Whole Genome Sequencing and Comparative Genomics of Extended-Spectrum Beta-Lactamase-Producing *Klebsiella Pneumoniae*

Dr Lim May Chin

Dissertation Submitted in Partial Fulfilment of The Requirements for The Degree of Master of Medicine (Anaesthesiology and Critical Care)

UNIVERSITI SAINS MALAYSIA

2022

ACKNOWLEDGEMENTS

Immeasurable appreciation and profound gratitude for the professional guidance and support extended to my thesis supervisors - Dr Kamaruddin Ibrahim and Dr Nik Yusnoraini, the knowledgeable medical laboratory technologist - Ms Mawaddah Binti Mohd Azlan and the dedicated laboratory assistant - Mdm Haslizai Binti Hassan Noor, without whom, I would not have been able to complete this dissertation.

A debt of gratitude to my academic supervisor, Dr Ariffin Marzuki Mohktar, for his wisdom, advice, and continuous encouragement for the study. And heartiest thanks to my out-campus academic supervisors, Dr Maseeda Mohamed Yusuf and Dr Khoo Teik Hooi, for their exemplary tutorage during the first two years in Hospital Pulau Pinang. Thank you for all your selfless help and sea-vast patience in guiding me in my academic journey towards my Master's degree, which is well reflected in this dissertation.

This study would have been challenging to execute without my colleagues' help, whose effort and time spent could never be repaid. I am forever indebted to all of you.

To my dearest parents, Mdm Tan Siew Eng and Mr Lim Yong Teh, to my kindest of comrades, Dr Ng Li En, Dr Tang Yi Chuan, Dr Ung Xiu Ming and Dr Wee Siok Choon, I would like to thank all of you for your unwavering support, encouragement, sacrifice, and endless love for me throughout my studies. I dedicate this work to you.

TABLE OF CONTENTS

LIST OF ABBREVIATION

AMR	Antibiotic resistance or Antimicrobial resistance
ANI	Average nucleotide identity
ARO	Antibiotic Resistance Ontology
BLASR	Basic Local Alignment with Successive Refinement
BLAST	Basic Local Alignment Search Tool
bp	Base pair
CPS	Capsular polysaccharide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphates
gANI	Genome-wide average nucleotide identity
GC content	Guanine-cytosine content
INFORMM	Institute for Research in Molecular Medicine
ICU	Intensive Care Unit
iss	Increased serum survival
E. coli	Escherichia coli
ESBL	Extended-spectrum beta-lactamase
ESBL-KP	Extended-spectrum beta-lactamase *K. pneumoniae*
ENT	Ear, Nose & Throat
FM	Ferragina-Manzini
HGAP	Hierarchical genome assembly process

HUSM	Hospital Universiti Sains Malaysia
K. pneumoniae	*Klebsiella pneumoniae*
LB broth	Luria-Bertani broth or lysogeny broth
lpfA	Long polar fimbriae
LPS	Lipopolysaccharide
MDR	Multidrug-resistant
MGE	Mobile genetic elements
MLST	Multilocus sequence typing
MR	Methyl Red
NA	Nutrient Agar
NCBI	National Center for Biotechnology Information
NSAR	National Surveillance of Antibiotic Resistance
P. aeruginosa	Pseudomonas aeruginosa
PCR	Polymerase chain reaction
PHASTER	PHAge Search Tool
QUAST	Quality assessment tool for genome assemblies
RAST	Rapid Annotation using Subsystem Technology
RGI	Resistance Gene Identifier
RNase	Ribonuclease
RND	Resistance-nodulation-cell division
Roary	Rapid large-scale prokaryote pan-genome analysis
rRNA	Ribosomal ribonucleic acid
RT-PCR	Real-time polymerase chain reaction

SMRT	Single-molecule real-time sequencing
ST	Sequence type
T6SS	Type VI secretion system
tRNA	Transfer ribonucleic acid
TSI	Triple Iron Sugar
UV	Ultraviolet
VF(s)	Virulence factor(s)
VFDB	Virulence Factor Database
WGS	Whole Genome Sequencing

ABSTRAK

Tajuk: Penjujukan genom keseluruhan dan perbandingan genomik bakteria *Klebsiella pneumoniae* yang menghasilkan beta-lactamase-spektrum-lanjutan

Latar belakang: ESBL *Klebsiella pneumoniae* (ESBL-KP) telah muncul sebagai punca penting jangkitan kuman dalam hospital dengan akibat yang teruk. Kajian ini menjalankan pencirian molekul gen rintangan antibiotik, faktor virulensi dan *multilocus sequence typing (MLST)* ke atas lima sampel ESBL-KP daripada Hospital Universiti Sains Malaysia (HUSM) dengan menggunakan *whole genome sequencing (WGS).*

Metodologi: Lima pengasingan bakteria tulen yang diperolehi dari INFORMM BioBank telah dikulturkan, sebelum pengekstrakan DNA genomik boleh dilakukan. Kuantifikasi DNA genomik yang telah diekstrak, ditentukan pada spektrofotometer NanoDrop dan juga pengesahan dengan elektroforesis gel agarose. WGS dilakukan dengan menggunakan perkhidmatan jujukan *MyGS Microbial Whole Genome Nanopore Research*. Bacaan mentah yang dihasilkan, ditapis mengikuti kepanjangan dan kualiti, dan *de novo assembly* dilakukan. Analisis perbandingan dicapai dengan menggunakan carian persamaan nukleotida berasaskan "*Basic Local Alignment Search Tool"* (BLAST) dengan rujukan yang didapati dari GenBank.

Keputusan: Gen-gen beta-laktamase yang dominan, dikesan dalam empat (80%) sampel bakteria yang diuji ialah $bla_{CTX\text{-}M\text{-}15}$ dan $bla_{TEM\text{-}1}$. Gen-gen beta-laktamase lain yang ditemui dalam sampel bakteria dalam kajian ini ialah $bla_{OXA\text{-}1}$ dan $bla_{SHV\text{-}187}$, sebanyak 40%; $bla_{OXA\text{-}9}$, $bla_{SHV\text{-}2}$, $bla_{SHV\text{-}11}$, $bla_{SHV\text{-}182}$, sebanyak 20%. Gen $bla_{SHV\text{-}2}$ ditemui dalam salah satu sampel tempatan yang tidak ditemui dalam sampel-sampel rujukan. Pelbagai gen-gen pam efluks yang ketara - *AcrA*, *AcrAB*, *mdtK*, *TolC*, dan gen-

gen *siderophores – entB, iutA, fur, ybtS* telah dikesan dalam kesemua sampel bakteria yang diuji. Sequence type (ST) yang didapati dari lima sampel bakteria adalah berbeza, ST29, ST147, ST656, ST3146 dan satu ST yang tidak diketahui.

Kesimpulan: Kajian kami menggunakan lima sampel bakteria ESBL-KP yang diperolehi dalam HUSM dan ini telah menambah data DNA *K. pneumoniae* yang bernilai. Ia telah menunjukkan kelebihan WGS dalam mencirikan rintangan antimikrob, dan perbandingan genomik dengan rujukan lain dalam GenBank. Ini dapat membantu pelan rawatan klinikal dalam persekitaran kritikal dan tidak kritical dalam HUSM.

Kata kunci: Penjujukan genom keseluruhan, genetik perbandingan, spektrum lanjutan beta-laktamase, *Klebsiella pneumoniae*, Malaysia

ABSTRACT

Background: Extended-spectrum beta-lactamase *Klebsiella pneumoniae* (ESBL-KP) has emerged as an important cause of hospital-acquired infections with severe consequences. This study undertook the molecular characterisation of antibiotic resistance (AMR) genes, virulence factors and multilocus sequence type of five clinical isolates of ESBL-KP from Hospital Universiti Sains Malaysia (HUSM) using whole genome sequencing (WGS).

Methods: Five pure bacterial isolates obtained from INFORMM BioBank were cultured, and genomic DNA extraction was done. The quantification of extracted DNA was determined on a NanoDrop spectrophotometer with verification by agarose gel electrophoresis. WGS was done using MyGS Microbial Whole genome Nanopore Research-based sequence service. The resulting raw reads were quality-filtered and length-filtered, and *de novo* assembly was done. Comparative analysis was accomplished using a BLAST-based nucleotide similarity search against references from GenBank.

Results: The predominant beta-lactamases were $bla_{CTX\text{-}M\text{-}15}$ and $bla_{TEM\text{-}1}$, detected in four local isolates (80%). Other beta-lactamases discovered were $bla_{OXA\text{-}1,}$ and $bla_{SHV\text{-}187}$ (40%); $bla_{SHV\text{-}2}$, $bla_{SHV\text{-}11}$, and $bla_{SHV\text{-}182}$ (20%). The $bla_{SHV\text{-}2}$ was discovered in one local isolate and not found in reference isolates. Multiple efflux-pump genes, *AcrA*, *AcrAB*, *mdtK*, *TolC*, siderophores – *entB*, *iutA*, *fur*, *ybtS* were detected in all local isolates tested. Four isolates were affiliated with different sequence types, ST29, ST147, ST656, ST3416, and one isolate showed an unknown ST.

Conclusion: Our study adds significant DNA sequence data on *K. pneumoniae* strains of local isolates and has demonstrated the value of WGS in characterising AMR and

genomic comparison with other references in GenBank. This helps guide health practitioners' treatment plans in both critical and non-critical settings within HUSM.

Keywords: Whole genome sequencing, comparative genetics, extended-spectrum beta-lactamase (ESBL), *Klebsiella pneumoniae*, Malaysia

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

The emergence of extended-spectrum β-lactamase-producing *Klebsiella pneumoniae* (ESBL-KP) represents a serious clinical concern in both healthcare and community settings (1–3). ESBL-producing *Enterobacteriaceae,* including *K. pneumoniae,* are listed as pathogens of critical priority for the research and development of antibiotics by the World Health Organization in 2017 (4). ESBL enzymes confer resistance to a large spectrum of β-lactam antibiotics, including penicillin and third-generation cephalosporins, thus limiting effective therapeutic options, increasing morbidity, mortality and hospital costs. In particular, to prevent selection of carbapenem-resistant *Enterobacteriaceae*, which is a severe problem in the world, it should be necessary to keep up surveillance and control of ESBL-producing gram-negative organisms (5).

In Malaysia, a recent study done in a public hospital in Johor Bahru on 97 *K. pneumoniae* strains isolated in clinical samples (sputum, endotracheal aspiration, blood, pleural cavity tap and urine) showed different resistance genes with a majority prevalence of bla_{TEM}, bla_{SHV}, and $bla_{CTX\text{-}M\text{-}1}$. (6). Similarly, several studies conducted in Peninsular Malaysia have shown that various TEM-, SHV-, and CTX-M types were commonly detected as clinical isolates of *K. pneumoniae* (7,8). Internationally, a study on the complexity and diversity of 25 *K. pneumoniae* strains isolated in clinical samples showed similar prevalence on TEM and SHV beta-lactamases. Additionally, various TEM and SHV types (SHV-19, 20, 21 and 22, and TEM-1, 53 and 63) were observed in *K. pneumoniae* isolates in KwaZulu-Natal and Western Cape region of South Africa (9–

11). Nosocomial outbreaks of ESBL-KP affecting both adult and paediatric populations were reported internationally, as well as in Malaysia by Palasubramaniam et al., 2005 (4,9,12). The propensity of *K. pneumoniae* to act as a nosocomial pathogen and the plasticity related to the acquisition of resistance genes stressed the need for further investigation in different Malaysian settings.

This study undertook the molecular characterisation of antibiotic resistance (AMR) genes, virulence factors (VFs) and evolutionary relationship of clinical isolates of ESBL-KP from patients in the Hospital Universiti Sains Malaysia (HUSM) using whole genome sequencing (WGS). Our study will add new and significant deoxyribonucleic acid (DNA) sequence data on *K. pneumoniae* strains of local isolates to demonstrate the value of WGS in characterising antibiotic resistance and genomic comparison with other references in GenBank.

1.2 PROBLEM STATEMENT AND RESEARCH QUESTIONS

ESBL-KP remains a critical clinical concern worldwide. The aim of this study was to characterise ESBL-KP detected in HUSM using WGS. WGS-based typing of bacterial pathogens could provide unprecedented resolution in discriminating even highly related lineages. By sequencing the entire genome (chromosome and mobile genetic elements), WGS immediately provides information on pathogen detection and identification, epidemiological typing, and drug susceptibility, which is crucial information that in conventional outbreak management is achievable only through the use of multiple methods.

To approach controlling such a bacterium, we first must define what it is and how it varies genetically. Therefore, the determination of the DNA sequence of ESBL-KP local isolates in comparison with other references *K. pneumoniae* in GenBank will shed light on genetic diversity, including variation within shared sequences and gain and loss of whole genes. Of additional importance is that resistance and virulence genes detected via WGS might not be expressed under conditions of phenotypic testing in vitro or, for that matter, in vivo.

The detection of such pathogenicity features via WGS could help clinicians identify potential nosocomial transmission events earlier and manage bacterial outbreaks before conventional phenotypic tests can detect them. Furthermore, the significant pathogenicity information of ESBL-KP would lead to the accurate development of pathogen detection technology using lateral flow, optical biosensors and electrochemical sensors in patients.

1.2.1 Research Questions

1. What are the whole genome sequences of five isolates of pathogenic ESBL-producing *K. pneumoniae* in HUSM?

2. What are the complete annotated genome sequences of local ESBL-KP isolates and what are the AMR and VFs that could be found in the local ESBL-KP genome?

3. How different are *K. pneumoniae* isolates in HUSM compared to other available *K. pneumoniae* strains in the database?

1.3 JUSTIFICATION OF THE STUDY

WGS of ESBL-KP of isolates obtained in HUSM will not only elucidate the pathogenic sequence of organisms found in our hospital system but also add to the database along with other variants around the world. By comparatively analysing the genome sequence, we will be able to initiate the identification of high-risk clones of public health importance and, ultimately, management through implementing control strategies for the disease. Additionally, the genomic comparison with other references in GenBank will shed light on the genetic diversity of local isolates with other variants around the world. Prior to achieving the future development of detection kits and vaccines for ESBL-KP, a comprehensive understanding of the whole genome of this pathogen is critically essential.

1.4 LITERATURE REVIEW

K. pneumoniae is an essential Gram-negative bacillus associated with several clinical infections in humans (1). This pathogen is affiliated with the *Enterobacteriaceae* family. It is the fourth and fifth most common cause of pneumonia and bacteremia respectively, among intensive care units (ICU), newborn units and in immunocompromised patients (2,3). In hospital environments, *Klebsiella* species survive and multiply in wet environmental sites and colonise the human bowel, bladder, upper respiratory tract and skin (4). It was recognised initially as an agent of pneumonia and urinary tract infection; however, its epidemiology in the community had changed with the emergence of virulent capsular serotypes (5). Today *K. pneumoniae* is associated with a wide spectrum of infections, including a distinctive invasive *K. pneumoniae* syndrome (13). Despite gaining popularity as an emerging pathogen in the community, the vast majority of *K. pneumoniae* infections are nosocomial. It is reported to be amongst the ten most common nosocomial pathogens in various studies (14,15). The prevalence of ESBL-KP strains in hospitals, ranges from 5 to 25% in several parts of the world (9,10). *K. pneumoniae* has emerged as an important cause of hospital-acquired infections, especially among patients in the neonatal intensive-care unit and mortality rates can be as high as 70% (11).

This increased burden of *K. pneumoniae* infections observed in healthcare is compounded by increased ESBL-producing isolates globally (16,17). ESBL are plasmid-mediated enzymes that can hydrolyze a wide variety of penicillin and cephalosporins, including third-generation cephalosporins and monobactams. Thus, Gram-negative bacilli containing these plasmids are multidrug-resistant (MDR) (18). However, the ESBL-producing bacteria remains susceptible to cephamycin and

carbapenems. Furthermore, these plasmids are mobile genetic elements and can be transmitted between Gram-negative bacilli of different species in vivo (19). The ability of this *K. pneumoniae* to undergo chromosomal recombination and exchange plasmids enables them to readily alter the repertoire of VFs and AMR genes (20). Thus, the concern of ESBL-KP goes beyond healthcare settings to affect various ecological niches, including animals, food products, soil and wastewater. Humans may become colonised or infected by ESBL-KP upon contact with the blood, saliva, faeces and urine of ESBL carrier animals or consuming contaminated water or food products (21). Cross-transmission of these ESBL-producing bacteria in hospitalised settings has been implicated in nosocomial infections worldwide. The common modes of transmission are via contaminated hands or gloves of healthcare workers, direct contact with contaminated surfaces (22), splashing of pathogen-contaminated water on sterile goods (23), invasive procedures used on patients (24) and droplets for respiratory pathogens (25). The presence of these reservoirs in hospital environments may increase the risk of acquiring nosocomial infections.

The incidence of ESBL in *K. pneumoniae* differed geographically and was reported from 1997 to 2003 to be 12.3% for North America, 51.9% for South America, 16.7% for Northern Europe, 24.4% for Southern Europe, 58.7% for Eastern Europe and 28.2% for Asia Pacific region (26). In Malaysia, Palasubramaniam et al., 2005 reported an association of ESBL-KP with a nosocomial outbreak in a paediatric oncology unit in a Malaysian public hospital (12). The ESBL prevalence in Malaysia public hospitals that was studied by Mahdi et al. (2016) showed relative similarities when compared to that reported in the National Surveillance of Antibiotic Resistance (NSAR) report for 2014 (27), in which the prevalence of ESBLs ranged between 47% and 70% for *K.*

pneumoniae (28–30). In 2017, according to the NSAR's report, ESBL-KP resistance to cephalosporin drugs such as cefuroxime, ceftazidime and cefotaxime, the aminoglycoside amikacin, as well as ciprofloxacin had demonstrated increment in contrast to the decreasing trend observed in the previous year (2016) (31).

In conventional methods, phenotypic identification of resistance bacteria is based on different clinical tests, including bacterial growth inhibition in disk diffusion or dilution tests (E test). These methods often do not give positive results, take a prolonged time to complete, and may lead to wrong prescriptions (32–35). Acquisition of antibiotic resistance in clinically significant bacteria from the hospital environment is directly correlated to the overuse and indiscriminate prescription of antibiotics, thus increasing antibiotic selective pressure in bacteria (11). The incidence of the emergence of carbapenem resistance due to metallo-beta-lactamase enzyme production in *K. pneumoniae* is being reported worldwide (36). Furthermore, the ability of *K. pneumoniae* to form mixed-species biofilm with *Pseudomonas aeruginosa (P. aeruginosa)* poses further challenges in managing infections, particularly chronic wounds and mechanical ventilation-associated infections. The emergence of these high-risk isolates and the global spread of these isolates has left clinicians with very few therapeutic options. The main concerns with ESBL are limited treatment options, resistance to multiple classes of antimicrobials leading to MDR ESBL phenotype, increased mortality and cost of management (18). Hence, it is critical to have protocols in place for antimicrobial stewardship and enhanced surveillance control efforts to limit the spread of MDR and biofilm-forming *K. pneumoniae* strains (37).

Therefore, continuous surveillance is essential to monitor the ESBL-KP in HUSM. To the best of our knowledge, this is the first study investigating ESBL-KP

using WGS to investigate infections in patients caused by this organism in our hospital system. By analysing the genome sequence of pathogenic ESBL-KP, we will be able to initiate the identification of high-risk clones of public health importance and, ultimately, management through the implementation of control strategies for the disease. Additionally, the genomic comparison with other references in GenBank will shed light on the genetic diversity of local isolates with other variants around the world. Prior to achieving the future development of detection kits and vaccines of ESBL-KP, a comprehensive understanding of the whole genome of this pathogen is critically important.

1.5 OBJECTIVES

1.5.1 General Objective

In this study, we aim to report on the molecular characterisation through WGS of ESBL-KP, isolated from patients admitted at the HUSM in Kelantan.

1.5.2 Specific objectives

1. To sequence the whole genome of five isolated pathogenic ESBL-KP.
2. To predict and annotate the genes of local ESBL-KP isolates and to identify AMR and VFs in sequencing data.
3. To compare the molecular characterisation of isolates with genome references from GenBank using comparative genomics analysis.

CHAPTER 2: STUDY PROTOCOL

2.1 STUDY PROTOCOL SUBMITTED FOR ETHICAL APPROVAL

Research Methodology

1. Bacterial strains: The pure bacterial isolates will be obtained from INFORMM BioBank. All samples will be cultured on MacConkey agar supplemented with 2 mg/L cefotaxime and incubated for 18-24 hours at 37°C in normal atmosphere.

2. Genomic DNA extraction: The DNA isolation from pure bacteria culture will be done using commercially available DNA extraction kit, Genomic Tip 100/G (Qiagen, USA) and Genomic Buffer Set (Qiagen, USA) following the protocol by the manufacturer. Firstly, the bacteria culture from LB broth (20 ml) will be centrifuged at 3,000-5,000 x g for 5-10 minutes to collect the bacteria pellet. The bacteria pellet will be suspended in 3.5 ml of Buffer B1 containing 100 mg/ml RNase A by vortexing at high speed until the suspension becomes homogenous. Qiagen Proteinase K stock solution (100 µL) will be added and incubated at 37°C for at least 30 minutes. If the suspension is not homogenous after vortexing, a longer incubation is recommended to avoid clogging the Qiagen Genomic Tip. Buffer B2 (1.2 ml) will be added to the suspension after incubation and mixed by inverting the tube several times and vortexing for a few seconds. The suspension will be incubated at 50°C for 30 minutes. Before loading the lysates into the Genomic Tip 100/G column, the column will be equilibrated with 4 ml of Buffer QBT, and the gravity flow will empty the column. The lysates from the previous step will be vortexed for 10 seconds at maximum speed and apply it to the

equilibrated column. The column will be washed with 2 x 7.5 ml of Buffer QC. Buffer QC will be emptied from the column by gravity flow. The bacterial DNA will bind to the column while the other cell constituent will pass through. The washing step will be done to remove any remaining contaminants. Pure and high molecular weight DNA will be eluted using the 5 ml pre-warmed elution Buffer QF at 50°C. The quantification of extracted genomic DNA was determined on a NanoDrop spectrophotometer with verification by agarose gel electrophoresis.

3. Genome sequencing will be done using PacBio technology. It will be sent for PacBio sequencing as it will be outsourced to the company.

4. Sequencing data analysis:

i) The resulting raw reads will be checked for quality and trimmed using CLC Genomics Workbench version 10 (CLC, Bio-QIAGEN, Aarhus, Denmark). *De novo* assembly will be done using PacBio's Single-molecule real-time sequencing (SMRT) Portal (v2 .3.0), the hierarchical genome assembly process (HGAP) with default settings and a seed read cut-off length of 4,000 bp for an accurate assembly across all isolates' genomes. Assemblies will be performed multiple times with the combination of three SMRT cells of sequencing reads, and these will be done in HGAP. Higher-quality preassembled reads with around 25-30 x coverage will be generated by aligning shorter reads against longer seed reads. The preassembled reads will then be fed to the celera assembler (38) to obtain a draft assembly, followed by a consensus polishing procedure called quiver. Basic local alignment with successive refinement (BLASR) (39) will be

used for aligning candidate overlaps, which are identified using a Ferragina-Manzini (FM)-index search and clustering of k-mer hits.

ii) The *de novo* assembled reads will be uploaded in GenBank and annotated using NCBI prokaryotic genome annotation pipeline and RAST 2.0 server (40). RAST 2.0 server will identify encoding proteins, rRNA and tRNA, assign functions to the genes and predict subsystems represented in the genome.

iii) Comparative genomics will be done using RAST SEED viewer and BLAST Ring Image Generator (BRIG). Multilocus sequence typing (MLST) will be predicted using MLST 2.0 program (https://cge.cbs.dtu.dk/services/MLST/). The references from GenBank will also be included for the genomics comparison, including sequence-based and function-based analysis (41).

iv) The bacterial analysis pipeline of the GoSeqit tool will also be used to annotate and identify known acquired antibiotic-resistant genes via ResFinder, VFs (including capsular polysaccharide (CPS), lipopolysaccharide (LPS), adhesin, long polar fimbriae (*lpfA*), increased serum survival (*iss*), enterobactin siderophore receptor protein (*iroN*), biofilm, lipase, ABC transporter protein *MchF*, and gelatinase using VirulenceFinder and mobile genetic elements (MGEs) through PlasmidFinder (42,43).

v) Phylogenetic analysis will be performed using the Rapid large-scale prokaryote pan-genome analysis (Roary) (44). The maximum likelihood phylogenetic will be generated, edited and visualised using FastTree version 2.1.7.

vi) The RAST SEED viewer will be used to identify the presence of transposases and integrons flanking the - lactamase genes. PHAge Search Tool (PHASTER) server will be used to identify, annotate, and visualise prophage sequences (45).

Study area

Institute for Research in Molecular Medicine (INFORMM), USM Health Campus

Sampling method and subject recruitment

The samples of positive ESBL-KP isolates were archived from glycerol stock in INFORMM Biobank, USM Health Campus.

Study population

Archived samples of ESBL-KP isolates from INFORMM, Biobank.

Research tool

Name	**Description**	**Existing**	
		USM	**External**
Specialised equipment	Thermal cycler Incubator Incubator shaker	√	
Facility	Biological safety class 2	√	

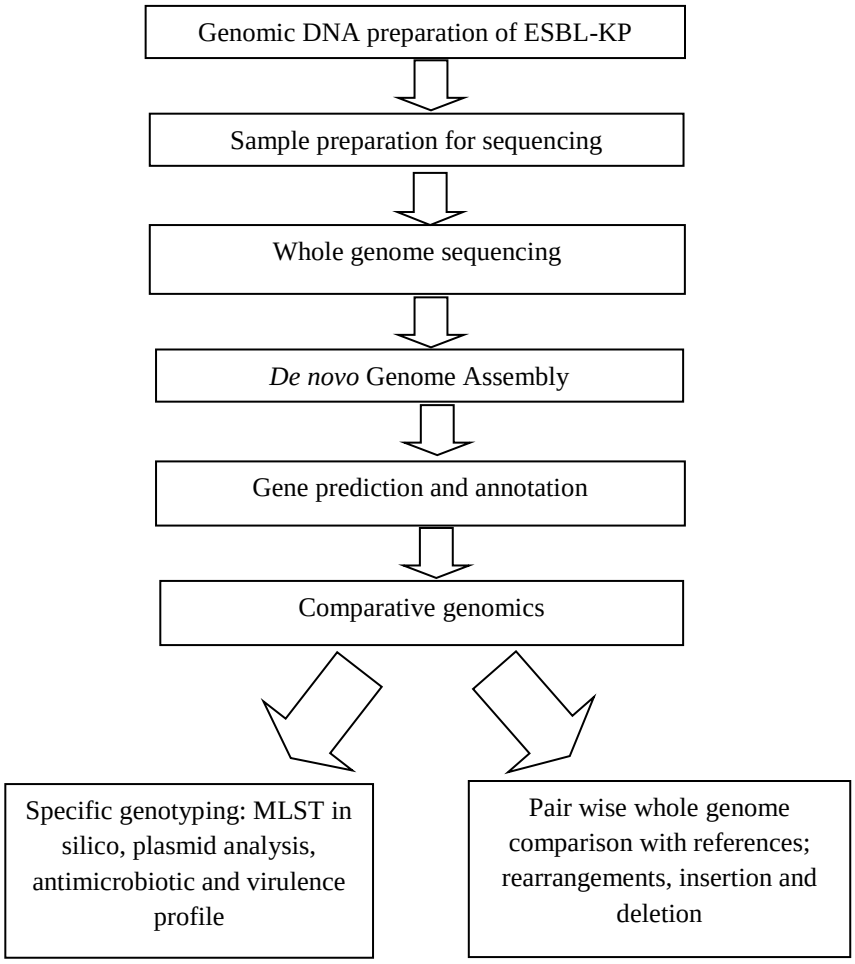
Data collection method

The samples will be sourced from glycerol stock in INFORMM Biobank. Prior to use for research purposes, the samples will be cultured overnight for DNA extraction. DNA extraction will be performed on isolates using commercially available kits according to the manufacturer's protocol. The DNA samples will be outsourced for whole genome sequencing. The sequence data will be collected and analysed using several appropriate bioinformatics tools (as mentioned in the Data analysis section). The data collection will be tabulated as below:

1. Anonymous genome identity
2. Genome size
3. Total number of predicted genes
4. Total number of predicted AMR genes
5. Total number of predicted virulence genes

Study flowchart

Whole genome sequencing and comparative genomics of ESBL-KP



Data analysis

Analysis	Software	Objectives
Genome assembly	SPAdes	To assemble the genome
Genome annotation and prediction	Prodigal, BLAST & InterPro	To annotate and predict the gene
Sequence Type	MLST 2.0	To predict the sequence type
Antimicrobial resistance	ResFinder	To predict antimicrobial resistance genes
Virulence gene	Virulence Database	To predict virulence genes

Expected result(s)

No	Publication	Year 1
1	Publications in citation-indexed journals (ISI/Scopus)	1

Ethical Consideration(s)

Ethical approval was obtained from the Human and Research Ethics Committee of University Sains Malaysia (JEPeM Code: USM/JEPeM/22010067).

Subject vulnerability

There are no vulnerable people involved in this study.

Declaration of absence of conflict of interest

There is no conflict of interest.

Privacy and confidentiality

The information will be kept confidential by the researchers. It will not be made publicly available unless disclosure is required by law. The samples will be anonymous/non-identifiable (i.e. personal identifiers will not be kept with the sample, and the sample will not have a code number that can be used to identify the subject) or coded. The samples are considered de-identified (i.e. any identifying information such as name will be replaced with a code, and only a few authorised people will have access to this code to link samples and data back to personal identifiers). All the bacteria isolates will be kept in INFORMM BioBank in glycerol stock. Only the principal investigator and team will have access to the isolates and information such as identity/registration number, source of infection, source of isolates and antibiotics profile. The isolates obtained during this study will be stored and analysed only for this study for a period not exceeding one year. They will be destroyed after the completion of the study. Data obtained from this study does not identify the subjects individually and only will be published for knowledge purposes. The researchers and the Ethical Review Board for this study may review the original medical records of the subjects.

Community sensitivities and benefits

This study does not possess community sensitivities, but it will benefit the community by improving methods of future antibiotic creation per any genetic variation found in the local isolates.

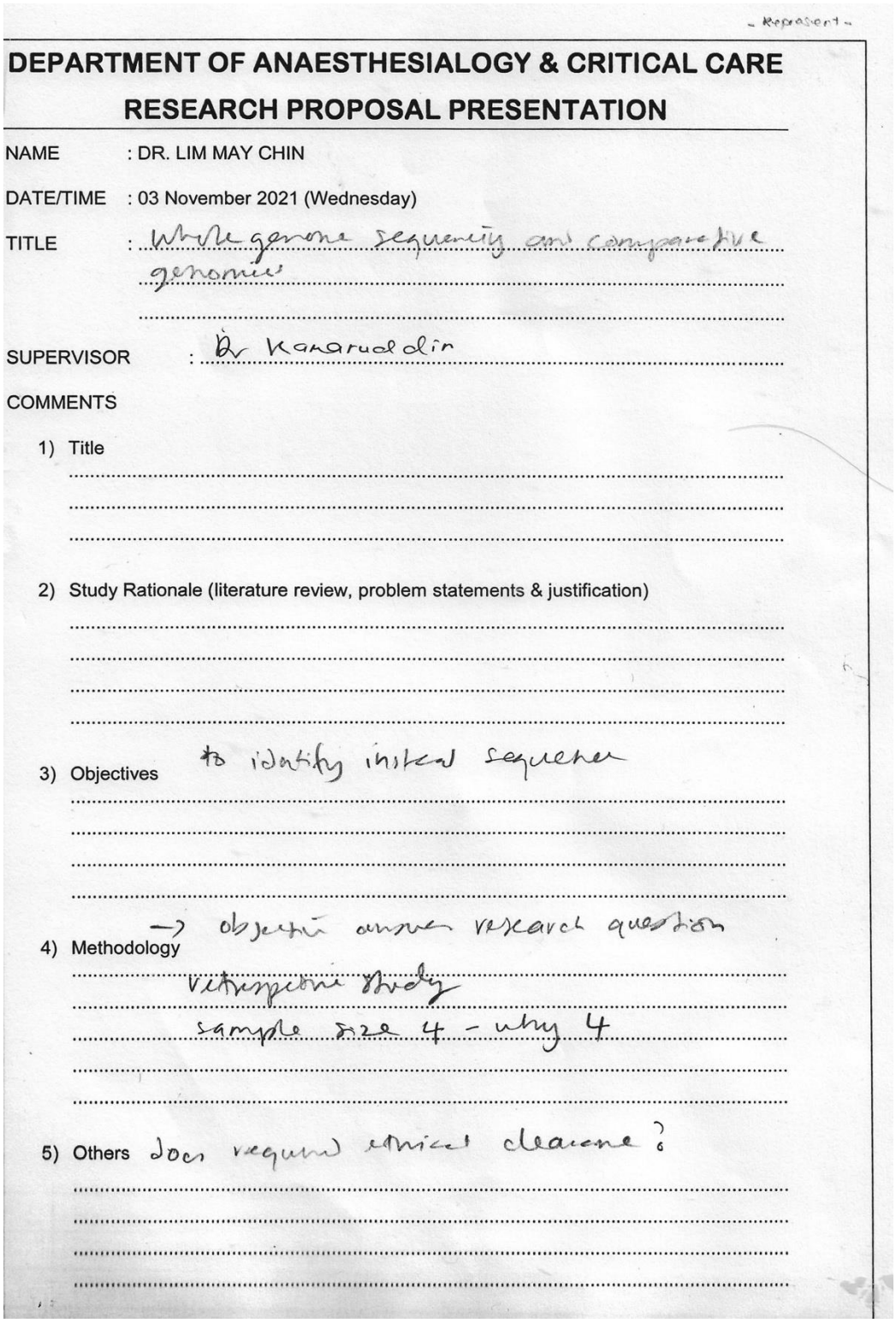
Gantt Chart

Tasks/Month	1	2	3	4	5	6	7	8	9	10	11	12
Culturing and sample preparation of ESBL-KP												
Whole genome sequencing												
Genome assembly												
Genome prediction and identification												
Genomic comparison												

Milestones

Milestone	Description	Month	Duration (month)
M1	Completion of culturing and confirmation of ESBL-KP	Nov 2021	2
M2	Completion of whole genome sequencing	January 2022	2
M3	Completion of genome assembly	March 2022	1
M4	Completion of genome prediction, annotation, antimicrobial resistance and virulence factor identification	April 2022	3
M5	Completion of comparative genomics with other references in GenBank	July 2022	4
M6	Data interpretation and dissertation write up	October 2022	1

2.2 RESEARCH PROPOSAL PRESENTATION



- Represent -

DEPARTMENT OF ANAESTHESIALOGY & CRITICAL CARE

RESEARCH PROPOSAL PRESENTATION

NAME : DR. LIM MAY CHIN

DATE/TIME : 03 November 2021 (Wednesday)

TITLE : Whole genome sequencing and comparative genomics

SUPERVISOR : Dr Kamarul Adin

COMMENTS

1) Title

2) Study Rationale (literature review, problem statements & justification)

3) Objectives → to identify invitro sequence
→ objective answer research question

4) Methodology
Retrospective study
sample size 4 - why 4

5) Others does require ethical clearance?