reference standards at the species level in the resolved Nanopore 16S and Sanger 16S results, possibly due to the high similarity (>97%) in 16S rRNA gene sequences between the inferred species and the reference taxa (Table 2-4). Traditionally, 16S rRNA gene sequences with >95% identity represent the same genus, while sequences with >97% identity represent closely related species (228). Studies have reported that the 16S rRNA gene cannot discriminate between closely related species in certain genera, including *Bacillus*, *Burkholderia*, *Acinetobacter baumannii-calcoaceticus complex*, *Achromobacter*, *Actinomyces*, *Staphylococcus*, and *Enterobacterales* (140, 142), which is in line with our findings.

Although WGS was conducted to obtain a definite identity for the eight samples with completely discordant results from the three 16S workflows, seven of them demonstrated low query coverage (<70%) and low ANI (<94%) between the consensus sequence and the best-matched genome, indicating the lack of a well-matched reference genome in the database for these samples. This highlights the need for regular updates of reference databases in order to ensure accurate identification in 16S rRNA gene sequencing. Further investigations are required to confirm that these samples are novel bacterial species. However, none of the 16S workflows accurately identified the species of R062. One sample, R062, was confirmed as *Klebsiella michiganensis* (ANI = 98.71%) using WGS. This discrepancy may be due to the high similarity between the 16S rRNA gene of *Klebsiella michiganensis* and those of closely related taxa identified in the

workflows: Sanger 16S (*Klebsiella grimontii*; 99.20%), Illumina 16S (*Enterobacter cloacae*; 97.07%), and Nanopore 16S (*Yokenella regensburgei*; 98.56%).

To further investigate the discrepancies in the taxonomic assignment of R062, sequencing reads from the three workflows were aligned to the reference 16S rRNA gene sequence of *Klebsiella michiganensis* (NR_118335.1), with alignments visualized using the Integrative Genomics Viewer (IGV) (Appendix 2). Two distinct groups of 16S rRNA gene sequences were observed in both Illumina 16S and Nanopore 16S workflows. BLAST analysis using the NCBI 16S rRNA database revealed that both variants showed the best match to members of the *Klebsiella oxytoca* complex or other closely related species, indicating that R062 carries multiple copies of the 16S rRNA gene. Notably, when the NCBI core nucleotide database was used for BLAST analysis instead of the NCBI 16S rRNA database, both Sanger 16S and Nanopore 16S workflows showed the best match with *Klebsiella michiganensis*. This underscores the influence of reference databases on the accuracy of taxonomic assignments. For Illumina 16S, even when the NCBI core nucleotide database was used, V3–V4 regions were too short to reliably distinguish between members of the *Klebsiella oxytoca* complex and other closely related species.

Apart from the higher diagnostic accuracy, Nanopore 16S offers several advantages over Illumina 16S for rapid diagnosis in clinical laboratories in terms of shorter turnaround time, lower running cost, and more flexible sample sizes. The turnaround time of Nanopore 16S workflow was 8.25 hours, compared to 78 hours on Illumina 16S workflow. A lengthy quantification process (qPCR and bioanalyzer) is necessary for Illumina sequencing to obtain precise library concentration which can greatly affect the subsequent cluster generation process. While over-clustering lowers base accuracy, under-clustering reduces data output in Illumina sequencing. In contrast, library concentration is simply quantified using qubit fluorometer in the Nanopore 16S workflow. The

sample size of Nanopore 16S is more flexible than Illumina 16S. To be more cost-effective, a large sample size (up to 384 samples per run) is preferred on Illumina 16S workflow. With reusable flow cell and smaller sample capacity (up to 24 samples per batch), Nanopore 16S allows rapid diagnosis of samples with clinical emergencies. Moreover, the running cost of Nanopore 16S (US $17.7 per sample) was lower than that of Illumina 16S (US $21.6 per sample).

This study has several limitations. First, only the V3–V4 sub-regions of the 16S rRNA gene were sequenced using the commercially available 16S sequencing workflow from Illumina. While it is possible to sequence the full-length 16S rRNA gene using Illumina MiSeq with a custom protocol (142), which may improve diagnostic accuracy, this adds complexity in analysis and requires additional steps that cannot be performed by MSR. Second, it should be noted that, except for the eight samples displaying discordant results, the taxonomic classification of the isolates was solely established through 16S rRNA sequencing, which may not provide a definitive representation of their actual taxa. Third, in the WGS analysis, taxonomic assignments were made based on consensus sequences generated through *de novo* assembly. It is important to highlight that the construction of complete, circular, and gap-free bacterial genomes was not achieved in this study.

## 2.5 Conclusion

In this chapter, the performance of Nanopore sequencing was compared with traditional Sanger sequencing and the short-read Illumina sequencing. Despite the relatively lower read accuracy, the long-read Nanopore 16S workflow demonstrated higher concordance to the Sanger 16S and higher diagnostic accuracy than the Illumina workflow. With the higher flexibility in sample size and sequencing time, lower running cost and comparable diagnostic accuracy with Sanger 16S, it is suggested that Nanopore 16S is a potential alternative for reliable bacterial identification in clinical

settings. Additionally, this chapter underscored the capacity of sequencing technologies to identify clinical isolates that failed to be identified by MALDI-TOF MS and their ability to uncover novel species. Sequencing-based tests not only allow rapid diagnosis, but also detect rare or novel species. However, further investigations were required to confirm the identity of proposed novel species.

# Chapter 3: Characterization of a novel bacterial species *Scrofimicrobium appendicitidis*

## 3.1 Introduction

In the previous chapter, two samples, namely R121 and R131, displayed a remarkably low query coverage of approximately 6% when compared to the best-matched reference genome in the WGS analysis. The substantial divergence from existing reference genomes suggests the possibility of these samples representing novel bacterial species within a previously unclassified genus. As R121 and R131 were bacterial isolates that originated from the same patient and have highly similar 16S rRNA gene sequence (99.66% of nucleotide identity), they were considered the same clinical isolate. To simplify the analysis, only the strain R131 was investigated in this study. The objective of this chapter was to verify the identity of this possibly novel species through a comprehensive analysis involving phenotypic characterization and phylogenetic analysis.

The strain R131 was initially isolated from a patient admitted to Pamela Youde Nethersole Eastern Hospital after experiencing right lower quadrant abdominal pain for three days, along with a one-day episode of fever. Following a contrast computed tomography (CT) scan of the abdomen and pelvis, he was diagnosed with acute appendicitis complicated by peri-appendiceal abscess formation. The patient subsequently underwent a laparoscopic appendicectomy procedure, which included drainage of the appendiceal abscess. Peritoneal swabs taken during the intervention revealed the presence of *Escherichia coli*, *Bacteroides thetaiotaomicron*, and an unidentified gram-positive bacillus (R131) through culture-based analysis.

Considering the relatively low read accuracy of Nanopore sequencing, WGS of R131 was performed using Illumina sequencing in addition to nanopore sequencing. While the long-read

nanopore sequencing generated the backbone of the bacterial genome, the high accuracy Illumina reads were used to further polish the nanopore reads. A complete, circular, and highly contiguous bacterial genome was constructed for the subsequent phylogenetic and phylogenomic analyses. Moreover, pan-genome analysis, digital DNA-DNA hybridization analysis, and percentage of conserved proteins analysis were conducted to compare R131 with its closest species. Additionally, transmission electron microscopy, biochemical tests, and antimicrobial susceptibility test were performed to examine the phenotypic properties of the strain.

## 3.2 Material and methods

### 3.2.1 Investigation of physical characteristics and biochemical properties

Gram staining and transmission electron microscopy were performed to examine cell wall composition and morphology of R131. Biochemical properties were determined using the bioMérieux VITEK® 2 Systems (bioMérieux) with the VITEK® 2 Anaerobic and Corynebacteria identification card (ANC). Additionally, manual tests were conducted for catalase and oxidase activities.

### 3.2.2 Antimicrobial susceptibility test

Disc diffusion test was used to determine the susceptibility of R131 to fourteen gram-positive spectrum antibiotics, including ampicillin (AMP 10), ceftriaxone (CRO 30), tetracycline (TE 30), vancomycin (VA 30), gentamicin (CN 10), meropenem (MEM 10), cefoxitin (FOX 30), chloramphenicol (C 30), erythromycin (E 15), co-trimoxazole (SXT 25), ciprofloxacin (CIP 5), ceftazidime (CAZ 30), cefepime (FEP 30), and clindamycin (DA 2).

### 3.2.3 WGS, genome assembly, and taxonomic identification

For Nanopore-based WGS, library was constructed using the transposase-based rapid barcoding kit (SQK-RBK110.96) and sequenced for 48 hours in high-accuracy base calling mode. For Illumina-based WGS, library was prepared using NEBNext® Ultra™ II FS DNA Library Prep Kit (NEB, Ipswich, Massachusetts, USA) coupled with NEBNext® Multiplex Oligos for Illumina® (96 Unique Dual Index Primer Pairs). The library was sequenced on Illumina MiSeq system with MiSeq Reagent Kit V2.

For quality control, fastp v0.22.0 (229) was used to filter Illumina short reads, and Filtlong v0.2.1 was used to filter nanopore long reads. The assembly process employed a long-read-first approach with Trycycler v0.5.0 (230). The long reads were divided into 12 subsets and assembled using Flye (231), minasm (232)+minipolish (233), and raven (234), resulting in 12 different genome assemblies. These assemblies were merged into a consensus using Trycycler and polished with Medaka v1.4.4. Polishing with short reads was performed using Polypolish (235) and further refined with POLCA (236). To capture small plasmids potentially missed by long-read sets, a short-read-first hybrid assembly was conducted using Unicycler v0.5.0 (237). The quality and completeness of the genome assembly were evaluated using BUSCO v5.2.2 (bacteria_odb10) (238).

The taxonomic classification of the strain was determined in two approaches: analysis of the 16S and 23S genes, and analysis of marker genes across the entire genome. The 16S and 23S genes were analyzed using SILVA ACT (Alignment, Classification and Tree Service) with SINA 1.2.12 (239) and SILVA SSU and LSU databases 138.1 (240). The "Classify Microbes with GTDB-Tk - v1.7.0" app (241) in KBase (242) was utilized to identify marker genes and assign taxonomic

classifications based on the whole genome. The Orthologous Average Nucleotide Identity Tool (OAT) (243) was used to estimate genomic similarity with related species within the same genus.

### 3.2.4 Phylogenetic and phylogenomic analysis

For phylogenetic analysis, 16S rRNA gene of the R131 was aligned with sequences from *Scrofimicrobium* and other *Actinomycetaceae* species, as well as an *Escherichia coli* outgroup, using SINA v1.7.2. The phylogenetic tree was constructed using RAxML v8.2.12 (244) with the GTRGAMMA substitution model and a bootstrap of 800 replicates under the AutoMRE option. The resulting tree was visualized and annotated using iTOL v6 (245).

A phylogenomic tree was constructed using GTDB-Tk v2.3.0 (246) based on GTDB release 08-RS214.0 (247). The tree was visualized and annotated using iTOL v6 for a comprehensive representation of the genome similarity between the novel bacterium R131 and bacteria from different genera within the family *Actinomycetaceae*.

### 3.2.5 Genome annotation and antimicrobial resistance prediction

Based on the classification of R131 in the genus *Scrofimicrobium*, the genome was annotated using RASTtk (248) on the RAST server (249) and the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) build6771 (250). The potential genes predicted from PGAP were analyzed for functional annotation using BlastKOALA (251) based on KEGG Orthology. The genome map was visualized using Proksee (252). Antimicrobial resistance was predicted using ResFinder (253), ResFinderFG 2.0 (254), and AMRFinderPlus v3.11.26 (255).

### 3.2.6 Pan-genome analysis

Four genomes classified as *Scrofimicrobium* based on GTDB were downloaded from NCBI and annotated using PGAP. Pan-genome analysis was performed using get_homologues (256), including clustering of orthologous genes with OrthoMCL, calculation of ANI, and generation of a pan-genome matrix. A pangenome tree was created using IQ-TREE 2 (257) and visualized with FigTree. Core genes were annotated using BlastKOALA with the "Prokaryotes" database.

### 3.2.7 Digital DNA-DNA hybridization (dDDH) analysis

Digital DNA-DNA hybridization (dDDH) analysis was also performed to evaluate the genome similarity between R131 and the four *Scrofimicrobium* genomes. In general, a dDDH value of 70% defines species boundaries, while a value of 79% delimits subspecies (258). The five genomes were submitted to GGDC (http://ggdc.dsmz.de/ggdc.php, GGDC Genome-to-Genome Distance Calculator 3.0) (259) for dDDH analysis.

### 3.2.8 Percentage of conserved proteins (POCP) analysis
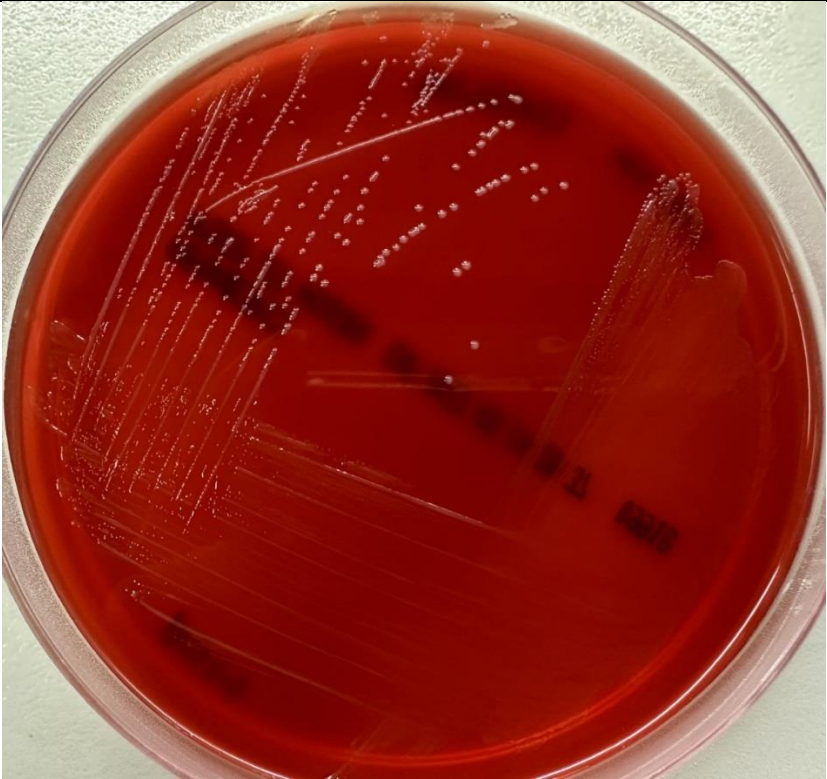
In addition to the dDDH analysis, a Percentage of Conserved Proteins (POCP) analysis was conducted to assess the genomic similarity between two microbial genomes by evaluating the shared protein content. While dDDH analysis aided in species delineation, POCP was utilized for genus demarcation. A threshold of 50% indicates that two strains fall within the same genus (260). POCP values for R131, the four *Scrofimicrobium* genomes, and the type species of *Actinomyces* (*Actinomyces bovis* NCTC 11535) and *Schaalia* (*Schaalia odontolytica* NCTC9935), were calculated using POCP-nf (https://github.com/hoelzer/pocp) (261).

## 3.3 Results

### 3.3.1 Phenotypic characteristics

As shown in Table 3-1, R131 presents as small (0.5-1.0 mm), grey, semi-translucent, and alpha-haemolytic colonies on blood agar. It is a facultative anaerobe, capable of growing under both anaerobic and aerobic conditions (with or without 5% $CO_2$) at a temperature of 37 °C. The growth of R131 is highly dependent on blood media; growth was observed on 5% horse blood agar but not on brain-heart infusion agar. Isolated colonies typically appear on solid media after approximately 48 hours. Gram staining reveals the presence of gram-positive coccobacilli. The strain was non-motile, and the cell size was about 0.5-0.6 μm wide and 0.7-1.0 μm long.

**Table 3-1. Colony, cell morphology, and MALDI-TOF MS spectrum of R131**

| Colony morphology on anaerobic blood agar |  |
| --- | --- |

| | |
|---|---|
| Alpha-haemolysis on blood agar |  |
| Gram stain |  |

| | |
|---|---|
| Transmission electron microscopic photo |  |
| MALDI-TOF MS spectrum |  |
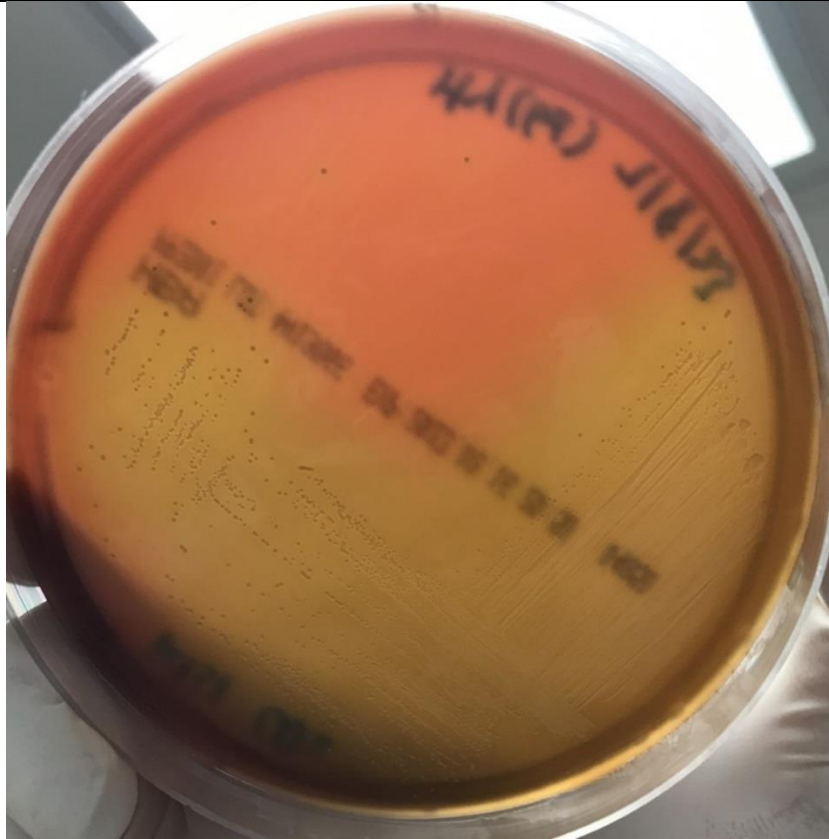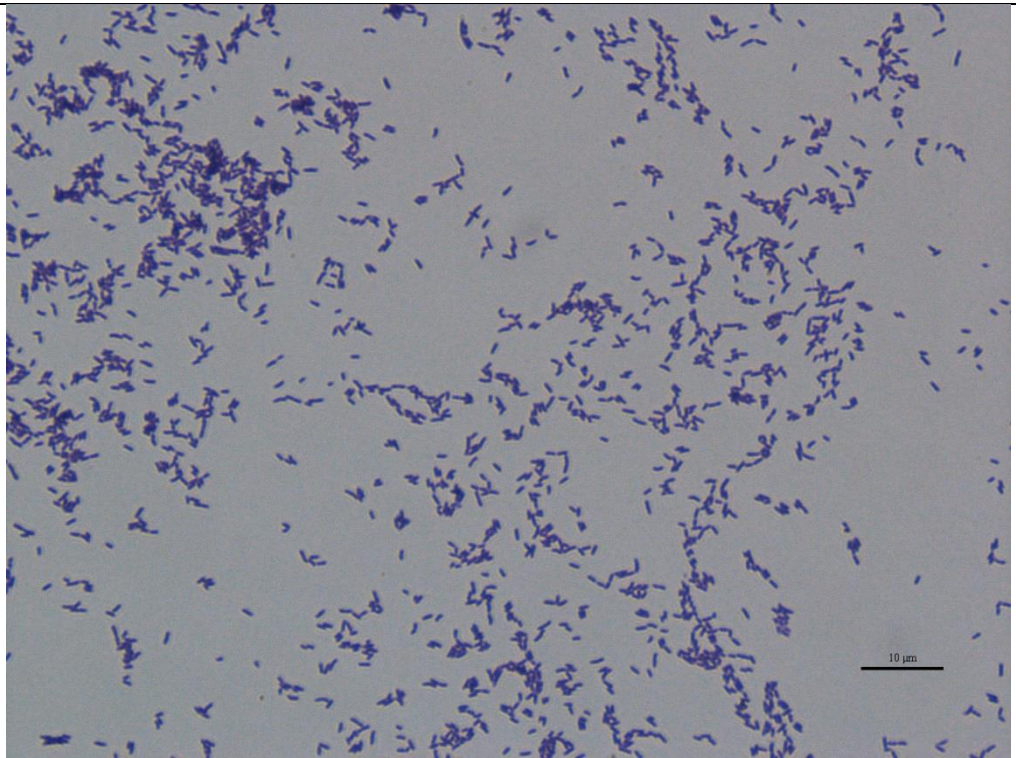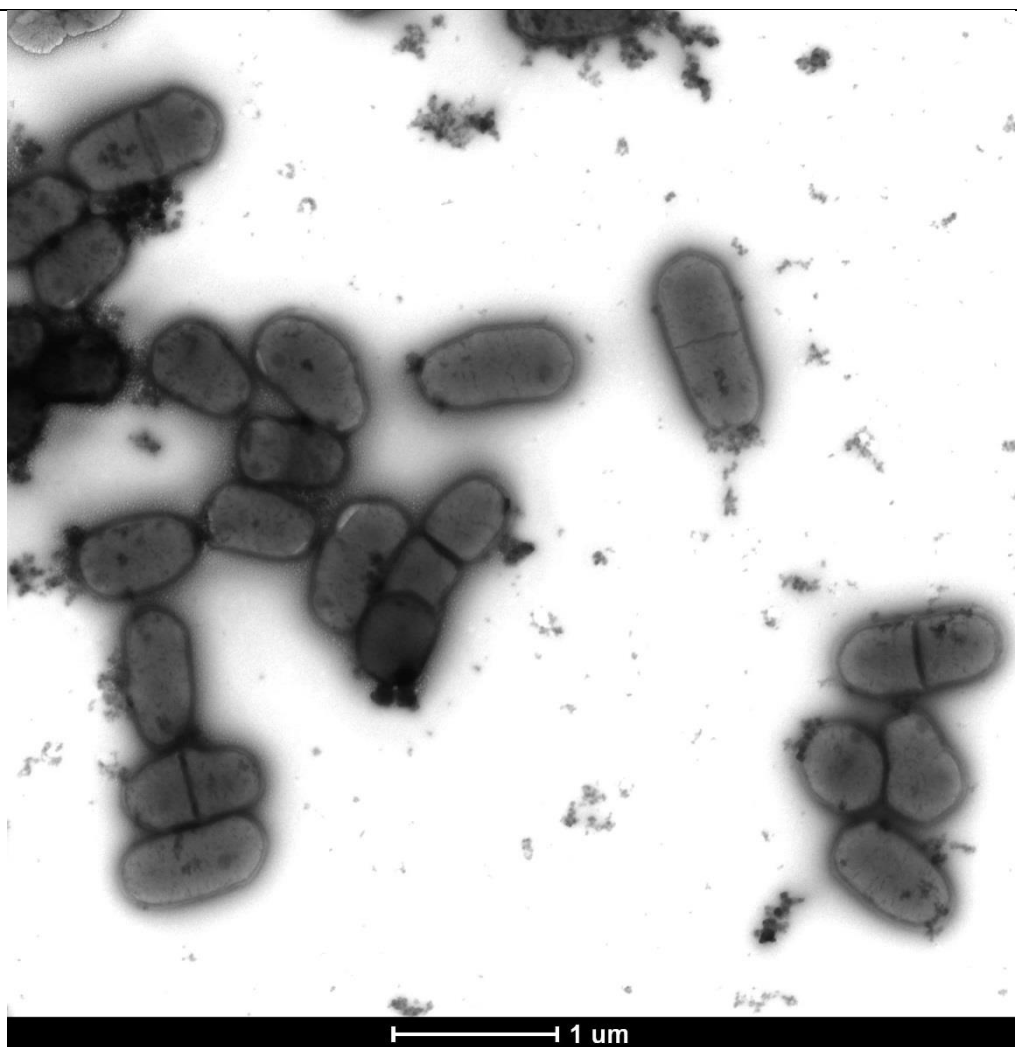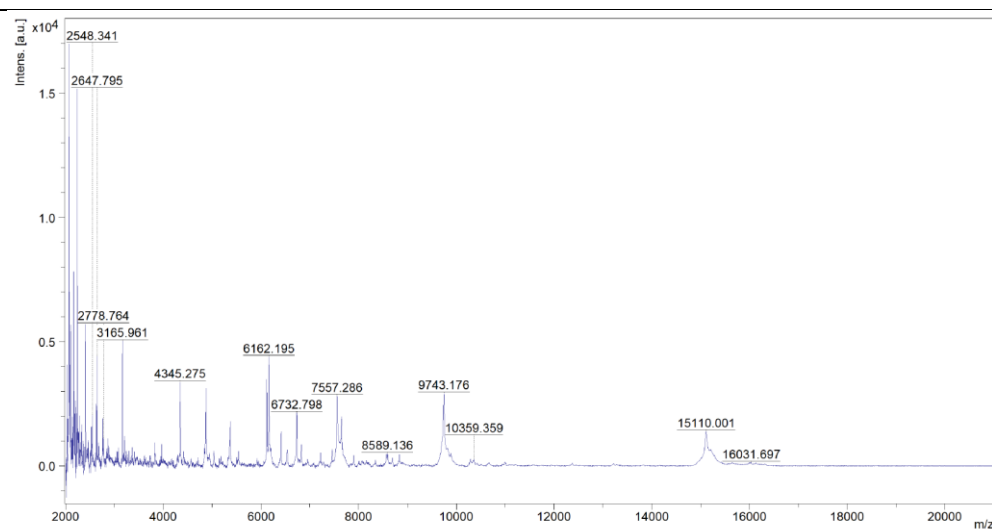
## 3.3.2 Biochemical properties and AST profile

While the biochemical properties of R131 are summarized in Table 3-2, the susceptibility of R131 to fourteen gram-positive spectrum antibiotics is summarized in Table 3-3. Indole production, catalase activity, and oxidase activity were found to be negative. R131 demonstrated the ability to utilize certain carbohydrates, including D-galactose, D-cellobiose, D-glucose, D-mannose, D-ribose, maltotriose, L-arabinose, and D-xylose. However, it did not utilize D-maltose and sucrose. Metabolism of N-Acetyl-D-glucosamine, pyruvate, and phenylphosphonate was observed, while the hydrolysis of esculin and arginine was not detected. R131 exhibited the production of various enzymes, such as leucine arylamidase, phenylalanine arylamidase, L-proline arylamidase, tyrosine arylamidase, Ala-Phe-Pro-arylamidase, beta-glucuronidase, beta-galactosidase, alpha-arabinosidase, beta-D-fucosidase, alpha-L-fucosidase, and urease. With reference to the breakpoints of other gram-positive bacteria based on Clinical and Laboratory Standards Institute (CLSI), R131 was susceptible to most of the tested antibiotics (12 out of 14), but it could have intermediate resistance to ciprofloxacin and clindamycin. However, it should be noted that this was only the inferred, not the confirmed, antimicrobial resistance profile of R131.

**Table 3-2. The biochemical properties of R131**

| Tests | Result |
| --- | --- |
| D-GALACTOSE | + |
| Leucine ARYLAMIDASE | + |
| ELLMAN | + |
| Phenylalanine ARYLAMIDASE | + |
| L-Proline ARYLAMIDASE | + |
| L-Pyrrolidonyl-ARYLAMIDASE | - |
| D-CELLOBIOSE | + |
| Tyrosine ARYLAMIDASE | + |
| Ala-Phe-Pro-ARYLAMIDASE | + |
| D-GLUCOSE | + |
| D-MANNOSE | + |
| D-MALTOSE | - |
| SACCHAROSE/SUCROSE | - |
| ARBUTIN | - |
| N-ACETYL-D-GLUCOSAMINE | + |
| 5-Bromo-4-chloro-3-indoxyl-beta-glucoside | - |
| UREASE | + |
| 5-Bromo-4-chloro-3-indoxyl-beta-glucuronide | + |
| BETA-GALACTOPYRANOSIDASE Indoxyl | + |
| ALPHA-ARABINOSIDASE | + |
| 5-Bromo-4-chloro-3-indoxyl-alpha-galactoside | - |
| BETA-MANNOSIDASE | - |
| ARGININE GP | - |
| PYRUVATE | + |
| MALTOTRIOSE | + |
| ESCULIN hydrolysis | - |
| BETA-D-FUCOSIDASE | + |
| 5-Bromo-4-chloro-3-indoxyl-beta-N-acetyl-glucosamide | (-) |
| 5-Bromo-4-chloro-3-indoxyl-alpha-mannoside | - |
| ALPHA-L-FUCOSIDASE | + |
| PHOSPHATASE | - |
| L-ARABINOSE | + |
| d-Ribose 2 | + |
| Phenylphosphonate | + |
| ALPHA-L-ARABINOFURANOSIDE | - |
| D-XYLOSE | + |
| Indole | - |
| Catalase | - |
| Oxidase | - |

**Table 3-3. AST profile of R131 to 14 antibiotics**

| Drug name | Reference breakpoint for susceptible (mm)* | Measured zone diameter (mm) | Inferred susceptibility |
|---|---|---|---|
| ampicillin (AMP 10) | ≥17-29 | 53 | susceptible |
| ceftriaxone (CRO 30) | ≥21-24 | 44 | susceptible |
| tetracycline (TE 30) | ≥19-23 | 47 | susceptible |
| vancomycin (VA 30) | ≥17 | 34 | susceptible |
| gentamicin (CN 10) | ≥15 | 28 | susceptible |
| meropenem (MEM 10) | ≥16 | 38 | susceptible |
| cefoxitin (FOX 30) | ≥22-25 | 43 | susceptible |
| chloramphenicol (C 30) | ≥18-24 | 38 | susceptible |
| erythromycin (E 15) | ≥21-23 | 32 | susceptible |
| co-trimoxazole (SXT 25) | ≥16-19 | 27 | susceptible |
| ciprofloxacin (CIP 5) | ≥21 | 20 | Intermediate |
| ceftazidime (CAZ 30) | ≥18 | 18 | susceptible |
| cefepime (FEP 30) | ≥18-24 | 39 | susceptible |
| clindamycin (DA 2) | ≥19-21 | 15 | Intermediate |

*Due to the lack of a standard breakpoint for the novel species, a reference breakpoint based on the CLSI guidelines was used. Since breakpoints vary among different species, the range of inhibition zone diameters indicating susceptibility for different gram-positive species is shown.

### 3.3.3 The quality of the assembly and the genome characteristics

A total of 238,874 reads with an average read length of 5,767.5 base pairs were obtained from Nanopore whole-genome sequencing, while 2,645,473 reads with an average read length of 247.5 bps (for each read 1 and read 2) were obtained from Illumina whole-genome sequencing. The assembly (CP138335.1), estimated to be approximately 2.22 Mbps in size, was accomplished with a genome coverage of 2,450x. No small plasmids were detected using the Unicycler hybrid assembly approach. Evaluation of the genome assembly using BUSCO (Benchmarking Universal Single-Copy Orthologs) analysis, which determines the presence of a set of highly conserved and

evolutionarily widespread genes in a given genome assembly (262). The BUSCO analysis revealed the presence of 122 complete and single copy BUSCOs (98.4%) and two fragmented BUSCOs (1.6%), indicating a high level of assembly completeness. The G+C content of the genome was determined to be 64 mol%.

Annotation of the genome using RASTtk identified a total of 2054 features, including 2001 coding sequences (CDSs) and 53 RNAs. RASTtk assigned 28% of the genes to subsystems, with 19 genes associated with the "Virulence, Disease, and Defense" subsystem. Among these, 10 genes were involved in resistance to antibiotics and toxic compounds, while nine genes were associated with invasion and intracellular resistance.

Annotation using PGAP identified a total of 2027 genes/pseudogenes, including 1970 CDSs and 57 RNAs. BlastKOALA analysis of the 1957 protein sequences encoded by the non-pseudo protein coding genes revealed that 60.3% were successfully annotated. The KEGG Mapper Reconstruction results from BlastKOALA included 198 pathways, 36 BRITE categories, and 33 complete modules.

Figure 3-1 illustrates the genetic map of the circular chromosome of R131, based on PGAP annotation, with selected features labeled from RAST and BlastKOALA annotations. Although RAST predicted genes associated with resistance to antibiotics and toxic compounds, and BlastKOALA assigned genes to KEGG pathways and BRITE categories related to antimicrobial resistance, no complete KEGG module for antimicrobial resistance was found. Furthermore, analyses using ResFinder, ResFinderFG 2.0, and AMRFinderPlus predicted no antimicrobial resistance in R131. However, resistance-associated point mutations could not be determined due to the lack of reference sequences in the database. Based on bioinformatics predictions and the

results of the antimicrobial susceptibility test, it can be concluded that R131 is susceptible to most

of the common antibiotics.

**Figure 3-1. Genetic map of the chromosome of R131.**
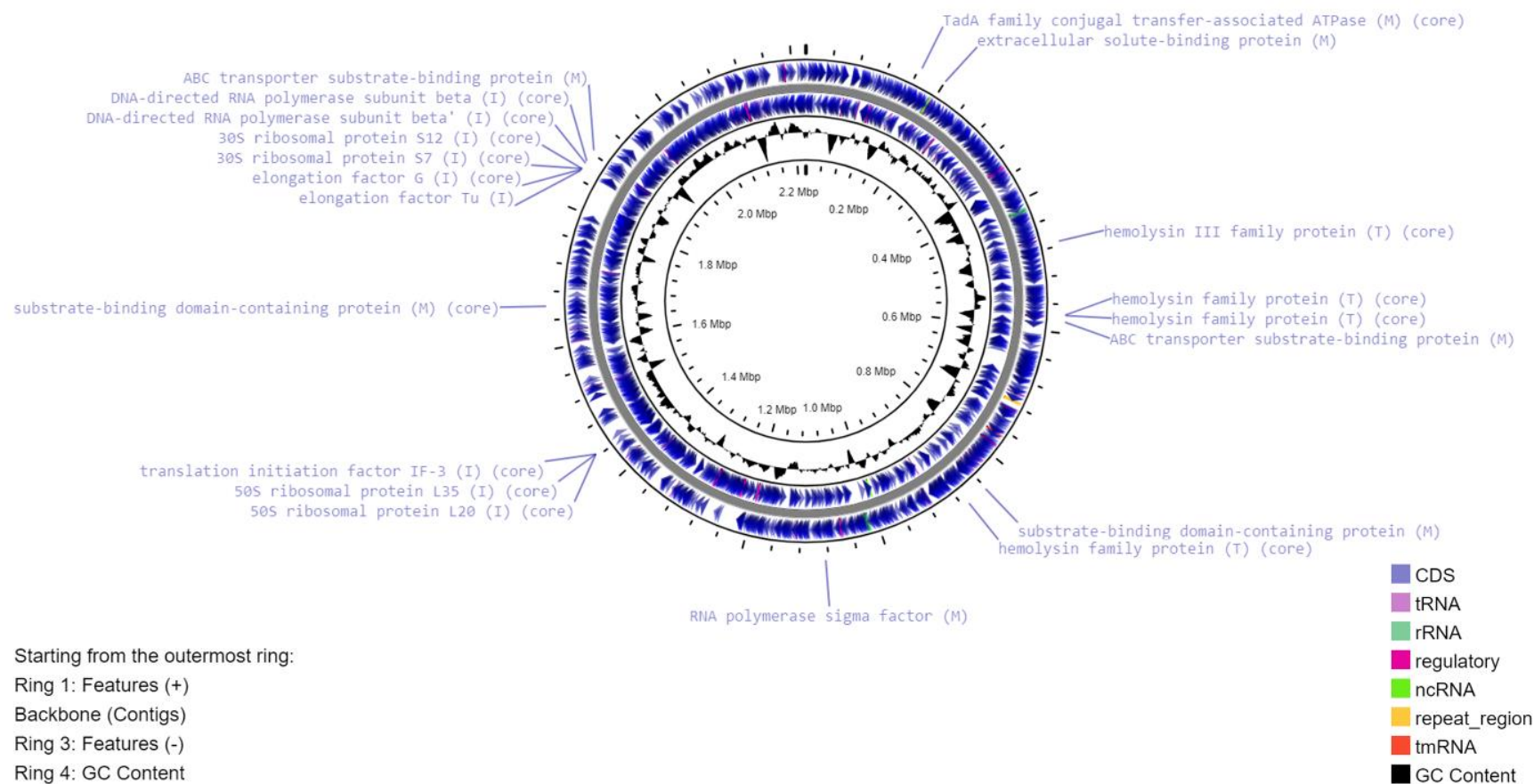


Ring 1 and Ring 3 represent the features annotated using PGAP. Selected CDS features are labeled with the corresponding protein product names and the subcategory: (I) = "Invasion and intracellular resistance" by RAST, (T) = Brite "Bacterial toxins", (M) = Brite "Bacterial motility proteins" or the "Cell motility" by BlastKOALA, and (core) = core genes from the pan-genome analysis

### 3.3.4 Taxonomic assignment of R131

Analysis of the 23S rRNA gene sequence of R131 against the SILVA LSU database yielded an "unclassified" status, while comparison with the SILVA SSU database assigned it to the genus *Actinomyces* based on the 16S rRNA gene sequence (deposited in GenBank under the accession number OR652275). Utilizing GTDB-Tk in KBase, R131 was initially classified as a member of the *Actinomyces*_I genus in GTDB release 202, which has since been revised to "*Scrofimicrobium*" in GTDB release 207. Notably, both the SILVA ACT and GTDB-Tk tools failed to identify the specific species of R131. Nevertheless, consistent results from both methods placed R131 within the *Actinomycetaceae* family, prompting focused phylogenetic and phylogenomic analyses comparing R131 to other *Actinomycetaceae* members.

In the 16S rRNA tree (Figure 3-2), R131 clustered with *Scrofimicrobium canadense*, and in the phylogenomic tree (Figure 3-3), it grouped alongside other members of the *Scrofimicrobium* genus. However, assessing the OrthoANI values between R131's genome and other *Scrofimicrobium* genomes revealed percentages ranging from 67.1% to 69.1%, falling below the 95% species threshold. Consequently, these findings indicate that R131 does not correspond to any known species within the *Scrofimicrobium* genus. Thus, it can be inferred that R131 represents a novel species within the genus *Scrofimicrobium*.

**Figure 3-2. Phylogenetic tree of the family *Actinomytaceae* based on 16S rRNA gene sequences.**
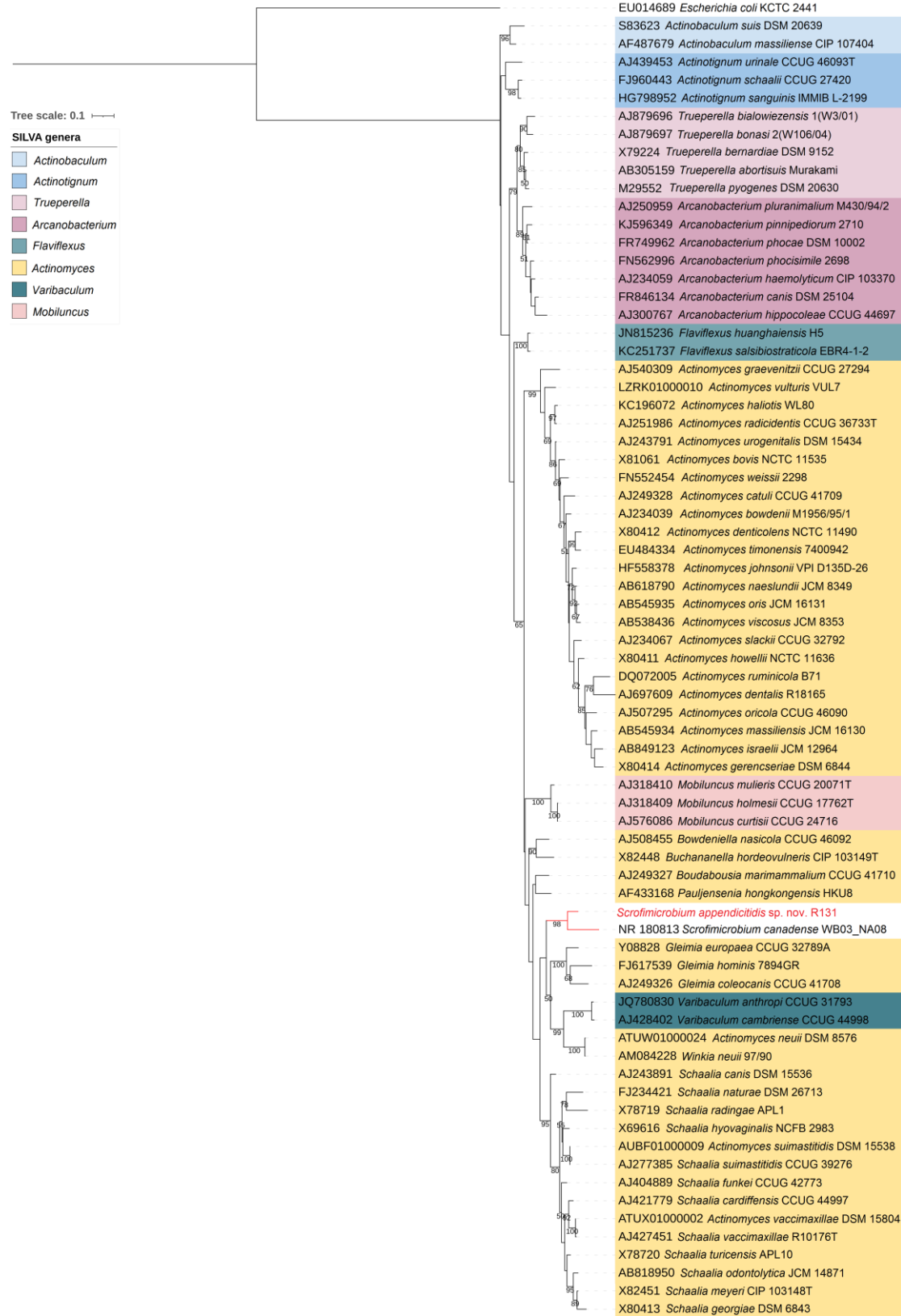
**Figure 3-3.** Phylogenomic tree of the family *Actinomytaceae* based on the 120 phylogenetically informative markers protein of bacteria in GTDB.
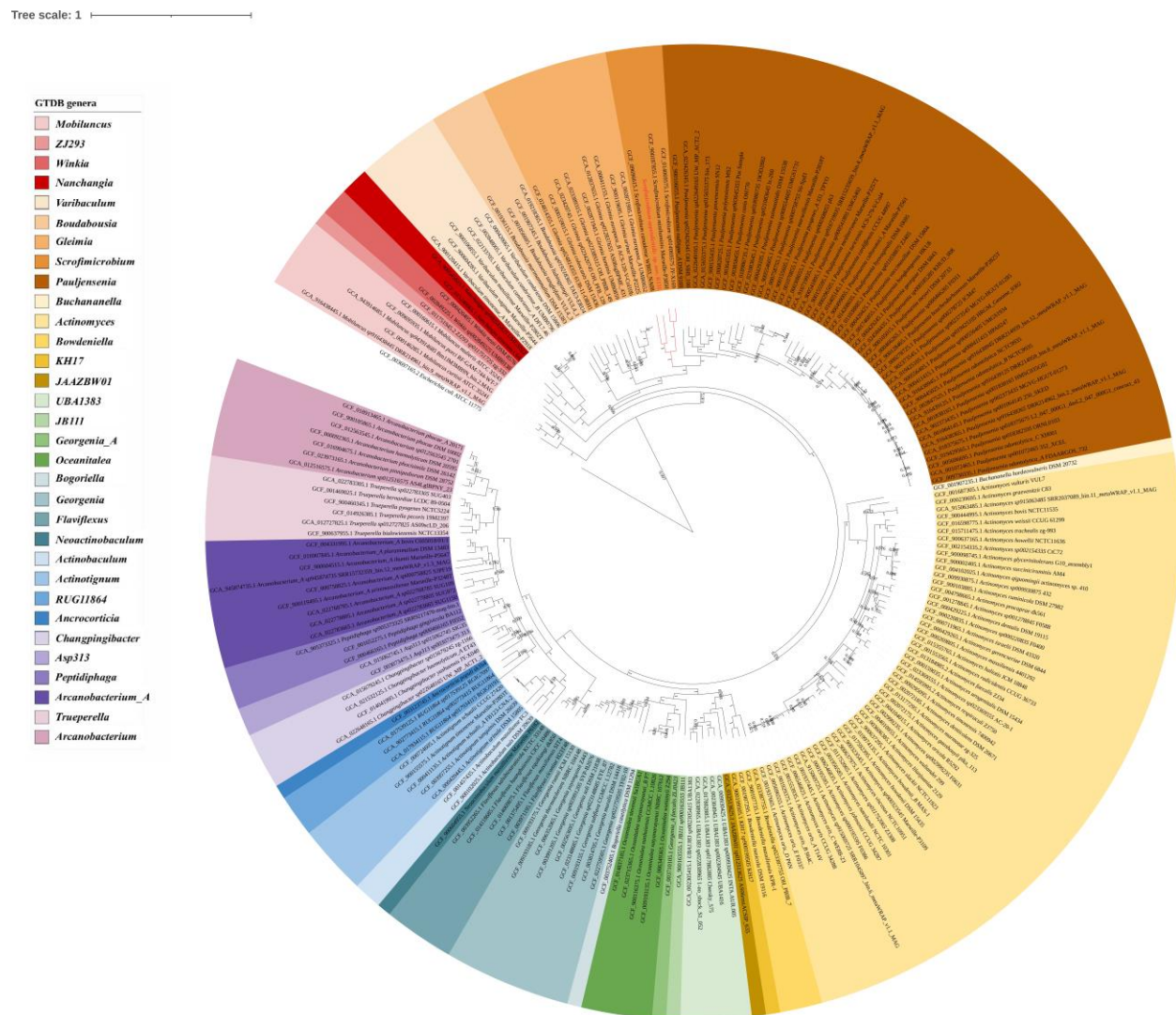
### 3.3.5 Result of pan-genome analysis

In the Genome Taxonomy Database (GTDB), three additional genomes are classified as *Scrofimicrobium*, namely *Actinomyces* (*Scrofimicrobium*) *minihominis* (GCF_900187855.1), *Schaalia* (*Scrofimicrobium*) sp JY-X159 (GCF_014525425.1), and *Schaalia* (*Scrofimicrobium*) sp JY-X169 (GCF_014069575.1). Therefore, these three genomes were examined alongside R131 and *Scrofimicrobium canadense* (GCF_009696615.1). A total of 3994 gene clusters were identified, with 986 core gene clusters shared among the five genomes (Table 3-3). For each core gene cluster, a representative sequence in R131 was subjected to annotation using BlastKOALA, and the summarized results can be found in Supplementary 3.

The cloud genes, which are genes present in two or fewer genomes, varied from 361 to 1105 genes across the genomes. Analysis of the ANI values of the clustered sequences (Table 3-4) and the phylogenetic tree (Figure 3-4) demonstrated that R131 exhibited considerable distance from the other four genomes, providing evidence for its classification as a novel species within the genus *Scrofimicrobium*.

**Table 3-3. Number of gene clusters in the 5 genomes.**

|  | Cloud | Shell | Soft-core | Core |
|---|---|---|---|---|
| *Scrofimicrobium canadense* | 1105 | 114 | 1155 | 986 |
| R131 | 601 | 109 | 1190 | 986 |
| *Scrofimicrobium minihominis* | 361 | 204 | 1175 | 986 |
| *Scrofimicrobium* sp JY-X159 | 434 | 220 | 1209 | 986 |
| *Scrofimicrobium* sp JY-X169 | 467 | 250 | 1245 | 986 |

**Table 3-4. ANI values\* of the clustered sequences in the 5 genomes**

| | *Scrofimicrobium canadense* | R131 | *Scrofimicrobium minihominis* | *Scrofimicrobium sp* JY-X159 | *Scrofimicrobium sp* JY-X169 |
|---|---|---|---|---|---|
| *Scrofimicrobium canadense* | 100 | 62.51 | 59.7 | 59.34 | 58.68 |
| R131 | 62.51 | 100 | 60.13 | 60.31 | 59.78 |
| *Scrofimicrobium minihominis* | 59.7 | 60.13 | 100 | 74.91 | 74.51 |
| *Scrofimicrobium sp* JY-X159 | 59.34 | 60.31 | 74.91 | 100 | 96.82 |
| *Scrofimicrobium sp* JY-X169 | 58.68 | 59.78 | 74.51 | 96.82 | 100 |

\*ANI values above 95% indicate that the two genomes likely belong to the same species (263).

**Figure 3-4. The pan-genomic phylogenetic tree of the 5 genomes**
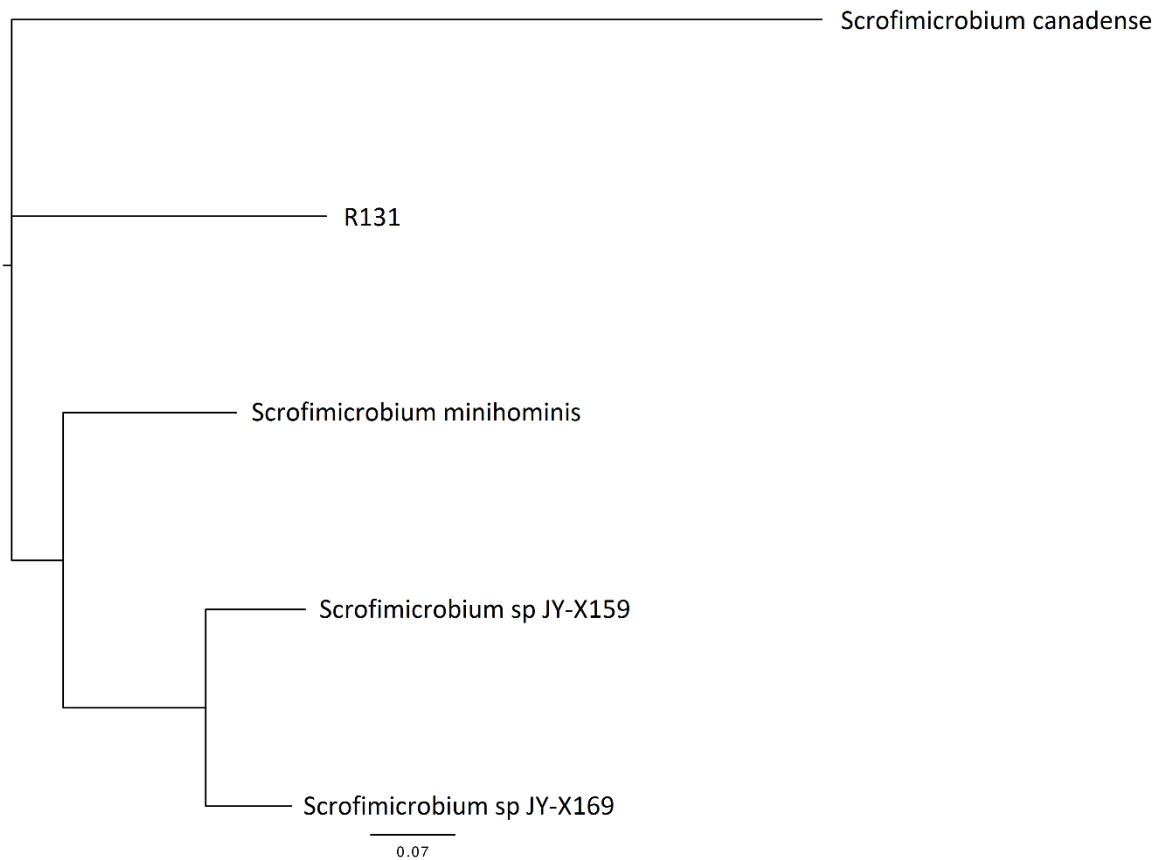
### 3.3.6 Result of dDDH analysis

A comparison with the four genomes and R131 resulted in a dDDH value below 70% (Table 3-5),

suggesting that R131 represents a distinct species from *Scrofimicrobium canadense* WB03_NA08,

*Scrofimicrobium minihominis* Marseille-P3850, *Scrofimicrobium* sp. JY-X159, and

*Scrofimicrobium* sp. JY-X169. Furthermore, the dDDH value of 84.1% between *Scrofimicrobium*

sp. JY-X159 and *Scrofimicrobium* sp. JY-X169 indicated that they belong to the same species.

**Table 3-5. The dDDH values obtained through a comparison of the 5 genomes using GGDC 3.0, formula 2 (DDH calculated based on identities/HSP length).**

|  | *Scrofimicrobium canadense* WB03_NA08 | *Scrofimicrobium minihominis* Marseille-P3850 | *Scrofimicrobium* sp. JY-X159 | *Scrofimicrobium* sp. JY-X169 |
|---|---|---|---|---|
| R131 | 22.20% | 19.20% | 19.70% | 22.30% |
| *Scrofimicrobium canadense* WB03_NA08 |  | 20.90% | 20.20% | 21.50% |
| *Scrofimicrobium minihominis* Marseille-P3850 |  |  | 19.00% | 19.20% |
| *Scrofimicrobium* sp JY-X159 |  |  |  | 84.1% |

### 3.3.7 Result of POCP analysis

Considering the discrepancies in genus assignments between NCBI and GTDB for *Actinomyces*

(*Scrofimicrobium*) *minihominis* Marseille-P3850, *Schaalia* (*Scrofimicrobium*) sp. JY-X159, and

*Schaalia* (*Scrofimicrobium*) sp. JY-X169, POCP analysis was conducted for genus assignment of

R131. The type species of *Actinomyces* and *Schaalia*, namely *Actinomyces bovis* NCTC 11535

(GCF_900444995.1) and *Schaalia odontolytica* NCTC9935 (GCF_900445025.1), were also included in the POCP analysis. The result of POCP analysis was summarized in Table 3-6.

When compared with the four *Scrofimicrobium* genomes based on GTDB, R131 exhibited a POCP value exceeding 50%, suggesting that they belong to the same genus. Conversely, the POCP value between *Actinomyces bovis* NCTC 11535 and *Actinomyces* (*Scrofimicrobium*) *minihominis* Marseille-P3850, as well as those between *Schaalia odontolytica* NCTC9935 and the two *Schaalia* (*Scrofimicrobium*) species (JY-X159 and JY-X169), fell below the 50% threshold, indicating that they are from different genera.

**Table 3-6. POCP values of the studied genomes.**

| | *Scrofimicrobium canadense* WB03_NA08 | R131 | *Scrofimicrobium minihominis* Marseille-P3850 | *Scrofimicrobium* sp. JY-X159 | *Scrofimicrobium* sp. JY-X169 | *Actinomyces bovis* NCTC 11535 | *Schaalia odontolytica* NCTC 9935 |
|---|---|---|---|---|---|---|---|
| *Scrofimicrobium canadense* WB03_NA08 | 100 | 57.48 | 53.86 | 50.96 | 51.38 | 38.22 | 42.50 |
| R131 | 57.48 | 100 | 60.94 | 58.82 | 59.12 | 44.13 | 49.16 |
| *Scrofimicrobium minihominis* Marseille-P3850 | 53.86 | 60.94 | 100 | 71.95 | 71.12 | 40.81 | 44.22 |
| *Scrofimicrobium* sp. JY-X159 | 50.96 | 58.82 | 71.12 | 100 | 87.17 | 40.73 | 45.26 |
| *Scrofimicrobium* sp. JY-X169 | 51.39 | 59.12 | 71.95 | 87.17 | 100 | 39.84 | 44.21 |
| *Actinomyces bovis* NCTC 11535 | 38.22 | 44.13 | 40.81 | 40.73 | 39.84 | 100 | 47.44 |
| *Schaalia odontolytica* NCTC 9935 | 42.50 | 49.16 | 44.22 | 45.26 | 44.21 | 47.44 | 100 |

## 3.4 Discussion and conclusion

The clinical isolate, R131, was confirmed as a novel bacterial species based on phenotypic, phylogenetic, phylogenomic, pan-genome, dDDH, and POCP analyses. The type strain was deposited in two recognized culture collections: the Japan Collection of Microorganisms with accession number JCM 36615$^T$ and the Belgian Coordinated Collections of Microorganisms with accession number LMG 33627$^T$. Considering the discovery of this novel species in patient with appendicitis, R131 was officially described and published in the International Journal of Systematic and Evolutionary Microbiology (IJSEM) under the proposed name *Scrofimicrobium appendicitidis* (264).

Phylogenetic analysis revealed that R131 exhibited the closest relationship to *Scrofimicrobium canadense*, sharing a 94.15% nucleotide identity with its 16S rRNA gene. Notably, *Scrofimicrobium* is a relatively recent genus proposed by Wylensek et al. in 2021 (265) and is classified under the *Actinomycetaceae* family. According to the NCBI taxonomy database, *Scrofimicrobium canadense* is the sole species currently classified within this genus, initially isolated from the pig intestine in Canada (265). However, in the previous chapter, the best-matched species for R131 based on Sanger 16S and Nanopore 16S sequencing was not *Scrofimicrobium canadense*, primarily due to the absence of the reference 16S sequence of *Scrofimicrobium canadense* in the reference database during the study period. The 16S rRNA gene of *Scrofimicrobium canadense* (NR_180813.1) was recorded in NCBI in November 2022, while the study in the previous chapter was published in the Journal of Clinical Microbiology in January 2022 (216).

Traditional culture-based identification methods often fall short in discovering novel species, whereas sequencing-based diagnostic tools like 16S rRNA gene sequencing can offer genetic

insights into potential novel species. In addition to rapid identification, sequencing can be utilized to elucidate the evolutionary relationships among various species or groups of organisms. This chapter demonstrated the utilization of WGS and phylogenomics for the taxonomic assignment of R131, which was a proposed novel bacterial species based on the sequencing findings in chapter two. Moreover, it highlights the importance of regularly updating reference databases for accurate species identification in clinical microbiology.

# Chapter 4: The clinical utility of Nanopore 16S for direct bacterial identification in invasive bacterial infections

## 4.1 Introduction

Chapter 2 has demonstrated that the diagnostic accuracy of Nanopore 16S is comparable to Sanger 16S in identifying clinical isolates. However, unlike Sanger 16S, Nanopore 16S has the added advantage of being able to identify mixed 16S sequences from multiple species within a single sample. This capability allows for direct bacterial identification from clinical specimens without the need for lengthy incubation times, thereby further reducing the sample-to-report time.

While several studies have shown the utility of MALDI-TOF MS for direct identification of monomicrobial blood cultures without pure isolated colonies (266-269), the confidence score of MALDI-TOF MS on direct specimens is often lower compared to pure cultures due to lower microbial concentration and the presence of competing proteins (266, 270). The reported correct identification rates at the species level, with a confidence score ≥2.0, range from 35.1% to 87% for monomicrobial blood cultures (266-269, 271). Additionally, identifying polymicrobial samples using MALDI-TOF MS is challenging as the peaks of multiple species are merged into a single spectrum (267, 272). Hence, MALDI-TOF MS tends to report only the predominant species in polymicrobial samples (272). In a study by Chien et al., MALDI-TOF MS correctly identified one of the isolated species in 33 out of 40 (82.5%) polymicrobial blood cultures, with only two samples had two species correctly identified by MALDI-TOF MS (271). Therefore, MALDI-TOF MS still relies on culture for accurate identification, especially for polymicrobial samples.

Considering the comparable diagnostic accuracy to Sanger 16S, the culture-independent workflow, the capability to identify multiple species in a sample, and real-time analysis, Nanopore 16S is

suggested for rapid bacterial identification in IBIs. This study aimed to assess the clinical utility of Nanopore 16S in accurately identifying bacterial pathogens directly from 213 normally sterile body fluids, with reference to the culture results (273). The sequencing reads were classified using three analysis pipelines with different algorithms, including Epi2me, NanoCLUST, and Emu, to reduce classification bias. Moreover, a threshold of relative abundance will be determined using receiver operating characteristic (ROC) analysis to discriminate potential pathogens from sequencing noises.

## 4.2 Materials and methods

### 4.2.1 Sample collection and preparation

A total of 213 residual body fluids, accompanied by their corresponding culture results, were collected from the clinical microbiology laboratories of four public hospitals in Hong Kong: Pamela Youde Nethersole Eastern Hospital, Prince of Wales Hospital, Princess Margaret Hospital, and Tuen Mun Hospital. Among the collected samples, 193 were culture-positive, comprising 128 cases of monomicrobial infections and 65 cases of polymicrobial infections. Additionally, there were 20 culture-negative samples. The specimen type and culture results of each sample was summarized in Supplementary 4. DNA extraction was performed using the QIAamp BiOstic Bacteremia DNA Kit (Qiagen, Hilden, Germany).

### 4.2.2 Nanopore 16S and sequencing data analysis

Nanopore 16S was performed using 16S Barcoding Kit 1 – 24 (SQK-16S024) from ONT, with some modifications. To account for the challenge of quantifying bacterial DNA in clinical

specimens with a high abundance of human DNA, 15 µl of DNA extract was utilized for the 16S PCR instead of the suggested 10 ng DNA input. Additionally, the PCR cycle was increased from 25 to 35 cycles to enhance the assay's sensitivity. Following normalization, the pooled library was subjected to sequencing on the GridION sequencer for a maximum of 24 hours, employing the super-accuracy basecalling model. To minimize index misassignment, "mid-read barcode filtering", "barcode both ends", and a minimum barcoding score of 85 were adopted in the nanopore sequencing.

The sequencing reads were uploaded on Epi2me for real-time analysis using the FASTQ 16S workflow (v2021.09.09) with a minimum QSCORE of 10 was applied. To address the lower read accuracy of nanopore sequencing, only reads between 1000 to 2000 base pairs were included, and a minimum coverage and identity of 90% were adopted. Additionally, the reads underwent further analysis using Emu (205) and NanoCLUST (204) pipelines with default parameters.

### 4.2.3 Data and statistical analysis

For culture-positive samples, the concordance between culture and the Nanopore 16S, along with the three analysis pipelines, was determined by (number of cultured species detected by Nanopore 16S)/ (total number of cultured species). Regarding culture-negative samples, if clinically important species were identified using Nanopore 16S, the results were correlated with the patients' clinical manifestations and medical history. All statistical analyses were conducted using GraphPad Prism (v9.5.0). A significance threshold of $p < 0.05$ was applied to determine statistical significance. The Mann-Whitney U test was employed to assess the statistical differences between populations of sequencing reads and classified species.

### 4.2.4 Determination of threshold of relative abundance for detecting potential pathogens

The relative abundance of a species in a sample was determined by (number of reads of a species in a sample)/ (total number of classified reads of a sample). To establish the threshold of relative abundance ($T_{RA}$) for detecting potential pathogens in body fluids using Nanopore 16S, ROC curves were generated for each analysis pipeline. The ROC curves were based on the relative abundance of true positives (detected by both culture and Nanopore 16S) and false positives (detected only by Nanopore 16S). Separate ROC analyses were conducted for monomicrobial and polymicrobial samples. The $T_{RA}$ for detecting pathogens in each analysis pipeline was determined by identifying the optimal point on the ROC curve with the maximum Youden's index. The average value of the $T_{RA}$ obtained from the three pipelines was considered the $T_{RA}$ for Nanopore 16S.

### 4.2.5 Limit of detection (LOD)

The LOD of Nanopore 16S was determined using simulated bacteremic blood samples. Two reference strains, *Staphylococcus aureus* BAA-3114 and *Klebsiella pneumoniae* BAA-3079, were spiked into EDTA-blood at final concentrations of 150, 100, 50, and 10 CFU/ml, respectively. The spike-in samples were processed, sequenced, and analyzed as the clinical samples. The experiment was performed three times, from spike-in to sequencing, to minimize random errors introduced during experimental procedures, such as spike-in preparation, DNA extraction, and sequencing. The average relative abundances of each species in Nanopore 16S were plotted against bacterial concentrations, and a trend line was used to identify the minimal bacterial concentration meeting the $T_{RA}$.

## 4.3 Results

### 4.3.1 Commonly encountered bacterial pathogens in body fluids

In the analysis of 128 monomicrobial body fluids, the most frequently identified pathogen in culture was coagulase-negative *Staphylococci*, detected in 21 out of 128 samples. This was followed by *Staphylococcus aureus* (20/128), *Escherichia coli* (12/128), and *Enterococcus faecalis* (12/128).

Among the 65 polymicrobial body fluids, the most commonly isolated pathogen in culture was *Escherichia coli*, found in 42 out of 65 samples. This was followed by *Klebsiella pneumoniae* (16/65) and *Bacteroides fragilis* (14/65).

The results obtained from Nanopore 16S analysis are summarized in Supplementary 5, which provides information on the twelve most abundant species and their corresponding relative abundances in each sample as classified by the three analysis pipelines.

### 4.3.2 Statistics of sequencing reads

Following 24-hour sequencing, the monomicrobial samples yielded an average of 76,551 reads per sample (s.d. ±91,867), with a median of 32,699 reads. In contrast, the polymicrobial samples exhibited a substantially higher number of reads per sample compared to the monomicrobial samples ($p < 0.0001$), with an average of 160,172 reads (s.d. ±168,470) and a median of 127,662 reads per sample.

Significant differences were observed between the culture-negative and culture-positive samples, with the culture-negative samples displaying a significantly lower number of reads ($P < 0.0001$).

The average number of reads in the culture-negative samples was 23,442 (s.d. ±54,352), with a median of 1,373 reads per sample.

### 4.3.3 Comparison of the speed and resources requirement of the three analysis pipelines

Among the local analysis tools, NanoCLUST demonstrates faster analysis times compared to Emu, as it clusters highly similar reads prior to taxonomic assignment. Furthermore, NanoCLUST leverages multiple CPU threads to parallelize the classification of different clusters, significantly reducing execution time. In contrast, Emu processes individual reads by aligning them directly to the reference database, followed by an error correction step, which results in longer analysis times. The computation time varies based on sequencing output; in this study, NanoCLUST generally completed the analysis of a sample within a minute, whereas Emu required several minutes per sample. However, NanoCLUST requires higher computational resources, with its clustering step demanding up to 32–36 GB of RAM. In contrast, Emu has lower memory requirements, as demonstrated in the study by Curry et al. (205), where Emu utilized 5–22.7 GB of RAM compared to NanoCLUST's 16–33.8 GB of RAM.

### 4.3.4 Concordance between culture and Nanopore 16S

Figure 4-1 provides an overview of the concordance between culture and Nanopore 16S results, while detailed records of the concordance for each monomicrobial and polymicrobial sample were summarized in Supplementary 6 and 7, respectively. In both monomicrobial and polymicrobial samples, Nanopore 16S coupled with the Emu analysis pipeline demonstrated the highest concordance. Among the 128 monomicrobial samples, Emu correctly identified the taxon of 125 samples (97.7%), compared to 109 samples (85.2%) and 102 samples (79.7%) identified by Epi2me and NanoCLUST, respectively. Among the 65 polymicrobial samples, Emu correctly

identified all cultured species in 35 samples (53.8%), compared to 19 samples (29.2%) and 9 samples (13.8%) in Epi2me and NanoCLUST, respectively.

Traditionally, a 97% sequence similarity threshold in the 16S rRNA gene was commonly used to differentiate between two species (140, 227, 274). However, some closely related species can exhibit sequence similarities of 98% or higher (274-276). Table 4-1 provides a list of closely related species with a sequence similarity of ≥ 98% found in this study. Considering these closely related species as concordant classifications increased the number of correctly detected samples for Epi2me. In this case, the number of correctly detected samples for Epi2me, Emu, and NanoCLUST in monomicrobial samples increased to 127 (99.2%), 126 (98.4%), and 119 (93.0%), respectively. For the 65 polymicrobial samples, Epi2me successfully detected all cultured species in 50 samples (76.9%), compared to 37 samples (56.9%) and 30 samples (46.2%) in Emu and NanoCLUST, respectively. Partially concordant samples were defined as those where Nanopore 16S failed to detect all, but at least one, of the cultured species. In Epi2me, Emu, and NanoCLUST, a total of 15 samples (23.1%), 28 samples (43.1%), and 33 samples (50.8%) of the polymicrobial samples were partially concordant with the culture results, respectively.

Figure 4-2 presents the concordance between culture and Nanopore 16S in the 230 species cultured from the 65 polymicrobial body fluids. When closely related species were not considered, Nanopore 16S combined with Emu showed the highest concordance (81.7%), compared to 75.7% for Epi2me and 54.3% for NanoCLUST. If closely related species were considered concordantly classified, Epi2me had the highest concordance (91.7%), followed by Emu (83.0%), and NanoCLUST (72.6%).

Figure 4-3 presents the overall concordance of Nanopore 16S with culture in the 193 culture-positive body fluids. When closely related species were not considered, the overall concordance

was 79.1% for Epi2me, 87.4% for Emu, and 63.4% for NanoCLUST. Considering closely related species, the overall concordance would be 94.4% for Epi2me, 88.6% for Emu, and 79.9% for NanoCLUST.

There were cases where Nanopore 16S failed to detect the expected species based on the culture results. Among the monomicrobial samples, the number of discordant samples was 1 (0.8%) for Epi2me, 2 (1.6%) for Emu, and 6 (4.7%) for NanoCLUST. Notably, three monomicrobial samples were unable to undergo analysis by NanoCLUST, possibly due to insufficient reads (<120 reads) for cluster generation. Regarding the polymicrobial samples, two samples exhibited complete discordance exclusively in NanoCLUST, as none of the cultured species were detected. Conversely, both Epi2me and Emu successfully detected at least one of the cultured species in each sample.

**Figure 4-1. An overview of concordance between culture and Nanopore 16S coupled with the three analysis pipelines.**
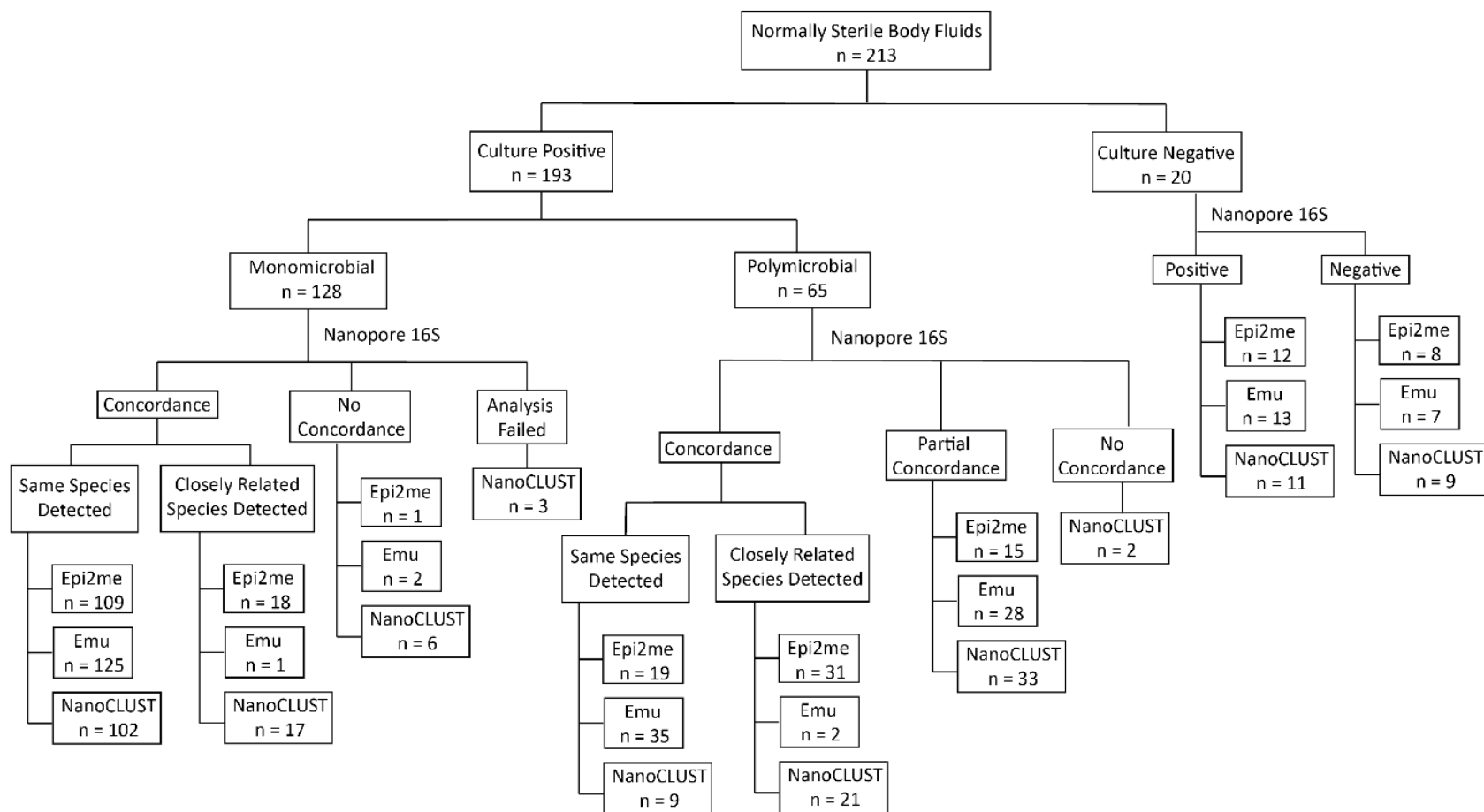
**Table 4-1: The closely related species with sequence similarity ≥ 98% in 16S rRNA gene.**

| Culture species | Epi2me (%identity) | Emu (%identity) | Nanoclust (%identity) |
|---|---|---|---|
| *Acinetobacter ursingii* | *Acinetobacter septicus* (99.1%) | | *Acinetobacter septicus* (99.1%) |
| *Burkholderia cepacia complex* | *Burkholderia cepacia* (99.28%, 99.52%) | *Burkholderia cenocepacia* (99.28%, 99.36%) | *Burkholderia metallica* (99.52%, 99.36%) |
| *Burkholderia pseudomallei* | | | *Burkholderia mallei* (99.8%) |
| *Campylobacter jejuni* | | | *Campylobacter subantarcticus* (99.18%) |
| *Citrobacter freundii complex* | *Citrobacter murliniae* (99.53%) | *Citrobacter freundii* (99.53%) | *Citrobacter murliniae* (99.53%) |
| *Comamonas kerstersii* | *Comamonas terrigena* (97.87%) | | *Comamonas terrigena* (97.87%) |
| *Enterococcus avium* | *Enterococcus raffinosus* (99.44%) | *Enterococcus gilvus* (99.39%) | |
| *Enterococcus raffinosus* | | *Enterococcus gilvus* (99.79%)/ *Enterococcus avium* (99.44%) | *Enterococcus avium* (99.44%) |
| *Escherichia coli* | *Shigella sonnei* (99.1%)/ *Escherichia fergusonii* (99.09%)/ *Shigella flexneri* (99.24%)/ *Escherichia marmotae* (98.21%)/ *Shigella dysenteriae* (98.4%)/ *Shigella boydii* (98.81%) | | *Escherichia fergusonii* (99.09%)/ *Escherichia marmotae* (98.21%)/ *Shigella flexneri* (99.24%) |
| *Klebsiella aerogenes* | | *Klebsiella variicola* (98.16%)/ *Klebsiella pneumoniae* (98.24%) | *Klebsiella pneumoniae* (98.24%) |
| *Klebsiella variicola* | | | *Klebsiella pneumoniae* (99.72%) |
| *Listeria monocytogenes* | *Listeria welshimeri* (98.1%) | | *Listeria welshimeri* (98.1%) |
| *Neisseria macacae* | | | *Neisseria sicca* (99.87%) |
| *Pantoea septica* | *Pantoea stewartii* (98.06%) | | *Pantoea stewartii* (98.06%) |
| *Paraclostridium bifermentans* | *Paraclostridium benzoelyticum* (99.39%) | | |
| *Proteus penneri* | | *Proteus vulgaris* (99.49%) | *Proteus mirabilis* (98.99%) |
| *Staphylococcus saprophyticus* | *Staphylococcus edaphicus* (99.94%) | | *Staphylococcus edaphicus* (99.94%) |

**Figure 4-2: The concordance between culture and Nanopore 16S coupled with (a) Epi2me, (b) Emu and (c) NanoCLUST, in 230 cultured species among the 65 polymicrobial body fluids.**
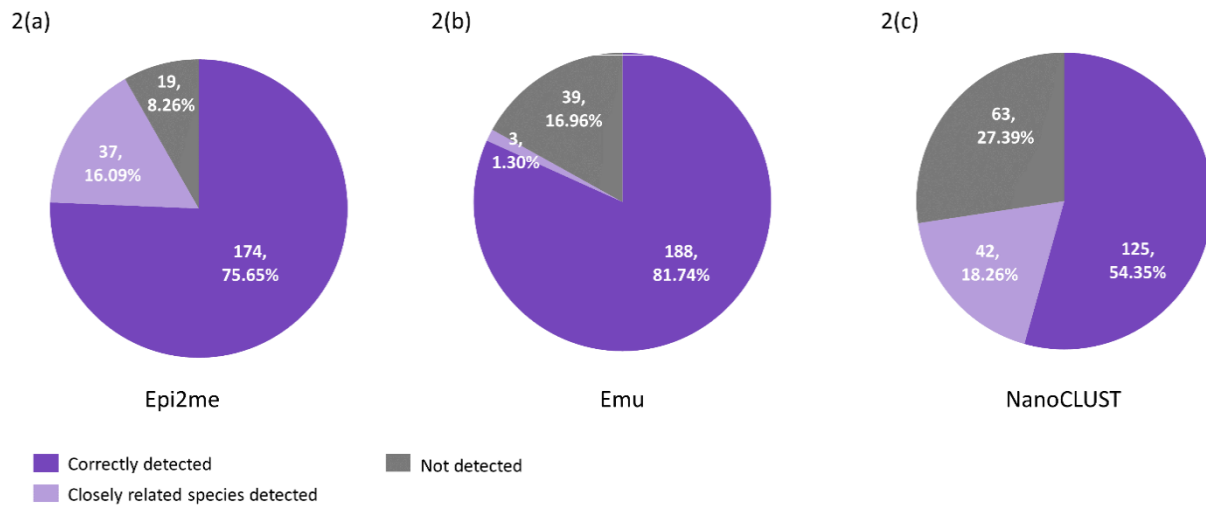


**Figure 4-3: The concordance between culture and Nanopore 16S coupled with (a) Epi2me, (b) Emu and (c) NanoCLUST, in 358 cultured species among the 193 culture-positive body fluids.**
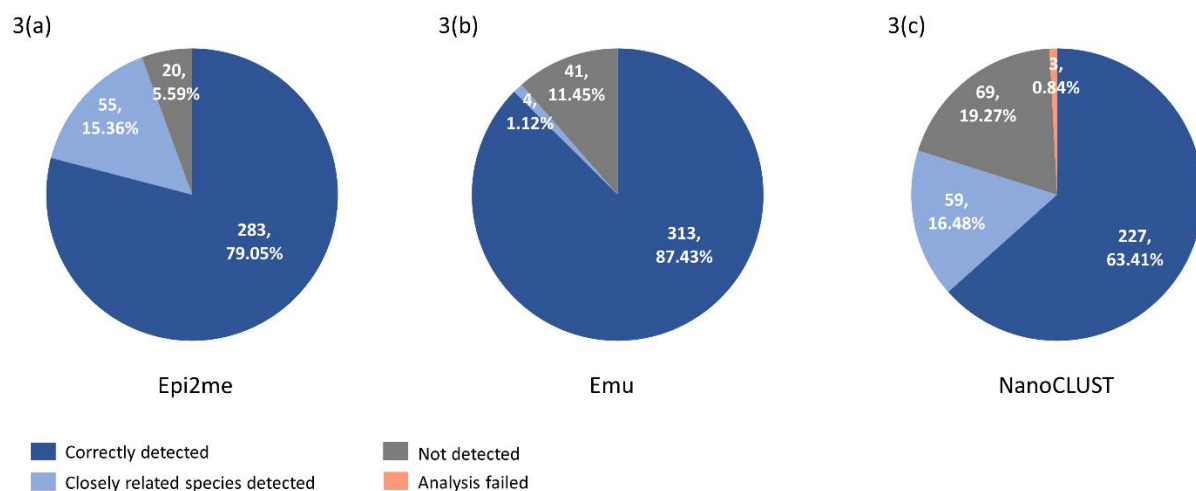
### 4.3.5 Average number of classified species across the three analysis pipelines and various specimen types

Figure 4-4 summarizes the number of classified species per sample in each analysis pipeline. In monomicrobial samples, even though only a single species was cultured, Nanopore 16S often classified multiple species (Figure 4a), regardless of the analysis pipelines. Epi2me, Emu, and NanoCLUST detected additional organisms, distinct from the cultured species, in 128 (100%), 122 (95.3%), and 106 (82.8%) samples, respectively. In monomicrobial samples, the average number of classified species per sample was 83.9, 15.1, and 7.7 for Epi2me, Emu, and NanoCLUST, respectively.

It is noteworthy that the cultured species did not always represent the most prevalent organisms in the correctly detected samples. In Epi2me, out of the 109 samples that were correctly detected, 27 samples exhibited cultured species that were not the most abundant among the classified species, including 3 samples with extremely low relative abundances (<0.1%). Among the 125 correctly detected samples in Emu, 37 samples did not have the most abundant species among the classified species, and one sample had an extremely low relative abundance (<0.1%). In the case of NanoCLUST, 25 out of the 102 correctly detected samples did not contain the most abundantly classified species.

Regarding the polymicrobial samples, regardless of the taxonomic classifiers used (Figure 4b), the number of classified species per sample was significantly higher (p < 0.0001) compared to monomicrobial samples. Among the three analysis pipelines, Epi2me exhibited the highest average number of classified species per sample, with 153 species classified on average, while Emu and NanoCLUST classified 62 and 30 species, respectively. Additional organisms other than expected were detected in 65 (100%), 64 (98.5%), and 62 (95.4%) samples by Epi2me, Emu, and

NanoCLUST, respectively. It is worth noting that among the 50 concordant samples in Epi2me, 32 samples (64.0%) had species with extremely low relative abundances (<0.1%), compared to 12 out of 37 (32.4%) concordant samples in Emu, and 1 out of 30 (3.33%) concordant samples in NanoCLUST. This indicates that Epi2me exhibited a higher sensitivity in detecting species with extremely low abundances.

The average number of classified species derived from Nanopore 16S coupled with the three analysis pipelines across different types of body fluids is summarized in Table 4-2. Notably, ascitic fluids and peritoneal fluids exhibited a significantly higher average number of classified species compared to other fluid types, with a predominant representation of gut microbiota species. Also, most of polymicrobial samples collected in this study were ascitic fluids and peritoneal fluids. However, these gut microbes are frequently overlooked in culture-based approaches and attempting to identify numerous gut microbes through culture methods is also impractical.

**Figure 4-4: The number of classified species per sample by the three analysis pipelines in (a) monomicrobial samples and (b) polymicrobial samples.**

**(a)**
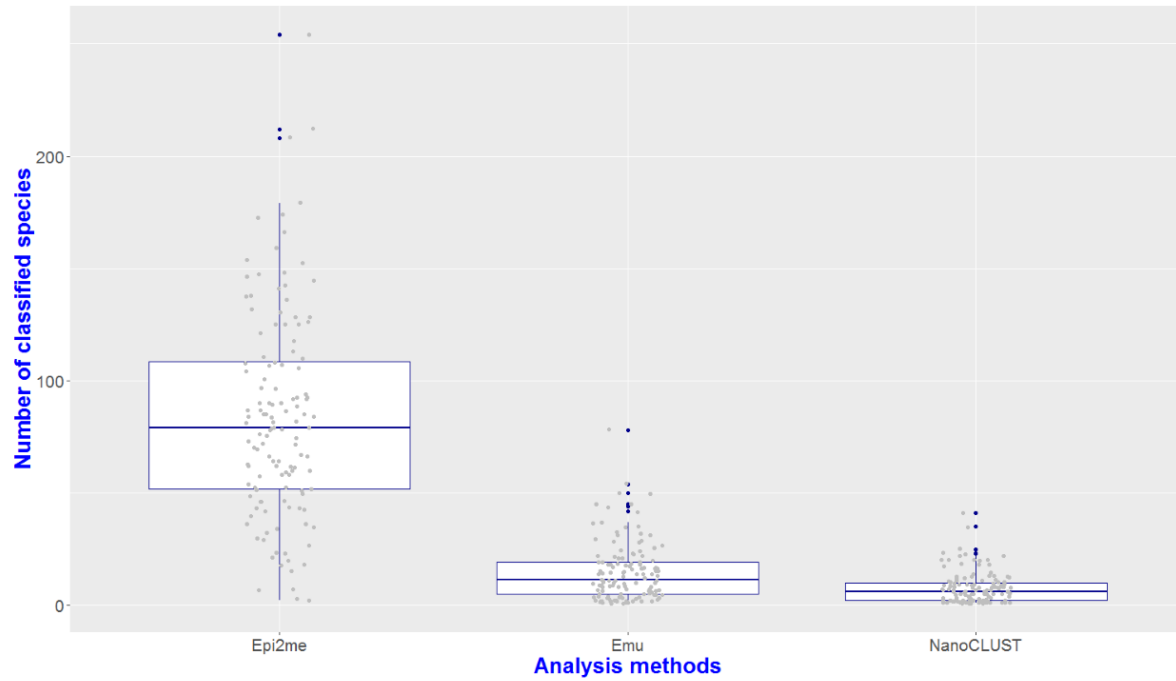


**(b)**
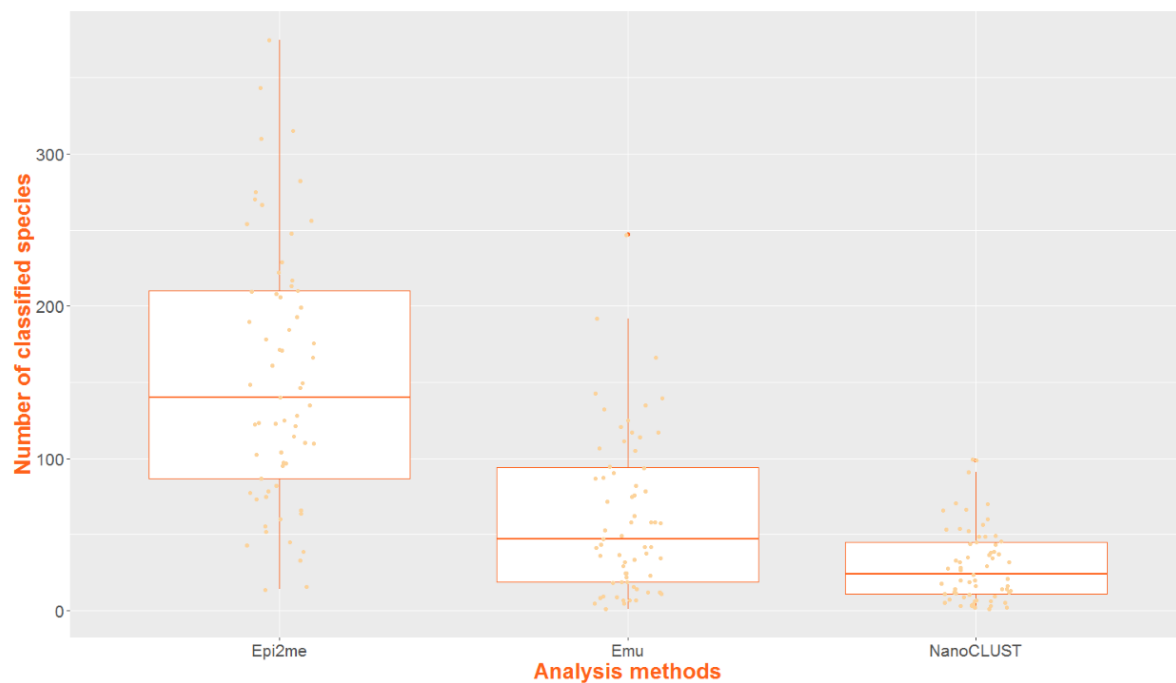
**Table 4-2: The average number of classified species by the three analysis pipelines in different types of body fluids.**

| Specimen type | Epi2me | Emu | NanoCLUST |
|---|---|---|---|
| Abscess (n=2) | 102 | 5.5 | 3 |
| Ascitic fluid (n=20) | 124.5 | 43.15 | 19.3 |
| Bile (n=1) | 190 | 25 | 14 |
| Cerebrospinal fluid (n=3) | 78 | 6 | 1.5 |
| Intrauterine Fluid (n=1) | 123 | 9 | 5 |
| Joint fluid / aspirate (n=24) | 74.1 | 14 | 5.7 |
| Midline laparotomy fluid (n=1) | 248 | 38 | 16 |
| Miscellaneous (n=1) | 178 | 36 | 19 |
| Pericardial fluid (n=2) | 23 | 18 | 9 |
| Peritoneal dialysis fluid (n=79) | 84.8 | 15.2 | 7.9 |
| Peritoneal fluid (n=55) | 134.2 | 59.6 | 30.4 |
| Pleural fluid / aspirate (n=24) | 80.7 | 16.8 | 8.3 |

## 4.3.6 Differentiation of *E. coli* and *Shigella* by the three pipelines

The higher concordance in Emu was attributed to its ability to differentiate *E. coli* from *Shigella* and other *Escherichia* species. In contrast, both Epi2me and NanoCLUST misidentified *E. coli* as other closely related species. Epi2me misidentified *E. coli* in 12 monomicrobial samples and 37 polymicrobial samples, while NanoCLUST misidentified it in 12 monomicrobial samples and 35 polymicrobial samples. In Epi2me, most of the reads assigned to *E. coli* were classified at the family level (*Enterobacteriaceae*), resulting in a lower relative abundance at the species level compared to Emu and NanoCLUST.

It is well-known that *E. coli* and *Shigella* are closely related species with highly similar 16S rRNA gene sequences. Although Emu showed better performance in identifying *E. coli* compared to the other pipelines, its ability to identify *Shigella* species remained uncertain due to the lack of *Shigella* species in the collected body fluids. To assess their capability in distinguishing between

*E. coli* and *Shigella*, Nanopore 16S was performed using isolates of *E. coli*, *S. sonnei*, and *S. flexneri*, and the resulting sequencing reads were analyzed using the three pipelines.

Table 4-3 presents the taxonomic classifications provided by the three analysis pipelines for the isolates of *E. coli* and *Shigella* species. Both Epi2me and NanoCLUST were unable to distinguish between *E. coli* and the two *Shigella* species, classifying them all as *E. fergusonii*. Additionally, Epi2me had a notably lower relative abundance of *E. fergusonii*, mainly because most of its reads were classified only at the family level. In contrast, Emu demonstrated accurate differentiation of all three species.

**Table 4-3: The taxonomic classification by the three analysis pipelines in differentiating *E. coli* and *Shigella* species.**

| Bacterial isolates | Total no. of reads | Epi2me | | Emu | | NanoCLUST | |
|---|---|---|---|---|---|---|---|
| | | 1st classified species | relative abundance | 1st classified species | relative abundance | 1st classified species | relative abundance |
| *E. coli* | 29,464 | *E. fergusonii* | 0.217 | *E. coli* | 0.999 | *E. fergusonii* | 1 |
| *S. sonnei* | 29,701 | *E. fergusonii* | 0.205 | *S. sonnei* | 0.749 | *E. fergusonii* | 1 |
| *S. flexneri* | 23,757 | *E. fergusonii* | 0.180 | *S. flexneri* | 0.986 | *E. fergusonii* | 1 |

### 4.3.7 Determination of $T_{RA}$ for detecting potential pathogens

ROC analyses were carried out to determine the relative abundance threshold for detecting potential pathogens in Nanopore 16S (Figure 4-5). In monomicrobial samples, $T_{RA}$ for Epi2me, Emu, and NanoCLUST were 0.043, 0.066, and 0.066, respectively, achieving ≥90% sensitivity and specificity. The average threshold of 0.058 across the three analysis pipelines was designated as the $T_{RA}$ for Nanopore 16S. Applying this threshold correctly identified 107 (83.6%), 117 (91.4%), and 114 (91.2%) monomicrobial samples for Epi2me, Emu, and NanoCLUST,

respectively. Notably, this included 20 (15.6%), 29 (22.7%), and 23 (18.0%) samples that were not identified as the most abundant species by Epi2me, Emu, and NanoCLUST, respectively.

In the case of polymicrobial samples, $T_{RA}$ was indeterminable since Nanopore 16S behaved as a random classifier. There were numerous species exclusively detected by Nanopore 16S being classified as "false positives" when compared to the culture results. However, these "false positives" may indeed be present in the body fluids but overlooked in culture-based methods. Application of the 0.058 threshold to the polymicrobial samples decreased the total number of detected cultured species to 46 (20.0%), 59 (25.7%), and 64 (27.8%) for Epi2me, Emu, and NanoCLUST, respectively. This indicates that most species exhibited a relative abundance below the 0.058 threshold in polymicrobial samples, possibly due to the dominance of certain species or the high complexity of the bacterial population in the sample.

**Figure 4-5: The ROC curves of monomicrobial samples based on (a) Epi2me, (b) Emu, (c) NanoCLUST analysis and the ROC curves of polymicrobial samples based on (d) Epi2me, (e) Emu, (f) NanoCLUST analysis.**
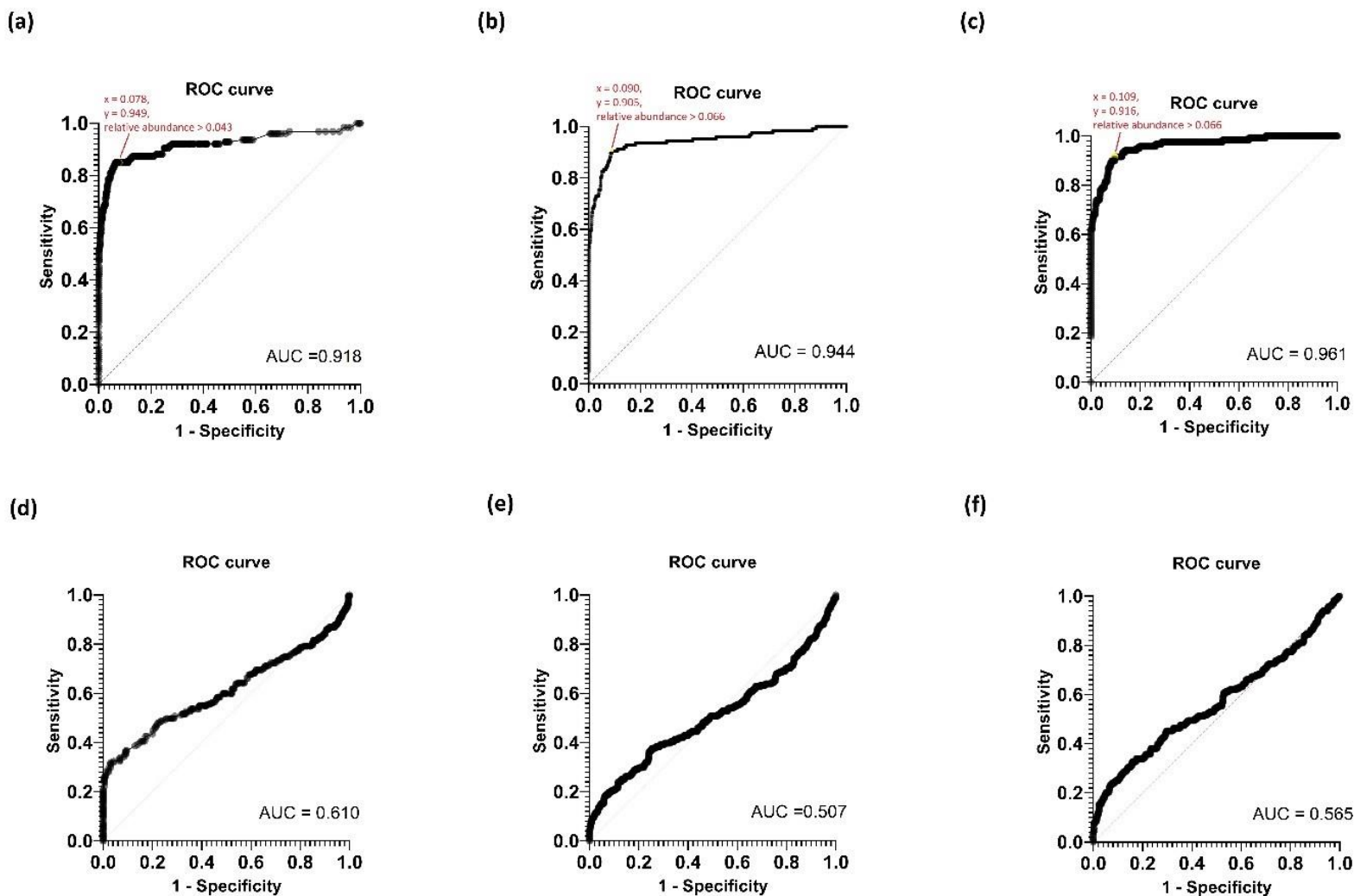
### 4.3.8 Clinically important species in culture-negative samples

Applying a $T_{RA}$ of 0.058 revealed clinically significant species in 13 of the 20 culture-negative body fluids, including opportunistic pathogens. While the sequencing results are summarized in Supplementary 8, the patients' medical histories for the 13 Nanopore 16S-positive samples are shown in Table 4-4. The species identified by Nanopore 16S showed alignment with the medical histories of the patients in two cases (samples 19MB090618 and 20MB068970). In sample 19MB090618, *Streptococcus pneumoniae* identified by Nanopore 16S was in line with the detection of pneumococcal antigen in urine samples. Similarly, the identification of *Prevotella* by Nanopore 16S in sample 20MB068970 matched the previous isolation of *Prevotella*. The primary reason for the failure to detect the Nanopore 16S-inferred species in culture was the administration of empirical antibiotics. Eleven out of thirteen cases had received antibiotic treatment before sample collection. For the two samples where antibiotics were not administered (21M2019392 and 21M2019576), the positive Nanopore 16S outcomes likely caused by skin flora contamination. Also, the detection of *E. coli* was expected in the case of a duodenal ulcer in sample 21M2019576.

**Table 4-4: The clinical background of thirteen culture-negative samples with positive Nanopore 16S result.**

| Sample ID | Specimen type | Bacteria detected by Nanopore 16S | Clinical detail/ Diagnosis | Previous culture result | Pre-treatment before sample collection |
|---|---|---|---|---|---|
| 19B2153436 | Pleural aspirate | *Streptococcus intermedius* *Parvimonas micra* | Right pleural effusion and empyema, cough, SOB and chest discomfort for 2 weeks. | No history of the isolate | Yes. Tazocin and Ornidazole. |
| 19B2153759 | Pleural aspirate | *Parvimonas micra* *Streptococcus milleri/ Streptococcus constellatus* *Escherichia coli* | Right pleural effusion, under chemotherapy, right breast cancer with high suspicion of multiple vertebral metastases. | No history of the isolate | Yes. Tazocin and Vancomycin. |
| 20MP2015461 | Knee joint fluid (left) | *Enterococcus cecorum* *Staphylococcus aureus* | Gout | No history of the isolate | Yes. IV Augmentin. |
| 20MP2015462 | Pleural fluid | *Facklamia hominis* | TB-pericarditis | No history of the isolate, but AFB positive in pericardial fluid | Yes. PO/IV Augmentin. |
| 21M2019392 | Ascitic fluid | *Staphylococcus hominis* | Malignancy-related ascites | No history of the isolate | No. |
| 21M2019256 | Peritoneal dialysis fluid | *Moraxella osloensis* | Fluid overload | No history of the isolate | Yes. IV Augmentin. |
| 21M2019465 | Peritoneal fluid | *Cutibacterium acnes* *Staphylococcus hominis* | Fever, malignancy-related ascites | No history of the isolate | Yes. IV Augmentin. |
| 21M2019576 | Knee joint fluid (left) | *Moraxella osloensis* *Escherichia coli* | Fever, duodenal ulcer | No history of the isolate | No. |
| 19MB068751 | Joint fluid (right shoulder) | *Staphylococcus epidermidis* *Staphylococcus hominis* *Staphylococcus cohnii* | Septic arthritis | No history of the isolate | Yes. IV Augmentin. |
| 18MB084305 | Joint fluid | *Streptococcus canis* | Prosthetic joint infection | No history of the isolate | Yes. IV Tazocin. |
| 19MB090618 | Lung abscess aspirate | *Streptococcus pneumoniae* | Community acquired pneumonia with lung abscess | No history of isolate, but urine pneumococcal antigen positive. | Yes. IV Augmentin. |
| 19MB069360 | Ascitic fluid | *Escherichia coli* | Peritonitis | No history of the isolate | Yes. IV Augmentin. |
| 20MB068970 | Pleural fluid | *Prevotella oris* *Streptococcus anginosus* | Perforated esophagus with lung empyema | Previous pleural fluid culture: *Enterococcus faecalis*, *Lactobacillus* species, *Candida albicans, Prevotella* species | Yes. IV vancomycin, IV meropenem |

## 4.3.9 LOD of Nanopore 16S

To evaluate the LOD for Nanopore 16S, two pathogens commonly encountered in infected body fluids, gram-positive *S. aureus* and gram-negative *K. pneumoniae*, were spiked into EDTA-blood at varying dilutions and the results were illustrated in Figure 4-6. During the LOD assessment, Nanopore 16S exhibited sensitivity to environmental contamination, particularly in the sample with extremely low bacterial concentration. As a result, incidental detection of environmental bacteria such as *Pelomonas saccharophila*, *Cutibacterium acnes*, and *Stenotrophomonas rhizophila* occurred. By applying the $T_{RA}$ determined from the ROC analysis (0.058), the LOD of Nanopore 16S for *S. aureus* and *K. pneumoniae* was calculated at 89.32 CFU/ml and 14.47 CFU/ml, respectively. Despite variations in LOD across species, a conservative approach was taken in this study by selecting the higher LOD of *S. aureus* as the benchmark for Nanopore 16S. Consequently, the overall LOD for Nanopore 16S was estimated to be approximately 90 CFU/ml.

**Figure 4-6: The relative abundance of spiked species in Nanopore 16S against the bacterial concentration in simulated bacteremic blood.**
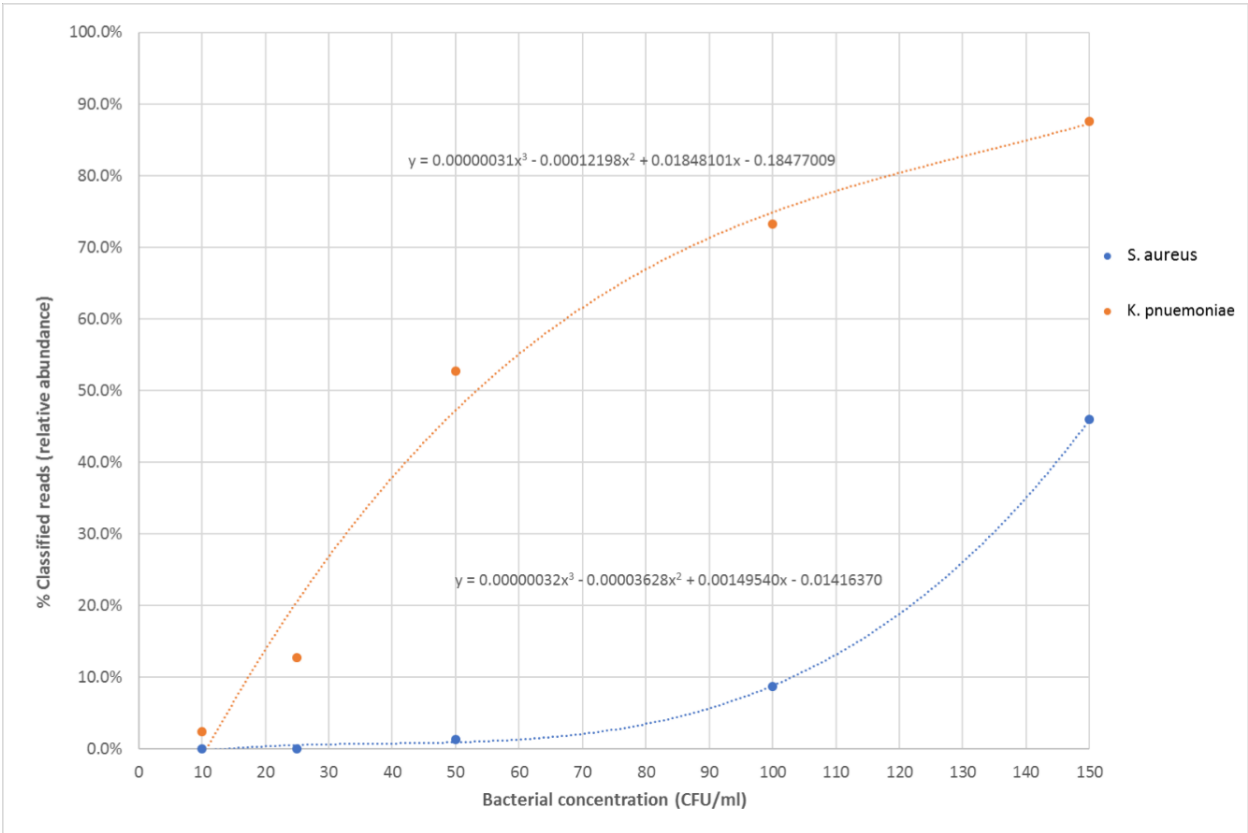
### 4.3.10 Workflow of Nanopore 16S

The workflow of Nanopore 16S and traditional culture was illustrated in Figure 4-7. While traditional culture-based identification typically takes at least 24 hours (and longer for fastidious bacteria), Nanopore 16S with Epi2me provides initial results within 6 hours. Among the 127 concordant monomicrobial samples, Epi2me identified all targeted or closely related within the first hour of sequencing (Supplementary 9). For polymicrobial samples, 31 out of 50 concordant samples (62%) showed detection of all species within an hour. Given the substantial variation in bacterial load across the samples, a sequencing time of 24 hours was designated in this study to enhance sequencing depth for samples with markedly low DNA concentrations. Nonetheless, the sequencing duration can be notably reduced, as most samples demonstrate the ability to detect targeted species within the initial hours of sequencing.

**Figure 4-7: The workflow of Nanopore 16S and traditional culture.**
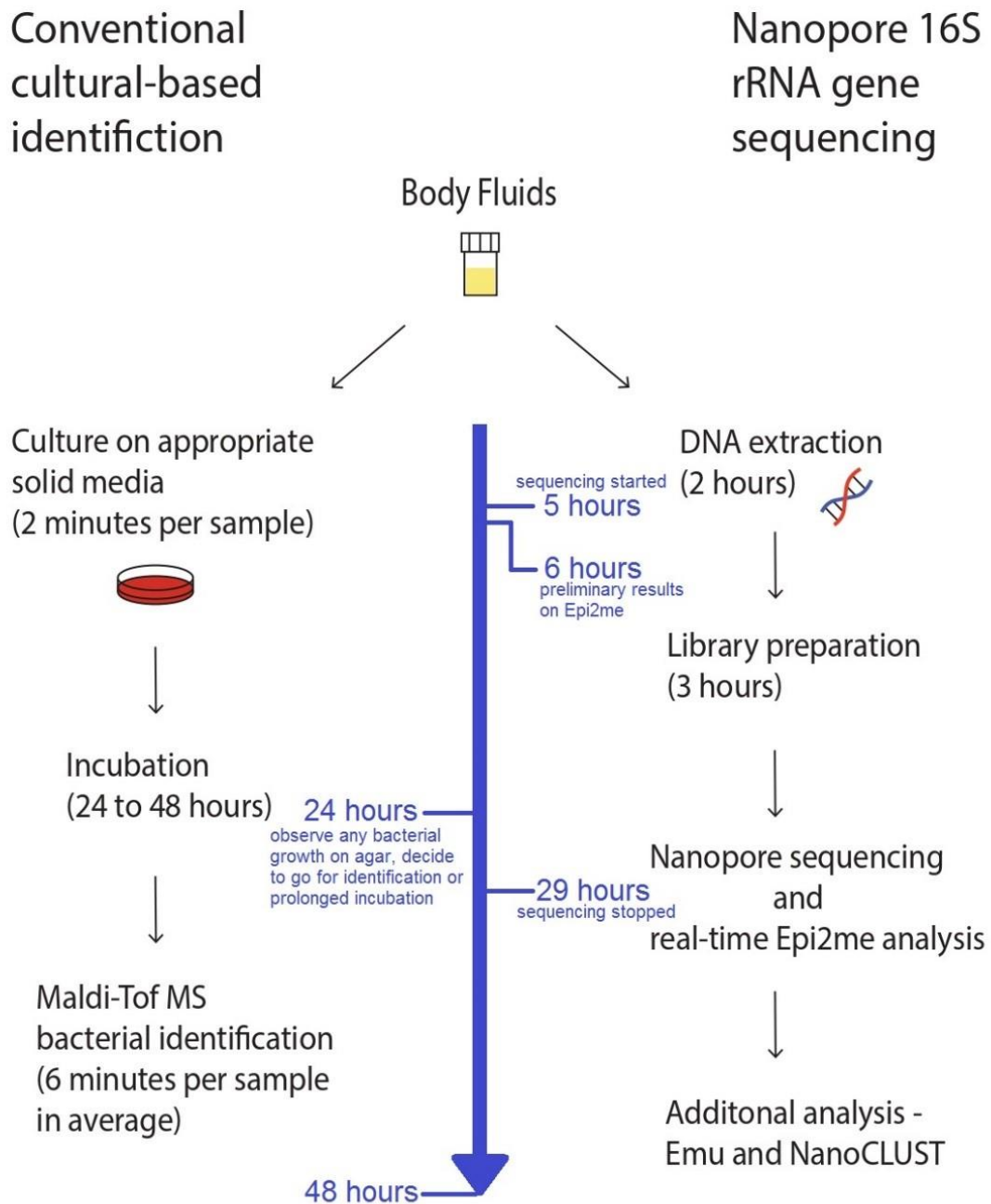
## 4.4 Discussion

In this study, the efficacy of Nanopore 16S for bacterial identification in direct body fluids was assessed, with reference to culture standards. In addition to the default ONT analysis platform (Epi2me), the sequencing reads were further analyzed using two external analysis pipelines, Emu and NanoCLUST, to mitigate classification bias. Results showed that Nanopore 16S exhibited optimal concordance when paired with Emu, achieving concordance rates of 97.7% and 81.7% for monomicrobial and polymicrobial samples, respectively. However, Epi2me demonstrated superior concordance if closely related species were considered concordant, with concordance rates of 99.2% and 90.0% for monomicrobial and polymicrobial samples, respectively. NanoCLUST exhibited the lowest concordance among the three analysis pipelines, with concordance rates of 93.0% for monomicrobial samples and 79.89% for polymicrobial samples, encompassing closely related species.

The read-by-read classification approach in Epi2me offers higher sensitivity, yet, it is susceptible to sequencing errors in individual reads, potentially leading to misassignments of taxa to closely related species and an augmented count of classified species per sample. Conversely, NanoCLUST exhibited the lowest sensitivity due to its clustering strategy for highly similar sequences, resulting in diminished sensitivity in detecting closely related and low-abundance species. Emu employs read alignment with reference sequences and a subsequent error-correction phase grounded on an expectation-maximization algorithm, allowing the effective differentiation of closely related species (205) and presenting a more balanced approach in terms of sensitivity and the count of classified species. Although Epi2me allows real-time analysis for preliminary result, it is recommended to further confirm the results with other analysis pipeline.

In general, Nanopore 16S detected a higher number of classified species per sample compared to culture, particularly in polymicrobial samples. Moreover, the most abundant species identified by Nanopore 16S did not necessarily align with the cultured species in monomicrobial samples. This discrepancy could be due to the detection of environmental contaminants in Nanopore 16S or the under-detection of microorganisms in culture. Clinical specimens with low bacterial loads were prone to environmental DNA contamination, especially in cases where sterilization procedures, such as autoclaving, were inadequate in removing DNA (277, 278). Several environmental bacteria were detected from the body fluids in this study, including *Aquitalea magnusonii*, *Delftia acidovorans*, and *Deinococcus geothermalis*. Additionally, anaerobes were frequently overlooked in culture. Certain anaerobic gut microbiota species identified by Nanopore 16S were rarely reported in culture-based identification, such as *Parvimonas micra*, *Peptostreptococcus stomatis*, *Faecalibacterium prausnitzii*, *Fusobacterium nucleatum*, and *Filifactor alocis*. On the contrary, some species were solely detected via culture and not by Nanopore 16S, potentially due to the masking effect of other abundant species in the samples.

Owing to Nanopore 16S's susceptibility to contamination, ROC analysis was conducted to determine the $T_{RA}$ for differentiating potential pathogens from background noise. According to the ROC analysis of monomicrobial samples, a threshold of 0.058 was proposed for Nanopore 16S. Using this threshold, the LOD for Nanopore 16S was determined as 89.32 CFU/ml for *S. aureus* and 14.47 CFU/ml for *K. pneumoniae*. Thus, the overall LOD of Nanopore 16S was about 90 CFU/ml. The higher LOD for gram-positive species than gram-negative species could be due to the thicker cell wall of gram-positive bacteria, rendering them more resistant to lysis during extraction (279). Besides the application of $T_{RA}$, the integration of non-template controls can aid in identifying contaminant species in Nanopore 16S results.

There were certain limitations in this study. Due to the lack of a perfect reference method to reveal the bacterial profile of body fluids, the performance of Nanopore 16S was evaluated solely based on the traditional culture. Hence, ROC analysis failed to detect the relative abundance threshold for polymicrobial samples because of the presence of numerous "false positive" species. Besides, uneven sample sizes were noted for different types of body fluids, with a significant portion comprising peritoneal fluids or peritoneal dialysis fluids.

## 4.5 Conclusion

This chapter underscores the clinical utility of Nanopore 16S for direct bacterial detection in normally sterile body fluids. Nanopore 16S emerges as a rapid diagnostic tool, providing initial results within 6 hours when paired with Epi2me for real-time analysis, contrasting with the minimum 24-hour timeframe required for culture, making it advantageous in urgent medical scenarios. However, while offering a more comprehensive bacterial profile, especially in highly polymicrobial body fluids, Nanopore 16S merely suggests potential causative agents in patients. Interpretation of laboratory findings necessitates collaboration with the patient's medical history, clinical manifestations, and the expertise of clinicians to achieve an accurate diagnosis. Normal flora presence in body fluids could result from sample collection contamination or opportunistic pathogens. Thus, it is essential to consider the broader clinical context when analyzing results. Concurrent culture-based identification with Nanopore 16S can aid in detecting low-abundance organisms below Nanopore 16S's detection limit.

# Chapter 5: The development and evaluation of Nanopore-based targeted and unbiased metagenomic workflows for pathogen identification and AMR detection in clinical samples

## 5.1 Introduction

Although Nanopore 16S allows rapid identification of bacterial pathogens, it provides no information about AMR profile of bacteria or the detection of other common pathogens such as fungi in invasive infections. To address these limitations, nanopore targeted sequencing (NTS) and nanopore metagenomic sequencing (NMgS) present as potential solutions. NTS involves the selective enrichment of targeted sequences, offers higher sensitivity for pathogen identification and AMR genes detection. In contrast, NMgS sequences all the genetic content in a sample, providing more comprehensive and unbiased information. However, NMgS has a relatively lower sensitivity due to the presence of abundant host DNA in clinical samples. This study aimed to optimize NTS and NMgS workflows for invasive infections and to compare their performance in microbial and drug resistance detection.

For NTS workflow, the 16S rRNA gene and ITS regions were amplified for bacterial and fungal detection respectively. Additionally, a set of 19 pairs of primers were used to detect the clinically important antimicrobial resistance associated with the ESKAPE pathogens, which are the plasmid mediated AMR genes for β-lactams and vancomycin. The multiplex PCR was validated using bacterial isolates with known drug resistance profile based on WGS or the reference strains from ATCC. Additionally, the NTS workflow was validated using simulated bacteremic and fungemic samples spiked with known microorganisms. An in-house analysis pipeline was used for

taxonomic classification. To evaluate the clinical utility of NTS workflow, a total of 138 normally

sterile body fluids were sequenced using NTS workflow and the results were compared to culture.

A NMgS workflow was optimized and validated using blood samples spiked with four

microorganisms with different cell wall compositions (*Staphylococcus aureus*, *Escherichia coli*,

*Mycobacterium marinum*, and *Candida krusei*) with a final concentration of 1000 CFU/ml.

Considering the limited computing resources in clinical laboratories, Kraken 2 (202) coupled with

Bracken (280) was used for metagenomic analysis, which was shown to be a time and

computational efficient alternative of Blastn (281, 282). The optimal confidence threshold for

Kraken 2 in classifying low microbial biomass clinical samples was also determined. To evaluate

the clinical utility of NMgS, 138 body fluid samples were subjected to NMgS alongside NTS

workflows. The samples were sequenced for 4 hours in order to have rapid diagnosis, concordance

between culture and NMgS was then determined. Furthermore, to explore the effect of sequencing

time on NMgS sensitivity, 30 samples from the initial cohort underwent extended sequencing

periods of up to 48 hours. The concordance with culture results was evaluated after 24 hours and

48 hours of sequencing.

## 5.2 Materials and methods

### 5.2.1 NTS workflow

#### 5.2.1.1 Multiplex PCR

A set of ID primers was utilized to amplify bacterial and fungal pathogens, while a set of AMR

primers was employed to amplify clinically important AMR genes for the detection of ESKAPE

pathogens. Detailed primer sequences are provided in Table 5-1. Nanopore adapter sequences 5'-

TTTCTGTTGGTGCTGATATTGC-3' and 5'-ACTTGCCTGTCGCTCTATCTTC-3' were added

to the forward and reverse primer respectively. Two PCR reactions per sample were prepared as in Table 5-2. Both reactions were amplified under the same condition: initial denaturation at 95°C for 3 minutes, followed by 20 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds and extension at 65°C for 1 minute and 40 seconds; final extension was performed at 65°C for 5 minutes. After PCR, the two reactions of each sample were pooled together and purified with 0.7x AMPure XP beads (AXP) and eluted in 24 μl of nuclease-free water.

**Table 5-1: Primer sequences of ID and AMR primer set.**

| Primer set | Primer name | Sequence (5' to 3') | Amplicon size (bp) | Targeted AMR genes | Working concentration (μM) | References |
|---|---|---|---|---|---|---|
| ID primer set | 16S-27F-YM | AGAGTTTGATYMTGGCTCAG | | | | |
| | 16S-1492R | ACGGYTACCTTGTTACGACTT | 1500 | Bacterial 16S rRNA | 2 | (283) |
| | ITS1-27F | TACGTCCCTGCCCTTTGTAC | | | | |
| | ITS4 | TCCTCCGCTTATTGATATGC | 200-500 | Fungal ITS | 1 | (284, 285) |
| AMR primer set | MecA-F | GGCTATCGTGTCACAATCGTT | | | | |
| | MecA-R | TCACCTTGTCCGTAACCTGA | 689 | MecA | 1 | (286) |
| | VanA-F | TCTGCAATAGAGATAGCCGC | | | | |
| | VanA-R | GGAGTAGCTATCCCAGCATT | 377 | VanA | 1 | (287) |
| | VanB-F | CATCGCCGTCCCCGAATTTCAAA | | | | |
| | VanB-R | GATGCGGAAGATACCGTGGCT | 297 | VanB | 1 | (288) |
| | CTX-MA1 | SCSATGTGCAGYACCAGTAA | | | | |
| | CTX-MA2 | CCGCRATATGRTTGGTGGTG | 450 | blaCTX-M | 4 | (289) |
| | CTX-M Gp8/25-F | AACRCRCAGACGCTCTAC | | blaCTX-M-8, blaCTX-M-25, blaCTX-M-26 and blaCTX-M-39 to blaCTX-M-41 | | |
| | CTX-M Gp8/25-R | TCGAGCCGGAASGTGTYAT | 326 | | 2 | (290) |
| | TEM-F | CATTTCCGTGTCGCCCTTATTC | | | | |
| | TEM-R | CGTTCATCCATAGTTGCCTGAC | 800 | blaTEM | 1 | (291) |
| | SHV-F | AGCCGCTTGAGCAAATTAAAC | | | | |
| | SHV-R | ATCCCGCAGATAAATCACCAC | 713 | blaSHV | 1 | (291) |
| | OXA-F | GGCACCAGATTCAACTTTCAAG | | | | |
| | OXA-R | GACCCCAAGTTTCCTGTAAGTG | 564 | blaOXA-1 | 1 | (291) |

| | | | | | |
|---|---|---|---|---|---|
| KPC-F | TGTCACTGTATCGCCGTC | | | | |
| KPC-R | CTCAGTGCTCTACAGAAAACC | 900 | blaKPC | 1 | (292) |
| IMP-F | GAAGGCGTTTATGTTCATAC | | | | |
| IMP-R | GTACGTTTCAAGAGTGATGC | 587 | blaIMP | 1 | (292) |
| VIM-F | GTTTGGTCGCATATCGCAAC | | | | |
| VIM-R | AATGCGCAGCACCAGGATAG | 389 | blaVIM | 1 | (292) |
| NDM-F | GCAGCTTGTCGGCCATGCGGGC | | | | |
| NDM-R | GGTCGCGAAGCTGAGCACCGCAT | 782 | blaNDM | 1 | (292) |
| OXA-48-F | GCTTGATCGCCCTCGATT | | | | |
| OXA-48-R | GATTTGCTCCGTGGCCGAAA | 281 | blaOXA-48 | 1 | (293) |
| ACC-F | AACAGCCTCAGCAGCCGGTTA | | | | |
| ACC-R | TTCGCCGCAATCATCCCTAGC | 346 | blaACC | 1 | (294) |
| FOX-F | GCCGAGGCTTACGGGATCAAG | | | | |
| FOX-R | CAAAGCGCGTAACCGGATTGG | 247 | blaFOX-1 to 9 | 1 | (295) |
| MOX-F | GCAACAACGACAATCCATCCT | | blaMOX-1, blaMOX-2, blaCMY-1, blaCMY-8 to blaCMY-11 and blaCMY-19 | | |
| MOX-R | GGGATAGGCGTAACTCTCCCAA | 895 | | 1 | (290) |
| DHA-F | AACTTTCACAGGTGTGCTGGGT | | blaDHA-1, blaDHA-2 | | |
| DHA-R | CCGTACGCATACTGGCTTTGC | 405 | | 1 | (294) |
| CIT-F | CGAAGAGGCAATGACCAGAC | | blaLAT-1 to blaLAT-3, blaBIL-1, blaCMY-2 to blaCMY-7, blaCMY-12 to blaCMY-18 and blaCMY-21 to blaCMY-23 | | |
| CIT-R | ACGGACAGGGTTAGGATAGY | 538 | | 1 | (290) |
| EBC-F | TCGGTAAAGCCGATGTTGCGG | | | | |
| EBC-R | CTTCCACTGCGGCTGCCAGTT | 302 | blaMIR-1, blaACT-1 | 1 | (294) |

**Table 5-2: Preparation of multiplex PCR.**

| Components | Reaction 1 | Reaction 2 |
|---|---|---|
| LongAmp Hot Start Taq 2X Master Mix | 12.5 µl | 12.5 µl |
| ID primer set | 2.5 µl | - |
| AMR primer set | - | 2.5 µl |
| DNA | 10 µl | 10 µl |

### 5.2.1.2 Library preparation

Library preparation was performed using the ligation sequencing kit (SQK-LSK110) and PCR Barcoding Expansion 1-96 (EXP-PBC096) with some modifications. In brief, Barcoding PCR was performed by adding 25 µl of LongAmp Taq 2x master mix and 1 µl of PCR barcode to 24 µl of purified multiplex PCR product with cycling condition: initial denaturation at 95°C for 3 minutes, followed by 15 cycles of denaturation at 95°C for 15 seconds, annealing at 62°C for 15 seconds and extension at 65°C for 2 minute; and a final extension at 65°C for 5 minutes. The amplicons were purified with 0.8x AXP and eluted in 11 µl of nuclease-free water. After DNA quantification and normalization, the pooled library (up to 24 samples per batch) was end-repaired, adapter ligated and sequenced for 4 hours on GridION. During the sequencing, "mid-read barcode filtering", "barcode both ends", and "trim barcodes" were adopted with the minimum barcoding score 85.

### 5.2.1.3 Sequencing data analysis

The sequencing data was analyzed using an in-house analysis pipeline. Briefly, the reads were separated into two pools based on the read length – 1300bp to 1700bp (16S) and 100bp to 1200bp (ITS + AMR), using nanofilt v2.8.0 (296). The 16S reads were classified using emu v3.4.5 (205). The ITS + AMR reads were classified using BLAST+ (297) with the ITS database built based on NCBI Fungal ITS RefSeq Targeted Loci Project (PRJNA177353) and the NCBI AMR database Reference Gene Catalog, (255), respectively. An e-value of 1e-5 and percentage identity above 90% were applied in the blastn search (298).

### 5.2.2 Validation of multiplex PCR of NTS workflow

To validate multiplex PCR of AMR genes, isolates of 13 bacterial strains with targeted AMR genes were sequenced using NTS workflow. The primer for FOX genes could not be validated due to the lack of reference strain. For each primer set, 10ng of DNA from each isolate was used for amplification. Additionally, to validate the multiplex PCR in polymicrobial samples, three simulated polymicrobial samples were prepared by mixing the DNA from (1) two *Candida* species, (2) one bacterial species and one *Candida* species, and (3) three bacterial species. The presence of targeted genes and microorganisms indicates efficient amplification in the multiplex PCR.

### 5.2.3 Validation NTS workflow

The NTS workflow was validated using simulated bacteremic and fungemic blood samples. Four microorganisms, each characterized by distinct cell wall structures (Table 5-3), were individually introduced into whole blood at a final concentration of 100 CFU/ml. Briefly, the colonies from an overnight culture of each microorganism (except 7-day incubation for *M. marinum*) were resuspended in saline with a concentration of 0.5 McFarland (~$10^8$ CFU/ml for bacteria (299) or ~$10^6$ CFU/ml for *Candida* (300)) using a densitometer. The microbial suspensions underwent serial dilution to achieve a concentration of $10^4$ CFU/ml with saline. After that, 10µl of $10^4$ dilution was added to 990µl of EDTA blood obtained from a healthy individual, resulting in a final concentration of 100 CFU/ml. The suspension for each microorganism was prepared in triplicate to minimize the effect of random error. DNA extraction was performed using the QIAamp BiOstic Bacteremia DNA Kit. NTS workflow was performed to determine the presence of spiked species by applying the $T_{RA}$ of 0.058, along with targeting AMR genes in the antimicrobial resistant strains (*S. aureus* and *E. coli*).

**Table 5-3: Distinct cell wall structures of 4 spiked microorganisms.**

| Microorganism | Characteristics of cell wall | References |
|---|---|---|
| *Staphylococcus aureus* (ATCC BAA-3114) | gram-positive bacterial cell wall contains thick peptidoglycan layer, lack of outer membrane | (78, 301) |
| *Escherichia coli* (ATCC BAA-3054) | gram-negative bacterial cell wall contains thin peptidoglycan layer surrounded by lipopolysaccharide outer membrane | (78, 301) |
| *Mycobacterium marinum* (ATCC BAA-535) | mycobacterial cell wall contains thin layers of peptidoglycan and arabinogalactan, and a thick layer of mycolic acids; lack of outer membrane | (301) |
| *Candida krusei* | fungal cell wall contains layers of chitin, β-glucan and mannan | (301) |

## 5.2.4 Evaluation of NTS workflow using normally sterile body fluids

A total of 138 normally sterile body fluids alongside with the culture results were collected from three public hospitals in Hong Kong, including United Christian Hospital, Pamela Youde Nethersole Eastern Hospital, and Tuen Mun Hospital (Supplementary 10). There were 98 monomicrobial and 40 polymicrobial samples included in the study. An extraction control (ETC), comprising 1ml of nuclease-free water, was included in every extraction batch to detect potential environmental contaminants. After DNA extraction, NTS was performed as described previously. The concordance between culture and NTS results were determined.

## 5.2.5 NMgS workflow

### 5.2.5.1 Host DNA depletion and DNA extraction

For blood samples, 200 µl of hetasep was added to 1 ml of whole blood samples to remove red blood cells. After mixing, the samples were centrifuged at $100 \times g$ for 3 minutes and incubated at 37°C for 5 minutes. The resulting supernatants were carefully transferred to a new tube and centrifuged again at $10,000 \times g$ at 4°C for 5 minutes. For other body fluid types, 1 ml of the fluid was directly centrifuged at $10,000 \times g$ at 4°C for 5 minutes without the treatment of hetasep.

Following the removal of the supernatant, the pellets were resuspended in 1 ml of water and incubated for 5 minutes to induce osmotic shock on human cells. The samples were then centrifuged at 10,000 × g at 4°C for 5 minutes. The pellets were collected and resuspended in 95 μl of 0.0125% saponin (302) and incubated at room temperature for 5 minutes to further lyse the human cells. To crosslink the human DNA, 5 μl of 0.4 mM PMAxx (20 μM) (303) was added to the samples and incubated in the dark for 10 minutes. Subsequently, the samples were horizontally placed on ice within 20 cm of a light source for 15 minutes, followed by a wash with 1 ml of saline. DNA extraction was conducted using QIAamp BiOstic Bacteremia DNA Kit.

## 5.2.5.2 Library preparation

Similar to NTS workflow, library preparation of NMgS workflow was performed using ligation sequencing kit (SQK-LSK110) and PCR Barcoding Expansion 1-96 (EXP-PBC096). In brief, fragmentation and A-tailing were performed by adding 7 μl of NEBNext Ultra II FS Reaction Buffer and 2 μl of NEBNext Ultra II FS Enzyme Mix to 26 μl of genomic DNA and incubated at 37°C for 30s and 65°C for 30 minutes. The samples were then purified with 35 μl of AXP and eluted in 15 μl of nuclease-free water. Barcode adapters were ligated to samples by adding 10 μl of barcode adapter and 25 μl of Blunt/TA Ligase Master Mix to 15 μl of DNA and incubated at room temperature for 10 minutes. The samples were purified with 25 μl of AXP and eluted in 24 μl of nuclease-free water. Barcode PCR was performed as described in the NTS workflow, but the PCR cycle was increased to 35 cycles. The amplicons were purified with 40 μl of AXP and eluted in 11 μl of nuclease-free water.

The required input for each sample for pooling was calculated by the formula (1000ng / total number of samples in a run). Then, the required volume of each sample was determined by

dividing the required input by the concentration of the sample. After pooling, 1 µl of DNA Control Sample, 7 µl of Ultra II End-prep Reaction Buffer and 3 µl of Ultra II End-prep Enzyme Mix were added to 49 µl of pooled library. The volume of reagents was scaled up accordingly if the volume of pooled library exceeded 49 µl. After incubation at 20°C for 5 minutes and 65°C for 5 minutes, the library was purified with 1x AXP and eluted in 60 µl of nuclease-free water. For adapter ligation, 5 µl of adapter mix, 25 µl of ligation buffer, and 10 µl of NEBNext quick T4 DNA ligase were added to 60 µl of the pooled library. After a 10-minute incubation at room temperature, the library was purified using 40 µl of AXP, subjected to two washes with 250 µl of short fragment buffer, and finally eluted in 13 µl of elution buffer.

Subsequently, the library was loaded on a flow cell FLO-MIN106D (R9.4.1) and sequenced on GridION with super accurate (SUP) model for 4 hours, following the parameters specified in the NTS workflow. In addition, adaptive sampling was turned on during sequencing to deplete the human reads that aligned with the reference human genome GRCh38.p13.

### 5.2.5.3 Sequencing data analysis

The sequencing reads were analyzed using an in-house analysis pipeline. Briefly, the sequencing reads were first filtered using NanoFilt v2.8.0 (296) to remove reads with length below 200bp, followed by removal of human reads using Kraken 2 v2.1.3 (202) with the human genome (GRCh38.p13) database. Then, the filtered reads were classified using Kraken 2 and with PlusPF database. The species abundance was re-estimated from the Kraken 2 results using Bracken v2.9 (280). For AMR genes identification, the filtered reads were classified using blast+ (297) with the NCBI Reference Gene Catalog (255), as described in the NTS workflow.

## 5.2.6 Validation and optimization of NMgS workflow

The NMgS workflow was validated using simulated bacteremic and fungemic blood samples prepared as described in the NTS workflow. Similarly, the four microorganisms with distinct cell wall components were spiked into EDTA-blood, respectively. To mimic low microbial biomass body fluids, the spiked microorganism concentration was set at 1000 CFU/ml. Duplicate preparations of the simulated bacteremic and fungemic blood samples were created for each suspension to assess the efficacy of host DNA depletion. Both treated and untreated control samples underwent hetasep treatment to eliminate red blood cells. Treated sample pellets were resuspended in nuclease-free water to induce osmotic shock, while control sample pellets were resuspended in saline. Host DNA depletion was carried out on the treated samples, followed by DNA extraction from both treated and control samples.

After library preparation, the pooled library was sequenced on GridION for 4 hours with SUP bascalling mode and adaptive sampling. To evaluate the optimal confidence threshold of Kraken 2 for classifying low microbial biomass clinical specimens, the reads were analyzed across a range of confidence thresholds from 0 to 1, with intervals of 0.05. Furthermore, the impact of the threshold for the number of reads in Bracken was assessed at values of 0, 5, and 10. The optimal threshold for both Kraken 2 and Bracken was defined as the minimum value that yielded the highest abundance of spiked organisms and the lowest background noise for the four spiked species. Additionally, a comparison was made between the relative abundance of human DNA in treated and control samples.

### 5.2.7 Evaluation of NMgS workflow using normally sterile body fluids

To evaluate the performance of NMgS workflow, the 138 normally sterile body fluids were subjected to NMgS in parallel with NTS. Host DNA depletion was performed before DNA extraction. After library preparation, the libraries were sequenced for 4 hours to have rapid diagnosis. Additionally, 30 samples underwent extended sequencing for up to 48 hours to explore potential sensitivity enhancements with prolonged sequencing. The sequencing data was analyzed using an in-house analysis pipeline and the results were compared with the reference culture results.

### 5.2.8 Statistical data analysis

All statistical analyses were performed using GraphPad Prism (v10.2.3). A significance threshold of $p < 0.05$ was employed to determine statistical significance. The Mann-Whitney U test was employed to assess the statistical differences between populations of sequencing reads derived from monomicrobial and polymicrobial samples.

## 5.3 Results

### 5.3.1 Validation of multiplex PCR in NTS workflow

All primers, except the primer pair designed for the FOX gene, successfully amplified the targeted genes in the multiplex PCR. The primers of FOX gene could not be validated due to the lack of a reference strain. For simulated polymicrobial samples, all the spiked microorganisms and their AMR genes were detected. The reference strains and the read counts of the targeted genes are detailed in Table 5-4.

**Table 5-4: Reference strains and their respective number of reads of targeted genes.**

| Reference stain ID | Species | 16S result | Read counts | Relative abundance | Targeted AMR genes | Detected AMR genes | Read counts |
|---|---|---|---|---|---|---|---|
| BAA-3114 | *Staphylococcus aureus* | *Staphylococcus aureus* | 1416 | 1 | MecA | MecA | 1381 |
| 23B3974577 | *Enterococcus faecium* | *Enterococcus faecium* | 2550 | 1 | VanA | vanA | 646 |
| 21MB801142 | *Enterococcus faecium* | *Enterococcus faecium* | 2905 | 1 | VanB | vanB | 662 |
| B52_C01_ESBL01 | *Escherichia coli* | *Escherichia coli* | 4919 | 1 | CTX-M-8 | blaCTX-M-8 | 1496 |
| BAA-3079 | *Klebsiella pneumoniae* | *Klebsiella pneumoniae* | 2906 | 1 | KPC-2 | blaKPC | 1629 |
| | | | | | MOX-2 | blaMOX | 1379 |
| | | | | | SHV-1 | | |
| | | | | | SHV-11 | blaSHV | 569 |
| | | | | | VIM-1 | blaVIM | 820 |
| QEH453981 | *Klebsiella michiganensis* | *Klebsiella michiganensis* | 2406 | 0.92 | IMP-4 | blaIMP | 17 |
| | | | | | OXY-1-2 | blaOXY-1 | 271 |
| B61_C01_CRE01 | *Escherichia coli* | *Escherichia coli* | 1314 | 1 | CTX-M-55 | blaCTX-M | 1049 |
| | | | | | NDM | blaNDM | 400 |
| | | | | | TEM | blaTEM | 104 |
| B44_P06_CRE02 | *Klebsiella pneumoniae* | *Klebsiella pneumoniae* | 3420 | 1 | OXA-48 | blaOXA-48_fam | 949 |
| | | | | | SHV-168 | blaSHV | 82 |
| | | | | | mcr-1.1 | mcr-1 | 1611 |
| B42_P03_ESBL04 | *Hafnia paralvei* | *Hafnia paralvei* | 6020 | 1 | ACC | blaACC | 1057 |
| B65_C01_ESBL03 | *Proteus mirabilis* | *Proteus mirabilis* | 1946 | 1 | CTX-M-65 | blaCTX-M | 103 |
| | | | | | DHA-1 | blaDHA | 920 |
| | | | | | OXA-1 | blaOXA-1_fam | 136 |
| BAA-3086 | | | 1838 | 1 | CMY-16 | blaCMY | 163 |

| | | | | | TEM-2 | blaTEM | 2058 |
|---|---|---|---|---|---|---|---|
| | *Proteus mirabilis* | *Proteus mirabilis* | | | VIM-1 | blaVIM | 562 |
| 230403_42 | *Enterobacter roggenkampii* | *Enterobacter roggenkampii* | 2448 | 0.975 | MIR | blaMIR | 713 |
| 230516_34 | *Enterobacter asburiae* | *Enterobacter asburiae* | 3180 | 0.997 | ACT-6 | blaACT | 419 |
| CA01 + CK01 | *Candida auris* | *Candida auris* | 1581 | 0.294 | N/A | N/A | N/A |
| | *Candida krusei* | *Candida krusei* | 3380 | 0.629 | N/A | N/A | N/A |
| CA01 + BAA-3114 | *Candida auris* + *Staphylococcus aureus* | *Staphylococcus aureus* | 1237 | 1 | MecA | mecA | 717 |
| | | *Candida krusei* | 2302 | 0.890 | N/A | N/A | N/A |
| BAA-3076 + BAA-3054 + BAA-3114 | *Klebsiella pneumoniae* + *Escherichia coli* + *Staphylococcus aureus* | *Klebsiella pneumoniae* | 1270 | 0.143 | blaCMY | blaCMY | 17 |
| | | | | | blaKPC | blaKPC | 20 |
| | | | | | blaSHV | blaSHV | 336 |
| | | | | | blaTEM | blaTEM | 244 |
| | | | | | blaVIM | blaVIM | 240 |
| | | *Escherichia coli* | 2732 | 0.308 | blaCTX-M | blaCTX-M | 5115 |
| | | | | | blaOXA-1_fam | blaOXA-1_fam | 1075 |
| | | *Staphylococcus aureus* | 4858 | 0.548 | mecA | mecA | 3044 |

## 5.3.2 Validation of NTS workflow

The LOD of Nanopore 16S was determined to be approximately 90 CFU/ml in chapter 4. Given that the NTS workflow shares similarities with the Nanopore 16S workflow, both involving the enrichment of specific genes, it was hypothesized that the LOD of NTS would align closely with that of Nanopore 16S. Hence, a validation of the NTS workflow for low microbial biomass body fluids was conducted by introducing four microorganisms with diverse cell wall compositions into whole blood at a final concentration of 100 CFU/ml.

The average relative abundance ± SD of *S. aureus*, *E. coli*, *M. marinum*, and *C. krusei*, was 0.495 ± 0.153, 0.305 ± 0.142, 0.174 ± 0.141, and 0.729 ± 0.114, respectively. All the spiked species reached the $T_{RA}$ of 0.058 determined in chapter 4. For the resistant ATCC reference strains, *S. aureus* (BAA-3114) and *E. coli* (BAA-3054), all the targeted AMR genes were detected in the NTS workflow with more than 10 reads. The mean reads ± SD of mecA gene in *S. aureus* was 18

± 2.94. The mean reads ± SD of CTX-M and OXA-1 genes in *E. coli* were 51.67 ± 30.92 and 12.33 ± 0.94, respectively.

### 5.3.3 Validation and optimization of NMgS workflow

Similar to the NTS workflow, NMgS workflow was also validated using blood samples spiked with *S. aureus*, *E. coli*, *M. marinum*, and *C. krusei*. However, considering a lower sensitivity in NMgS workflow, a final concentration of 1000 CFU/ml was used instead of 100 CFU/ml. For each microbial suspension, two spiked blood samples were prepared: one went for the saponin-based host DNA depletion, while the other served as untreated control.

After 4 hours of sequencing, the untreated control samples yielded an average of 9036.58 reads, with a SD of 5132.72 and a median of 8399.50. For the host DNA depleted samples, the mean total number of reads ± SD was 5246.67 ± 2076.017, with a median of 4348. Only reads with lengths larger than 200bp were used for subsequent analysis in NMgS workflow. After filtering of reads, the mean number of reads ± SD for control and treated samples were 8548.17 ± 4952.59 and 3816.33 ± 1807.33, respectively. The median of number of filtered reads for control and treated samples were 7993.50 and 3059.50, respectively.

Figure 5-1 illustrates the relative abundance of human DNA in control and host DNA-depleted samples. On average, the percentage of human reads decreased from 99.5% to 49.1%, underscoring the efficiency of the saponin-PMA-based host DNA depletion method in eliminating human DNA from blood samples. The most significant reduction in human DNA content was observed in blood samples spiked with *M. marinum* (reduced from 99.6% to 24.0%), followed by those spiked with *E. coli* (reduced from 99.4% to 43.5%). The decrease in human DNA content in

samples spiked with *C. krusei* and *S. aureus* was comparable, decreasing from 99.6% to 63.7% and from 99.5% to 62.3%, respectively.

Kraken 2 was employed for the analysis in NMgS workflow due to its short computation time and low computer resource requirement. To determine the optimal confidence threshold of Kraken 2 for classifying low microbial biomass clinical samples, reads were evaluated over a range of thresholds from 0 to 1 in increments of 0.05. Moreover, the impact of the read number threshold in Bracken was examined at values of 0, 5, and 10. While raising the thresholds of Kraken 2 and Bracken effectively eliminated false positive bacteria with low precision, it also led to a decrease in the number of classified reads and therefore the sensitivity. The optimal thresholds for Kraken 2 and Bracken were identified as the lowest values that yielded the highest relative abundance of the spiked organism. Through the analysis of changes in the relative abundance of spiked organisms across different confidence thresholds of Kraken 2 and Bracken (Figure 5-2), the optimal thresholds were determined to be 0.1 for Kraken 2 and 10 for Bracken, respectively.

Figure 5-3 shows the differences of relative abundance of spiked organisms in control and host DNA depleted samples, with and without the application of optimal threshold in Kraken 2 and Bracken. Reads that were mapped to the reference human genome were filtered out using Kraken 2, the resulting non-human reads were subjected to taxonomic classification. In control samples, the average count of non-human reads was notably low, averaging 37.25 with a SD of 18.78 and a median of 30. Conversely, host DNA-depleted samples exhibited a substantially higher average count of non-human reads at 1899.75, with a SD of 1051.44 and a median of 1694. Following the application of the optimal threshold, the relative abundance of spiked organisms increased from an average of 57.9% to 98.6%.

**Figure 5-1: The relative abundance of human DNA in control and treated samples.**

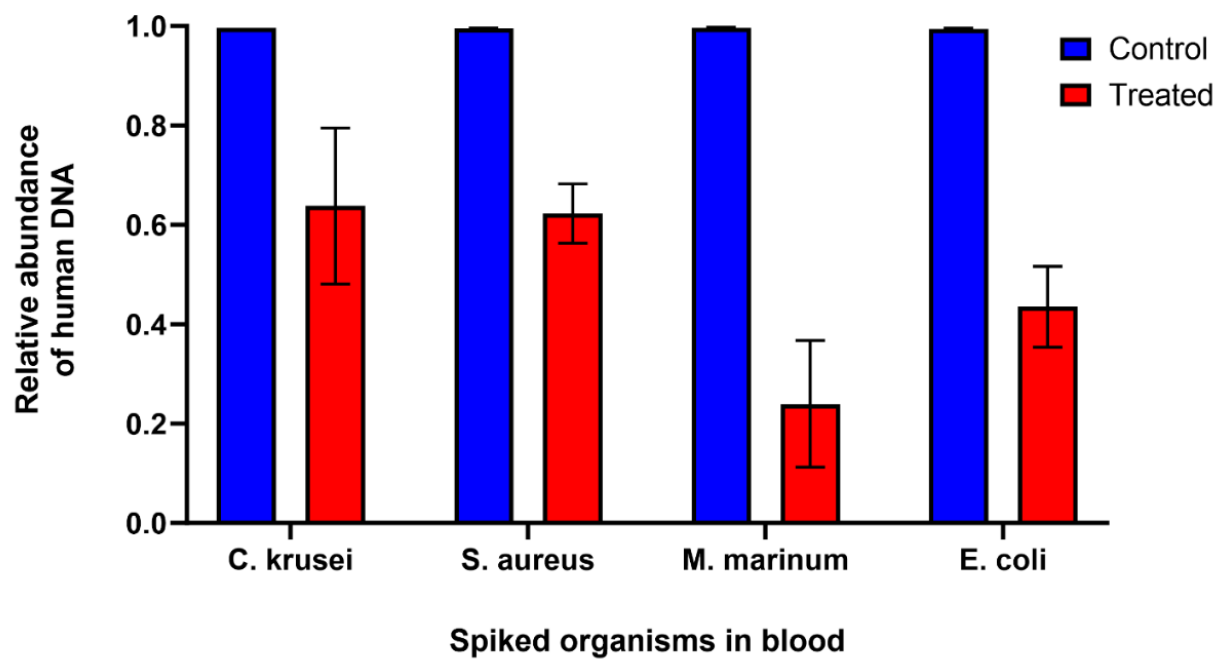**Figure 5-2: The change in relative abundance of each spiked organism across a range of confidence thresholds of Kraken 2 and threshold of Bracken at (a) t =0, (b) t = 5, and (c) t = 10; (d) shows the average relative abundance of the four organisms varies against confidence thresholds of Kraken 2 and threshold of Bracken.**
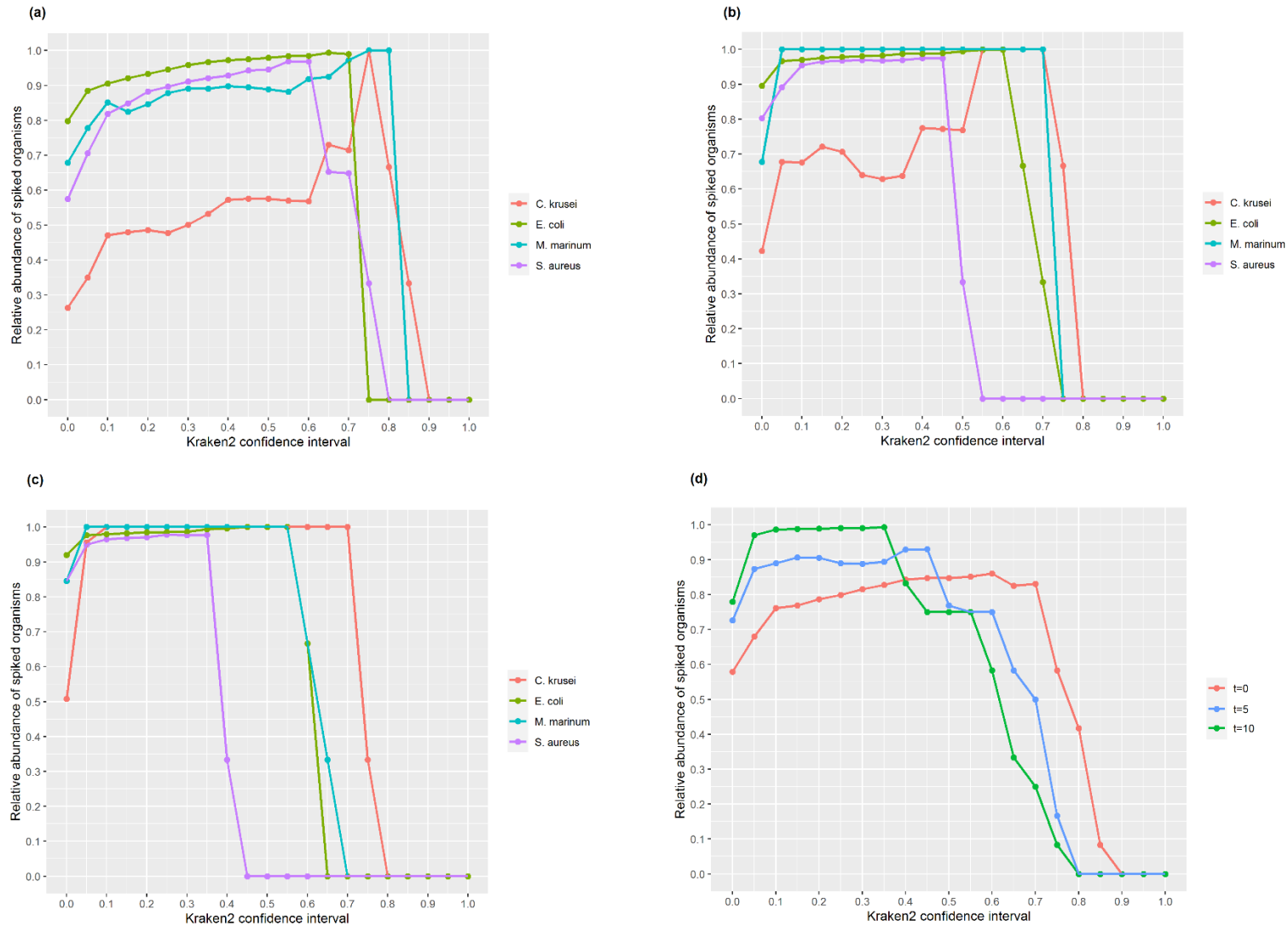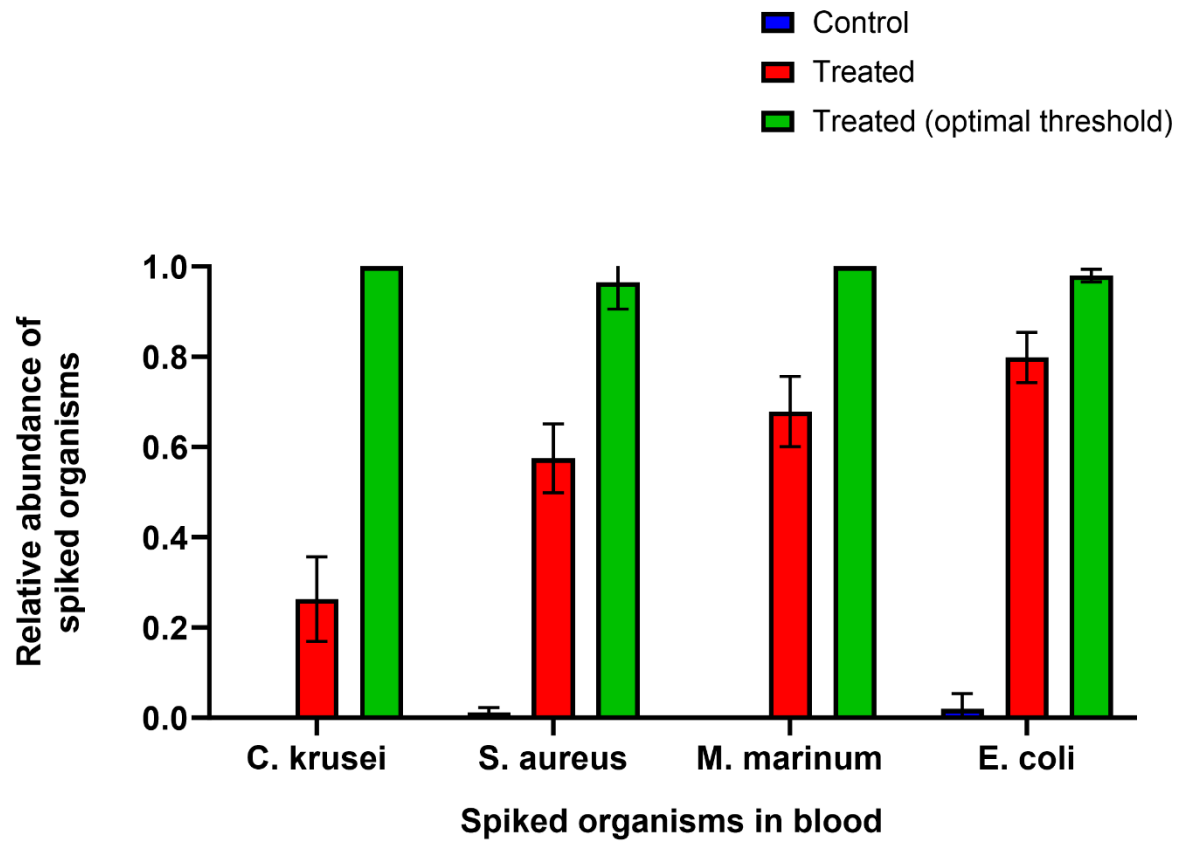
155

**Figure 5-3: The relative abundance of spiked organisms after removal of human reads.**

### 5.3.4 The statistic of sequencing reads in NTS of 138 body fluids

Following 4-hour sequencing, NTS workflow yielded an average of 1.69 Gb per sequencing run, with 1.28Gb passing the QSCORE threshold of 10. Monomicrobial samples had a mean total number of reads ± SD of 8297.89 ± 12109.92 and a median of 3149.50. In contrast, polymicrobial samples showed a mean total number of reads ± SD of 17892.33 ± 17429.62 and a median of 12094.50. Notably, polymicrobial samples displayed a significantly higher number of reads compared to monomicrobial samples (P=0.0004). The substantial standard deviation values suggest a considerable disparity in the number of reads across the samples.

The sequencing reads were divided into two groups based on read length for subsequent analysis. Reads falling between 100-1200bp were subjected to AMR and ITS detection, while those between 1300-1700bp were subjected to taxonomic classification based on the 16S rRNA gene. In monomicrobial samples, the mean number of reads ± SD and median for reads between 100-1200 bp were 2594.94 ± 6018.79 and 684.50, respectively. For reads between 1300-1700 bp, the corresponding values were 4543.81 ± 7935.80 and 873. For polymicrobial samples, the mean number of reads ± SD and median for reads between 100-1200 bp were 5663.30 ± 9161.68 and 2718.50, while for reads between 1300-1700 bp, they were 11559.60 ± 14581.52 and 3628. Likewise, polymicrobial samples exhibited a notably greater number of reads for both the 100-1200 bp range (P=0.0004) and the 1300-1700 bp range (P=0.0003) when compared with monomicrobial samples.

### 5.3.5 The statistic of sequencing reads in NMgS of 138 body fluids

After 4 hours of sequencing, NMgS workflow yielded an average of 1.92Gb per sequencing run, with an average of 1.55Gb passing the QSCORE threshold. The mean N50 of read length was

1006.4 bases. The mean number of reads ± SD per sample was 17284.96 ± 18877.93, with a median of 10750.50. Following filtering out the reads below 200bp, the mean number of reads ± SD per sample was 16323.27 ± 18243.49, with a median of 10519. For monomicrobial samples, the mean number of filtered reads ± SD and median were 15449.89 ± 18157.39 and 8988, respectively. Comparatively, the mean number of reads ± SD and median of polymicrobial samples were 18540.31 ± 18273.72 and 12468, respectively. There was no significant difference in number of filtered reads between monomicrobial and polymicrobial samples (P=0.1429).

A total of 30 samples were sequenced up to 48 hours to study the impact of prolonged sequencing on the sensitivity of the assay. After 48 hours of sequencing, the NMgS workflow yielded an average of 15.30Gb per sequencing run, with an average of 10.82 Gb meeting the QSCORE threshold. The average N50 read length was 756.50 bases. Following 24 hours of sequencing, the total number of reads ± SD per sample was 107723.17 ± 121597.42, with a median of 60177. The number of filtered reads ± SD per sample was 102959.63 ± 120478.91, with a median of 56454. After 48 hours of sequencing, the total number of reads ± SD per sample was 169769.33 ± 201218.11, with a median of 77760. The number of filtered reads ± SD per sample was 159755.73 ± 195031.23, with a median of 72355.

### 5.3.6 The concordance between NTS and culture in pathogen identification

The concordance between NTS and culture for pathogen identification in each sample is summarized in Supplementary 11. The overall concordance between NTS and culture for pathogen identification in monomicrobial samples and polymicrobial samples is illustrated in Figure 5-4.

In the 98 monomicrobial samples, there were 95 samples of bacterial infections and 3 samples of fungal infections. NTS detected at least one read of the targeted species with reference to culture in 89 samples (90.08%). When the $T_{RA}$ of 0.058 was applied, targeted species in 87 samples (88.78%) were detected. Among the 89 concordant samples, there were 10 samples that had very low classified reads (<10 reads) of the targeted species, one of them had a relative abundance below $T_{RA}$. In the 9 discordant samples (9.18%), two samples have completely no classified reads, while 7 of them had at least one classified read for species other than expected. In 6 out of these 7 samples, the classified species were mainly environmental contaminants, such as *Pelomonas saccharophila*, *Ralstonia insidiosa*, *Burkholderia stabilis*, and *Sphingomonas lacus*. The mean number of classified reads ± SD of the targeted species was 4278.749 ± 7739.223, with a median of 677. The mean number of classified bacterial species by Emu ± SD was 4.04 ± 6.16, with a median of 2. For ITS analysis, there were 42 samples (42.86%) had at least one classified read, 24 of them had very low total classified reads (<10 reads). Among the 3 samples with fungal infections, NTS detected two of the targeted fungal species, one of them had classified reads below 10.

Regarding the 40 polymicrobial samples, there were 35 samples of bacterial infections, 1 sample of fungal infections, and 4 samples of bacterial and fungal co-infections. NTS detected all the targeted species in 16 samples (40%), with at least one targeted species detected in the remaining 24 samples (60%). A total of 131 species were detected from the 40 polymicrobial samples by culture. Among which, NTS detected 95 targeted species (72.52%), with 3 of them had very low classified reads (<10 reads). The mean number of classified reads ± SD of the targeted species was 2839.31 ± 6686.79, with a median of 232, in polymicrobial samples. An average of 18.38 bacterial species were classified by Emu in the 40 polymicrobial samples, with a SD of 23.93 and a median

of 8. A total of 11 samples (27.50%) showed at least one classified read in the ITS analysis, and 6 of them had reads below 10. NTS detected 4 out of 6 fungal species in the polymicrobial samples, with one targeted species with classified reads below 10.

The overall concordance between NTS workflow and culture in 138 body fluids was 80.35% (184/229), when targeted species with at least one read were considered detected. If $T_{RA}$ of 0.058 was applied to the monomicrobial samples, the overall concordance became 79.48% (182/229). If a threshold of detection of ten reads was further applied, the overall concordance became 74.24% (170/229).

**Figure 5-4: The concordance between NTS and culture for pathogen identification in (a) monomicrobial samples and (b) polymicrobial samples.**
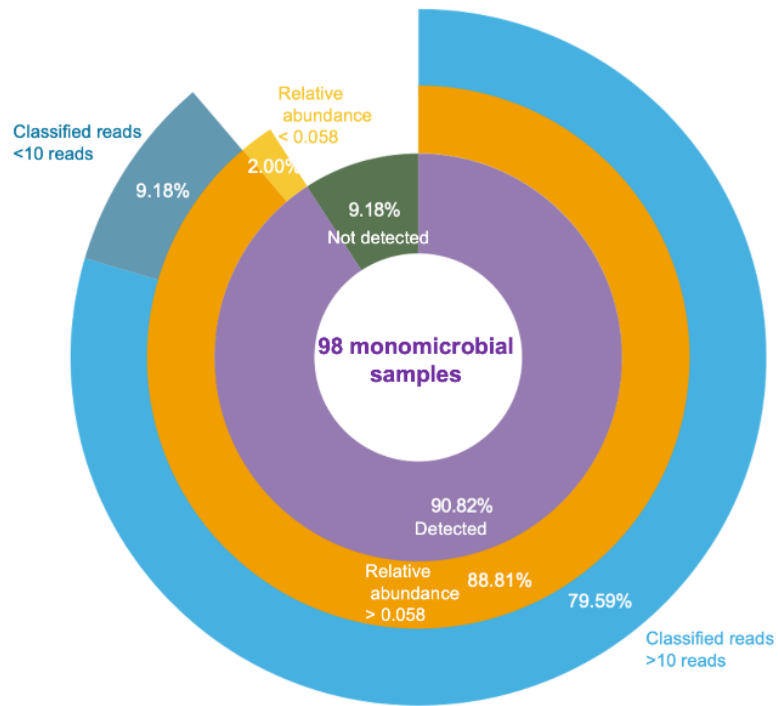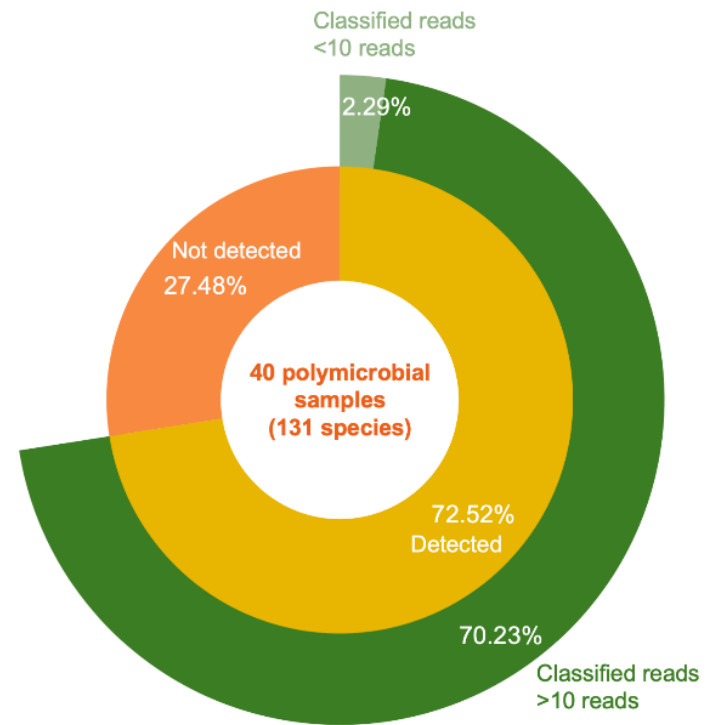
### 5.3.6 The concordance between NMgS and culture in pathogen identification

The concordance between NMgS and culture for pathogen identification in each sample was detailed in Supplementary 12. An overview of the overall concordance between NMgS and culture for pathogen identification in both monomicrobial and polymicrobial samples is illustrated in Figure 5-5.

For monomicrobial samples, NMgS detected 37 out of 98 samples (37.76%) after 4 hours of sequencing, utilizing an optimal threshold of 10 reads per species in Bracken. The mean number of reads for the targeted species was 9377.95, with a SD of 19095.03 and a median of 1388. Among the 98 monomicrobial samples, 58 samples (59.18%) did not yield any classified reads post threshold application. In the remaining 40 samples, the mean number of classified species was 2.23, with a SD of 2.04 and a median of 1.

In a set of 40 polymicrobial samples, NMgS successfully identified all targeted species in 6 samples (15.0%) and at least one but not all targeted species in 22 samples (55.0%) after four hours of sequencing. Among the 12 samples with no matches at all (30.0%), 11 yielded no classified reads under the 10-read threshold in Bracken. The average number of classified species in the remaining 29 samples was 8.21, with a standard deviation of 9.11 and a median of 5. Out of the 131 species cultured from the 40 polymicrobial samples, NMgS detected 58 species (44.27%).

Without a threshold in Bracken, NMgS detected 53 out of 98 targeted species (54.08%) in monomicrobial samples. Furthermore, 79 samples (80.61%) had at least one classified read, and the mean number of classified species per sample increased to 7.97, with a SD of 11.91 and a median of 4 in monomicrobial samples.

In the case of polymicrobial samples, NMgS detected all the targeted species in 14 out of 40 samples (35.0%) and at least one but not all targeted species in 18 samples (45.0%), without the threshold in Bracken. Among the 8 completely unmatched samples (20.0%), 4 still had no classified reads without the threshold. The mean number of classified species in the resulting 36 samples was 24.50, with a SD of 27.96 and a median of 14. Out of the 131 species in the 40 polymicrobial samples, NMgS detected 86 species (65.65%).

In a 4-hour sequencing period with a Bracken threshold set at 10, the overall concordance between the NMgS workflow and culture across 138 body fluids was 41.48% (95/229). The concordance rate increased to 60.70% (139/229) when no threshold was applied in Bracken.

A set of 30 samples were sequenced up to 48 hours to investigate the effect of prolonged sequencing time on sensitivity of NMgS workflow. Out of 40 species that were cultured from the 30 samples, NMgS detected 15 of the targeted species (37.50%) after 4 hours of sequencing. Subsequently, it detected 5 additional targeted species (50.0%) after 24 hours of sequencing and 9 more targeted species (60.0%) after 48 hours of sequencing, in comparison to the initial results following the 4-hour mark. If threshold was not employed in Bracken, NMgS detected 26 out of 40 species (65.0%) after 4 hours of sequencing. This number increased to 30 species (75.0%) after 24 hours of sequencing and further to 31 species (77.5%) after 48 hours of sequencing.

**Figure 5-5: The concordance between NMgS and culture for pathogen identification in (a) monomicrobial samples and (b) polymicrobial samples.**
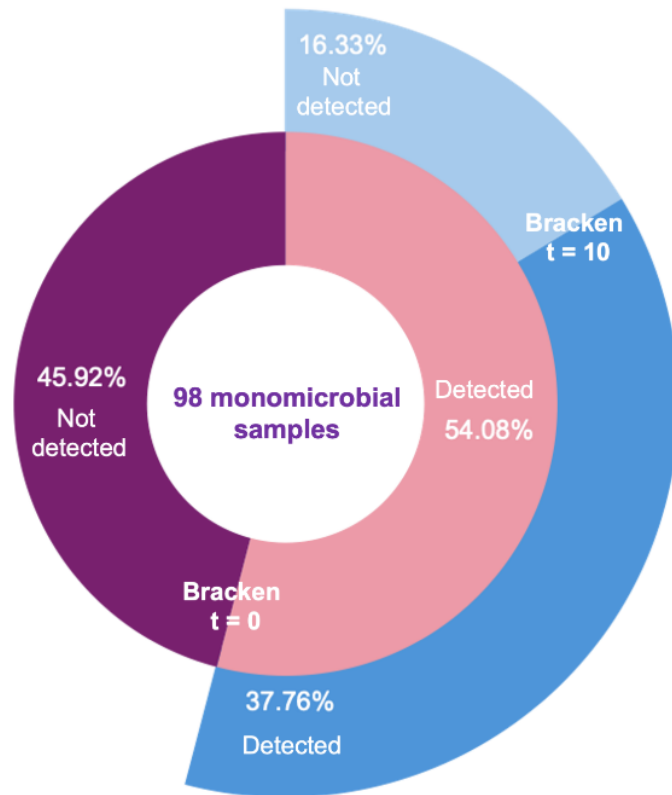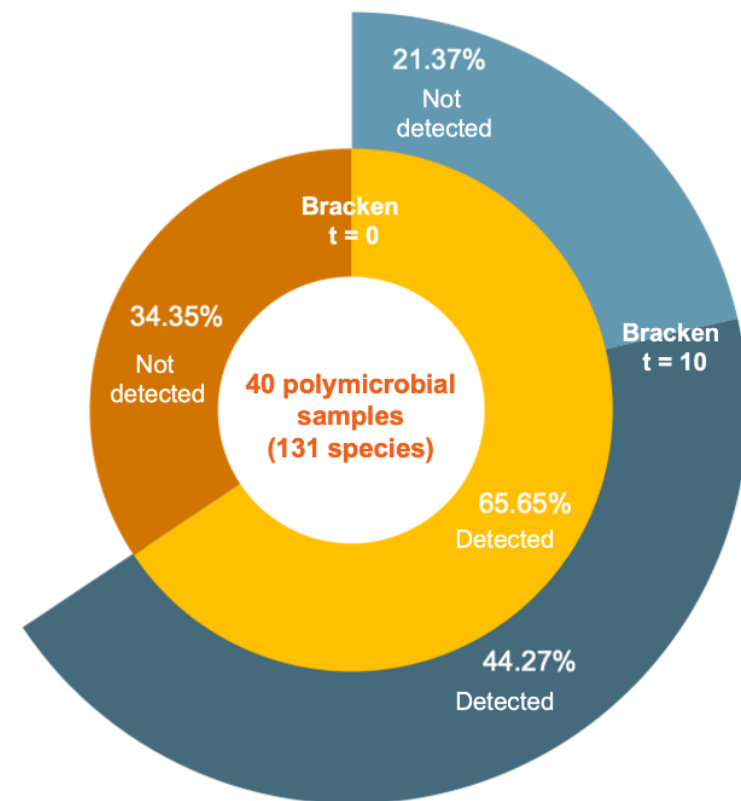
### 5.3.7 The concordance between NTS, NMgS and culture in AMR identification

A total of 24 antimicrobial resistant ESKAPE pathogens were identified in 20 out of 138 samples through culture-based methods, encompassing 12 ESBL-producing *Enterobacterales*, 8 methicillin-resistant *S. aureus*, and 4 carbapenem-resistant organisms. No vancomycin-resistant *Enterococcus* was found in this study. An overview of the concordance in the identification of AMR ESKAPE pathogens across NTS, NMgS, and culture is illustrated in Figure 5-6. The AMR genes detected in NTS and NMgS are detailed in Supplementary 13.

### 5.3.7.1 Detection of ESBL and AmpC genes

While NTS detected only the presence of targeted AMR genes in a sample, it provided no information about the origin of the AMR gene. Among the twelve ESBL-PE strains detected in 9 samples via culture, NTS identified 11 species in 8 samples, and ESBL genes were detected in 7 samples (77.78%). However, one sample had only two reads for the blaCTX-M gene. The most prevalent ESBL gene was blaCTX-M, followed by blaTEM, and blaOXA, which were detected in 7, 5, and 1 sample out of 8 samples, respectively.

For NMgS workflow, it detected 8 targeted species in 6 samples. However, only 4 ESBL-PE strains (33.33%) were detected in 3 samples (33.33%), including three *E. coli* stains and one *K. pneumoniae* strain. While blaTEM genes were found in three samples, blaCTX-M and blaOXA genes were detected in one sample, respectively. All the ESBL genes detected in NMgS were also detected in NTS.

In a polymicrobial sample (24MB017347) containing ESBL-producing *E. coli*, both NTS and NMgS identified blaAmpC genes, specifically blaMOX and blaCMY. Notably, NMgS revealed

that the blaMOX gene originated from another pathogen, *Aeromonas caviae*, present within the polymicrobial sample. *Aeromonas* species are known to harbor chromosomal blaAmpC genes and are hypothesized to be the source of CMY-1/MOX-family enzymes (304). Moreover, NMgS identified the blaCMY gene in both *Aeromonas caviae* and *E. coli*. Additionally, NMgS exclusively detected blaEC and blaLAT in *E. coli*. This highlights that while NTS solely confirmed the existence of AMR genes, NMgS had the capability to trace the source of these AMR genes.

### 5.3.7.2 Detection of mecA genes

Among the 8 MRSA strains identified through culture, *S. aureus* was detected in all 8 samples by NTS, though one sample had fewer than 10 classified reads. The mecA gene was identified in 7 out of the 8 samples (87.50%) by NTS, but it was not detected in the sample with low classified reads for *S. aureus*.

For NMgS, *S. aureus* was found in 4 out of the 8 samples, with two samples only being detected when no threshold was applied in Bracken. Similarly, the mecA gene was only detected in the two samples (25.0%) with classified reads exceeding 10 (the Bracken threshold) in NMgS.

### 5.3.7.3 Detection of carbapenemase genes

A total of 4 carbapenem-resistant organisms were identified by culture, including one *Klebsiella pneumoniae*, one *Pseudomonas aeruginosa*, and two *Acinetobacter baumannii*. Although all species were detected in NTS, no carbapenemase genes were found. Instead, ESBL or blaAmpC genes were identified in samples containing *Klebsiella pneumoniae* and *Pseudomonas aeruginosa* by NTS. Specifically, blaTEM, blaSHV (<10 reads), and blaDHA (<10 reads) were detected in

the sample with *Klebsiella pneumoniae*, while blaTEM (<10 reads) was found in the sample with *Pseudomonas aeruginosa*.

For NMgS, only one *Pseudomonas aeruginosa* and one *Acinetobacter baumannii* were identified among the 4 carbapenem-resistant organisms. Furthermore, *Acinetobacter baumannii* was detected with classified reads below 10. Similar to NTS, no carbapenemase genes were detected in all samples by NMgS.

For the carbapenem-resistant *Klebsiella pneumoniae*, PCR tests were conducted by the clinical laboratory to validate the presence of carbapenemase genes, including blaGES, blaIMI, blaKPC, blaNmcA, blaSME, blaGIM, blaIMP, blaNDM, blaSIM, blaSPM, blaVIM, and blaOXA-48-like. However, none of the above carbapenemase genes were detected, consistent with the sequencing results. This suggests that carbapenem-resistant strains can acquire resistance to carbapenems through mechanisms other than carbapenemase production.

**Figure 5-6: The concordance in identifying AMR ESKAPE pathogens among NTS, NMgS, and culture.**
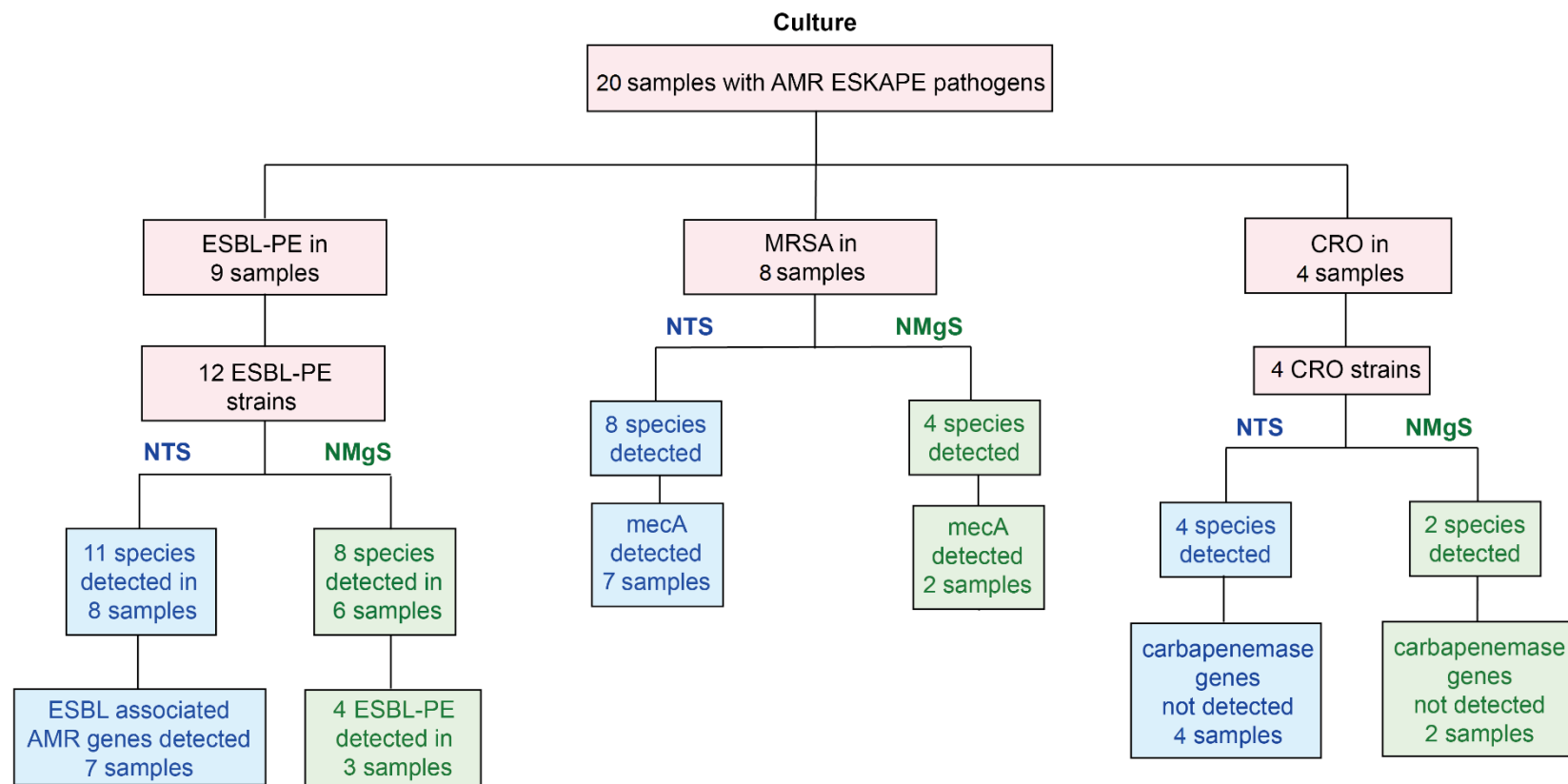
### 5.3.8 Turnaround time of NTS and NMgS workflows

The workflow and required time for each step in both NTS and NMgS are outlined in Figure 5-7.

Compared to the 24-48 hours of incubation time needed for culture, the turnaround time for both

NTS and NMgS workflows was significantly shorter, at 10.75 hours and 12.82 hours, respectively.

Similar to the 16S Barcoding Kit 1-24 (SQK-16S024), a batch of 24 samples were sequenced per

run in NTS workflow. In contrast, NMgS workflow accommodated up to 16 samples per run to

enhance sequencing depth per sample and sensitivity. In this study, both NTS and NMgS

workflows were sequenced up to 4 hours in order to have rapid diagnosis. Notably, nanopore

sequencing enables real-time sequencing, allowing the extraction of fastq files at any point during

the process for immediate analysis. Furthermore, nanopore flow cell is reusable and can be used

up to 72 hours, enabling multiple runs with a single flow cell. The analysis for both workflows

could be completed within an hour in general, depending on factors such as the number of reads

per sequencing run and available computing resources.

**Figure 5-7: The workflow of NTS and NMgS.**
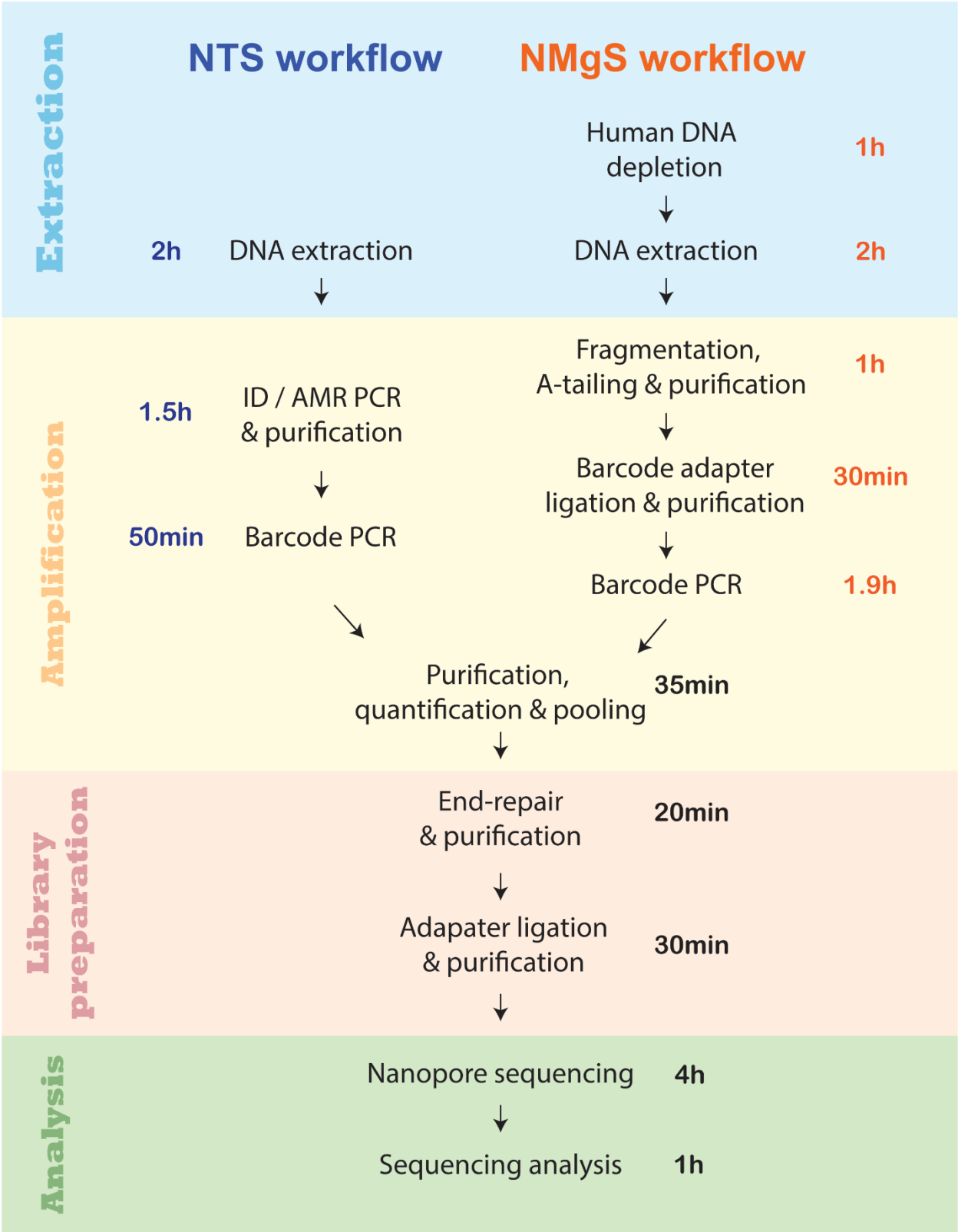
## 5.4 Discussion

In clinical microbiology, a main challenge with sequencing-based diagnosis is the lack of standardized protocols and guidelines for data interpretation. This study aimed to establish an optimized protocol for nanopore sequencing-based diagnosis. Two workflows with different sequencing approaches were developed, namely nanopore targeted sequencing (NTS) and unbiased nanopore metagenomic sequencing (NMgS). A pilot study was conducted to evaluate their performance for pathogen identification and antimicrobial resistance (AMR) detection in 138 normally sterile body fluids with reference to traditional culture-based methods. This chapter provides insights into developing standardized sequencing-based protocols and explores the strengths and limitations of targeted and unbiased metagenomic approaches in clinical diagnosis.

In the NTS workflow, identification and classification of bacterial and fungal pathogens in samples were achieved by amplifying the bacterial 16S rRNA gene and the fungal ITS gene, respectively. Notably, the NTS detection panel did not encompass viruses, primarily due to the absence of a universal DNA barcode capable of accommodating the vast genetic diversity inherent in viral genomes. Multiple DNA barcodes are necessary for detecting viruses of distinct species or lineages. Additionally, an extra reverse transcription step is essential for amplifying RNA viruses, adding complexity to the workflow. Alternatively, rapid detection for targeted viral pathogens could be achieved through real-time PCR or immunoassays. Viral culture is not routinely performed in clinical laboratories. For AMR detection a multiplex PCR approach was employed to amplify a set of clinically significant AMR primers commonly linked with ESKAPE pathogens. Despite its high sensitivity, the detection of AMR genes in NTS is limited by the scope of the detection panel, and the host organisms of the AMR genes could not be determined.

In contrast, unbiased metagenomic sequencing allows the detection of all kinds of pathogens and AMR genes within a sample. Given the exclusion of samples with viral infections in this study, reverse transcription was omitted from the NMgS workflow. Host DNA depletion is essential in NMgS to enhance the sequencing depth of microbial DNA, a saponin-PMA based host DNA depletion step was incorporated in NMgS workflow. Saponin is a natural detergent that can differentially disrupt the lipid bilayer of human cell membrane with little effect on bacterial and fungal cell wall. It is an effective and economic method commonly used for host DNA depletion, however, the working concentration of saponin varies from 0.0125% to 2.2% in different studies (180, 302). Notably, higher saponin concentrations may impact gram-negative bacteria, potentially skewing the profiling towards gram-positive bacteria (305). To mitigate such effects on the microbial profile within a sample, a saponin concentration of 0.0125% was employed in the NMgS workflow. Following saponin treatment, propidium monoazide (PMA), a photo-reactive DNA-binding dye capable of penetrating dead cells and binding to double-stranded nucleic acids upon light exposure (306), was utilized. PMA is a more efficient and economical method for degrading host DNA compared to DNases (307).

In the validation study of NMgS workflow, the combination of 0.0125% saponin and 20μM PMA effectively reduced human DNA in blood samples spiked with low microbial content (1000 CFU/ml) from an average of 99.5% to 49.1%. However, the host DNA depletion efficiency was species-specific, with the highest efficiency observed in blood samples spiked with *M. marinum*, followed by *E. coli*, *C. krusei*, and *S. aureus*. The depletion efficiency was influenced by multiple factors, including genome size and cell wall composition. Microorganisms with larger genome sizes were more likely to be detected due to higher DNA yields after extraction. Among the four spiked species, *C. krusei* had the largest genome, approximately 11 Mb, compared to 6.64Mb for

*M. marinum*, 5Mb for *E. coli*, and 2.81Mb for *S. aureus*. Furthermore, variations in cell wall components among microorganisms impacted their resistance to saponin. Microorganisms with lower tolerance to saponin were more likely to experience cell death and DNA cross-linking by PMA dye, resulting in reduced DNA yield.

Bioinformatic tools and databases can significantly affect the results for taxonomic classification and AMR identification, emphasizing the necessity for standardizing analysis methods when incorporating sequencing-based diagnosis into routine clinical microbiology practices. In both NTS and NMgS workflow, bioinformatic pipelines with good balance between accuracy, speed, and computing resources requirement were employed. A shell script was used for automated execution of multiple commands in sequencing analysis of each workflow.

To speed up the classification process in NTS analysis, reads were divided into two groups based on their lengths. Reads falling within the range of 100-1200bp underwent ITS and AMR analysis, while those within 1300-1700bp were subjected to 16S analysis. For the taxonomic classification of 16S rRNA genes, Emu was employed due to its superior ability to distinguish closely related species and its higher concordance with culture, as shown in chapter 4. BLAST+, a widely recognized gold standard, was utilized for the classification of reads in ITS and AMR analysis. This involved comparing reads against databases constructed from the NCBI ITS RefSeq Targeted Loci project (PRJNA177353) and the NCBI Reference Gene Catalog, respectively. A percentage identity of 95 was employed in all BLAST+ analysis. Given the use of a relatively compact database, the computational time required for BLAST+ analysis remained within acceptable limits. The overall sequencing analysis time for a batch of 24 samples was about 30 minutes to 1 hour, depending on the number of reads.

In the NMgS workflow, adaptive sampling was employed during the sequencing to deplete human reads in addition to host DNA depletion before DNA extraction. Various studies have demonstrated the effectiveness of adaptive sampling in depleting human reads and enriching microbial reads (308-310). Sequencing reads shorter than 200bp were filtered out, and human reads were removed using Kraken 2 against a human database. Subsequently, the remaining reads underwent taxonomic classification with Kraken 2, followed by species abundance re-estimation with Bracken. This combination was found to be more accurate and less susceptible to misclassification of clinical metagenomic samples compared to using Kraken 2 alone, particularly when dealing with long reads (200). While mini databases for Kraken 2 offer advantages such as reduced processing time and lower computing resource requirement, utilizing standard database enhances species and genus level classification accuracy. To include the taxonomic classification of fungal species, the PlusPF database, which encompasses the standard database along with RefSeq protozoa and fungi, was employed in the NMgS analysis. Additionally, setting an optimal threshold in Kraken 2 and Bracken is essential to minimize false positive species with low precision (280, 311). Using spiked blood samples with low microbial biomass (1000 CFU/ml), this study established the optimal confidence threshold for Kraken 2 at 0.1 and identified the optimal threshold for Bracken as 10. Similarly, the AMR detection in the NMgS workflow also uses BLAST+ coupled with NCBI Reference Gene Catalog. To identify the possible host of AMR genes, the classified reads in AMR analysis were extracted and re-classified using BLAST+ with the NCBI nucleotide database (nt). The overall analysis pipeline takes approximately 30 minutes for a batch of 16 samples following 4 hours of sequencing.

The NTS workflow showed an overall concordance rate of 80.35% (184/229) with culture for pathogen identification in 138 normally sterile body fluids. Notably, a higher concordance with

culture was observed in monomicrobial samples (90.82%) compared to polymicrobial samples (72.82%). In contrast, the NMgS workflow had an overall concordance rate of 60.70%, with 54.08% of targeted species detected in monomicrobial samples and 65.65% in polymicrobial samples. The lower sensitivity in the NMgS workflow was expected due to the difficulty in complete removal of human DNA, leading to microbial reads being masked by abundant human DNA, especially in samples with very low microbial biomass. To assess the rapid diagnostic capability of NMgS, the sequencing time was set to 4 hours, matching that of the NTS workflow. While extending the sequencing time to 24 or 48 hours could enhance the sensitivity of the NMgS workflow, it was impractical for routine clinical diagnosis, as culture results are typically available within 24-48 hours. Since sequencing output directly impacts the sensitivity of the NMgS workflow, utilizing an ultra-high-throughput Nanopore sequencer, such as PromethION, could enhance sensitivity within a shorter timeframe. For comparison, each MinION flow cell can generate up to 48 Gb of data, whereas a PromethION flow cell can produce up to 200 Gb.

While applying a detection threshold can help to distinguish potential pathogens from the environmental contaminants, it can also decrease the sensitivity of the assay. In the NTS workflow, utilizing a TRA of 0.058 (established in chapter 4) for monomicrobial samples decreased the concordance rate from 90.82% to 88.81%, leading to an overall concordance rate of 79.48%. In the NMgS workflow, an optimal confidence threshold of 0.1 for Kraken 2 and a read threshold of 10 for Bracken were determined. Upon implementation of these thresholds, the overall concordance rate decreased from 60.70% to 41.48%, with 37.76% of targeted species detected in monomicrobial samples and 44.27% in polymicrobial samples. Despite target amplification in the NTS workflow, certain targeted species still exhibited fewer than 10 reads in both monomicrobial (9.18%) and polymicrobial samples (2.29%). When a read threshold of 10 was applied in NTS,

the concordance rates for monomicrobial and polymicrobial samples became 79.59% and 70.23%, respectively, resulting in an overall concordance rate of 74.24%.

NTS also demonstrated a higher sensitivity than NMgS in detecting AMR ESKAPE pathogens. A total of 20 samples were found to contain AMR ESKAPE pathogens in culture, including 12 ESBL-PE, 8 MRSA, and 4 CRO. One main limitation of NTS in AMR detection is its inability to determine the hosts of AMR genes. Culture identified 12 ESBL-PE species across 9 samples, with 3 samples harboring 2 AMR ESKAPE pathogens each. Although ESBL genes were detected in these 3 samples, indicating the presence of ESBL-producing strains, the exact hosts remained unidentified. As a result, ESBL genes were detected in 7 out of 9 samples by NTS, but with one sample below the threshold of 10 reads. Comparatively, NMgS offers the potential to identify the likely hosts of AMR genes by classifying neighboring flanking sequences of these genes. Although a threshold of 10 reads provided more reliable results in AMR detection, support from just one relevant read was acceptable in nanopore metagenomic sequencing due to the long-read lengths (312). Therefore, in this study, the sample was considered positive for antimicrobial resistance with the presence of at least one relevant read. Among the 12 ESBL-PE species, 4 were detected by NMgS. In the case of MRSA strains, NTS showed high concordance with culture, with the mecA gene detected in 7 out of 8 samples. However, NMgS displayed lower sensitivity, detecting only 2 MRSA strains. Regarding carbapenem-resistant organisms, neither NTS nor NMgS identified carbapenemase genes in the 4 cases. Previous PCR testing in the clinical laboratory that aimed to verify the presence of carbapenemase in the carbapenem-resistant *K. pneumoniae* also yielded negative results. The absence of carbapenemase genes implies that carbapenem resistance might have been acquired through alternative mechanisms. This underscores the challenge of inferring phenotypic antimicrobial resistance from genotypic findings. Discrepancies between

genotypic and phenotypic resistance have also been reported by other studies (313-315), indicating that the presence of AMR genes does not always correspond to phenotypic resistance, and vice versa.

This study has several limitations. The optimization of NTS and NMgS workflows relied on blood samples spiked with microorganisms at known concentrations, as blood is the most accessible sterile body fluid from healthy individuals. However, no patient blood samples were collected due to the practice of direct drawing into blood culture bottles to minimize contamination risks in suspected bloodstream infections. Consequently, there were no residual blood samples available for collection from hospitals. Additionally, each blood sample was spiked with only one species to simplify the analysis, meaning polymicrobial samples were not prepared and validated in NTS and NMgS workflows. The use of spiked blood samples may not perfectly mimic the biological matrix of real clinical samples. For example, the number of white blood cells would differ between samples from healthy individuals and patients, and different specimen types have varying compositions. Consequently, spiked samples might not capture the full complexity of actual clinical samples, potentially leading to an overestimation of the performance of the diagnostic workflows.

Moreover, most of the samples were associated with bacterial infections, with only a few showing fungal infections. The performance of NTS and NMgS in fungal identification may not be accurately reflected. Furthermore, the evaluation of NTS and NMgS performance was solely based on comparisons with routine culture results. However, traditional culture may not fully represent the microbial profile in clinical samples, as it often overlooks anaerobic bacteria, particularly obligate anaerobes (316). Notably, some obligate anaerobes, such as *Parvimonas micra*, were detected in NTS and NMgS but not in culture. Additionally, nanopore R9.4.1 flow cells were used

in this study. The performance of the latest released R10.4 flow cells, which claimed to have higher read accuracy, was not investigated.

## 5.5 Conclusion

This study developed and compared two nanopore-based sequencing workflows: NTS and NMgS workflows for pathogen identification and antimicrobial resistance detection, referencing culture methods. With target amplification, the NTS workflow demonstrated higher sensitivity and concordance with culture in pathogen and AMR detection. However, its detection panel limitations hindered identifying pathogens and AMR genes beyond the panel. On the other hand, NMgS showed a broad pathogen and AMR gene detection range in samples, but the presence of abundant human DNA limited assay sensitivity. Complete elimination of human DNA in the host DNA depletion process was unattainable, and a single protocol might not be optimized for all specimen types. Although increasing sequencing time and reducing sample size could enhance sensitivity in NMgS, these strategies are impractical for routine clinical diagnosis.

With higher sensitivity, shorter turnaround time, and lower costs, NTS is advantageous over NMgS for routine rapid diagnosis in acute invasive infections. In contrast, NMgS is more suitable for detecting infections involving unknown or novel pathogens, such as the discovery of SARS-CoV-2. Nevertheless, sequencing-based tests still cannot substitute traditional culture; instead, they complement by providing rapid preliminary results. Culture remains essential for samples with extremely low microbial biomass falling below the detection limits of sequencing-based diagnosis and confirming the phenotypic antimicrobial resistance.

# Chapter 6: Overall conclusion and recommendations for future studies

Long-read sequencing has unlocked the potential for rapid pathogen identification and antimicrobial resistance detection, potentially revolutionizing the field of clinical diagnosis. This study demonstrated that long-read nanopore sequencing offers a higher resolution in taxonomic identification of clinical isolates than the short-read Illumina sequencing, using 16S rRNA gene sequencing. Additionally, nanopore sequencing is advantageous over traditional Sanger sequencing in directly identifying pathogens from normally sterile body fluids. The elimination of lengthy incubation periods can significantly reduce sample-to-report times. The real-time sequencing, reusable flow cells, flexible sample size and sequencing time, and low set-up cost of nanopore sequencing facilitate rapid diagnosis in clinical laboratories.

The primary challenge hindering the implementation of sequencing-based diagnosis in clinical laboratories is the lack of standardized protocols and guidelines for data interpretation. This study developed two nanopore sequencing workflows: nanopore targeted sequencing (NTS) and nanopore metagenomic sequencing (NMgS), aimed at providing insights for establishing a standardized direct sequencing protocol for clinical samples. For normally sterile body fluids, any presence of microorganisms is regarded as pathogens in general. However, sequencing is highly susceptible to environmental DNA contamination. While establishing a detection threshold can help to distinguish possible pathogens from false positive species, the optimal detection threshold may differ based on the sequencing approach, analysis pipeline, and sample type. Lower detection thresholds might be necessary for samples with extremely low microbial biomass and consequently low sequencing output. To prevent contaminations in sequencing-based diagnosis, it is crucial to regularly clean the work environment, establish pre- and post-PCR areas, and include extraction and non-template controls.

This study demonstrated the clinical utility of Nanopore sequencing for rapid identification and revealing novel species. While NTS is recommended for routine diagnosis in invasive infections due to its heightened sensitivity and cost-effectiveness, NMgS aids in diagnosing infections of unknown origins. However, further studies are required for enhancing the diagnostic accuracy of NTS and NMgS workflows and facilitating the adoption of nanopore sequencing –based tests in clinical laboratories.

Recommendations for future studies:

(1) Expand the sample size and optimization for diverse sample types

While this study incorporated normally sterile body fluids, there was an uneven distribution of samples across different specimen types, with peritoneal fluid being the predominant specimen. Further research is advised to broaden the sample pool and to assess the performance and applicability of nanopore sequencing workflows across various specimen types. Moreover, a uniform extraction method may not be suitable for all specimen types. It is suggested to refine nanopore sequencing workflows for diverse sample varieties, including tissues and viscous pus samples, to ensure robust performance across a spectrum of clinical specimens.

(2) Onsite prospective evaluation

In this retrospective study, residual body fluids were collected to evaluate the NTS and NMgS workflows in comparison to traditional culture methods. An onsite prospective evaluation, where NTS and NMgS workflows are performed in parallel with routine culture, is recommended to further assess the clinical utility of NTS and NMgS workflows, focusing on sample-to-report time

and sensitivity. This approach will also aid in standardizing the workflows and establishing interpretation guidelines for sequencing-based diagnosis in clinical laboratories.

(3) Genotypic-Phenotypic Antimicrobial Resistance Correlation

Discrepancies between genotypic and phenotypic antimicrobial profiles were observed in this study, aligning with findings from other research groups (313, 315, 317). Further exploration of the correlation between genotypic and phenotypic antimicrobial resistance profiles in antimicrobial resistant bacteria is essential for precise antimicrobial resistance prediction within NTS and NMgS workflows. The Hi-C technique can be utilized in metagenomic sequencing to link plasmid-mediated AMR genes with their host genomes within the same cell, thereby identifying the host of the AMR genes. Subsequently, the AMR genes detected in an organism can be compared with its phenotypic antimicrobial resistance.

(4) Regular Updates and Maintenance

The significance of regularly updating databases for precise pathogen identification was underscored in chapter 2. While ONT persists in enhancing the read accuracy and throughput of Nanopore sequencing, various researchers are continually developing new analysis tools. Therefore, future efforts should encompass regular updates of databases and the maintenance of analysis pipelines to keep abreast of advancements in the field.

# Appendices

**Appendix 1.** Ethical approval for sample collection in the 5 public hospitals

| Name of hospital | Ethics committee | Reference number |
|---|---|---|
| Princess Margaret Hospital | The Kowloon West Cluster Research Ethics Committee (KWC-REC) | KW/EX-22-076(176-04) |
| Prince of Wales Hospital | Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee (CUHK-NTEC CREC) | CREC Ref. No.: 2022.381 |
| Pamela Youde Nethersole Eastern Hospital | Hong Kong East Cluster Research Ethics Committee (HKEC REC) | HKECREC-2021-053 |
| Tuen Mun Hospital | New Territories West Cluster Research Ethics Committee (NTWC REC) | NTWC/REC/22087 |
| United Christian Hospital | Research Ethics Committee (Kowloon Central/ Kowloon East) | KC/KE-22-0146/ER-1 |

**Appendix 2.** The alignments of R062 reads from Sanger 16S, Illumina 16S, and Nanopore 16S sequencing workflows to the reference 16S rRNA gene sequence of *Klebsiella michiganensis* (NR_118335.1).
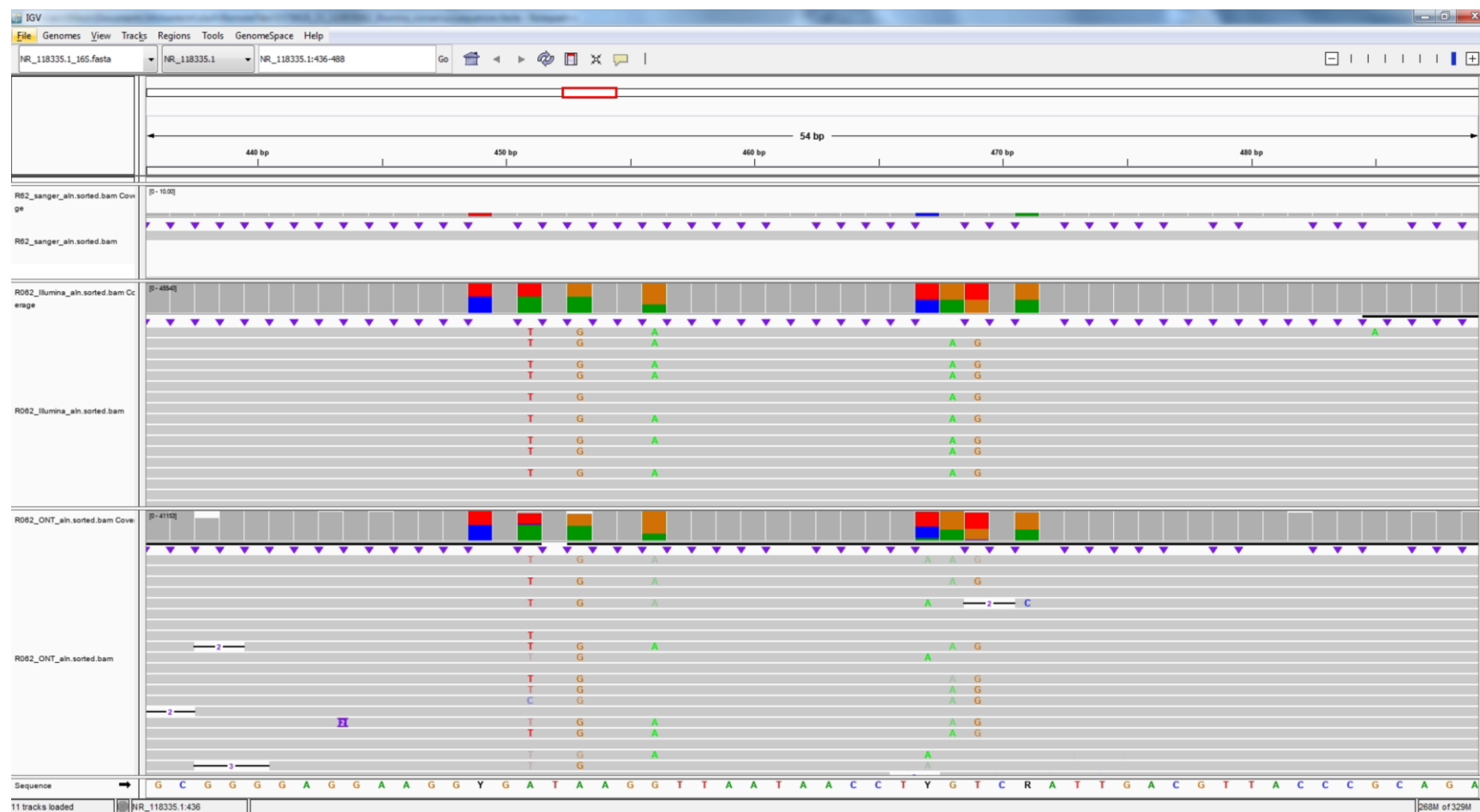
# References

1.	Balloux F, van Dorp L. Q&A: What are pathogens, and what have they done to and for us? BMC Biol. 2017;15(1):91.
2.	WHO. The top 10 causes of death 2024 [Available from: https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death.
3.	Collaborators GBDAR. Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2022;400(10369):2221-48.
4.	Antimicrobial Resistance C. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399(10325):629-55.
5.	Lee JH, Cho HK, Kim KH, Kim CH, Kim DS, Kim KN, et al. Etiology of invasive bacterial infections in immunocompetent children in Korea (1996-2005): a retrospective multicenter study. J Korean Med Sci. 2011;26(2):174-83.
6.	Hotchkiss RS, Moldawer LL, Opal SM, Reinhart K, Turnbull IR, Vincent JL. Sepsis and septic shock. Nat Rev Dis Primers. 2016;2:16045.
7.	Nasir N, Jamil B, Siddiqui S, Talat N, Khan FA, Hussain R. Mortality in Sepsis and its relationship with Gender. Pak J Med Sci. 2015;31(5):1201-6.
8.	Luo Q, Lu P, Chen Y, Shen P, Zheng B, Ji J, et al. ESKAPE in China: epidemiology and characteristics of antibiotic resistance. Emerg Microbes Infect. 2024;13(1):2317915.
9.	De Oliveira DMP, Forde BM, Kidd TJ, Harris PNA, Schembri MA, Beatson SA, et al. Antimicrobial Resistance in ESKAPE Pathogens. Clin Microbiol Rev. 2020;33(3).
10.	Miller WR, Arias CA. ESKAPE pathogens: antimicrobial resistance, epidemiology, clinical impact and therapeutics. Nat Rev Microbiol. 2024.
11.	Akova M. Epidemiology of antimicrobial resistance in bloodstream infections. Virulence. 2016;7(3):252-66.
12.	Ueda T, Takesue Y, Nakajima K, Ichiki K, Ishikawa K, Yamada K, et al. Correlation between Antimicrobial Resistance and the Hospital-Wide Diverse Use of Broad-Spectrum Antibiotics by the Antimicrobial Stewardship Program in Japan. Pharmaceutics. 2023;15(2).
13.	Leibovici L, Shraga I, Drucker M, Konigsberger H, Samra Z, Pitlik SD. The benefit of appropriate empirical antibiotic treatment in patients with bloodstream infection. J Intern Med. 1998;244(5):379-86.
14.	Chinemerem Nwobodo D, Ugwu MC, Oliseloke Anie C, Al-Ouqaili MTS, Chinedu Ikem J, Victor Chigozie U, et al. Antibiotic resistance: The challenges and some emerging strategies for tackling a global menace. J Clin Lab Anal. 2022;36(9):e24655.
15.	Ohnuma T, Chihara S, Costin B, Treggiari MM, Bartz RR, Raghunathan K, et al. Association of Appropriate Empirical Antimicrobial Therapy With In-Hospital Mortality in Patients With Bloodstream Infections in the US. JAMA Netw Open. 2023;6(1):e2249353.
16.	Xu S, Song Z, Han F, Zhang C. Effect of appropriate empirical antimicrobial therapy on mortality of patients with Gram-negative bloodstream infections: a retrospective cohort study. BMC Infect Dis. 2023;23(1):344.
17.	Retamar P, Portillo MM, Lopez-Prieto MD, Rodriguez-Lopez F, de Cueto M, Garcia MV, et al. Impact of inadequate empirical therapy on the mortality of patients with bloodstream infections: a propensity score-based analysis. Antimicrob Agents Chemother. 2012;56(1):472-8.
18.	Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med. 2006;34(6):1589-96.
19.	De la Rosa-Riestra S, Martinez Perez-Crespo PM, Perez Rodriguez MT, Sousa A, Goikoetxea J, Reguera Iglesias JM, et al. Mortality impact of further delays in active targeted antibiotic therapy in bacteraemic patients that did not receive initial active empiric treatment: Results from the prospective, multicentre cohort PROBAC. Int J Infect Dis. 2024;145:107072.

20.     Proulx N, Fréchette D, Toye B, Chan J, Kravcik S. Delays in the administration of antibiotics are associated with mortality from adult acute bacterial meningitis. Qjm-Int J Med. 2005;98(4):291-8.

21.     Ayobami O, Brinkwirth S, Eckmanns T, Markwart R. Antibiotic resistance in hospital-acquired ESKAPE-E infections in low- and lower-middle-income countries: a systematic review and meta-analysis. Emerg Microbes Infect. 2022;11(1):443-51.

22.     Arbune M, Gurau G, Niculet E, Iancu AV, Lupasteanu G, Fotea S, et al. Prevalence of Antibiotic Resistance of ESKAPE Pathogens Over Five Years in an Infectious Diseases Hospital from South-East of Romania. Infect Drug Resist. 2021;14:2369-78.

23.     Salam MA, Al-Amin MY, Salam MT, Pawar JS, Akhter N, Rabaan AA, et al. Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. Healthcare (Basel). 2023;11(13).

24.     Organization WH. Global antimicrobial resistance and use surveillance system (GLASS) report. 2022.

25.     Bush K, Bradford PA. beta-Lactams and beta-Lactamase Inhibitors: An Overview. Cold Spring Harb Perspect Med. 2016;6(8).

26.     Narendrakumar L, Chakraborty M, Kumari S, Paul D, Das B. beta-Lactam potentiators to re-sensitize resistant pathogens: Discovery, development, clinical use and the way forward. Front Microbiol. 2022;13:1092556.

27.     Gaynes R. The Discovery of Penicillin-New Insights After More Than 75 Years of Clinical Use. Emerging Infectious Diseases. 2017;23(5):849-53.

28.     Mora-Ochomogo M, Lohans CT. beta-Lactam antibiotic targets and resistance mechanisms: from covalent inhibitors to substrates. RSC Med Chem. 2021;12(10):1623-39.

29.     Bradford PA. Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. Clin Microbiol Rev. 2001;14(4):933-51, table of contents.

30.     Ibadin EE, Enabulele IO, Muinah F. Prevalence of mecA gene among staphylococci from clinical samples of a tertiary hospital in Benin City, Nigeria. Afr Health Sci. 2017;17(4):1000-10.

31.     Idrees MM, Saeed K, Shahid MA, Akhtar M, Qammar K, Hassan J, et al. Prevalence of mecA- and mecC-Associated Methicillin-Resistant Staphylococcus aureus in Clinical Specimens, Punjab, Pakistan. Biomedicines. 2023;11(3).

32.     El Aila NA, Al Laham NA, Naas T. Prevalence of mecA and Panton-Valentine Leukocidin Genes in Staphylococcus aureus Clinical Isolates from Gaza Strip Hospitals. Microorganisms. 2023;11(5).

33.     Askari E, Soleymani F, Arianpoor A, Tabatabai SM, Amini A, Naderinasab M. Epidemiology of mecA-Methicillin Resistant Staphylococcus aureus (MRSA) in Iran: A Systematic Review and Meta-analysis. Iran J Basic Med Sci. 2012;15(5):1010-9.

34.     Gittens-St Hilaire MV, Chase E, Alleyne D. Prevalence, molecular characteristics and antimicrobial susceptibility patterns of MRSA in hospitalized and nonhospitalized patients in Barbados. New Microbes New Infect. 2020;35:100659.

35.     Elhassan MM, Ozbak HA, Hemeg HA, Elmekki MA, Ahmed LM. Absence of the mecA Gene in Methicillin Resistant Staphylococcus aureus Isolated from Different Clinical Specimens in Shendi City, Sudan. Biomed Res Int. 2015;2015:895860.

36.     Centre for Health Protection. Statistics on Antimicrobial Resistance Control - Methicillin resistant Staphylococcus aureus (MRSA) 2024 [Available from: https://www.chp.gov.hk/en/statistics/data/10/100044/6864.html.

37.     Tooke CL, Hinchliffe P, Bragginton EC, Colenso CK, Hirvonen VHA, Takebayashi Y, et al. beta-Lactamases and beta-Lactamase Inhibitors in the 21st Century. J Mol Biol. 2019;431(18):3472-500.

38.     Sawa T, Kooguchi K, Moriyama K. Molecular diversity of extended-spectrum beta-lactamases and carbapenemases, and antimicrobial resistance. J Intensive Care. 2020;8:13.

39.     Rawat D, Nair D. Extended-spectrum beta-lactamases in Gram Negative Bacteria. J Glob Infect Dis. 2010;2(3):263-74.

40.     Bajpai T, Pandey M, Varma M, Bhatambare GS. Prevalence of TEM, SHV, and CTX-M Beta-Lactamase genes in the urinary isolates of a tertiary care hospital. Avicenna J Med. 2017;7(1):12-6.

41.     Song W, Lee H, Lee K, Jeong SH, Bae IK, Kim JS, et al. CTX-M-14 and CTX-M-15 enzymes are the dominant type of extended-spectrum beta-lactamase in clinical isolates of Escherichia coli from Korea. J Med Microbiol. 2009;58(Pt 2):261-6.

42.     Brolund A. Overview of ESBL-producing Enterobacteriaceae from a Nordic perspective. Infect Ecol Epidemiol. 2014;4.

43.     Ramatla T, Mafokwane T, Lekota K, Monyama M, Khasapane G, Serage N, et al. "One Health" perspective on prevalence of co-existing extended-spectrum β-lactamase (ESBL)-producing Escherichia coli and Klebsiella pneumoniae: a comprehensive systematic review and meta-analysis. Ann Clin Microb Anti. 2023;**22**(1).

44.     Cheng VC, Wong SC, Ho PL, Yuen KY. Strategic measures for the control of surging antimicrobial resistance in Hong Kong and mainland of China. Emerg Microbes Infect. 2015;4(2):e8.

45.     Centre for Health Protection. Statistics on Antimicrobial Resistance Control - Escherichia coli with reduced susceptibility to cefotaxime (CTX) or ceftazidime (CAZ) or ceftriaxone (CRO) 2024 [Available from: https://www.chp.gov.hk/en/statistics/data/10/100044/6890.html.

46.     Centre for Health Protection. Statistics on Antimicrobial Resistance Control - Klebsiella species with reduced susceptibility to cefotaxime (CTX) or ceftazidime (CAZ) or ceftriaxone (CRO) 2024 [Available from: https://www.chp.gov.hk/en/statistics/data/10/100044/6891.html.

47.     Tang F, Lee CH, Li X, Jiang S, Chow KH, Tse CW, et al. Evaluation of Two Tests for the Rapid Detection of CTX-M Producers Directly in Urine Samples. Antibiotics (Basel). 2023;12(11).

48.     Tekele SG, Teklu DS, Tullu KD, Birru SK, Legese MH. Extended-spectrum Beta-lactamase and AmpC beta-lactamases producing gram negative bacilli isolated from clinical specimens at International Clinical Laboratories, Addis Ababa, Ethiopia. PLoS One. 2020;15(11):e0241984.

49.     Imkamp F, Kolesnik-Goldmann N, Bodendoerfer E, Zbinden R, Mancini S. Detection of Extended-Spectrum beta-Lactamases (ESBLs) and AmpC in Class A and Class B Carbapenemase-Producing Enterobacterales. Microbiol Spectr. 2022;10(6):e0213722.

50.     Philippon A, Arlet G, Jacoby GA. Plasmid-determined AmpC-type beta-lactamases. Antimicrob Agents Chemother. 2002;46(1):1-11.

51.     Tamma PD, Doi Y, Bonomo RA, Johnson JK, Simner PJ, Antibacterial Resistance Leadership G. A Primer on AmpC beta-Lactamases: Necessary Knowledge for an Increasingly Multidrug-resistant World. Clin Infect Dis. 2019;69(8):1446-55.

52.     Rizi KS, Mosavat A, Youssefi M, Jamehdar SA, Ghazvini K, Safdari H, et al. High prevalence of bla(CMY) AmpC beta-lactamase in ESBL co-producing Escherichia coli and Klebsiella spp. clinical isolates in the northeast of Iran. J Glob Antimicrob Resist. 2020;22:477-82.

53.     Queenan AM, Bush K. Carbapenemases: the versatile beta-lactamases. Clin Microbiol Rev. 2007;20(3):440-58, table of contents.

54.     Papp-Wallace KM, Endimiani A, Taracila MA, Bonomo RA. Carbapenems: past, present, and future. Antimicrob Agents Chemother. 2011;55(11):4943-60.

55.     Halat DH, Moubareck CA. The Current Burden of Carbapenemases: Review of Significant Properties and Dissemination among Gram-Negative Bacteria. Antibiotics-Basel. 2020;9(4).

56.     Tzouvelekis LS, Markogiannakis A, Psichogiou M, Tassios PT, Daikos GL. Carbapenemases in Klebsiella pneumoniae and other Enterobacteriaceae: an evolving crisis of global dimensions. Clin Microbiol Rev. 2012;25(4):682-707.

57.     Centre for Health Protection. Statistics on Antimicrobial Resistance Control - Escherichia coli with reduced susceptibility to carbapenems (imipenem, meropenem or ertapenem) 2024 [Available from: https://www.chp.gov.hk/en/statistics/data/10/100044/6908.html.

58.     Centre for Health Protection. Statistics on Antimicrobial Resistance Control - Klebsiella species with reduced susceptibility to carbapenems (imipenem, meropenem or ertapenem) 2024 [Available from: https://www.chp.gov.hk/en/statistics/data/10/100044/6909.html.

59.     Nordmann P, Poirel L. Epidemiology and Diagnostics of Carbapenem Resistance in Gram-negative Bacteria. Clin Infect Dis. 2019;69(Suppl 7):S521-S8.

60.     Centre for Health Protection. Statistics on Antimicrobial Resistance Control - Acinetobacter species with reduced susceptibility to carbapenems (imipenem, meropenem or doripenem) 2024 [Available from: https://www.chp.gov.hk/en/statistics/data/10/100044/6910.html.

61.     Stogios PJ, Savchenko A. Molecular mechanisms of vancomycin resistance. Protein Sci. 2020;29(3):654-69.

62.     Rubinstein E, Keynan Y. Vancomycin revisited - 60 years later. Front Public Health. 2014;2:217.

63.     Sinha Roy R, Yang P, Kodali S, Xiong Y, Kim RM, Griffin PR, et al. Direct interaction of a vancomycin derivative with bacterial enzymes involved in cell wall biosynthesis. Chem Biol. 2001;8(11):1095-106.

64.     Shrestha S, Kharel S, Homagain S, Aryal R, Mishra SK. Prevalence of vancomycin-resistant enterococci in Asia-A systematic review and meta-analysis. J Clin Pharm Ther. 2021;46(5):1226-37.

65.     Weiner LM, Webb AK, Limbago B, Dudeck MA, Patel J, Kallen AJ, et al. Antimicrobial-Resistant Pathogens Associated With Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011-2014. Infect Control Hosp Epidemiol. 2016;37(11):1288-301.

66.     Brinkwirth S, Ayobami O, Eckmanns T, Markwart R. Hospital-acquired infections caused by enterococci: a systematic review and meta-analysis, WHO European Region, 1 January 2010 to 4 February 2020. Euro Surveill. 2021;26(45).

67.     Fujiya Y, Harada T, Sugawara Y, Akeda Y, Yasuda M, Masumi A, et al. Transmission dynamics of a linear vanA-plasmid during a nosocomial multiclonal outbreak of vancomycin-resistant enterococci in a non-endemic area, Japan. Scientific Reports. 2021;11(1):14780.

68.     Centre for Health Protection. Statistics on Antimicrobial Resistance Control - Vancomycin resistant enterococcus (VRE) 2024 [Available from: https://www.chp.gov.hk/en/statistics/data/10/100044/6863.html.

69.     Nasaj M, Mousavi SM, Hosseini SM, Arabestani MR. Prevalence of Virulence Factors and Vancomycin-resistant Genes among Enterococcus faecalis and E. faecium Isolated from Clinical Specimens. Iran J Public Health. 2016;45(6):806-13.

70.     Arredondo-Alonso S, Top J, Corander J, Willems RJL, Schurch AC. Mode and dynamics of vanA-type vancomycin resistance dissemination in Dutch hospitals. Genome Med. 2021;13(1):9.

71.     Cetinkaya Y, Falk P, Mayhall CG. Vancomycin-resistant enterococci. Clin Microbiol Rev. 2000;13(4):686-707.

72.     Palmieri A, Martinelli M, Pellati A, Carinci F, Lauritano D, Arcuri C, et al. Prevalence of Enterococci and Vancomycin Resistance in the Throat of Non-Hospitalized Individuals Randomly Selected in Central Italy. Antibiotics (Basel). 2023;12(7).

73.     Laupland KB, Valiquette L. The changing culture of the microbiology laboratory. Can J Infect Dis Med Microbiol. 2013;24(3):125-8.

74.     Houpikian P, Raoult D. Traditional and molecular techniques for the study of emerging bacterial diseases: one laboratory's perspective. Emerg Infect Dis. 2002;8(2):122-31.

75.     Church D, Melnyk E, Unger B. Quantitative gram stain interpretation criteria used by microbiology laboratories in Alberta, Canada. J Clin Microbiol. 2000;38(11):4266-8.

76.     Vijayakumar T, Divya B, Vasanthi V, Narayan M, Kumar AR, Krishnan R. Diagnostic Utility of Gram Stain for Oral Smears - A Review. J Microsc Ultrastruct. 2023;11(3):130-4.

77.     Beveridge TJ. Use of the Gram stain in microbiology. Biotech Histochem. 2001;76(3):111-8.

78.     Silhavy TJ, Kahne D, Walker S. The bacterial cell envelope. Cold Spring Harb Perspect Biol. 2010;2(5):a000414.

79.     Tripathi N, Sapra A. Gram Staining.  StatPearls. Treasure Island (FL)2024.

80.     Altun O, Almuhayawi M, Ullberg M, Ozenci V. Rapid identification of microorganisms from sterile body fluids by use of FilmArray. J Clin Microbiol. 2015;53(2):710-2.

81.     Ohst C, Saschenbrecker S, Stiba K, Steinhagen K, Probst C, Radzimski C, et al. Reliable Serological Testing for the Diagnosis of Emerging Infectious Diseases. Adv Exp Med Biol. 2018;1062:19-43.

82. Mohammadi SF, Patil AB, Nadagir SD, Nandihal N, Lakshminarayana SA. Diagnostic value of latex agglutination test in diagnosis of acute bacterial meningitis. Ann Indian Acad Neurol. 2013;16(4):645-9.

83. Amidu N, Antuamwine BB, Addai-Mensah O, Abdul-Karim A, Stebleson A, Abubakari BB, et al. Diagnosis of bacterial meningitis in Ghana: Polymerase chain reaction versus latex agglutination methods. PLoS One. 2019;14(1):e0210812.

84. Tate J, Ward G. Interferences in immunoassay. Clin Biochem Rev. 2004;25(2):105-20.

85. Bhaskar Shenoy SB. Latex Agglutination Test (LAT) – For rapid diagnosis of acute bacterial meningitis. Pediatric Infectious Disease. 2014;**6**(4):150-2.

86. Tarafdar K, Rao S, Recco RA, Zaman MM. Lack of sensitivity of the latex agglutination test to detect bacterial antigen in the cerebrospinal fluid of patients with culture-negative meningitis. Clin Infect Dis. 2001;33(3):406-8.

87. Padmanaban V, Ranganathan UDK. CRISPR-Cas system and its use in the diagnosis of infectious diseases. Microbiol Res. 2022;263:127100.

88. Abayasekara LM, Perera J, Chandrasekharan V, Gnanam VS, Udunuwara NA, Liyanage DS, et al. Detection of bacterial pathogens from clinical specimens using conventional microbial culture and 16S metagenomics: a comparative study. BMC Infect Dis. 2017;17(1):631.

89. Akcam FZ, Yayli G, Uskun E, Kaya O, Demir C. Evaluation of the Bactec microbial detection system for culturing miscellaneous sterile body fluids. Res Microbiol. 2006;157(5):433-6.

90. Awulachew E, Diriba K, Awoke N. Bacterial Isolates from CSF Samples and Their Antimicrobial Resistance Patterns Among Children Under Five Suspected to Have Meningitis in Dilla University Referral Hospital. Infect Drug Resist. 2020;13:4193-202.

91. Singhal N, Kumar M, Kanaujia PK, Virdi JS. MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. Front Microbiol. 2015;6:791.

92. Halperin AV, Del Castillo Polo JA, Cortes-Cuevas JL, Cardenas Isasi MJ, Ampuero Morisaki M, Birch R, et al. Impact of Automated Blood Culture Systems on the Management of Bloodstream Infections: Results from a Crossover Diagnostic Clinical Trial. Microbiol Spectr. 2022;10(5):e0143622.

93. Tjandra KC, Ram-Mohan N, Abe R, Hashemi MM, Lee JH, Chin SM, et al. Diagnosis of Bloodstream Infections: An Evolution of Technologies towards Accurate and Rapid Identification and Antibiotic Susceptibility Testing. Antibiotics (Basel). 2022;11(4).

94. Opota O, Croxatto A, Prod'hom G, Greub G. Blood culture-based diagnosis of bacteraemia: state of the art. Clin Microbiol Infect. 2015;21(4):313-22.

95. Giuliano C, Patel CR, Kale-Pradhan PB. A Guide to Bacterial Culture Identification And Results Interpretation. P T. 2019;44(4):192-200.

96. Harris AM, Bramley AM, Jain S, Arnold SR, Ampofo K, Self WH, et al. Influence of Antibiotics on the Detection of Bacteria by Culture-Based and Culture-Independent Diagnostic Tests in Patients Hospitalized With Community-Acquired Pneumonia. Open Forum Infect Dis. 2017;4(1):ofx014.

97. Ter SK, Rattanavong S, Roberts T, Sengduangphachanh A, Sihalath S, Panapruksachat S, et al. Molecular Detection of Pathogens in Negative Blood Cultures in the Lao People's Democratic Republic. Am J Trop Med Hyg. 2021;104(4):1582-5.

98. Nagy E, Boyanova L, Justesen US, Infections ESGoA. How to isolate, identify and determine antimicrobial susceptibility of anaerobic bacteria in routine laboratories. Clin Microbiol Infect. 2018;24(11):1139-48.

99. Jorgensen JH, Ferraro MJ. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. Clin Infect Dis. 2009;49(11):1749-55.

100. Benkova M, Soukup O, Marek J. Antimicrobial susceptibility testing: currently used methods and devices and the near future in clinical practice. J Appl Microbiol. 2020;129(4):806-22.

101. Balouiri M, Sadiki M, Ibnsouda SK. Methods for in vitro evaluating antimicrobial activity: A review. J Pharm Anal. 2016;6(2):71-9.

102. Llor C, Bjerrum L. Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. Ther Adv Drug Saf. 2014;5(6):229-41.

103.     Khan A, Arias CA, Abbott A, Dien Bard J, Bhatti MM, Humphries RM. Evaluation of the Vitek 2, Phoenix, and MicroScan for Antimicrobial Susceptibility Testing of Stenotrophomonas maltophilia. J Clin Microbiol. 2021;59(9):e0065421.

104.     Carroll KC, Patel R. Systems for Identification of Bacteria and Fungi.  Manual of Clinical Microbiology2015. p. 29-43.

105.     Nielsen LE, Clifford RJ, Kwak Y, Preston L, Argyros C, Rabinowitz R, et al. An 11,000-isolate same plate/same day comparison of the 3 most widely used platforms for analyzing multidrug-resistant clinical pathogens. Diagn Microbiol Infect Dis. 2015;83(2):93-8.

106.     Kulah C, Aktas E, Comert F, Ozlu N, Akyar I, Ankarali H. Detecting imipenem resistance in Acinetobacter baumannii by automated systems (BD Phoenix, Microscan WalkAway, Vitek 2); high error rates with Microscan WalkAway. BMC Infect Dis. 2009;9:30.

107.     Gerace E, Mancuso G, Midiri A, Poidomani S, Zummo S, Biondo C. Recent Advances in the Use of Molecular Methods for the Diagnosis of Bacterial Infections. Pathogens. 2022;11(6).

108.     Debnath M, Prasad, G. B. K. S., & Bisen, P. S. Molecular Microbiological Testing. Molecular Diagnostics: Promises and Possibilities. 2009:227-43.

109.     Aggarwal D, Kanitkar T, Narouz M, Azadian BS, Moore LSP, Mughal N. Clinical utility and cost-effectiveness of bacterial 16S rRNA and targeted PCR based diagnostic testing in a UK microbiology laboratory network. Sci Rep. 2020;10(1):7965.

110.     Schmitz JE, Stratton CW, Persing DH, Tang YW. Forty Years of Molecular Diagnostics for Infectious Diseases. J Clin Microbiol. 2022;60(10):e0244621.

111.     Valones MA, Guimaraes RL, Brandao LA, de Souza PR, de Albuquerque Tavares Carvalho A, Crovela S. Principles and applications of polymerase chain reaction in medical diagnostic fields: a review. Braz J Microbiol. 2009;40(1):1-11.

112.     Lleo MM, Ghidini V, Tafi MC, Castellani F, Trento I, Boaretti M. Detecting the presence of bacterial DNA by PCR can be useful in diagnosing culture-negative cases of infection, especially in patients with suspected infection and antibiotic therapy. FEMS Microbiol Lett. 2014;354(2):153-60.

113.     Miringu G, Musyoki A, Muriithi B, Wandera E, Waithiru D, Odoyo E, et al. Development of two multiplex PCR assays for rapid detection of eleven Gram-negative bacteria in children with septicemia. Trop Med Health. 2024;52(1):40.

114.     Lapa SA, Klochikhina ES, Miftakhov RA, Zolotov AM, Zasedatelev AS, Chudinov AV. Multiplex PCR for Identification of Bacterial Pathogens of Infectious Pneumonia. Russ J Bioorg Chem. 2020;46(5):859-61.

115.     Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology (N Y). 1993;11(9):1026-30.

116.     Kralik P, Ricchi M. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. Front Microbiol. 2017;8:108.

117.     Hauner A, Onwuchekwa C, Arien KK. Sample-to-result molecular diagnostic platforms and their suitability for infectious disease testing in low- and middle-income countries. Expert Rev Mol Diagn. 2024;24(5):423-38.

118.     Hlousek L, Voronov S, Diankov V, Leblang AB, Wells PJ, Ford DM, et al. Automated high multiplex qPCR platform for simultaneous detection and quantification of multiple nucleic acid targets. Biotechniques. 2012;52(5):316-24.

119.     Gunson RN, Bennett S, Maclean A, Carman WF. Using multiplex real time PCR in order to streamline a routine diagnostic service. J Clin Virol. 2008;43(4):372-5.

120.     Wolff N, Geiss AF, Barisic I. Crosslinking of PCR primers reduces unspecific amplification products in multiplex PCR. J Microbiol Methods. 2020;178:106051.

121.     Poritz MA, Blaschke AJ, Byington CL, Allen L, Nilsson K, Jones DE, et al. FilmArray, an automated nested multiplex PCR system for multi-pathogen detection: development and application to respiratory tract infection. PLoS One. 2011;6(10):e26047.

122.     Peker N, Couto N, Sinha B, Rossen JW. Diagnosis of bloodstream infections from positive blood cultures and directly from blood samples: recent developments in molecular approaches. Clin Microbiol Infect. 2018;24(9):944-55.

123.     Bhatti MM, Boonlayangoor S, Beavis KG, Tesic V. Evaluation of FilmArray and Verigene systems for rapid identification of positive blood cultures. J Clin Microbiol. 2014;52(9):3433-6.

124.     Garaizar J, Rementeria A, Porwollik S. DNA microarray technology: a new tool for the epidemiological typing of bacterial pathogens? FEMS Immunol Med Microbiol. 2006;47(2):178-89.

125.     Donatin E, Drancourt M. DNA microarrays for the diagnosis of infectious diseases. Med Mal Infect. 2012;42(10):453-9.

126.     Miller MB, Tang YW. Basic concepts of microarrays and potential applications in clinical microbiology. Clin Microbiol Rev. 2009;22(4):611-33.

127.     Asmare Z, Erkihun M. Recent Application of DNA Microarray Techniques to Diagnose Infectious Disease. Pathol Lab Med Int. 2023;15:77-82.

128.     Hou Y, Zhang X, Hou X, Wu R, Wang Y, He X, et al. Rapid pathogen identification using a novel microarray-based assay with purulent meningitis in cerebrospinal fluid. Sci Rep. 2018;8(1):15965.

129.     Jarvinen AK, Laakso S, Piiparinen P, Aittakorpi A, Lindfors M, Huopaniemi L, et al. Rapid identification of bacterial pathogens using a PCR- and microarray-based assay. BMC Microbiol. 2009;9:161.

130.     Spiess B, Seifarth W, Hummel M, Frank O, Fabarius A, Zheng C, et al. DNA microarray-based detection and identification of fungal pathogens in clinical samples from neutropenic patients. J Clin Microbiol. 2007;45(11):3743-53.

131.     Uva P, de Rinaldis E. CrossHybDetector: detection of cross-hybridization events in DNA microarray experiments. BMC Bioinformatics. 2008;9:485.

132.     Kim JS, Kang GE, Kim HS, Kim HS, Song W, Lee KM. Evaluation of Verigene Blood Culture Test Systems for Rapid Identification of Positive Blood Cultures. Biomed Res Int. 2016;2016:1081536.

133.     Lebovitz EE, Burbelo PD. Commercial multiplex technologies for the microbiological diagnosis of sepsis. Mol Diagn Ther. 2013;17(4):221-31.

134.     Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333-51.

135.     Hebert PDN, Cywinska A, Ball SL, DeWaard JR. Biological identifications through DNA barcodes. P Roy Soc B-Biol Sci. 2003;270(1512):313-21.

136.     Hoang MTV, Irinyi L, Hu YH, Schwessinger B, Meyer W. Long-Reads-Based Metagenomics in Clinical Diagnosis With a Special Focus on Fungal Infections. Frontiers in Microbiology. 2022;12.

137.     Kress WJ, Erickson DL. DNA barcodes: Genes, genomics, and bioinformatics. P Natl Acad Sci USA. 2008;105(8):2761-2.

138.     Antil S, Abraham JS, Sripoorna S, Maurya S, Dagar J, Makhija S, et al. DNA barcoding, an effective tool for species identification: a review. Mol Biol Rep. 2023;50(1):761-75.

139.     Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for. P Natl Acad Sci USA. 2012;109(16):6241-6.

140.     Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. J Clin Microbiol. 2007;45(9):2761-4.

141.     Buetas E, Jordan-Lopez M, Lopez-Roldan A, D'Auria G, Martinez-Priego L, De Marco G, et al. Full-length 16S rRNA gene sequencing by PacBio improves taxonomic resolution in human microbiome samples. BMC Genomics. 2024;25(1):310.

142.     Church DL, Cerutti L, Gurtler A, Griener T, Zelazny A, Emler S. Performance and Application of 16S rRNA Gene Cycle Sequencing for Routine Identification of Bacteria in the Clinical Microbiology Laboratory. Clin Microbiol Rev. 2020;33(4).

143.     Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, Garcia-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. J Biotechnol. 2017;243:16-24.

144. Winand R, Bogaerts B, Hoffman S, Lefevre L, Delvoye M, Braekel JV, et al. Targeting the 16s Rrna Gene for Bacterial Identification in Complex Mixed Samples: Comparative Evaluation of Second (Illumina) and Third (Oxford Nanopore Technologies) Generation Sequencing Technologies. Int J Mol Sci. 2019;21(1).

145. Flurin L, Hemenway JJ, Fisher CR, Vaillant JJ, Azad M, Wolf MJ, et al. Clinical Use of a 16S Ribosomal RNA Gene-Based Sanger and/or Next Generation Sequencing Assay to Test Preoperative Synovial Fluid for Periprosthetic Joint Infection Diagnosis. mBio. 2022;13(6):e0132222.

146. Heikema AP, Horst-Kreft D, Boers SA, Jansen R, Hiltemann SD, de Koning W, et al. Comparison of Illumina versus Nanopore 16S rRNA Gene Sequencing of the Human Nasal Microbiota. Genes (Basel). 2020;11(9).

147. Chakravorty S, Helb D, Burday M, Connell N, Alland D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods. 2007;69(2):330-9.

148. Larin AK, Klimina KM, Veselovsky VA, Olekhnovich EI, Morozov MD, Boldyreva DI, et al. An improved and extended dual-index multiplexed 16S rRNA sequencing for the Illumina HiSeq and MiSeq platform. BMC Genom Data. 2024;25(1):8.

149. Adewale BA. Will long-read sequencing technologies replace short-read sequencing technologies in the next 10 years? Afr J Lab Med. 2020;9(1):1340.

150. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021;39(11):1348-65.

151. Weirather JL, de Cesare M, Wang Y, Piazza P, Sebastiano V, Wang XJ, et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. F1000Res. 2017;6:100.

152. Zhang H, Jain C, Aluru S. A comprehensive evaluation of long read error correction methods. BMC Genomics. 2020;21(Suppl 6):889.

153. Ma X, Shao Y, Tian L, Flasch DA, Mulder HL, Edmonson MN, et al. Analysis of error profiles in deep next-generation sequencing data. Genome Biol. 2019;20(1):50.

154. Cheng C, Xiao P. Evaluation of the correctable decoding sequencing as a new powerful strategy for DNA sequencing. Life Sci Alliance. 2022;5(8).

155. Szoboszlay M, Schramm L, Pinzauti D, Scerri J, Sandionigi A, Biazzo M. Nanopore Is Preferable over Illumina for 16S Amplicon Sequencing of the Gut Microbiota When Species-Level Taxonomic Classification, Accurate Estimation of Richness, or Focus on Rare Taxa Is Required. Microorganisms. 2023;11(3).

156. Stevens BM, Creed TB, Reardon CL, Manter DK. Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. Sci Rep. 2023;13(1):9323.

157. Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, et al. Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. mSphere. 2021;6(1).

158. Hongoh Y, Yuzawa H, Ohkuma M, Kudo T. Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. FEMS Microbiol Lett. 2003;221(2):299-304.

159. Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. Microb Inform Exp. 2012;2(1):3.

160. Batool M, Galloway-Pena J. Clinical metagenomics-challenges and future prospects. Front Microbiol. 2023;14:1186424.

161. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G, Professional Practice C, et al. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. Arch Pathol Lab Med. 2017;141(6):776-86.

162. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med. 2014;370(25):2408-17.

163. Naccache SN, Peggs KS, Mattes FM, Phadke R, Garson JA, Grant P, et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. Clin Infect Dis. 2015;60(6):919-23.

164.     Mongkolrattanothai K, Naccache SN, Bender JM, Samayoa E, Pham E, Yu G, et al. Neurobrucellosis: Unexpected Answer From Metagenomic Next-Generation Sequencing. J Pediatric Infect Dis Soc. 2017;6(4):393-8.

165.     Salzberg SL, Breitwieser FP, Kumar A, Hao H, Burger P, Rodriguez FJ, et al. Next-generation sequencing in neuropathologic diagnosis of infections of the nervous system. Neurol Neuroimmunol Neuroinflamm. 2016;3(4):e251.

166.     Qian L, Shi Y, Li F, Wang Y, Ma M, Zhang Y, et al. Metagenomic Next-Generation Sequencing of Cerebrospinal Fluid for the Diagnosis of External Ventricular and Lumbar Drainage-Associated Ventriculitis and Meningitis. Front Microbiol. 2020;11:596175.

167.     Blauwkamp TA, Thair S, Rosen MJ, Blair L, Lindner MS, Vilfan ID, et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. Nat Microbiol. 2019;4(4):663-74.

168.     Hong DK, Blauwkamp TA, Kertesz M, Bercovici S, Truong C, Banaei N. Liquid biopsy for infectious diseases: sequencing of cell-free plasma to detect pathogen DNA in patients with invasive fungal disease. Diagn Microbiol Infect Dis. 2018;92(3):210-3.

169.     Zhang H, Liang R, Zhu Y, Hu L, Xia H, Li J, et al. Metagenomic next-generation sequencing of plasma cell-free DNA improves the early diagnosis of suspected infections. BMC Infect Dis. 2024;24(1):187.

170.     Kalantar KL, Neyton L, Abdelghany M, Mick E, Jauregui A, Caldera S, et al. Integrated host-microbe plasma metagenomics for sepsis diagnosis in a prospective cohort of critically ill adults. Nat Microbiol. 2022;7(11):1805-16.

171.     Qin C, Zhang S, Zhao Y, Ding X, Yang F, Zhao Y. Diagnostic value of metagenomic next-generation sequencing in sepsis and bloodstream infection. Front Cell Infect Microbiol. 2023;13:1117987.

172.     Langelier C, Zinter MS, Kalantar K, Yanik GA, Christenson S, O'Donovan B, et al. Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic Cellular Transplant Patients. Am J Respir Crit Care Med. 2018;197(4):524-8.

173.     Huang J, Jiang E, Yang D, Wei J, Zhao M, Feng J, et al. Metagenomic Next-Generation Sequencing versus Traditional Pathogen Detection in the Diagnosis of Peripheral Pulmonary Infectious Lesions. Infect Drug Resist. 2020;13:567-76.

174.     Langelier C, Kalantar KL, Moazed F, Wilson MR, Crawford ED, Deiss T, et al. Integrating host response and unbiased microbe detection for lower respiratory tract infection diagnosis in critically ill adults. Proc Natl Acad Sci U S A. 2018;115(52):E12353-E62.

175.     Wang H, Lu Z, Bao Y, Yang Y, de Groot R, Dai W, et al. Clinical diagnostic application of metagenomic next-generation sequencing in children with severe nonresponding pneumonia. PLoS One. 2020;15(6):e0232610.

176.     Ivy MI, Thoendel MJ, Jeraldo PR, Greenwood-Quaintance KE, Hanssen AD, Abdel MP, et al. Direct Detection and Identification of Prosthetic Joint Infection Pathogens in Synovial Fluid by Metagenomic Shotgun Sequencing. J Clin Microbiol. 2018;56(9).

177.     Asmerom B, Drobish I, Winckler B, Chiang L, Farnaes L, Beauchamp-Walters J, et al. Detection of Neisseria gonorrhoeae from Joint Aspirate by Metagenomic Sequencing in Disseminated Gonococcal Infection. J Pediatric Infect Dis Soc. 2021;10(3):367-9.

178.     Thoendel M, Jeraldo P, Greenwood-Quaintance KE, Chia N, Abdel MP, Steckelberg JM, et al. A Novel Prosthetic Joint Infection Pathogen, Mycoplasma salivarium, Identified by Metagenomic Shotgun Sequencing. Clin Infect Dis. 2017;65(2):332-5.

179.     Huang Y, Ma Y, Miao Q, Pan J, Hu B, Gong Y, et al. Arthritis caused by Legionella micdadei and Staphylococcus aureus: metagenomic next-generation sequencing provides a rapid and accurate access to diagnosis and surveillance. Ann Transl Med. 2019;7(20):589.

180.     Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol. 2019;37(7):783-92.

181.    Horiba K, Torii Y, Aizawa Y, Yamaguchi M, Haruta K, Okumura T, et al. Performance of Nanopore and Illumina Metagenomic Sequencing for Pathogen Detection and Transcriptome Analysis in Infantile Central Nervous System Infections. Open Forum Infect Dis. 2022;9(10):ofac504.

182.    Liu Y, Xu Y, Xu X, Chen X, Chen H, Zhang J, et al. Metagenomic identification of pathogens and antimicrobial-resistant genes in bacterial positive blood cultures by nanopore sequencing. Front Cell Infect Microbiol. 2023;13:1283094.

183.    Deng Q, Cao Y, Wan X, Wang B, Sun A, Wang H, et al. Nanopore-based metagenomic sequencing for the rapid and precise detection of pathogens among immunocompromised cancer patients with suspected infections. Front Cell Infect Microbiol. 2022;12:943859.

184.    Jia X, Hu L, Wu M, Ling Y, Wang W, Lu H, et al. A streamlined clinical metagenomic sequencing protocol for rapid pathogen identification. Sci Rep. 2021;11(1):4405.

185.    Ali J, Johansen W, Ahmad R. Short turnaround time of seven to nine hours from sample collection until informed decision for sepsis treatment using nanopore sequencing. Sci Rep. 2024;14(1):6534.

186.    Gu W, Deng X, Lee M, Sucu YD, Arevalo S, Stryke D, et al. Rapid pathogen detection by metagenomic next-generation sequencing of infected body fluids. Nat Med. 2021;27(1):115-24.

187.    Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn LJ, et al. Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis. Front Microbiol. 2019;10:1277.

188.    Shi Y, Wang G, Lau HC, Yu J. Metagenomic Sequencing for Microbial DNA in Human Samples: Emerging Technological Advances. Int J Mol Sci. 2022;23(4).

189.    Rang FJ, Kloosterman WP, de Ridder J. From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biol. 2018;19(1):90.

190.    Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13:341.

191.    Hong M, Peng D, Fu A, Wang X, Zheng Y, Xia L, et al. The application of nanopore targeted sequencing in the diagnosis and antimicrobial treatment guidance of bloodstream infection of febrile neutropenia patients with hematologic disease. J Cell Mol Med. 2023;27(4):506-14.

192.    Han D, Yu F, Zhang D, Hu J, Zhang X, Xiang D, et al. Molecular rapid diagnostic testing for bloodstream infections: Nanopore targeted sequencing with pathogen-specific primers. J Infect. 2024;88(6):106166.

193.    Fu Y, Chen Q, Xiong M, Zhao J, Shen S, Chen L, et al. Clinical Performance of Nanopore Targeted Sequencing for Diagnosing Infectious Diseases. Microbiol Spectr. 2022;10(2):e0027022.

194.    Fu Y, Gu J, Chen LJ, Xiong M, Zhao J, Xiao X, et al. A prospective study of nanopore-targeted sequencing in the diagnosis of central nervous system infections. Microbiol Spectr. 2024;12(3):e0331723.

195.    Harris PNA, Bauer MJ, Luftinger L, Beisken S, Forde BM, Balch R, et al. Rapid nanopore sequencing and predictive susceptibility testing of positive blood cultures from intensive care patients with sepsis. Microbiol Spectr. 2024;12(2):e0306523.

196.    Taxt AM, Avershina E, Frye SA, Naseer U, Ahmad R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. Sci Rep. 2020;10(1):7622.

197.    Kanaujia R, Sharma V, Biswal M, Singh S, Ray P, Angrup A. Microbial cell-free DNA detection: Minimally invasive diagnosis of infectious diseases. Indian J Med Microbiol. 2023;46:100433.

198.    Zhao M, Zhang Y, Chen L, Yan X, Xu T, Fu M, et al. Nanopore sequencing of infectious fluid is a promising supplement for gold-standard culture in real-world clinical scenario. Front Cell Infect Microbiol. 2024;14:1330788.

199.    Neyton LPA, Langelier CR, Calfee CS. Metagenomic Sequencing in the ICU for Precision Diagnosis of Critical Infectious Illnesses. Crit Care. 2023;27(1):90.

200.    Govender KN, Eyre DW. Benchmarking taxonomic classifiers with Illumina and Nanopore sequence data for clinical metagenomic diagnostic applications. Microb Genom. 2022;8(10).

201.    Bazinet AL, Ondov BD, Sommer DD, Ratnayake S. BLAST-based validation of metagenomic sequence assignments. PeerJ. 2018;6:e4892.

202. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20(1):257.
203. Lu J, Rincon N, Wood DE, Breitwieser FP, Pockrandt C, Langmead B, et al. Metagenome analysis using the Kraken software suite. Nat Protoc. 2022;17(12):2815-39.
204. Rodriguez-Perez H, Ciuffreda L, Flores C. NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. Bioinformatics. 2021;37(11):1600-1.
205. Curry KD, Wang Q, Nute MG, Tyshaieva A, Reeves E, Soriano S, et al. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. Nat Methods. 2022;19(7):845-53.
206. Karger A. Current developments to use linear MALDI-TOF spectra for the identification and typing of bacteria and the characterization of other cells/organisms related to infectious diseases. Proteomics Clin Appl. 2016;10(9-10):982-93.
207. Hou TY, Chiang-Ni C, Teng SH. Current status of MALDI-TOF mass spectrometry in clinical microbiology. J Food Drug Anal. 2019;27(2):404-14.
208. Lau SK, Tang BS, Teng JL, Chan TM, Curreem SO, Fan RY, et al. Matrix-assisted laser desorption ionisation time-of-flight mass spectrometry for identification of clinically significant bacteria that are difficult to identify in clinical laboratories. J Clin Pathol. 2014;67(4):361-6.
209. Ge MC, Kuo AJ, Liu KL, Wen YH, Chia JH, Chang PY, et al. Routine identification of microorganisms by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: Success rate, economic analysis, and clinical outcome. J Microbiol Immunol Infect. 2017;50(5):662-8.
210. Garner O, Mochon A, Branda J, Burnham CA, Bythrow M, Ferraro M, et al. Multi-centre evaluation of mass spectrometric identification of anaerobic bacteria using the VITEK(R) MS system. Clin Microbiol Infect. 2014;20(4):335-9.
211. Knoester M, van Veen SQ, Claas EC, Kuijper EJ. Routine identification of clinical isolates of anaerobic bacteria: matrix-assisted laser desorption ionization-time of flight mass spectrometry performs better than conventional identification methods. J Clin Microbiol. 2012;50(4):1504.
212. Luo Y, Siu GKH, Yeung ASF, Chen JHK, Ho PL, Leung KW, et al. Performance of the VITEK MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system for rapid bacterial identification in two diagnostic centres in China. J Med Microbiol. 2015;64(Pt 1):18-24.
213. Bizzini A, Jaton K, Romo D, Bille J, Prod'hom G, Greub G. Matrix-assisted laser desorption ionization-time of flight mass spectrometry as an alternative to 16S rRNA gene sequencing for identification of difficult-to-identify bacterial strains. J Clin Microbiol. 2011;49(2):693-6.
214. Homem de Mello de Souza HAP, Dalla-Costa LM, Vicenzi FJ, Camargo de Souza D, Riedi CA, Filho NAR, et al. MALDI-TOF: a useful tool for laboratory identification of uncommon glucose non-fermenting Gram-negative bacteria associated with cystic fibrosis. J Med Microbiol. 2014;63(Pt 9):1148-53.
215. Ip CLC, Loose M, Tyson JR, de Cesare M, Brown BL, Jain M, et al. MinION Analysis and Reference Consortium: Phase 1 data release and analysis. F1000Res. 2015;4:1075.
216. Lao HY, Ng TT, Wong RY, Wong CS, Lee LK, Wong DS, et al. The Clinical Utility of Two High-Throughput 16S rRNA Gene Sequencing Workflows for Taxonomic Assignment of Unidentifiable Bacterial Pathogens in Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. J Clin Microbiol. 2022;60(1):e0176921.
217. Muyzer G, Teske A, Wirsen CO, Jannasch HW. Phylogenetic relationships of Thiomicrospira species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. Arch Microbiol. 1995;164(3):165-72.
218. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res. 2007;35(18):e120.
219. Andersson AF, Lindberg M, Jakobsson H, Backhed F, Nyren P, Engstrand L. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One. 2008;3(7):e2836.

220. Nossa CW, Oberdorf WE, Yang L, Aas JA, Paster BJ, Desantis TZ, et al. Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. World J Gastroenterol. 2010;16(33):4135-44.

221. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41.

222. Straub D, Blackwell N, Langarica-Fuentes A, Peltzer A, Nahnsen S, Kleindienst S. Interpretations of Environmental Microbial Community Studies Are Biased by the Selected 16S rRNA (Gene) Amplicon Sequencing Pipeline. Front Microbiol. 2020;11:550420.

223. Huang YT, Liu PY, Shih PW. Homopolish: a method for the removal of systematic errors in nanopore sequencing by homologous polishing. Genome Biol. 2021;22(1):95.

224. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110(10):1281-6.

225. Sierra MA, Li Q, Pushalkar S, Paul B, Sandoval TA, Kamer AR, et al. The Influences of Bioinformatics Tools and Reference Databases in Analyzing the Human Oral Microbial Community. Genes (Basel). 2020;11(8).

226. Park SC, Won S. Evaluation of 16S rRNA Databases for Taxonomic Assignments Using Mock Community. Genomics Inform. 2018;16(4):e24.

227. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun. 2019;10(1):5029.

228. Schloss PD, Handelsman J. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Environ Microbiol. 2005;71(3):1501-6.

229. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i90.

230. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Meric G, Vezina B, et al. Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biol. 2021;22(1):266.

231. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540-6.

232. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 2016;32(14):2103-10.

233. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. F1000Res. 2019;8:2138.

234. Vaser R, Sikic M. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci. 2021;1(5):332-6.

235. Wick RR, Holt KE. Polypolish: Short-read polishing of long-read bacterial genome assemblies. PLoS Comput Biol. 2022;18(1):e1009802.

236. Zimin AV, Salzberg SL. The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies. PLoS Comput Biol. 2020;16(6):e1007981.

237. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017;13(6):e1005595.

238. Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol Biol Evol. 2021;38(10):4647-54.

239. Pruesse E, Peplies J, Glockner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28(14):1823-9.

240. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.

241. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2019;36(6):1925-7.

242.	Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol. 2018;36(7):566-9.

243.	Lee I, Ouk Kim Y, Park SC, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol. 2016;66(2):1100-3.

244.	Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.

245.	Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293-W6.

246.	Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022;38(23):5315-6.

247.	Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Research. 2021;50(D1):D785-D94.

248.	Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, et al. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci Rep. 2015;5:8365.

249.	Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75.

250.	Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44(14):6614-24.

251.	Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. 2016;428(4):726-31.

252.	Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen CY, et al. Proksee: in-depth characterization and visualization of bacterial genomes. Nucleic Acids Res. 2023;51(W1):W484-W92.

253.	Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020;75(12):3491-500.

254.	Gschwind R, Ugarcina Perovic S, Weiss M, Petitjean M, Lao J, Coelho LP, et al. ResFinderFG v2.0: a database of antibiotic resistance genes obtained by functional metagenomics. Nucleic Acids Res. 2023;51(W1):W493-W500.

255.	Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep. 2021;11(1):12728.

256.	Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol. 2013;79(24):7696-701.

257.	Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37(5):1530-4.

258.	Wambui J, Cernela N, Stevens MJA, Stephan R. Draft Genome Sequence of Clostridium estertheticum CEST001, Belonging to a Novel Subspecies of C. estertheticum, Isolated from Chilled Vacuum-Packed Lamb Meat Imported to Switzerland. Microbiol Resour Announc. 2020;9(33).

259.	Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Goker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res. 2022;50(D1):D801-D7.

260.	Qin QL, Xie BB, Zhang XY, Chen XL, Zhou BC, Zhou J, et al. A proposed genus boundary for the prokaryotes based on genomic insights. J Bacteriol. 2014;196(12):2210-5.

261.	Holzer M. POCP-nf: an automatic Nextflow pipeline for calculating the percentage of conserved proteins in bacterial taxonomy. Bioinformatics. 2024;40(4).

262.	Huang N, Li H. compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics. 2023;39(10).

263.	Pearce ME, Langridge GC, Lauer AC, Grant K, Maiden MCJ, Chattaway MA. An evaluation of the species and subspecies of the genus Salmonella with whole genome sequence data: Proposal of type

strains and epithets for novel S. enterica subspecies VII, VIII, IX, X and XI. Genomics. 2021;113(5):3152-62.

264.    Lao H-Y, Wong AYP, Ng TT-L, Wong RY-L, Yau MC-Y, Lam JY-W, et al. Scrofimicrobium appendicitidis sp. nov., isolated from a patient with ruptured appendicitis. International Journal of Systematic and Evolutionary Microbiology. 2025;75(1).

265.    Wylensek D, Hitch TCA, Riedel T, Afrizal A, Kumar N, Wortmann E, et al. A collection of bacterial isolates from the pig intestine reveals functional and taxonomic diversity. Nat Commun. 2020;11(1):6389.

266.    Christner M, Rohde H, Wolters M, Sobottka I, Wegscheider K, Aepfelbacher M. Rapid identification of bacteria from positive blood culture bottles by use of matrix-assisted laser desorption-ionization time of flight mass spectrometry fingerprinting. J Clin Microbiol. 2010;48(5):1584-91.

267.    Gray TJ, Thomas L, Olma T, Iredell JR, Chen SC. Rapid identification of Gram-negative organisms from blood culture bottles using a modified extraction method and MALDI-TOF mass spectrometry. Diagn Microbiol Infect Dis. 2013;77(2):110-2.

268.    Jamal W, Saleem R, Rotimi VO. Rapid identification of pathogens directly from blood culture bottles by Bruker matrix-assisted laser desorption laser ionization-time of flight mass spectrometry versus routine methods. Diagn Microbiol Infect Dis. 2013;76(4):404-8.

269.    Leli C, Cenci E, Cardaccia A, Moretti A, D'Alo F, Pagliochini R, et al. Rapid identification of bacterial and fungal pathogens from positive blood cultures by MALDI-TOF MS. Int J Med Microbiol. 2013;303(4):205-9.

270.    Lagace-Wiens PR, Adam HJ, Karlowsky JA, Nichol KA, Pang PF, Guenther J, et al. Identification of blood culture isolates directly from positive blood cultures by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry and a commercial extraction system: analysis of performance, cost, and turnaround time. J Clin Microbiol. 2012;50(10):3324-8.

271.    Chien JY, Lee TF, Du SH, Teng SH, Liao CH, Sheng WH, et al. Applicability of an in-House Saponin-Based Extraction Method in Bruker Biotyper Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry System for Identification of Bacterial and Fungal Species in Positively Flagged Blood Cultures. Front Microbiol. 2016;7:1432.

272.    Florio W, Cappellini S, Giordano C, Vecchione A, Ghelardi E, Lupetti A. A new culture-based method for rapid identification of microorganisms in polymicrobial blood cultures by MALDI-TOF MS. BMC Microbiol. 2019;19(1):267.

273.    Lao HY, Wong LL, Hui Y, Ng TT, Chan CT, Lo HW, et al. The clinical utility of Nanopore 16S rRNA gene sequencing for direct bacterial identification in normally sterile body fluids. Front Microbiol. 2023;14:1324494.

274.    Beye M, Fahsi N, Raoult D, Fournier PE. Careful use of 16S rRNA gene sequence similarity values for the identification of Mycobacterium species. New Microbes New Infect. 2018;22:24-9.

275.    Kawamura Y, Hou XG, Sultana F, Miura H, Ezaki T. Determination of 16S rRNA sequences of Streptococcus mitis and Streptococcus gordonii and phylogenetic relationships among members of the genus Streptococcus. Int J Syst Bacteriol. 1995;45(2):406-8.

276.    Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Inbanathan FY, Veeraraghavan B. Accurate differentiation of Escherichia coli and Shigella serogroups: challenges and strategies. New Microbes New Infect. 2018;21:58-62.

277.    Yap JM, Goldsmith CE, Moore JE. Integrity of bacterial genomic DNA after autoclaving: possible implications for horizontal gene transfer and clinical waste management. J Hosp Infect. 2013;83(3):247-9.

278.    Suyama T, Kawaharasaki M. Decomposition of waste DNA with extended autoclaving under unsaturated steam. Biotechniques. 2013;55(6):296-9.

279.    Li X, Bosch-Tijhof CJ, Wei X, de Soet JJ, Crielaard W, Loveren CV, et al. Efficiency of chemical versus mechanical disruption methods of DNA extraction for the identification of oral Gram-positive and Gram-negative bacteria. J Int Med Res. 2020;48(5):300060520925594.

280.    Lu J, Breitwieser FP, Thielen P, Salzberg SL. Bracken: estimating species abundance in metagenomics data. Peerj Comput Sci. 2017.

281. Xu R, Rajeev S, Salvador LCM. The selection of software and database for metagenomics sequence analysis impacts the outcome of microbial profiling and pathogen detection. PLoS One. 2023;18(4):e0284031.

282. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking Metagenomics Tools for Taxonomic Classification. Cell. 2019;178(4):779-94.

283. Frank JA, Reich CI, Sharma S, Weisbaum JS, Wilson BA, Olsen GJ. Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Appl Environ Microbiol. 2008;74(8):2461-70.

284. Op De Beeck M, Lievens B, Busschaert P, Declerck S, Vangronsveld J, Colpaert JV. Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies. PLoS One. 2014;9(6):e97629.

285. Usyk M, Zolnik CP, Patel H, Levi MH, Burk RD. Novel ITS1 Fungal Primers for Characterization of the Mycobiome. mSphere. 2017;2(6).

286. de Melo DA, Coelho Ida S, da Motta CC, Rojas AC, Dubenczuk FC, Coelho Sde M, et al. Impairments of mecA gene detection in bovine Staphylococcus spp. Braz J Microbiol. 2014;45(3):1075-82.

287. Klare I, Heier H, Claus H, Reissbrodt R, Witte W. Vana-Mediated High-Level Glycopeptide Resistance in Enterococcus-Faecium from Animal Husbandry. Fems Microbiology Letters. 1995;125(2-3):165-71.

288. Bustamante W, Alpizar A, Hernandez S, Pacheco A, Vargas N, Herrera ML, et al. Predominance of vanA genotype among vancomycin-resistant Enterococcus isolates from poultry and swine in Costa Rica. Appl Environ Microbiol. 2003;69(12):7414-9.

289. Lartigue MF, Zinsius C, Wenger A, Bille J, Poirel L, Nordmann P. Extended-spectrum beta-lactamases of the CTX-M type now in Switzerland. Antimicrob Agents Chemother. 2007;51(8):2855-60.

290. Dallenne C, Da Costa A, Decré D, Favier C, Arlet G. Development of a set of multiplex PCR assays for the detection of genes encoding important β-lactamases in Enterobacteriaceae. J Antimicrob Chemoth. 2010;65(3):490-5.

291. Ibrahim ME, Algak TB, Abbas M, Elamin BK. Emergence of bla (TEM), bla (CTX-M), bla (SHV) and bla (OXA) genes in multidrug-resistant Enterobacteriaceae and Acinetobacter baumannii in Saudi Arabia. Exp Ther Med. 2021;22(6):1450.

292. Doyle D, Peirano G, Lascols C, Lloyd T, Church DL, Pitout JD. Laboratory detection of Enterobacteriaceae that produce carbapenemases. J Clin Microbiol. 2012;50(12):3877-80.

293. Mushi MF, Mshana SE, Imirzalioglu C, Bwanga F. Carbapenemase genes among multidrug resistant gram negative clinical isolates from a tertiary hospital in Mwanza, Tanzania. Biomed Res Int. 2014;2014:303104.

294. Perez-Perez FJ, Hanson ND. Detection of plasmid-mediated AmpC beta-lactamase genes in clinical isolates by using multiplex PCR. J Clin Microbiol. 2002;40(6):2153-62.

295. Geyer CN, Hanson ND. Multiplex high-resolution melting analysis as a diagnostic tool for detection of plasmid-mediated AmpC beta-lactamase genes. J Clin Microbiol. 2014;52(4):1262-5.

296. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018;34(15):2666-9.

297. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

298. Calderon-Franco D, Sarelse R, Christou S, Pronk M, van Loosdrecht MCM, Abeel T, et al. Metagenomic profiling and transfer dynamics of antibiotic resistance determinants in a full-scale granular sludge wastewater treatment plant. Water Res. 2022;219:118571.

299. Nix ID, Idelevich EA, Storck LM, Sparbier K, Drews O, Kostrzewa M, et al. Detection of Methicillin Resistance in Staphylococcus aureus From Agar Cultures and Directly From Positive Blood Cultures Using MALDI-TOF Mass Spectrometry-Based Direct-on-Target Microdroplet Growth Assay. Front Microbiol. 2020;11:232.

300. Guinea J, Recio S, Escribano P, Torres-Narbona M, Pelaez T, Sanchez-Carrillo C, et al. Rapid antifungal susceptibility determination for yeast isolates by use of Etest performed directly on blood samples from patients with fungemia. J Clin Microbiol. 2010;48(6):2205-12.

301. Brown L, Wolf JM, Prados-Rosales R, Casadevall A. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. Nature Reviews Microbiology. 2015;13(10):620-30.

302. Bruggeling CE, Garza DR, Achouiti S, Mes W, Dutilh BE, Boleij A. Optimized bacterial DNA isolation method for microbiome analysis of human tissues. Microbiologyopen. 2021;10(3):e1191.

303. Ganda E, Beck KL, Haiminen N, Silverman JD, Kawas B, Cronk BD, et al. DNA Extraction and Host Depletion Methods Significantly Impact and Potentially Bias Bacterial Detection in a Biological Fluid. mSystems. 2021;6(3):e0061921.

304. Ebmeyer S, Kristiansson E, Larsson DGJ. CMY-1/MOX-family AmpC β-lactamases MOX-1, MOX-2 and MOX-9 were mobilized independently from three

 species. J Antimicrob Chemoth. 2019;74(5):1202-6.

305. Longhi G, Argentini C, Fontana F, Tarracchini C, Mancabelli L, Lugli GA, et al. Saponin treatment for eukaryotic DNA depletion alters the microbial DNA profiles by reducing the abundance of Gram-negative bacteria in metagenomics analyses. Microbiome Res Rep. 2024;3(1):4.

306. Rogers GB, Cuthbertson L, Hoffman LR, Wing PAC, Pope C, Hooftman DAP, et al. Reducing bias in bacterial community analysis of lower respiratory infections. Isme J. 2013;7(4):697-706.

307. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome. 2018;6(1):42.

308. Marquet M, Zollkau J, Pastuschek J, Viehweger A, Schleussner E, Makarewicz O, et al. Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. Sci Rep. 2022;12(1):4000.

309. Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, Leggett RM. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. Genome Biol. 2022;23(1):11.

310. Ulrich JU, Epping L, Pilz T, Walther B, Stingl K, Semmler T, et al. Nanopore adaptive sampling effectively enriches bacterial plasmids. mSystems. 2024;9(3):e0094523.

311. Wright RJ, Comeau AM, Langille MGI. From defaults to databases: parameter and database choice dramatically impact the performance of metagenomic taxonomic classification tools. Microb Genom. 2023;9(3).

312. Cheng J, Hu H, Kang Y, Chen W, Fang W, Wang K, et al. Identification of pathogens in culture-negative infective endocarditis cases by metagenomic analysis. Ann Clin Microbiol Antimicrob. 2018;17(1):43.

313. Urmi UL, Nahar S, Rana M, Sultana F, Jahan N, Hossain B, et al. Genotypic to Phenotypic Resistance Discrepancies Identified Involving beta-Lactamase Genes, blaKPC, blaIMP, blaNDM-1, and blaVIM in Uropathogenic Klebsiella pneumoniae. Infect Drug Resist. 2020;13:2863-75.

314. Kong M, Liu C, Xu Y, Wang J, Jin D. Concordance between Genotypic and Phenotypic Drug-Resistant Profiles of Shigella Isolates from Taiyuan City, Shanxi Province, China, 2005 to 2016. Microbiol Spectr. 2023;11(3):e0011923.

315. Rasheed H, Ijaz M, Ahmed A, Javed MU, Shah SFA, Anwaar F. Discrepancies between phenotypic and genotypic identification methods of antibiotic resistant genes harboring Staphylococcus aureus. Microb Pathog. 2023;184:106342.

316. Ayesha BB, Gachinmath S, Sobia C. Isolation of obligate anaerobes from clinical samples received for routine bacterial culture and sensitivity: a cross sectional study. Iran J Microbiol. 2022;14(2):145-55.

317. Davis MA, Besser TE, Orfe LH, Baker KN, Lanier AS, Broschat SL, et al. Genotypic-phenotypic discrepancies between antibiotic resistance characteristics of Escherichia coli isolates from calves in management settings with high and low antibiotic use. Appl Environ Microbiol. 2011;77(10):3293-9.