COMPARATIVE GENOMICS OF MULTI-DRUG RESISTANT *Pseudomonas aeruginosa*

MUHAMMAD ADIB BIN ABD WAHAB

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

COMPARATIVE GENOMICS OF MULTI-DRUG RESISTANT *Pseudomonas aeruginosa*

MUHAMMAD ADIB BIN ABD WAHAB

DISSERTATION SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF BIOTECHNOLOGY

INSTITUTE OF BIOLOGICAL SCIENCE
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2019

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **MUHAMMAD ADIB BIN ABD WAHAB**

Matric No: **SGF150017**

Name of Degree: **MASTER OF BIOTECHNOLOGY**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

COMPARATIVE GENOMICS OF MULTI-DRUG RESISTANT *Pseudomonas aeruginosa*

Field of Study: **GENETIC, MOLECULAR MICROBIOLOGY AND BIOINFORMATIC**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

COMPARATIVE GENOMICS OF MULTI-DRUG RESISTANT

Pseudomonas aeruginosa

ABSTRACT

Pseudomonas aeruginosa is an opportunistic Gram-negative bacterium that inhabits diverse environments and is also one of the leading pathogens causing nosocomial infections in humans. Multidrug resistant (MDR) *P. aeruginosa* is of worldwide concern because of high rates of antibiotic resistance. The objectives of the study were to compare the genetic content, antimicrobial resistance genes and virulence genes of two clinical *P. aeruginosa* strains, PAC08 and PAC17, from a tertiary hospital in Kuala Lumpur. In addition, the intraspecies phylogenetic relationship between Malaysian *P. aeruginosa* and other geographically located strains was determined. The whole genome sequences of PAC08 and PAC17 were assembled using bioinformatics tools to determine the genomic content, antibiotic resistance and virulence-associated genes and phylogeny. Phylogenetic analysis amongst Malaysian and global strains were determined based on 16S rRNA gene, 23S rRNA gene, multilocus sequence typing (MLST) and whole-genome single nucleotide polymorphism (SNP). Genome assembly of PAC08 and PAC17 showed that both strains have similar genome size, 6.89 Mbp and 6.84 Mbp, respectively. MLST analysis showed that sequence type (ST) of PAC08 was ST1076 which is rarely found while the ST of PAC17 is ST235, a more common ST circulating in Europe, China, Japan and Brazil. PAC08, an urine strain has more bacteriophages while the blood strain PAC17 has more insertion sequence (IS). PAC17 which is a multidrug resistant strain, has high number of genes associated with drug resistance involving efflux pumps and antibiotic inactivation enzymes. A similar number of virulence genes were detected in PAC08 and PAC17. Clustered regularly interspaced short palindromic repeat (CRISPR) analysis revealed that PAC08 has three CRISPR loci and five questionable CRISPR loci while

PAC17 only has two questionable CRISPR loci. Pan genome analysis defined PAC08 and PAC17 as open-pangenome, which have large genomes and high horizontal rate of gene transfer. Through phylogenetic analysis based on 16S and 23S rRNA genes, MLST, and SNP, PAC17 was grouped together with a few MDR *P. aeruginosa* strains from Japan and United Kingdom while PAC08 was closely related with another MDR *P. aeruginosa* strain isolated from Malaysia. As a conclusion, even though PAC08 and PAC17 were isolated from different source, there was not much difference in term of genomic structures. Genomic contents of PAC08 and PAC17 are diverse with the presence of a variety mobile genetic elements such as insertion sequence, bacteriophages, and integrons. The major element contributing to the multidrug resistance in *P. aeruginosa* is multidrug efflux pumps which efficiently efflux out antibiotics. Moreover, the presence of resistance gene elements in PAC08 and PAC17 had increased resistance towards aminoglycoside, β-lactamase, Fosfomycin, Macrolide, Phenicol and sulphonamide. Various virulence factors were found in PAC08 and PAC17 which probably play roles in pathogenesis. This analysis had provided an insight into virulence and pathogenesis of MDR *P. aeruginosa* of Malaysia strains which will assist in preventing and controlling transmission of MDR *P. aeruginosa.*

Keywords: comparative genomics, multidrug resistant, phylogenetic analysis *Pseudomononas aeruginosa*, virulence genes.

PERBANDINGAN GENOM *Pseudomonas aeruginosa* YANG MEMPUNYAI RINTANGAN PELBAGAI UBAT

ABSTRAK

Pseudomonas aeruginosa merupakan bakteria gram-negatif oportunis yang didiami di pelbagai persekitaran dan juga merupakan salah satu patogen yang menyebabkan jangkitan nosokomial pada manusia. *P. aeruginosa* yang mempunyai rintangan pelbagai ubat ("*multidrug resistant" atau* MDR) telah menjadi salah satu kebimbangan di seluruh dunia kerana peningkatan peratusan kadar rintangan antibiotik yang tinggi. Objektif utama kajian ini adalah untuk mengumpul, menganotasi, dan memetakan data jujuran DNA daripada dua strain *P. aeruginosa*. Kemudian, objektif kedua kajian ini adalah untuk melaksanakan perbandingan kandungan genetik dan gen-gen yang berkaitan dengan kerintangan antibiotik dan kevirulenan didalam genom *P. aeruginosa* dari Malaysia iaitu PAC08 dan PAC17. Objektif ketiga kajian adalah untuk menentukan hubungan filogenetik antara strain *P. aeruginosa* dari Malaysia dan juga dari lokasi geografi yang lain. Penjujukan genom keseluruhan (WGS) PAC08 dan PAC17 telah digabung menggunakan perisian-perisian bioinformatik untuk mengetahui isi kandungan genom, gen-gen rintangan antibiotik dan gen-gen kevirulenan, serta mencari hubungan filogenetik. Analisis filogenetic strain dari Malaysia dan global telah ditentukan menggunakan gen 16S rRNA, gen 23S rRNA, *multilocus sequence typing* (MLST), dan *whole-genome single nucleotide polymorphism* (SNP). Hasil kajian menunjukkan gabungan genom strain PAC08 dan PAC17 menghasilkan saiz genom 6.89 Mbp bagi PAC08, and 6.84 Mbp bagi PAC17. Analisis MLST menunjukkan PAC08 dikategorikan di bawah garis keturunan ST1076 yang jarang dijumpai manakala PAC17 dikategorikan di bawah garis keturunan ST235 yang mana biasa dijumpai di Eropah, China, Jepun dan Brazil. PAC08 yang dijumpai didalam sampel urin menunjukkan banyak jujuran DNA dari *bacteriophages* manakala PAC17 yang dijumpai didalam sample darah menunjukkan

banyak *insertion sequence (IS)*.PAC17 yang merupakan strain yang mempunyai banyak rintangan kepada antibiotik, dan mempunyai banyak gen yang berkaitan dengan pam effluk serta enzim yang menghapuskan antibiotik. Selain itu, kedua-dua strain PAC08 dan PAC17 mempunyai jumlah gen kevirulenan yang sama. Analisis *Clustered regularly interspaced short palindromic repeat* (CRISPRs) menunjukkan PAC08 mempunyai tiga lokus CRISPRs dan lima lokus CRISPRs yang meragukan manakala PAC17 hanya mempunyai dua lokus CRISPRs yang meragukan. Analisis pan genom menklassifikasikan PAC08 dan PAC17 mempunyai pan genom terbuka, yang mempunyai genom yang besar dan mempunyai banyak pemindahan gen mendatar. Melalui analisis filogenetik yang berdasarkan 16 dan 23S rRNA gen, MLST dan SNP, PAC17 berada bersama-sama dengan beberapa strain *P. aeruginosa* dari Jepun dan United Kingdom sementara PAC08 adalah berkait rapat dengan MDR P. aeruginosa strain yang juga dari Malaysia. Kesimpulanya, walaupon PAC08 dan PAC17 dijumpai di lokasi berlainan, genom strukturnya tidak banyak berbeza antara kedua-dua strain. Kandungan genom PAC08 dan PAC17 lebih diversiti dengan hadirnya *mobile genetic element* seperti *insertion sequence (IS), bacteriophages,* dan *integrons*. Faktor utama yang menyumbang kepada rintangan kepada antibiotik adalah wujudnya pam effluk yang berfungsi untuk menyinkir keluar antibiotic. Selain itu, kehadiran gen kerintangan didalam PAC08 dan PAC17 juga meningkatkan lagi tahap kerintangan terhadap *aminoglycoside, β-lactamase, Fosfomycin, Macrolide, Phenicol* dan *Sulphonamide*. Banyak gen kevirulenan telah dijumpai di dalam PAC08 dan PAC17, yang mana gen-gen ini memainkan peranan penting dalam jangkitan *P. aeruginosa* kedalam hos.

Kata kunci: *Pseudomonoas aeruginosa*, rintangan pelbagai ubat, perbandingan genomic, analisis filogenetik, gen kevirulenan

ACKNOWLEDGEMENTS

First of all, Praise be to Allah for giving me strength, determination, and good health to complete my work for my master's degree. Secondly, I would like to thank Prof Thong Kwai Lin for her guidance, advices, constructive comments, concerns and most important, patience!. I'm so honored to be her student and I wish she could stay longer as one of lecturer in UM. Her contributions in medical microbiology for years are irreplaceable. A big thank to Dr Saharuddin in giving advices, opinions, and guidance especially in bioinformatic analysis part. His experiences in bioinformatic attracted me to seek more knowledge and change my perspective towards bioinformatic (I hated this topic during my bachelor).

Most important, I would like to thank my beloved wife, Sharifah Aina Afzan Syed Aminud-din, who also my classmate, for encouragement and support me to finish this thesis. To my parent, siblings and family in law, thank you for always being there and always being supportive especially in my hard time.

Also, I would like to thank my course mate especially Nava and Anis, whom spent time in discussion related to bioinformatic. Not to forget, my labmate in bioinformatic lab, Yosri (now pursue PhD) and Hamidah (now become Bioinformatician). Without these two persons, my journey in bioinformatic would be hard. Thank you for their wonderful friendhip, for great times, shared knowledge and support.

TABLE OF CONTENTS

ABSTRACT ..... **iii**

ABSTRAK ..... **v**

ACKNOWLEDGEMENTS ..... **vii**

TABLE OF CONTENTS ..... **viii**

LIST OF FIGURES ..... **xi**

LIST OF TABLES ..... **xii**

LIST OF SYMBOLS AND ABBREVIATIONS ..... **xiii**

LIST OF APPENDICES ..... **xv**

CHAPTER 1: INTRODUCTION **1**

1.1 Research Objective 3

CHAPTER 2: LITERATURE REVIEW **4**

2.1 *Pseudomonas aeruginosa* 4

2.2 Whole Genome Sequencing of *P. aeruginosa* 4

2.3 Multidrug Resistance of *P. aeruginosa* 5

2.3.1 Intrinsic Resistance of *P. aeruginosa* 5

2.3.2 Acquired Resistance of *P. aeruginosa* 6

2.4 Virulence Factors of *P. aeruginosa* 6

2.5 Mechanism of Pathogenesis 7

2.5.1 Motility, Attachment and Colonization 7

2.5.2 Toxin Secretion 7

2.6 Pan Genome Analysis 8

2.6.1 Core Genome 8

2.6.2 Accessory Genome 9

2.6.3 Species-Specific or Strain-Specific Genes 9
2.6.4 Pan genome of *P. aeruginosa* 9
2.7 Horizontal Gene Transfer 10
2.7.1 Mechanism of HGT 11
2.7.2 Transfer of Antibiotic Resistance Gene Through HGT 12
2.7.3 Virulence Gene Through HGT 12

CHAPTER 3: METHODOLOGY 14
3.1 Strains background 14
3.2 Whole Genome Sequencing 14
3.2.1 Genome Sequencing, Assembly and Annotation 14
3.2.2 Genome Analyses and Comparative Studies 15
3.2.3 Pan-Genome and Core Genome Analysis 16
3.2.4 Genbank Accession Number 16
3.3 Virulence Genes Identification 16
3.4 Antimicrobial Resistance Genes Determination 17
3.5 Phylogenetic Analysis of *P. aeruginosa* Isolates 17
3.5.1 16S rRNA Phylogenetic Analysis 18
3.5.2 23S rRNA Phylogenetic Analysis 18
3.5.3 *in-silico* MLST Phylogenetic Analysis 18
3.5.4 Whole Genome SNP-Based Phylogenetic Analysis 19

CHAPTER 4: RESULTS 20
4.1 Genome Features of *P. aeruginosa* 20
4.2 *In-silico* Multilocus Sequence Typing (MLST) 25
4.3 Pan-Genome and Core Genome Analysis 28
4.4 Mobile Genetic Elements 32

4.4.1 Insertion Sequence (IS) 32
4.4.2 Bacteriophages 33
4.4.3 Integron 35
4.5 CRISPR/CAS System 37
4.6 Phylogenetic Analysis of *P. aeruginosa* based on 16S rRNA, 23S rRNA, *In silico* MLST And Whole Genome SNP. 39
4.7 Virulence Factors in *P. aeruginosa* 45
4.8 Antibiotic Resistance 48
4.8.1 Efflux pumps 48
4.8.2 Resistance genes elements 50

CHAPTER 5: DISCUSSION 53
5.1 Genome Assembly and Genomic Features 53
5.2 Genetic diversity of *P. aeruginosa* based on MLST 57
5.3 Phylogenetic Analysis of Multidrug Resistant *P. Aeruginosa* 59
5.4 Resistance Genes 61

CHAPTER 6: CONCLUSION 66

REFERENCES 68
APPENDICES 83

LIST OF FIGURES

Figure 4.1:	Subsystem distribution based on RAST annotation server for PAC08	22
Figure 4.2:	Subsystem distribution based on RAST annotation server for PAC17	23
Figure 4.3:	Comparison of the number of cores, accessory and unique genes presented in 12 isolates of *P. aeruginosa* from Malaysia	28
Figure 4.4:	Graphical representation of core and pan genome from 13 isolates of multidrug resistant *P. aeruginosa* from Malaysia	20
Figure 4.5:	Percentage of insertion element in PAC08 and PAC17 based on insertion sequence family and Transposon family	31
Figure 4.6:	Gene cassette of complete Integron system in PAC08 consists of multiple antibiotic resistant gene, attC site, attI sites, two promoters and integron integrase	34
Figure 4.7:	Gene coordination and location of 9 gene cassettes found in the PAC17 genome	35
Figure 4.8:	Maximum-likehood phylogenetic tree based on 16S rRNA gene showing relationship between PAC08 and PAC17 with MDR isolates from other regions	40
Figure 4.9:	Maximum-likehood phylogenetic tree based on 23S rRNA gene showing relationship between PAC08 and PAC17 with MDR isolates from other regions	41
Figure 4.10:	Maximum-likehood tree based on MLST showing relationship between PAC08 and PAC17 with clinical isolated from other regions	42
Figure 4.11:	Maximum-likehood tree based on full genome SNP show relationship between PAC08 and PAC17 with MDR isolates from other regions	43

LIST OF TABLES

Table 4.1:	General genomic features of the whole genome sequence of the Malaysia *P. aeruginosa* PAC08 and PAC17	21
Table 4.2:	Allelic profiles and sequence typing (ST) assigned to the clinical Malaysian *P. aeruginosa* isolates according to the PubMLST protocol and MLST database.	25
Table 4.3:	Phage associated regions in PAC08 and PAC17 identified by PHAST	33
Table 4.4:	Distribution of different CRISPR loci among PAC08 and PAC17 isolates and their direct repeat sequence	37
Table 4.5:	Summary of all virulence genes detected in draft genome PAC08 and PAC17	45
Table 4.6:	Predicted multi-drug efflux pump from CARD and the presence or absence of genes associated with antibiotic resistance	48
Table 4.7:	Genes associated with antibiotic resistance present in the Malaysian and reference strains of *P. aeruginosa*	51

LIST OF SYMBOLS AND ABBREVIATIONS

Mbp	:	Mega base pair
ABC	:	ATP-Binding Cassette
AME	:	Aminoglycoside Modifying Enzyme
BAL	:	Bronchoalveolar lavage
BLAST	:	Basic Local Alignment Search Tool
Bp	:	Base pair
BPGA	:	Bacteria Pan Genome Analysis
BRIG	:	BLAST Ring Image Generator
CARD	:	Comprehensive Antibiotic Resistance Database
CDS	:	Coding sequence
COG	:	Cluster of Orthologous Group
CRISPR	:	Clustered Regularly Interspaced Short Palindromic Repeats
HGT	:	Horizontal Gene Transfer
IS	:	Insertion sequence
Kb	:	Kilobases
KEGG	:	Kyoto Encyclopedia of Genes and Genomes
LGT	:	Lateral gene transfer
MATE	:	Multidrug and toxic compound extrusion
MDR	:	Multidrug Resistant
MEGA	:	Molecular Evolutionary Genetics Analysis
MFS	:	Major Facilitator Superfamily
MGE	:	Mobile Genetic Element
ML	:	Maximum Likehood
MLST	:	Multi Locus Sequence Typing

MUSCLE	:	Multiple Sequence Comparison by Log-Expectation
NCBI	:	National Center for Biotechnology Information
NGS	:	Next Generation Sequencing
NSAR	:	National Surveillance of Antimicrobial Resistance
ORF	:	Open reading frame
PAI	:	Pathogenicity island
PHAST	:	Phage Search Tool
QUAST	:	Quality Assessment Tool for Genome Assemblies
RAST	:	Rapid Annotation using Subsystem Technology
RGI	:	Resistance Gene Identifier
RND	:	Resistance-Nodulation-cell Division
SMR	:	Small Multidrug Resistance
SMRT	:	Single-Molecule Real-Time sequencing
SNP	:	Single Nucleotide Polymorphism
UTI	:	Urinary tract infection
VFDB	:	Virulence Factor Database
WGS	:	Whole genome sequence

LIST OF APPENDICES

APPENDIX A	Genome statistic of *P. aeruginosa* strain PAC08 obtained from Quality Assesment Tools for Genome assembly (QUAST)	83
APPENDIX B	Genome statistic of *P. aeruginosa* strain PAC17 obtained from Quality Assesment Tools for Genome assembly (QUAST)	85
APPENDIX C	Genome statistic of MDR *P. aeruginosa* used in drawing phylogenetic tree	87

CHAPTER 1: INTRODUCTION

Pseudomonas aeruginosa is an aerobic Gram-negative, motile, and rod-shaped bacterium which inhabits diverse environments and capable infecting plants (Walker et al., 2004), insects (Jander et al., 2000), animals (Rahme et al., 1995; Wu et al., 2015) and nematodes (Mahajan-Miklos et al., 1999). *P. aeruginosa* has an uncanny ability to tolerate a wide range of environmental conditions, and survive in minimal requirement of nutrition, hence it could persist in many niches (Wu et al., 2015). In 2008, *P. aeruginosa* had caused an outbreak in one of the hospitals in Malaysia and was associated with severe contamination of respiratory care equipment, irrigating solution, catheter, dilute antiseptic, cleaning solution and soap (Suraiya et al., 2008).

P. aeruginosa is one of the causes of nosocomial infections particularly infecting immunocompromised, cancer, burn, cystic fibrosis and intensive care unit patients (Cramer et al., 2012; Driscoll et al., 2007). Infected patients commonly suffer pneumonia, urinary tract infection (UTI), surgical site infection, bloodstream infection and skin infection especially in burn injuries (Driscoll et al., 2007). Common antibiotics used in the treatment of *P. aeruginosa* related infections are β-lactam (penicillin, cephalosporin, and carbapenem) associated either with aminoglycoside (gentamicin, tobramycin and amikacin) or fluoroquinolone (ciprofloxacin and levofloxacin) (Hirsch & Tam, 2010; Mesaros et al., 2007).

Over the years, antibiotic treatment for *P. aeruginosa* infections such as ciprofloxacin and ceftazidime started to be less effective because of its development to resist multiple classes of antibiotics (Gooderham & Hancock, 2009). This occurrence of multidrug resistance was reported around the globe including the United State America (Obritsch et al., 2004), Europe (Goossens, 2003) and even Asia (Kang & Song, 2013). The infections

by multidrug resistant (MDR) *P. aeruginosa* had caused delay in treatment especially in immunocompromised patients, which often lead to mortality (Aloush et al., 2006).

In Malaysia, the National Surveillance of Antimicrobial Resistance (NSAR) reported that antibiotics such as piperacillin, ceftazidime, cefepime, imipenem, meropenem, amikacin, ciprofloxacin and collistin showed increased resistance rate from year 2015 to 2016. Amongst of these antibiotics, the resistance towards imipenem and meropenem were the most worrisome because the resistance rates had approached 10%. A recent study by Phoon et al. (2018) reported the resistance rate of *P. aeruginosa* to ceftazidime, cefepime, imipenem, and meropenem as 7.0%, 6.5%, 11.6% and 11.6%, respectively. A further verification done by Phoon et al. (2018) reported that the strains resistant to these antibiotics harboured multiple class one integron gene cassettes.

Since the first publication of completed sequence of *P. aeruginosa* strain PAO1, more than 100 complete genomes and thousands draft genomes of *P. aeruginosa* are available and have been submitted to National Center for Biotechnology Information (NCBI) database. In NCBI database, the genome sizes of *P. aeruginosa* range from 5 Mbp to 7.5 Mbp and are among the largest bacterial genome to be sequenced. Compared to other large bacterial genome, *P. aeruginosa* genomes contain large amount of genes encoded for outer membrane proteins (antibiotic efflux, motility, adhesion, virulence factor, and environment sensing by two-component system), transport systems, and enzymes which involve in nutrient uptake and metabolism (Kung et al., 2010).

1.1 Research Objective

In Malaysia, a few genome sequences *P. aeruginosa* have been reported but there are no complete genome sequences of clinical MDR *P. aeruginosa* strains. With the available genome sequences of local *P. aeruginosa*, genome content can be revealed therefore providing more information regarding to genome structure, genome evolution, antibiotic resistance and virulence factors of Malaysian strains.

Therefore, in this study, raw sequence data generated from the Illumina Miseq platform were retrieved to dissect the genome *P. aeruginosa*. The next generation sequencing was done by a local commercial company. Various bioinformatic softwares were used to analyze the genomic contents of *P. aeruginosa*. The objectives of this study were:

i. To assemble, annotate and map the sequence data of two *P. aeruginosa* strains
ii. To determine and compare the genetic content and various genes responsible for multidrug resistance and the virulence factors of two local *P. aeruginosa* strains
iii. To determine the phylogenomic relationship between two Malaysian *P. aeruginosa* strains and multiple representative strains of *P. aeruginosa* from North America, Europe, and Asia countries such as China, Japan, India, and Malaysia.

CHAPTER 2: LITERATURE REVIEW

2.1 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa (*P. aeruginosa*) is an ubiquitous Gram negative rod-shape bacterium that can cause disease in animals, plants and humans. *P. aeruginosa* can be identified by culture on the MacConkey medium agar and can be recognized by blue green colony or yellow-green fluorescent colonies depending on medium used (Kirisits et al., 2005). This coloured colony is formed because of *P. aeruginosa's* ability to secrete pyocyanin, pyoverdine and pyorubin (King et al., 1954). This species is recognised as an opportunistic pathogen and its ability in resisting antibiotics. *P. aeruginosa* is also associated with serious illness such as cystic fibrosis, pneumonia urinary tract infection and sepsis (CDC, 2013).

2.2 Whole Genome Sequencing of *P. aeruginosa*

In 2000, first genome sequence of *P. aeruginosa* strain PAO1 was completed and this has provided new information on the genetic composition, ecological versatility, mutational mechanism, genetic adaptation, outbreak analysis and study of pathogenicity and antibiotic resistance of this organism (Marvig et al., 2015; Stover et al., 2000). Since then, many *P. aeruginosa* complete genomes are available from National Centre for Biotechnology Information (NCBI) databases and this has provided a means to perform comparative genomic analysis among *P. aeruginosa* strains.

P. aeruginosa has a genome size ranged from 6.0 to 7.5 mega base pair (Mbp) with approximate 65% Guanine + Cytosine (G + C) content. The genome of *P. aeruginosa* is among the largest genome in bacterial world as compared to *Bacillus subtilis,* 4.2 Mbp; *Synechocytis,* 3.6 Mbp; and *Mycobacterium* tuberculosis, 4.4 Mbp. This offers *P. aeruginosa* a large genetic capacity and increases survival rate to environmental changes (Meletis & Bagkeri, 2013; Stover et al., 2000). *P. aeruginosa* has a single and supercoiled

circular chromosome. It is reported that *P. aeruginosa* genome carries a lot of mobile genetic elements especially plasmids, which have significance to *P. aeruginosa's* lifestyle as a pathogen. Through the genome annotation, it is reported that *P. aeruginosa* has 5500 to 6500 open reading frames depending on types of strains (Valot et al., 2015).

2.3 Multidrug Resistance of *P. aeruginosa*

The large genome of *P. aeruginosa* has significant effect with addition of genes acquired from transferrable genetic element. This has contributed to its ability to develop resistance against a plethora of antibiotics. Generally, mechanism of antibiotic resistance in *P. aeruginosa* comprise of intrinsic and acquired resistance. Intrinsic resistance is referred to its ability to resistant antibiotics through genetically-encoded mechanism, while acquired resistance is referred to resistance to antibiotics gained from additional mechanism such as mutational changes and horizontal gene transfer (Meletis & Bagkeri, 2013).

2.3.1 Intrinsic Resistance of *P. aeruginosa*

Outer membranes of Gram negative bacteria work as natural barriers which can prevent large hydrophilic molecules to diffuse and pass through porin channel (Meletis & Bagkeri, 2013). This outer membrane has been targeted by aminoglycoside as its mode of action to kill *P. aeruginosa.* Another exclusive characteristic of *P. aeruginosa* is that it has several efflux pumps that can expel drug from out of its cell. These pumps consist of 3 proteins which are protein transporter, periplasmic connective protein and outer membrane porin. Usually, these 3 proteins are named with Mex followed by a letter (eg: MexB and MexD) and for outer membrane porin called Opr (eg: OprM).

Besides that, *P. aeruginosa* has the ability to produce AmpC beta-lactamase which can hydrolyse most beta-lactams based drug. This capability is shared with other bacteria

belong to the SPICE group (*Serrati* sp., *P. aeruginosa,* Indole positive *Proteus*, *Citrobacter* sp, and *Enterobacter* sp.) (Meletis & Bagkeri, 2013).

2.3.2 Acquired Resistance of *P. aeruginosa*

Acquired antibiotic resistance in *P. aeruginosa* is a consequence of insertion of genetic elements containing specific resistance genes. Beta-lactamase encoded genes might be transferred to from one strain to another strain through plasmids, transposons or integrons. This is the reason for various types of transferable beta-lactamase found in *P. aeruginosa*. Example of mutational changes event that contribute to resistance is the loss function of OprD. Mutational changes of OprD has led to mechanism of resistance to carbapenem (Meletis & Bagkeri, 2013). This occurrence is associated with efflux pump and over-expression of AmpC (Cabot et al., 2016; Campana et al., 2017; Lambert, 2002; Morita et al., 2013; Oliver et al., 2015).

2.4 Virulence Factors of *P. aeruginosa*

Various virulence factors have been reported to play important roles in pathogenicity of *P. aeruginosa.* These reported virulence factors are mostly cell associated factors (type IV pili, flagella, lipopolysaccharide), exopolysaccharide alginate and secretion factors (toxin) (Gooderham & Hancock, 2009). Single polar flagellum and type IV pili-dependent are important factor involved in motility, colonization and surface attachment (Gooderham & Hancock, 2009). In chronic cystic fibrosis cases, lack of flagella in *P. aeruginosa* had been identified. Therefore, biofilm formation plays important role in surface attachment and cell colonization (Gooderham & Hancock, 2009). In secretion factor, Type III secretion system are important for mediated cytotoxic (exoenzyme) while Type II secretion responsible for secretion of toxin and another virulence-related enzyme such as alkaline proteases, elastases and protease IV (Driscoll et al., 2007; Gooderham & Hancock, 2009). In respiratory tract, phenazine secretion cause ciliary dysfunction thus

resulting in proinflammatory and oxidative stress which damage host cell (Driscoll et al., 2007). Similarly to phenazine, exotoxin A causes damage to host cell by inhibit eukaryotic elongation factor 2 responsible for protein synthesis (Driscoll et al., 2007).

2.5 Mechanism of Pathogenesis

2.5.1 Motility, Attachment and Colonization

The flagella enable *P. aeruginosa* to move and facilitate cell surface interaction between host cells. There are multiple types of IV pili available at the cell surface of *P. aeruginosa*, which play roles in adherence to the host cells (Driscoll et al., 2007). For example, in respiratory tract, *P. aeruginosa* binds to glycolipid asialo ganglioside (aGM1) which is expressed during cell repair process of injured epithelial cell (Driscoll et al., 2007). Therefore, *P. aeruginosa* has been shown to adhere only to these injured cells.

P. aeruginosa starts to grow when it adheres to cell and produces polysaccharide alginate (associated with mucoid morphology on culture plate). This compound has a major role in delaying bacteria clearance by scavenging free radical released by macrophage, provides physical barrier, and inhibits neutrophil chemotaxis (Driscoll et al., 2007). Most important, alginate will lead to the formation of biofilm. In the making of biofilm, *P. aeruginosa* will secrete exopolysaccharide resulting in production of matrix which *P. aeruginosa* colonies are separated by water channel. From this matrix, *P. aeruginosa* freely detach from biofilm and may spread to another site. Biofilm formation have important role in pathogenesis of disease especially in urinary infection, infective endocarditis and chronic osteomyelitis (Parsek & Singh, 2003).

2.5.2 Toxin Secretion

When *P. aeruginosa* binds to cells, type III secretion system is activated and exoenzyme is injected into the cell resulting in altered immune response, cell injury and cell death (Driscoll et al., 2007). There are four exoenzymes reported which are ExoS,

ExoT, ExoU and ExoY, and among of these exoenzymes, ExoU is highly toxigenic. Type III secretion system has been related to high mortality in patient with pneumonia, respiratory failure and sepsis (Driscoll et al., 2007).

2.6 Pan Genome Analysis

Pan-genome is the total number of genes present in the genomes of a group of species or strains. Within pan-genome analysis, there are core genome, accessory or dispensable genome and species-specific or strains-specific genes. In the literature, these terms are named differently but most have similar concept and meaning. Comparative study of multiple genomes of bacteria has becomes one of the approaches in understanding bacteria evolution as well as to determine gene function and coding region sequence. In 2005, the comparative studies of bacteria genomes evolved whereas first pan-genome analysis was done on *Streptococcus agalactiae* (Tettelin et al., 2005). With the increasing numbers of genomes available from databases, this has created opportunities to investigate pan genome characteristic in one or more species. Pan genome analysis has extended the idea of comparative study in understanding bacterial interaction, niche adaptation and population structure (Guimaraes et al., 2015). The information obtained from these analyses has applications in vaccine and drug designs as well as assisting in virulence genes identification.

2.6.1 Core Genome

Core genome represents all the genes present in all species or strains while accessory genome represents genes that are partially shared. Core genome consists of important genes required by the species or strains involved in cellular processes for example, genes related to replication, translation, regulatory roles and housekeeping functions (Guimaraes et al., 2015; Mira et al., 2010). The numbers of core genomes differed in every species and this indicates the genetic diversity amongst the species. However, a

study by Lawrence et al. (2005) indicated that increase in number of strains used results in smaller number of core genomes and phylogenetically related organism which share same genes, have large core genome.

2.6.2 Accessory Genome

Accessory genome is a group of genes that are shared by some organisms but not all and are present in each of the studied organism. Usually, accessory genomes are linked to the key genes for survival in specific environment and mostly linked to adaptation, virulence, and antibiotic resistances (Mira et al., 2010). The variation of accessory genomes could be due to paraphyletic evolution, mutation and gene duplication (Guimaraes et al., 2015).

2.6.3 Species-Specific or Strain-Specific Genes

Species-specific or strain-specific genes refer to genes exclusively present only in a particular species or strain which usually are obtained by horizontal gene transfer among the species or strains (Guimaraes et al., 2015; Mira et al., 2010). According to Guimaraes et al. (2015), the presence of genes in certain organisms offers advantages over the strain which lack of those genes. Moreover, these genes have been linked to virulence or pathogenicity in pathogenic organisms (Jordan et al., 2001)

2.6.4 Pan genome of *P. aeruginosa*

Pan genome analysis of *P. aeruginosa* has been done by Hilker et al. (2015), Li et al. (2016), Mosquera-Rendón et al. (2016); Ozer et al. (2014) and Valot et al. (2015). Each of this study used different number of strains and population of *P. aeruginosa* thus give variety of results. Study by Hilker et al. (2015) reported that the core genome of 20 strains of *P. aeruginosa* should consists of more than 4000 genes, and total of accessory genomes is 10000 genes. In another study by Li et al. (2016) which is used 27 of complete sequence

of *P. aeruginosa*, core genome and accessory genome are 4805 and 2603 genes, respectively.

The used of 12 *P. aeruginosa* reference strains in Ozer et al. (2014) study show the core genomes of *P. aeruginosa* is 5844 genes while average accessory genomes of *P. aeruginosa* are 727 genes. Lastly, the number of core genomes reported by Valot et al. (2015) is 5233 genes which used 17 reference strains for *P. aeruginosa*. Through of all these studies, the number of core genome and accessory genomes are mainly affected by a few factors. Large number of *P. aeruginosa* used in alignment, different population of isolates and use of incomplete genome (draft genomes) would affected the number of core genomes and accessory genomes for *P. aeruginosa* (Subedi et al., 2018).

2.7 Horizontal Gene Transfer

Transfer of foreign DNA frequently occurs in which the DNA originated from one organism are incorporated into another bacteria genome. This process is known as horizontal gene transfer (HGT) (also known as lateral gene transfer (LGT) in some literature) while the DNA mobilized into the genome is called mobile genetic elements (MGEs). MGEs have influences on bacterial genome and is one of factors contributing to the different genome size amongst same species and is an important factor for bacterial pathogenicity. The MGEs regions are indicated by the differences in G+C percentage when plotted and differences in codon usage (Ochman et al., 2000). Most of identified MGEs often carry important genes for survival in hosts, for example antimicrobial resistant genes and toxin genes. Therefore, HGT has been remarkably responsible for the versatility in bacteria especially in pathogen, where the gain and loss of MGEs lead to the survival of bacteria in wide range of environments. Moreover, MGEs had been identified as one of roles in bacteria evolution. To date, there are a few types MGEs identified which are insertion sequences (IS), bacteriophage, phage, plasmid, and pathogenicity island

(PAI). All of these elements have different characteristics and mechanisms of transfer of DNA.

2.7.1 Mechanism of HGT

In order to achieve successful transfer of MGEs into host genome, there have 3 requirements. First, the DNA of donors must be delivered into the host. Second, the transfer of DNA must be incorporated into the host's genome and lastly, the incorporated DNA must be expressed by the host as a way to show successful transfer.

To date, there are three mechanisms of transferring MGEs are transformation, transduction and conjugation. Transformation is a process involving the uptake of naked DNA from the environment in which the DNA comes from the dead or lysed bacteria. This mechanism allows the DNA transmitted from distantly relative organisms, for example from *Bacillus subtilis* to *Streptococcus pneumoniae*, and could possibly happen between related species such as from *Haemophilus parainfluenzae* to *Haemophilus influenzae* (Davison, 1999; Dubnau, 1999).

In transduction mechanism, new genetic material is introduced into a host by bacterial viruses called bacteriophage or phages, after invading a host cell. The phage-encoded proteins facilitate the transfer of genetic material into cytoplasm or integrated into bacterial chromosome. Some phages carried virulence genes that can be expressed at high level when phages start replicating in host cell (Brussow et al., 2004). The size of DNA that can be transferred into the host are limited by size of phage, which reported ranging up to 100 kilobases (Kb) (Ochman et al., 2000). Unlike the transformation, this mechanism is dependent on the bacterial host's receptor to which bacteriophage will bind to.

Lastly, conjugation mechanism involves contact between recipient and donor cell where genetic material is transferred by self-transmissible plasmids (Ochman et al., 2000). Besides plasmid, the transposon encoded protein also could be transferred into recipient by conjugative bridge formed by donor. Through this mechanism, the introduction of genetic material into recipient's cytoplasm did not promise successful transfer in which the genetic material could be 'rejected' by the recipient. In a study conducted by Ricchetti et al. (1999), the integration of genetic material can be mediated by bacteriophage integrase or transposases, and incorporation through double-strand break repair process.

2.7.2 Transfer of Antibiotic Resistance Gene Through HGT

According to Ochman et al. (2000), one of the traits introduced through HGT is antibiotic resistance which allows the microorganism to survive and expand its ecological niche. As a consequence, antibiotic resistance genes have been associated with mobile genetic elements. Plasmid which is rarely incorporated into the recipient, is readily mobilised between species might contain antibiotic resistance genes. Transposable element with a flanked resistant determinant, could form complex transposon to promote the transfer of resistance genes. This has been reported by Chalmers et al. (2000) where two IS10 flanked a tetracycline resistant determinant to form Tn10. Similar to a study conducted by Goryshin and Reznikoff, (1998) in which two IS50 flanked together with three genes conferred antibiotic resistance to kanamycin, bleomycin and streptomycin thus forming Tn5 complex.

2.7.3 Virulence Gene Through HGT

Unlike the acquisition of antibiotic resistance genes, virulence genes present in pathogenic bacteria have been discovered through Pathogenicity Islands in its genome. Pathogenicity islands consist of large cluster of virulence genes which could transform

avirulent bacteria into a pathogen. This pathogenicity island has been found and situated at tRNA and tRNA-like loci which has been identified as common site for integration of foreign sequence (Ochman et al., 2000). Inside the pathogenicity island, there are short direct repeats which are similar to mobile genetic element, and open reading frames are highly similar to bacteriophage integrases (Ochman et al., 2000). These characteristics have been found by Inouye et al. (1991) which suggested that pathogenicity island insertion near tRNA genes in *E. coli,* are acquired through phages P4 and R73 through transduction event. Besides bacteriophage, virulence plasmid has also been found to have impact on acquired virulence genes. Most interesting, the addition of virulence plasmid and two chromosomally encoded regions, as well as two gene deletion in *E. coli*, resulted in successful genetic event in contributing virulence gene in *Shigella* (Maurelli et al., 1998; Nakata et al., 1993)

CHAPTER 3: METHODOLOGY

3.1 Strains background

Two clinical multidrug resistant *P. aeruginosa* strains, PAC08 and PAC17 were previously isolated from patients from one local hospital in Kuala Lumpur, Malaysia. PAC08 was isolated from a urine sample of female patient who had advanced cervical cancer while PAC17 was isolated from blood sample of a male patient who had myeloid leukaemia (Phoon et al., 2018).

3.2 Whole Genome Sequencing

3.2.1 Genome Sequencing, Assembly and Annotation

Next generation sequencing of two *P. aeruginosa* strains were previously performed by a commercial vendor using Illumina Miseq version 2. The WGS data was then retrieved and analysed using basic bioinformatics tools. The generated reads were trimmed using Trimmomatic software (downloadable at: http://www.usadellab.org/cms/?page=trimmomatic) to filter and remove low quality read below Q30 of phred score (Bolger et al., 2014). Sequencing data generated were assembled with A5 pipeline using *de novo* approach (downloadable at: https://sourceforge.net/projects/ngopt/) (Coil et al., 2015). Contigs generated were preceded for scaffolding using SSPACE software (downloadable at: https://www.baseclear.com/genomics/bioinformatics/basetools/SSPACE) (Boetzer et al., 2011) and gap found in the scaffold were closed using GapFiller software (downloadable at: https://www.baseclear.com/genomics/bioinformatics/basetools/gapfiller) (Boetzer & Pirovano, 2012). The assemblies accuracy of these scaffold were evaluated using Quality Assessment Tool for Genome Assemblies (QUAST) (http://quast.bioinf.spbau.ru/) (Gurevich et al., 2013) by assessing assembly metric (N50, contigs number, length of largest contigs).The annotation of both draft genome sequences were performed using

Rapid Annotation using Subsystem Technology (RAST) automated service (http://rast.nmpdr.org/) (Overbeek et al., 2014). Sequence type (ST) of PAC08 and PAC17 were confirmed using *in-silico* multilocus sequence typing (MLST). The draft genome sequences of both strains were submitted to *P. aeruginosa* MLST Database (https://pubmlst.org/paeruginosa/) to identify their ST (Jolley & Maiden, 2010).

3.2.2 Genome Analyses and Comparative Studies

Comparative genomics had been carried out using RAST server which compares the genomes of PAC08 and PAC17 with several selected complete genomes of *P. aeruginosa*. By using PAO1 as the reference genome, the genomes were aligned using progressive Mauve (Darling et al., 2010) and MauveAligner (Rissman et al., 2009) (downloadable at: http://darlinglab.org/mauve/mauve.html). Similarity of amino acid in PAC08 and PAC17 genomes were used to generate circular map of gene using BLAST Ring Image Generator (BRIG) (downloadable at: http://brig.sourceforge.net/) (Alikhan et al., 2011). Insertion sequence (IS) elements were analyzed using IS Finder server (https://www-is.biotoul.fr/) (Siguier et al., 2006) and insertion sequences found were annotated and identify using ISsaga (http://issaga.biotoul.fr/issaga_index.php) (Varani et al., 2011). The bacteria phage sequences were predicted and analyzed using web server Phage Search Tool (PHAST) (http://phast.wishartlab.com/) and intact phage region were manually annotated through BLAST (Zhou et al., 2011). Integron Finder (downloadable at: https://github.com/gem-pasteur/Integron_Finder) were used to search integrons and it components in *P. aeruginosa* genome (Cury et al., 2016). The integron sequences were annotated through Comprehensive Antibiotic Resistance database (CARD) web server (Jia et al., 2016).

3.2.3 Pan-Genome and Core Genome Analysis

Core- and pan-genome analysis were carried out using Bacteria Pan Genome Analysis (BPGA) pipeline comprises of Pan-genome profile analysis, Pan- and Core phylogeny construction, COG and KEGG analysis and Pan genome statistic (downloadable at: https://sourceforge.net/projects/bpgatool/) (Chaudhari et al., 2016). Coding sequences of 30 clinical isolate of *P aeruginosa* included PAC08 and PAC17 were predicted using Prodigal gene prediction (downloadable at: https://github.com/hyattpd/Prodigal) (Hyatt et al., 2010). The coding sequences were submitted to CD-HIT (http://weizhongli-lab.org/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit) using 70% sequence identity cut-off for protein clustering (Li & Godzik, 2006). With default setting of BPGA, pan genome and core genome were calculated based on the exponential growth and decay model for every addition of *P aeruginosa* genomes. COG and KEGG analysis of *P aeruginosa* were also implemented in this BPGA pipeline.

3.2.4 Genbank Accession Number

The *P. aeruginosa* PAC08 and PAC17 were deposited in Genbank under accession numbers MPCQ00000000 and MPCR00000000.1, respectively.

3.3 Virulence Genes Identification

To identify virulence factors, two sets of virulence factors databases which are core dataset and full dataset, were downloaded from Virulence Factor Database (VFDB) (http://www.mgc.ac.cn/VFs/download.htm) (Chen et al., 2005). These databases were set up in BLAST software on local computer and virulence factors were searched using blastn function. The hit virulence factor genes were selected based on e-value ($>1x10^{-30}$) and nucleotide identity (>90%).

3.4 Antimicrobial Resistance Genes Determination

For antibiotic resistance gene annotation, the genome sequences were submitted to CARD web server using Resistance Gene Identifier (RGI) software (https://card.mcmaster.ca/analyze/rgi) (Jia et al., 2016). The hit antibiotic resistance genes were reviewed and selected based on hit value. For acquired antimicrobial resistance gene identification, the draft genome sequences were submitted to ResFinder webserver (https://cge.cbs.dtu.dk/services/ResFinder) (Zankari et al., 2012).

3.5 Phylogenetic Analysis of *P. aeruginosa* Isolates

To infer relationship of *P. aeruginosa* from Malaysia with other *P. aeruginosa* strains from other places, 28 genomes of clinical isolates *P. aeruginosa* comprising of complete genomes and high-quality draft genomes were obtained from NCBI Nucleotide Database (https://www.ncbi.nlm.nih.gov/nuccore). These include 39016 (CM001020.1), 19BR (AFXJ01000001), 213BR (AFXK01000001), Carb01 63 (CP011317.1), F30658 (CP008857), IOMTU 133 (AP017302.1), N15-01092 (CP012901.1), NCGM1900 (AP014622), NCGM1984 (AP014646), NCGM2.S1 (NC_017549), NCGM257 (AP014651.1), PA_D1 (CP012585.1), PA_D16 (CP012581.1), PA_D2 (CP012578.1), PA_D21 (CP012582), PA_D22 (CP012583.1), PA_D25 (CP012584.1), PA_D5 (CP012579), PA_D9 (CP012580.1), PA1088 (CP015001.1), PA11803 (CP015003.1), PA38182 (HG530068.1), PA7 (NC_009656.1), PA7790 (CP014999.1), PA8281 (CP015002.1), S86968 (CP008865), UM-01 (LCWH00000000) and VRFPA04 (CP008739). From these genomic sequences, phylogenetic analysis based on 16S rRNA, 23S rRNA, MLST and whole genome SNP-base phylogeny were carried out. For *in-silico* MLST phylogenetic analysis, the sequence typing (ST) profile of 32 *P. aeruginosa* strains from Phoon et al. (2018) were collected and included in this analysis.

3.5.1 16S rRNA Phylogenetic Analysis

DNA sequences of 28 clinical isolates *P. aeruginosa* were submitted to RNAmmer 1.2 server (http://www.cbs.dtu.dk/services/RNAmmer/) to predict the 16S rRNA sequence (Lagesen et al., 2007). The predicted sequences of 16S rRNA were cumulated and aligned using MUSCLE in MEGA7 software (Edgar, 2004; Kumar et al., 2016). Hasegawa-Kishino-Yano substitution model with discrete Gamma distribution (+G) and invariable fraction site (+I), had been selected as best-fit nucleotide substitution model to infer phylogenetic tree. The 16S rRNA phylogenetic tree was constructed using Maximum Likelihood (ML) with 500 bootstraps.

3.5.2 23S rRNA Phylogenetic Analysis

Similar to 16S rRNA phylogenetic analysis, genomic of 28 clinical isolates *P. aeruginosa* were submitted to RNAmmer 1.2 web-server (http://www.cbs.dtu.dk/services/RNAmmer/) to predict the 23S rRNA sequence (Lagesen et al., 2007). The predicted sequences of 23S rRNA were cumulated and aligned using MUSCLE in MEGA7 software (Edgar, 2004; Kumar et al., 2016). Tamura-Nei substitution model with discrete Gamma distribution (+G) had been selected as best-fit nucleotide substitution parameter to draw phylogenetic tree. The 23S rRNA phylogenetic tree was constructed using ML algorithm with 500 bootstraps.

3.5.3 *in-silico* MLST Phylogenetic Analysis

Seven housekeeping genes of loci, *acsA, aroE, guaA, mutL, nuoD, ppsA* and *trpE* were identified. The sequences of each locus were downloaded from PubMLST Database (https://pubmlst.org/). Each locus was concatenated under Linux environment using script provided by PubMLST Database (available at: https://pubmlst.org/software/scripts/concat/). Concatenated genes resulted in 2882 base pair (bp) for each isolate were aligned in MEGA7 using MUSCLE (Edgar, 2004; Kumar

et al., 2016). Phylogenetic tree was constructed using ML algorithm with Tamura-3-parameter substitute model with discrete Gamma distribution (+G) and invariable fraction site (+I) and bootstrap replication values as 500. The tree was viewed and each branched were annotated using Interactive Tree of Life (iTOL) server (https://itol.embl.de) by adding allelic profile, sequence typing (ST) and country (Letunic & Bork, 2016).

3.5.4 Whole Genome SNP-Based Phylogenetic Analysis

Whole genome SNP-based phylogenetic analysis was performed using software kSNP3.0. (Gardner, 2015). The genomes sequences were pre-loaded into kSNP3.0 software to create a list of genomes. Kchooser script, which is part of kSNP3.0, was used to identify optimum k-mers size for 28 isolates. From the Kchooser report, it is suggested the optimum k-mers size for 28 isolates was 27. Therefore, kSNP k-mer size as 27 was used to identify number of SNP in the genomic region. Phylogenetic tree was constructed based on core-SNP using Parsimony approach. The tree was viewed and each branched were annotated using Interactive Tree of Life (iTOL) server (https://itol.embl.de) by adding country and year of isolates (Letunic & Bork, 2016).

CHAPTER 4: RESULTS

4.1 Genome Features of *P. aeruginosa*

The assembled genome of strains PAC08 and PAC17 revealed genome size of 6891589 and 6841669 base pairs, respectively and both strains have G+C content of approximately 66%. The *de novo* assembler generated 143 and 125 contigs with average coverage of 54X and 62X for PAC08 and PAC17, respectively. For PAC08, the largest contigs consists of 527348 bp and the length of N50 is 163,610 bp. Meanwhile for PAC17, the largest contigs consists of 493,308 bp and the length of N50 is 205,210 bp. When aligning to reference strain, PAO1, both genomes covered 94.94% and 94.49% for PAC08 and PAC17, respectively. The properties and the statistics of the both genome and a few reference strains are summarized in Table 4.1.

In RAST server, the genome annotations were described and distributed into subsystem categories. The annotation revealed that PAC17 has 6336 predicted coding sequence (CDS) and 66 RNA genes, while PAC08 yield 6445 CDS and 64 RNA genes. Most of genes are abundant in Amino acids and derivative subsystem, Carbohydrate subsystem, and Membrane transport subsystem. The subsystem coverage for PAC08 consists of 49% of 6445 CDS which contribute to 573 subsystems. Meanwhile for PAC17, the subsystem coverage is 50% from 6336 CDS which contribute to 569 of subsystem. Detail distribution of subsystem is illustrated in Figure 4.1 and Fig. 4.2.

Even though PAC08 and PAC17 were isolated from different location, there were not much difference in term of genomic structure such as genome size, number of CDS, and number of subsystems between both strains. However, similar genome characteristic can be seen in PAC08, PAC17, VRFPAO4 and NCGM2.S1 where all of these strains have high number CDS and large genome size which is approximately 6.8 Mbp compared to reference genome PAO1 which is 6.2 Mbp. On the other hand, the G+C content of

PAC08, PAC17 and NCGM2.S1 are almost similar which is 66.0%, 66.1% and 66.14%, respectively. Overall, through this comparison, multidrug resistant strains were identified to have larger genome size and high number of CDS.

Table 4.1: General genomic features of the whole genome sequence of the Malaysia *P. aeruginosa* PAC08 and PAC17. Other sequenced strains are included for comparison.

Strain	Geographical location	Genome size (bp)	No of contigs (fold-coverage)	Sequence type (ST)	G+C content (%)	No of CDS	Structural RNA	No of Subsystem	Genbank Accession no.	Reference
PAC08	Malaysia	6891589	143	1076	66.0	6445	64	573	MPCQ00000000	This study
PAC17	Malaysia	6841669	125	235	66.1	6336	66	569	MPCR00000000	This study
PAO1	Australia	6264404	1	549	66.6	5584	75	563	NC_002516	(Stover et al., 2000)
VRFPAO4	India	6818030	1	823	66.48	6402	96	558	NZ_CP008739	(Murugan, *et a;*, 2016)
NCGM2.S1	Japan	6764661	1	235	66.14	6140	79	570	NC_017549	(Miyoshi-Akiyama, et al., 2011)

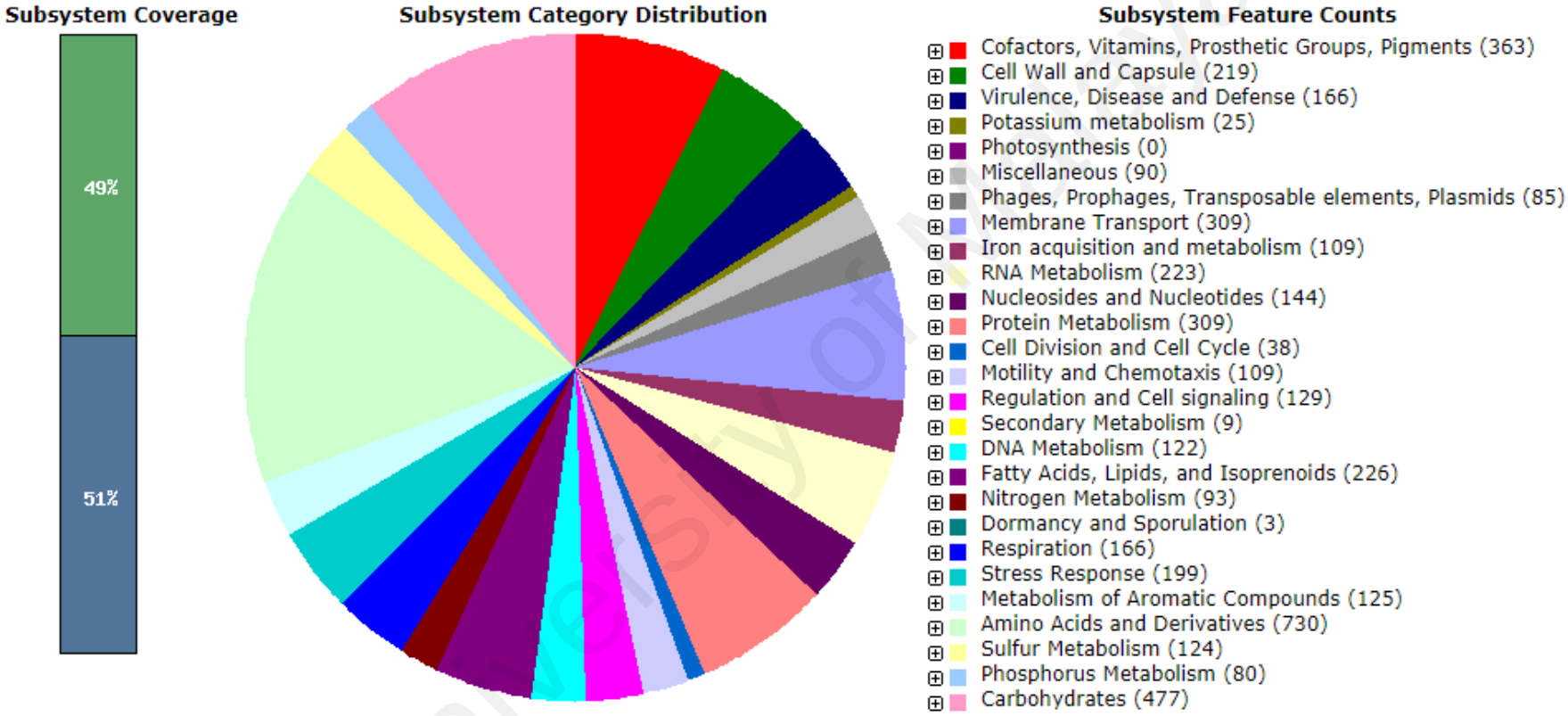


Figure 4.1: Subsystem distribution based on RAST annotation server for PAC08. The green bar of subsystem coverage indicates the percentage of CDS included in the subsystem, while the blue bar indicates the percentage of CDS not included in the system.

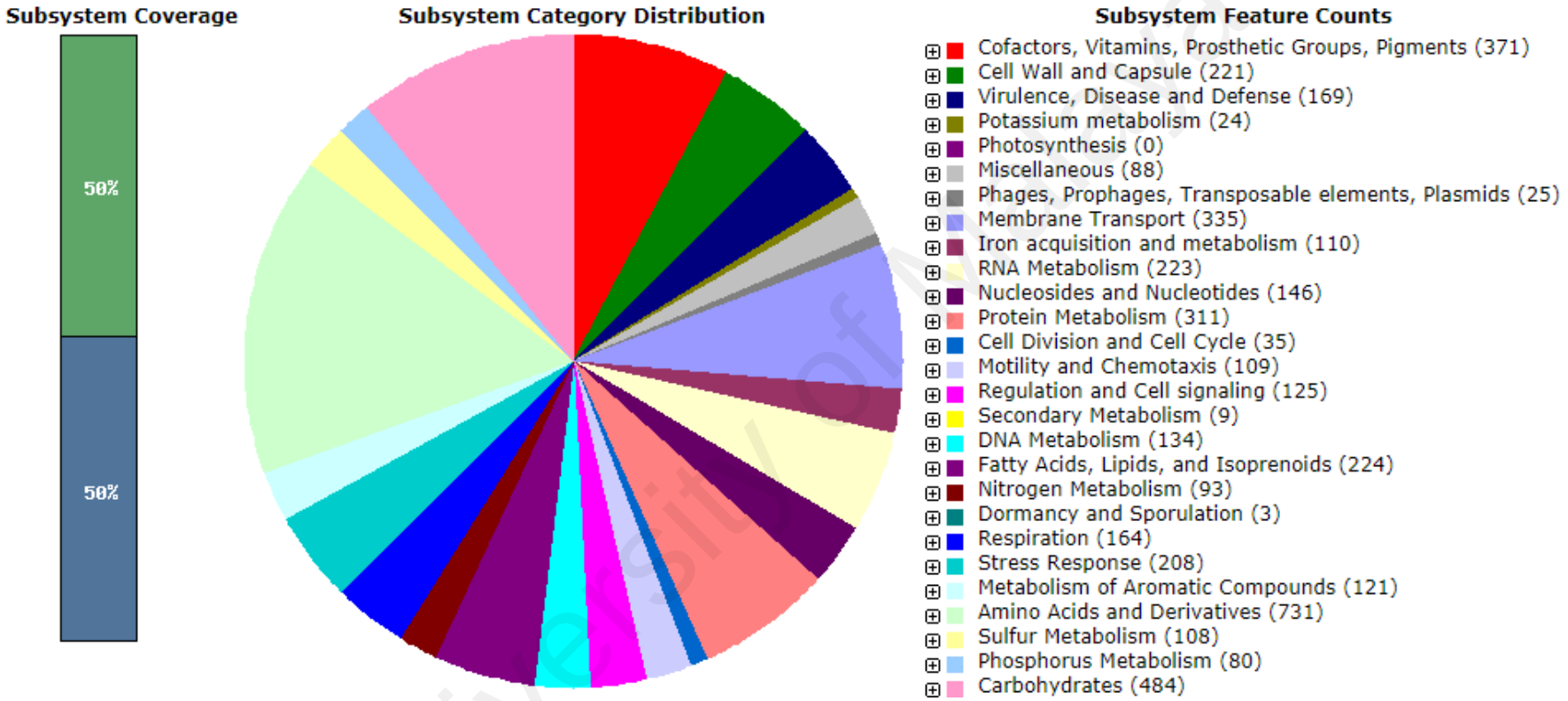


Figure 4.2: Subsystem distribution based on RAST annotation server for PAC17. The green bar of subsystem coverage indicates the percentage of CDS included in the subsystem, while the blue bar indicates the percentage of CDS not included in the system.

4.2 *In-silico* Multilocus Sequence Typing (MLST)

The sequence types (ST) of *P. aeruginosa* was analyzed using seven house-keeping genes which are *acsA* (acetyl coenzyme A synthetase), *aroE* (shikimate dehydrogenase), *guaA* (GMP synthetase), *mutL* (DNA repair protein), *nuoD* (NADH dehydrogenase I chain C,D), *ppsA* (phosphoenolpyruvate synthase) and *trpE* (anthralitessynthetase component I) (Curran et al., 2004; Jolley & Maiden, 2010). Results from *in silico* MLST analysis showed that that PAC08 and PAC17 were assigned to ST1076 (5-4-57-62-1-1-26) and ST235/ST2613 (38/172-11-3-13-1-2-4), respectively. Both strains did not belong to any clonal complex in the database. A comparison among STs of Malaysia strains was made and is summarized in Table 4.2.

From the comparison, the ST of PAC08 from urine, is similar to PAS2 (from wound), PAS3 (from urine), PAC36 (from urine) and PAS8 (from blood). A few strains from Ramanathan et al. (2017) (PAS1, PAS4, PAS5, PAS6, PAS7, PAS9 and PAS10) and Phoon et al. (2018) (PAC51 and PAC96) were found similar to ST of PAC17 even though these strains were isolated in 2009 and 2014. However, there are two ST numbers assigned for PAC17 which is either ST235 or ST2613 through PubMLST because two alleles of *acsA* were found in the genome. Such phenomenon was also observed in the reference strains NCGM2.S1 and other Malaysia strains such as PAS1, PAS4, PAS5, PAS6, PAS7, PAS9 and PAS10 that were uploaded in the NCBI database. From the MLST profiles of Malaysia strains (Table 4.2), ST235 and ST1076 representing PAC17 and PAC08, respectively, are the most prevalent and persistent genotypes from 2009 to 2014.

Table 4.2: Allelic profiles and sequence typing (ST) assigned to the clinical Malaysian *P. aeruginosa* strains according to the PubMLST protocol and MLST database.

			MLST Allelic profile								
Strains	Isolation Year	Isolation source	*acs*	*aro*	*gua*	*mut*	*nuo*	*pps*	*trp*	ST	Reference
PAS1	2009	Wound	38,172	11	3	13	1	2	4	235	(Ramanathan et al., 2017)
PAS4	2009	Urine	38,172	11	3	13	1	2	4	235	(Ramanathan et al., 2017)
PAS5	2009	Wound	38,172	11	3	13	1	2	4	235	(Ramanathan et al., 2017)
PAS6	2009	Urine	38,172	11	3	13	1	2	4	235	(Ramanathan et al., 2017)
PAS2	2009	Wound	5	4	57	62	1	1	26	1076	(Ramanathan et al., 2017)
PAS3	2009	Urine	5	4	57	62	1	1	26	1076	(Ramanathan et al., 2017)
PAS7	2010	Urine	38,172	11	3	13	1	2	4	235	(Ramanathan et al., 2017)
PAS9	2010	Wound	38,172	11	3	13	1	2	4	235	(Ramanathan et al., 2017)
PAS10	2010	Blood	38,172	11	3	13	1	2	4	235	(Ramanathan et al., 2017)
PAS8	2010	Blood	5	4	57	62	1	1	26	1076	(Ramanathan et al., 2017)
PAC28	2014	Blood	17	5	5	4	4	4	3	111	(Phoon et al., 2018)
PAC172	2014	Blood	47	4	5	33	1	6	40	207	(Phoon et al., 2018)
PAC51	2014	Tracheal Asp.	38	11	3	13	1	2	4	235	(Phoon et al., 2018)
PAC96	2014	Blood	38	11	3	13	1	2	4	235	(Phoon et al., 2018)
PAC17	2014	Blood	38,172	11	3	13	1	2	4	235	This study
PAC103	2014	Sputum	16	5	11	72	44	7	52	266	(Phoon et al., 2018)
PAC35	2014	Sputum	16	5	11	72	44	7	52	266	(Phoon et al., 2018)
PAC191	2014	Blood	23	5	11	7	1	12	7	274	(Phoon et al., 2018)
PAC165	2014	Blood	11	20	1	65	4	4	10	381	(Phoon et al., 2018)
PAC47	2014	Urine	11	20	1	65	4	4	10	381	(Phoon et al., 2018)

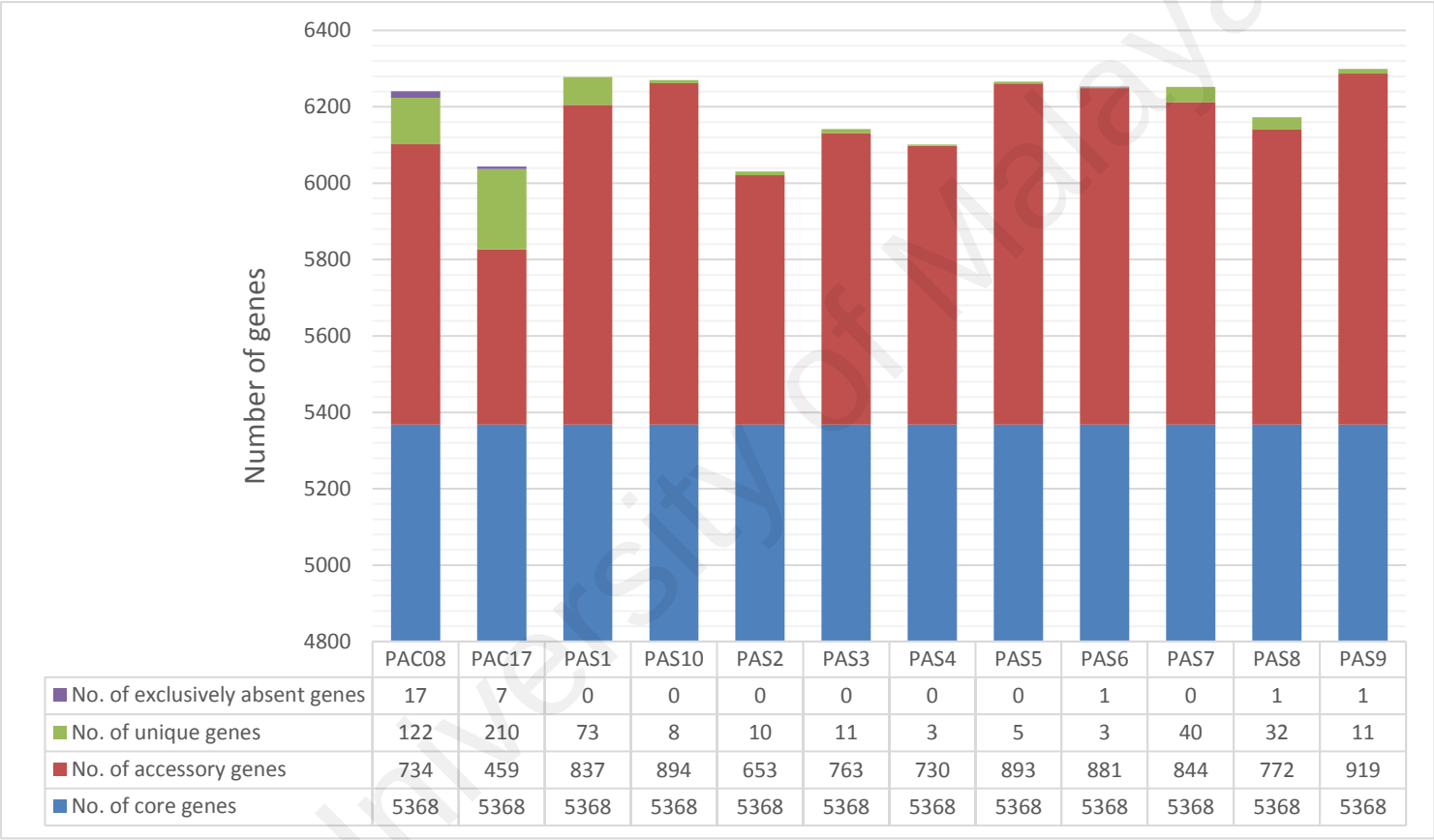
Table 4.2 continued.

Strains	Isolation Year	Isolation source	MLST Allelic profile								
			acs	*aro*	*gua*	*mut*	*nuo*	*pps*	*trp*	ST	Reference
PAC167	2014	Tracheal asp.	5	4	5	5	5	20	4	532	(Phoon et al., 2018)
PAC135	2014	BAL*	17	5	1	11	4	4	45	553	(Phoon et al., 2018)
PAC64	2014	BAL*	17	5	1	11	4	4	45	553	(Phoon et al., 2018)
PAC106	2014	Blood	11	3	11	3	1	4	60	708	(Phoon et al., 2018)
PAC107	2014	Tracheal Asp.	11	3	11	3	1	4	60	708	(Phoon et al., 2018)
PAC11	2014	Tissue	11	3	11	3	1	4	60	708	(Phoon et al., 2018)
PAC29	2014	Eye Swab	11	3	11	3	1	4	60	708	(Phoon et al., 2018)
PAC148	2014	Sputum	36	3	6	13	3	6	26	809	(Phoon et al., 2018)
PAC200	2014	Sputum	36	3	6	13	3	6	26	809	(Phoon et al., 2018)
PAC08	2014	Urine	5	4	57	62	1	1	26	1076	This study
PAC36	2014	Urine	5	4	57	62	1	1	26	1076	(Phoon et al., 2018)
PAC10	2014	Tissue	44	54	99	48	1	1	163	1400	(Phoon et al., 2018)
PAC90	2014	Sputum	16	10	11	85	4	4	10	1417	(Phoon et al., 2018)
PAC98	2014	Tracheal Asp.	16	10	11	85	4	4	10	1417	(Phoon et al., 2018)
PAC199	2014	Urine	15	5	30	72	3	6	68	2033	(Phoon et al., 2018)
PAC30	2014	Sputum	134	8	57	27	1	6	3	2329	(Phoon et al., 2018)
PAC54	2014	Blood	9	4	11	3	8	7	8	2335	(Phoon et al., 2018)
PAC60	2014	Pus Swab	30	202	11	4	4	4	7	2337	(Phoon et al., 2018)
PAC70	2014	Tracheal Asp.	16	5	1	54	58	7	19	2338	(Phoon et al., 2018)
PAC93	2014	Blood	11	5	11	5	1	6	2	2339	(Phoon et al., 2018)
PAC95	2014	Tissue	9	131	5	6	12	17	8	2340	(Phoon et al., 2018)
PAC108	2014	Blood	16	5	1	3	58	7	19	2341	(Phoon et al., 2018)

*BAL=Bronchoalveolar lavage

4.3 Pan-Genome and Core Genome Analysis

The analysis of Pan Genome and Core Genome through Bacterial Pan Genome Analysis (BPGA) automated tool based on 13 *P. aeruginosa* strains from Malaysia showed that the *P. aeruginosa* genomes from Malaysia contained 6165 gene on the average and the core genomes comprised of 5368 genes. The core genome is highly conserved average 87% of each of *P. aeruginosa* genome. The number of accessory genes in each of genomes ranging from 459 (PAC17) to 919 (PAS9) which comprise of 13% from total genes in a genome. Total number of unique genes found in this analysis varied in each strain. Only PAC08 and PAC17 have highest number of unique genes which is 122 and 210, respectively. The number of cores, accessory, unique and exclusively absence genes in each of the genomes were showed in Figure 4.3.



	PAC08	PAC17	PAS1	PAS10	PAS2	PAS3	PAS4	PAS5	PAS6	PAS7	PAS8	PAS9
No. of exclusively absent genes	17	7	0	0	0	0	0	0	1	0	1	1
No. of unique genes	122	210	73	8	10	11	3	5	3	40	32	11
No. of accessory genes	734	459	837	894	653	763	730	893	881	844	772	919
No. of core genes	5368	5368	5368	5368	5368	5368	5368	5368	5368	5368	5368	5368

Figure 4.3: Comparison of the number of cores, accessory and unique genes present in 12 strains of *P. aeruginosa* from Malaysia.

Through the power fit and exponential decay model (core-pan plot) as in Figure 4.4 (A), the changes of pan and core genes after addition of new *P. aeruginosa* can be exploited. Though pan genome graph (Figure 4.4 (A)), the expected size of gene repertoire was 7698 genes and will increase with addition of new genomes. Meanwhile for core genome, the extrapolation of exponential decay curve indicates that the core genome will reach minimum of 5223 genes with addition of new genomes and this probably will not change the core genome size significantly. In corresponding to power law regression model, ($f(x)=a \times x^{b,}$ where f(x) is expected number of genes for an increasing of genomes while a, is constant to fit the curve), the parameter b=0.085 estimate that *P. aeruginosa* pan genome is yet open but might be closed soon (Tettelin et al., 2008)

Functional classification of Cluster of Orthologous Group (COG) of *P. aeruginosa* gene repertoire showed that the core genes were mainly assigned to COG Category E (Amino acid transport and metabolism) and R (General function prediction only) whereas the accessory genes tend to be in Categories L (Replication, recombinant and repair), and K (Transcription). The unique genes for all strains show significant abundance which belongs to Categories R (General function prediction only), and L (Replication, recombinant and repair). The detail distribution of each COG functional categories amongst core, pan and unique gene in 13 strains were illustrated as in Figure 4.4 (C). Through classification of major COG categories as in Figure 4.4(C), the core genes of *P. aeruginosa* are abundant in Metabolism categories while accessory genes are abundant in Information storage and processing categories.

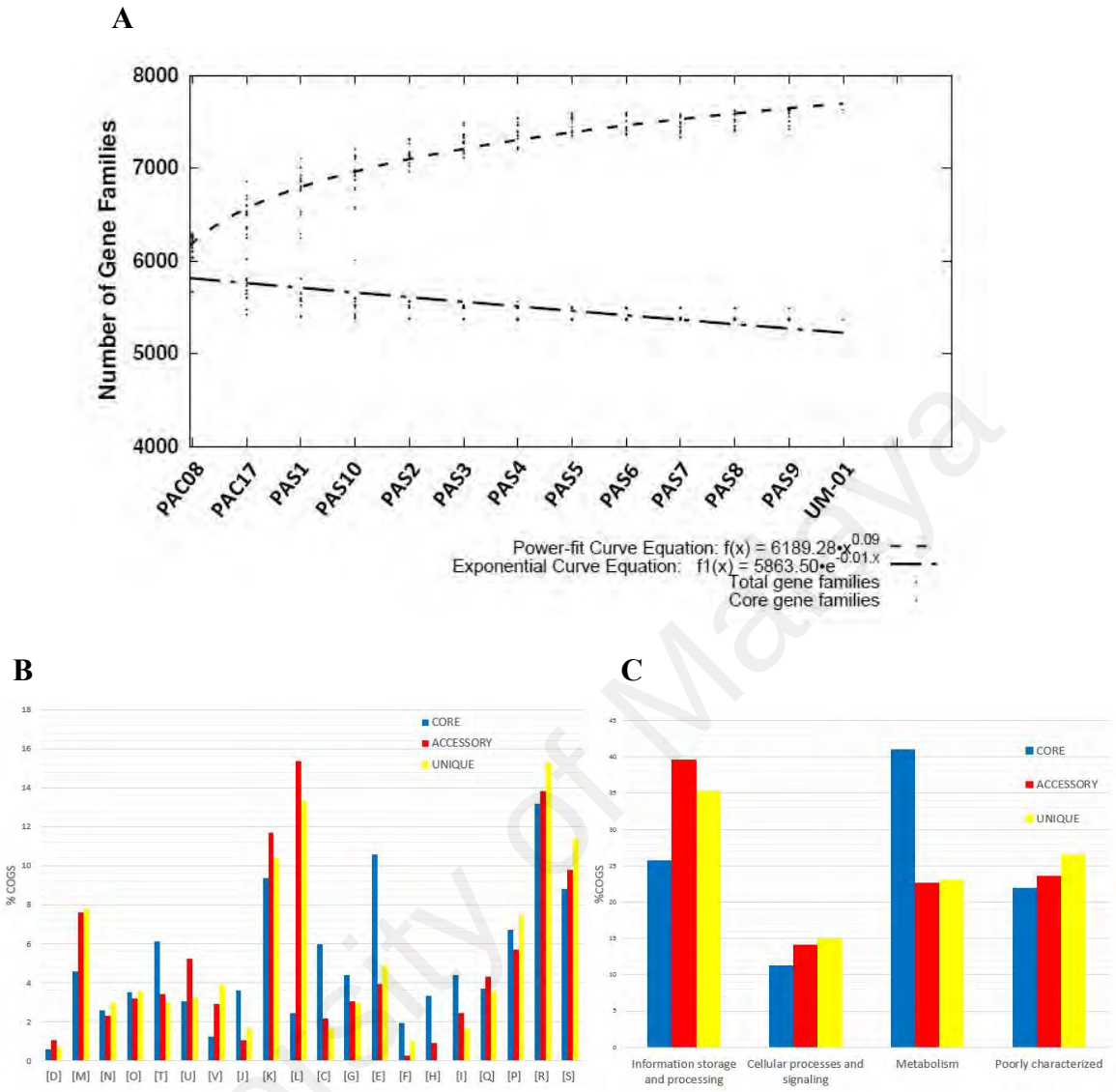
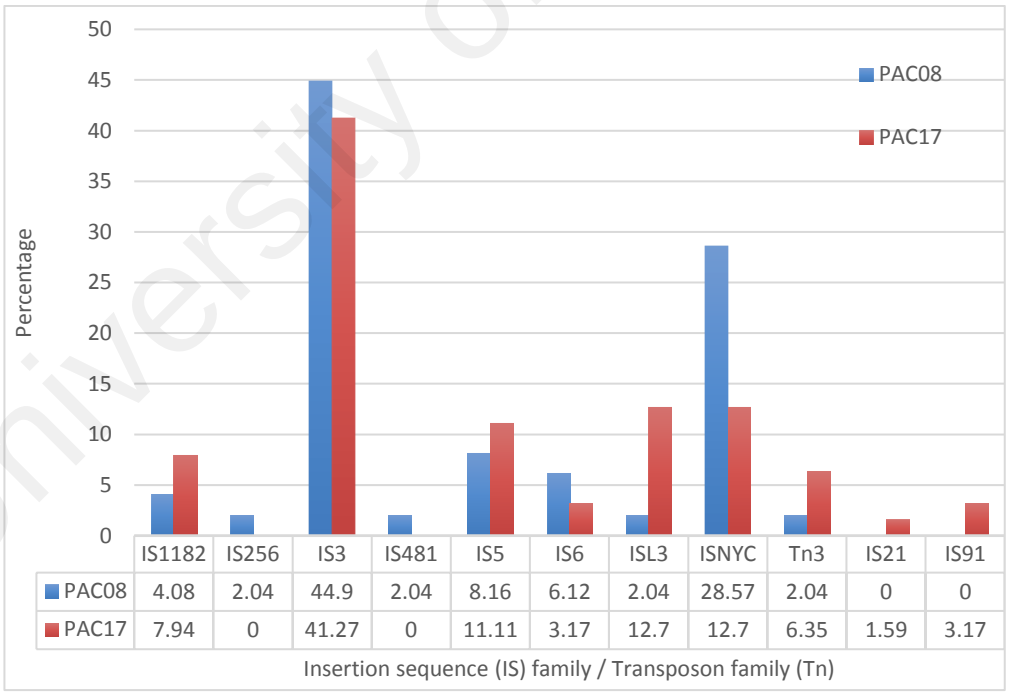


Figure 4.4: Graphical representation of core and pan genome from 13 strains of multidrug resistant *P. aeruginosa* from Malaysia. (A) Total number of pan genome and core genome according to the number of genomes sequentially added. (B) Distribution of COG functional categories amongst core, pan and unique gene in 13 isolates of multidrug resistant *P. aeruginosa* (D, Cell cycle control, cell division, chromosome partitioning; M, Cell wall/membrane/envelope biogenesis; N, Cell motility; O, Post-translational modification, protein turnover, chaperone functions; T, Signal Transduction; U, Intracellular trafficking and secretion; V, Defense mechanisms; K, Transcription; L, Replication, recombination and repair, C, Energy production and conversion; G, Carbohydrate transport and metabolism; E, Amino acid transport and metabolism; F, Nucleotide transport and metabolism; H, Coenzyme transport and metabolism; I, Lipid transport and metabolism; Q, Secondary metabolites biosynthesis, transport, and catabolism; P, Inorganic ion transport and metabolism; R, General function prediction only; S, Function unknown). (C) Classification of genes of multidrug resistant *P. aeruginosa* based on COG categories.

4.4 Mobile Genetic Elements

4.4.1 Insertion Sequence (IS)

From the ISFinder webserver, a total of 49 and 67 IS elements were identified in PAC08 and PAC17, respectively. A majority of IS elements found in PAC08 were members of the IS3 and ISNYC representing 44.9% and 28.57%, respectively. For PAC17, IS elements members from IS3 had high percentage (41.27%), indicating that IS3 is the predominant IS. A summary of percentage of IS elements found in PAC08 and PAC17 is illustrated in Figure 4.5 A total of 55 open reading frames (ORF) were found in PAC08 which are mostly transposes sequences. However, amongst of these numbers of ORF, only 5 ORFs IS elements are incomplete. The number of ORF in PAC17 was higher compared to PAC08. Sixty-seven ORFs were found in these IS element sequences and annotation of these ORF resulted in transposes sequences.



	IS1182	IS256	IS3	IS481	IS5	IS6	ISL3	ISNYC	Tn3	IS21	IS91
PAC08	4.08	2.04	44.9	2.04	8.16	6.12	2.04	28.57	2.04	0	0
PAC17	7.94	0	41.27	0	11.11	3.17	12.7	12.7	6.35	1.59	3.17

Figure 4.5: Percentage of insertion element in PAC08 and PAC17 based on insertion sequence family and Transposon family.

4.4.2 Bacteriophages

Based on the PHAST analysis, phages sequences ranging from 10.8 kb to 57.3 kb length were predicted in both PAC08 and PAC17. In PAC08, a total of 10 phage regions have been identified in which 3 regions are intact, 6 regions are incomplete and 1 region is questionable. For PAC17, there are only 2 phage regions in which 1 region is intact and another one is incomplete. The predicted phage sequences were searched through BLAST for sequence identity and annotation. Identity of 10 phage regions from PAC08 were from *Pseudomonas* phages while one of phage from PAC17 was from *Rastolnia* phage. Amongst these phage sequences, the Region 3 in PAC08 which was annotated as *Pseudomonas* phage phiCTX, showed a high possibility of the presence of complete sequence of *Pseudomonas* phage phiCTX with highest annotation identity (75%) and phage identity (95%). Moreover, the length of Region 3 is also similar to *Pseudomonas* phage phiCTX. Through the annotation, Region 3 contains 48 ORF where most of the ORF were identical to *Pseudomonas* phage phiCTX even though *Pseudomonas* phage phiCTX only has 47 ORF (Nakayama et al., 1999).

Detail analysis on open reading frames (ORF) in all predicted 10 prophage regions contained mainly hypothetical proteins. Besides hypothetical proteins, other ORFs predicted were integrase, transposase, terminase and protease. Multiple phage structural proteins such as tail and capsid were also detected. For PAC17, a total of 35 ORF were found in predicted prophage regions and a majority of these ORFs were hypothetical protein. Summary of identity phage regions in PAC08 and PAC17 is tabulated as in Table 4.3.

Table 4.3: Phage associated regions in PAC08 and PAC17 identified by PHAST.

Strain	Region	Concatenated Boundaries (bp)[1]	Size (kb)	PHAST Annotation, Accession No	Annotation Identity (Phage Identity)[2]	PHAST Prediction
PAC08	Region 1	1163972-1181784	17.8	*Pseudomonas* phage phi2, NC_030931	15% (99%)	incomplete
	Region 2	1407868-1451892	44	*Pseudomonas* phage phi3, NC_030940	3% (85%)	incomplete
	Region 3	**3235116-3270481**	**35.3**	***Pseudomonas* phage phiCTX, NC_003278.1**	**75% (96%)**	**intact**
	Region 4	**5073879-5119169**	**45.2**	***Pseudomonas* phage phi297, NC_016762.1**	**16% (94%)**	**intact**
	Region 5	6076978-6102904	25.9	*Pseudomonas* phage F10, NC_007805.1	4% (95%)	incomplete
	Region 6	**6468942-6508855**	**39.9**	***Pseudomonas* phage JD024, NC_024330.1**	**50%(96%)**	**intact**
	Region 7	6558620-6576302	17.6	*Pseudomonas* phage F10, NC_007805.1	26%(83%)	incomplete
	Region 8	6588370-6614209	25.8	*Pseudomonas* phage F10, NC_007805.1	33%(91%)	questionable
	Region 9	6702821-6710605	7.7	*Pseudomonas* phage JBD69, NC_030908.1	19%(96%)	incomplete
	Region 10	6757175-6783151	25.9	*Pseudomonas* phage MD8, NC_031091.1	10%(83%)	incomplete
PAC17	**Region 1**	**2724666-2743140**	**18.4**	***Pseudomonas* phage phi2, NC_030931**	**3% (72%)**	**intact**
	Region 2	6686336-6697256	10.9	*Ralstonia* phage p12J, NC_005131.2	4% (95%)	incomplete

[1]Identified regions are present in multiple contigs but were concatenated.
[2]Percentage in annotation identity and phage identity based on BLAST search result.

4.4.3 Integron

Integron prediction through Integon Finder involved detection of integron integrase gene, 59-base element recombinant site (*attC*), integron-associated recombinant site (*attI*), and open reading frame bounded by *attC* and *attI*. Between PAC08 and PAC17, Integron Finder had detected a complete class 1 integron in PAC08 consisting of integron integrase*, attC*, *attI*, promoter and a few open reading frames while in PAC17, only *attC* and a few open reading frames were found. These ORFs were submitted to Comprehensive antibiotic resistance database (CARD) for antibiotic resistance gene annotation. Analyst from CARD showed that the ORFs from PAC08 are *GES-13*, *AAC(6')-Ib10, emrE* and *ANT(3")-li-AAC(6')-IID fusion protein* which are responsible for β-lactamase and Aminoglycoside resistance. The combination of these ORFs with integron part formed a gene cassette in PAC08. Detail orientation of integron in PAC08 were illustrated as in Figure 4.6.

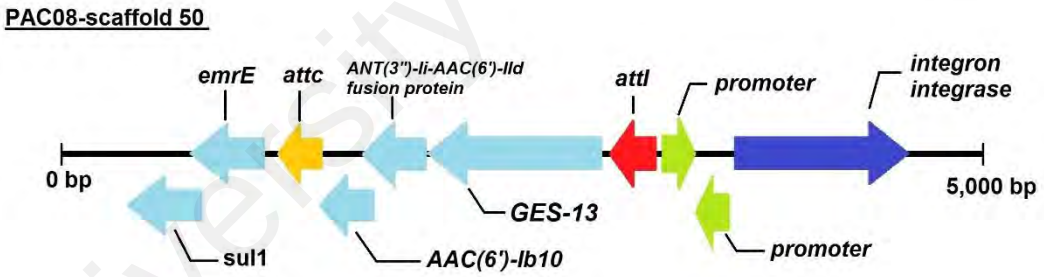


Figure 4.6: Gene cassette of complete Integron system in PAC08 consists of multiple antibiotic resistant gene, *attC* site, *attI* sites, two promoters and integron integrase. The *ACC(6')-Ib10* gene was overlapped with *attC* and another protein, *ANT(3")-li-AAC(6')-IID fusion protein.*

In PAC17, a total of 9 gene cassettes were found consisting of ORFs which contribute to antibiotic resistance. However, these gene cassettes are not complete without *attI* site and promoter. Since the integron integrase was not detected in this gene cassette, the class type of integron cannot be determined. The *attC* sites were found flanking at the ORFs.

Through annotation via CARD, the ORFs are genes responsible for β-lactamase, aminoglycoside, sulphonamide and chloramphenicol resistance. One interesting find was the gene encoded for carbapenem resistance *IMP-7* which is mostly harbored by ST235 as reported from Singapore and Japan study (Koh et al., 2010; Mano et al., 2015). Those ORFs which did not have any hit after annotation in CARD, were submit to BLAST server to find gene identity. A hypothetical protein was found in scaffold 30 through the BLAST, which belong to DUF1127 family. Detail of the gene cassettes found in PAC17 is illustrated as in Figure 4.7.

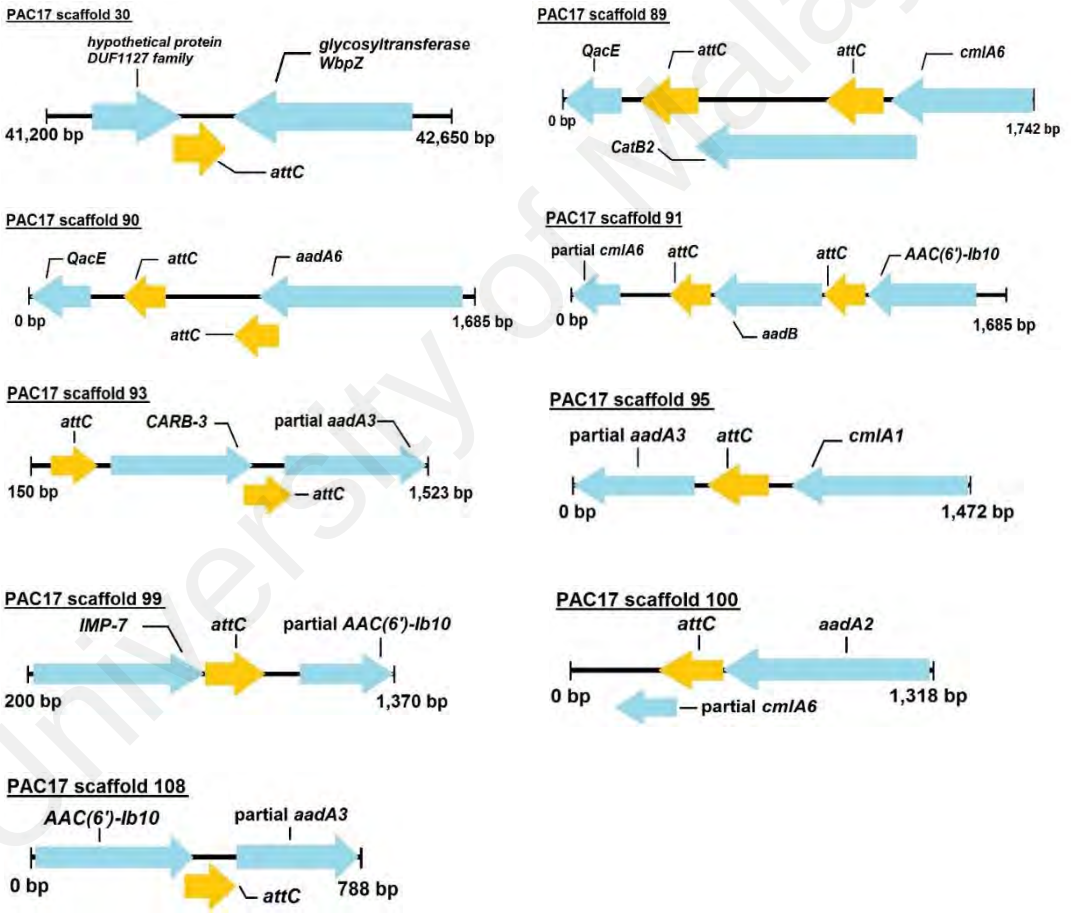


Figure 4.7: Gene coordination and location of 9 gene cassettes found in the PAC17 genome. Each of gene cassette contained different ORFs responsible for antibiotic resistance.

4.5 CRISPR/CAS System

CRISPR Finder web tool was used to identify CRISPR/Cas system in both strains. The CRISPR/Cas system was recognized by the difference in the direct repeat sequence previously studied by others researchers. In total, there are 8 CRISPR/Cas systems found in PAC08 and only 2 CRISPR/Cas systems had been found in PAC17. Amongst of these 10 CRISPR/Cas systems, only 3 were confirmed as CRISPRs (CRISPR 1, CRISPR 2 and CRISPR 3), which are found in PAC08, while the others were CRISPR-like regions (designated as questionable CRISPR by CRISPR Finder). CRISPR 1, CRISPR 2 and CRISPR 3 in PAC08 had 16, 7, and 5 different 32 bp length spacer, respectively. Interestingly, CRISPR 1 and CRISPR 3, and CRISPR 5 and CRISPR 7 have similar direct repeat sequence but found in different locations. The diversity of CRISPR/Cas system present in PAC08 and PAC17 is summarized in Table 4.4.

Table 4.4: Distribution of different CRISPR loci among PAC08 and PAC17 strains and their direct repeat sequence. Number of spacer and CRISPR length are included in the table.

No	Isolate	Direct Repeat Sequence	No of Spacer	Crispr Length (Bp)
CRISPR 1*	**PAC08**	**GTTCACTGCCGTATAGGCAGCTAAGAAA**	**16**	**987**
CRISPR 2*	**PAC08**	**TTTCTTAGCTGCCTACACGGCAGTGAAC**	**7**	**447**
CRISPR 3*	**PAC08**	**GTTCACTGCCGTATAGGCAGCTAAGAAA**	**5**	**328**
CRISPR 4	PAC08	TCATACCTTGCCCTCCAGTTCTTTGGCC	1	113
CRISPR 5	PAC08	AACGGGGTGCAACGGTTGCACCGG	1	85
CRISPR 6	PAC08	GCGGGCATTCGCTGCGGCCAGCTCGGC	1	110
CRISPR 7	PAC08	AACGGGGTGCAACGGTTGCACCGG	1	85
CRISPR 8	PAC08	GCTCGGGTAGGGCGAATAGCCGCTCGCGGCTATCCGCCGGGCGGGGTTCC	1	152
CRISPR 1	PAC17	TGCTTAGAAAACTAGAGTAGGGAAAATAAATCTGTCCCCTTT	1	139
CRISPR 2	PAC17	GGCCAAAGAACTGGAGGGCAAGGTATGA	1	113

*Confirmed CRISPR

4.6 Phylogenetic Analysis of *P. aeruginosa* based on 16S rRNA, 23S rRNA, *In silico* MLST And Whole Genome SNP.

The phylogenetic tree shown in Figure 4.8 is based on the 16S rRNA gene sequences extracted from the 38 *P. aeruginosa* genomes, and 2 other *Pseudomonas* species, *P. fluorescens* SBW25 and *P. putida* KT2440 were used as outgroups in this analysis. All *P. aeruginosa* strains including Malaysian strain, PAC08 and PAC17, were clustered close together in one clade because 16S rRNA gene is highly conserved. Therefore, the 16S rRNA tree could not discriminate these strains.

Figure 4.9 shows the phylogenetic tree based on 23S rRNA. It shows more variation that the species and strains were clustered according to sequence similarity, thus forming multiple clades. The outgroup, *P. fluorescens* SBW25 and *P. putida* KT2440 were well distinguished from the *P. aeruginosa*. PAC08 and PAC17 were clustered together with other *P. aeruginosa* strains from Malaysia. However, this tree did not show reliable relationship among the *P. aeruginosa* strains based on low support from the bootstrap value.

A clearer phylogenetic relationship amongst *P. aeruginosa* strains was displayed through MLST housekeeping genes phylogeny and whole genome single nucleotide polymorphism (SNP) phylogeny since the classification among the strains could not be determined through 16S rRNA gene and 23S rRNA gene. Figure 4.10 illustrates MLST tree integrated with strains, allelic profiles, ST and location of *P. aeruginosa* strains. All *P. aeruginosa* strains used in this analysis formed three distinct phylogenetic subclades labelled as A, B and C (as in Figure 4.10), and PA7 was outlier in *P. aeruginosa* strains as described by Roy et al. (2010). Based on the location of strains integrated into the tree, subclade A consists of strains from country in a different continent such as Canada, USA, Netherland, Brazil and even Malaysia strains were included in this tree. In subclade B, all

the Malaysian strains including PAC08 were grouped together in the same clade. Meanwhile, subclade C shows the strains from Asian countries such as Malaysia, Japan, India, China, and Nepal were clustered together and PAC17 was clustered with these strains. The bootstrap values which were high value (more than 70%) suggests reliability of this phylogenetic tree.

Through SNP phylogenetic tree as in Figure 4.11, the tree shows reliable result with support from bootstrap value and give information about phylogeny of PAC08 and PAC17 with other strains globally. In this tree, PAC08 was clustered with PAS2, PAS3 and PAS8 and this had supported the relationship as in MLST tree. For PAC17, this isolate was clustered with isolates from Japan, NCGM1900 and NCGM1984.

Strain	Year	Location
Pseudomonas fluorescens SBW25	2009	Finland
Pseudomonas putida KT2440	2002	n/a
Pseudomonas aeruginosa PAS10	2010	Malaysia
Pseudomonas aeruginosa PA D1	2014	China
Pseudomonas aeruginosa PA D2	2013	China
Pseudomonas aeruginosa PA D5	2014	China
Pseudomonas aeruginosa PA D9	2014	China
Pseudomonas aeruginosa PA D16	2014	China
Pseudomonas aeruginosa PA D21	2014	China
Pseudomonas aeruginosa PA D22	2014	China
Pseudomonas aeruginosa PA D25	2014	China
Pseudomonas aeruginosa PAS9	2010	Malaysia
Pseudomonas aeruginosa NCGM257	2004	Japan
Pseudomonas aeruginosa NCGM2.S1	2011	Japan
Pseudomonas aeruginosa NCGM1984	2012	Japan
Pseudomonas aeruginosa NCGM1900	2012	Japan
Pseudomonas aeruginosa IOMTU 133	2012	Nepal
Pseudomonas aeruginosa 19BR	2011	Canada
Pseudomonas aeruginosa 213BR	2011	Canada
Pseudomonas aeruginosa PA7	2007	Canada
Pseudomonas aeruginosa VRFPA04	2014	India
Pseudomonas aeruginosa PAC08	**2014**	**Malaysia**
Pseudomonas aeruginosa PAC17	**2014**	**Malaysia**
Pseudomonas aeruginosa Carb01 63	2015	Netherland
Pseudomonas aeruginosa F30658	2012	USA
Pseudomonas aeruginosa S86968	2012	USA
Pseudomonas aeruginosa N15-01092	2015	Canada
Pseudomonas aeruginosa 39016	2010	United Kingdom
Pseudomonas aeruginosa PA7790	2006	Brazil
Pseudomonas aeruginosa PA8281	2007	Brazil
Pseudomonas aeruginosa PA11803	2011	Brazil
Pseudomonas aeruginosa PA1088	1997	Brazil
Pseudomonas aeruginosa PA38182	2014	United Kingdom
Pseudomonas aeruginosa PAS1	2009	Malaysia
Pseudomonas aeruginosa PAS2	2009	Malaysia
Pseudomonas aeruginosa PAS4	2009	Malaysia
Pseudomonas aeruginosa PAS5	2009	Malaysia
Pseudomonas aeruginosa PAS6	2009	Malaysia
Pseudomonas aeruginosa PAS7	2010	Malaysia

Figure 4.8: Maximum-likehood phylogenetic tree based on 16S rRNA gene showing relationship between PAC08 and PAC17 with MDR strains from other regions. Bootstrap values are given as percentage to indicate the stability of the branching.

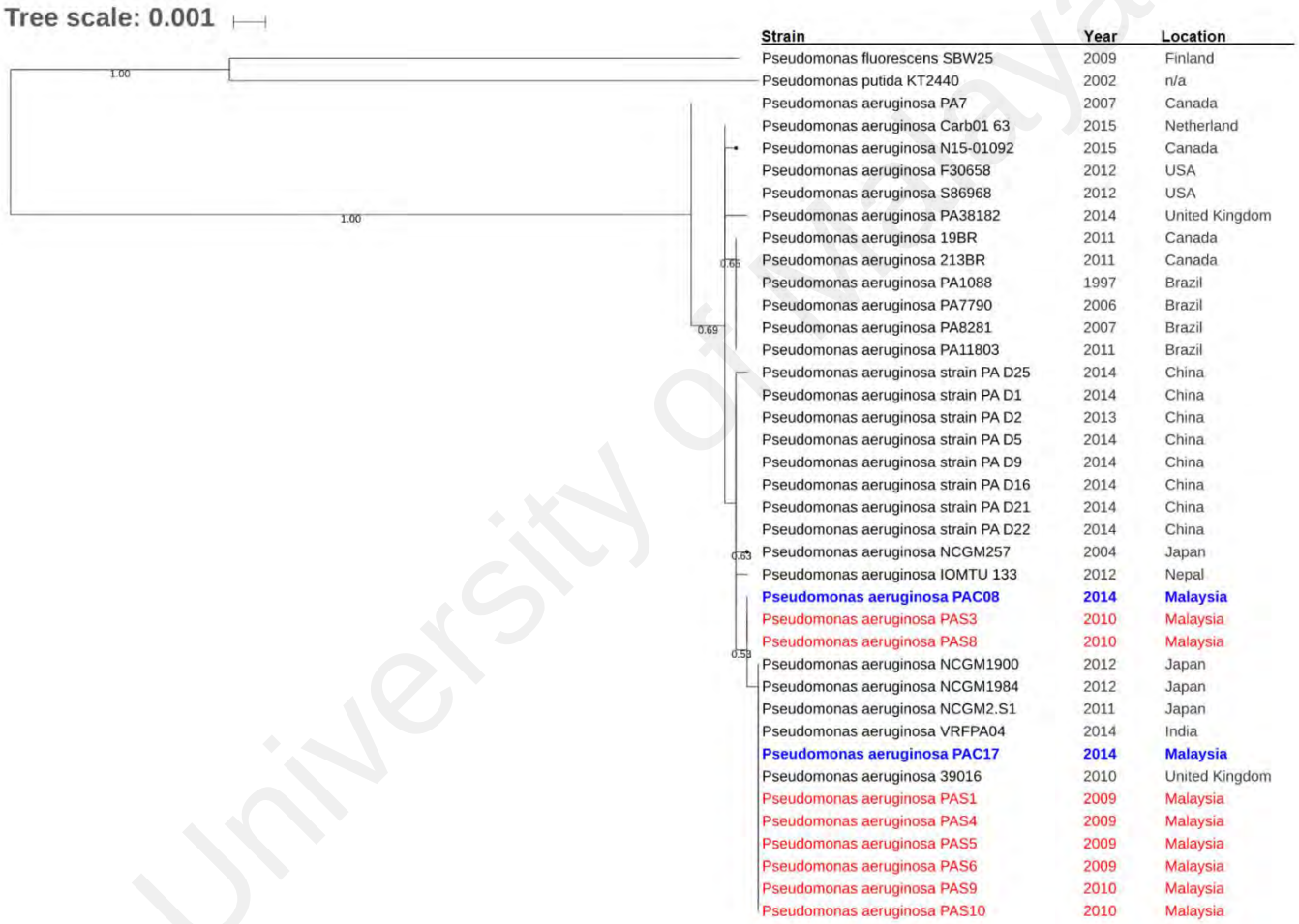


Figure 4.9: Maximum-likehood phylogenetic tree based on 23S rRNA gene showing relationship between PAC08 and PAC17 with MDR strains from other regions. Bootstrap values are given as percentage to indicate the stability of the branching.

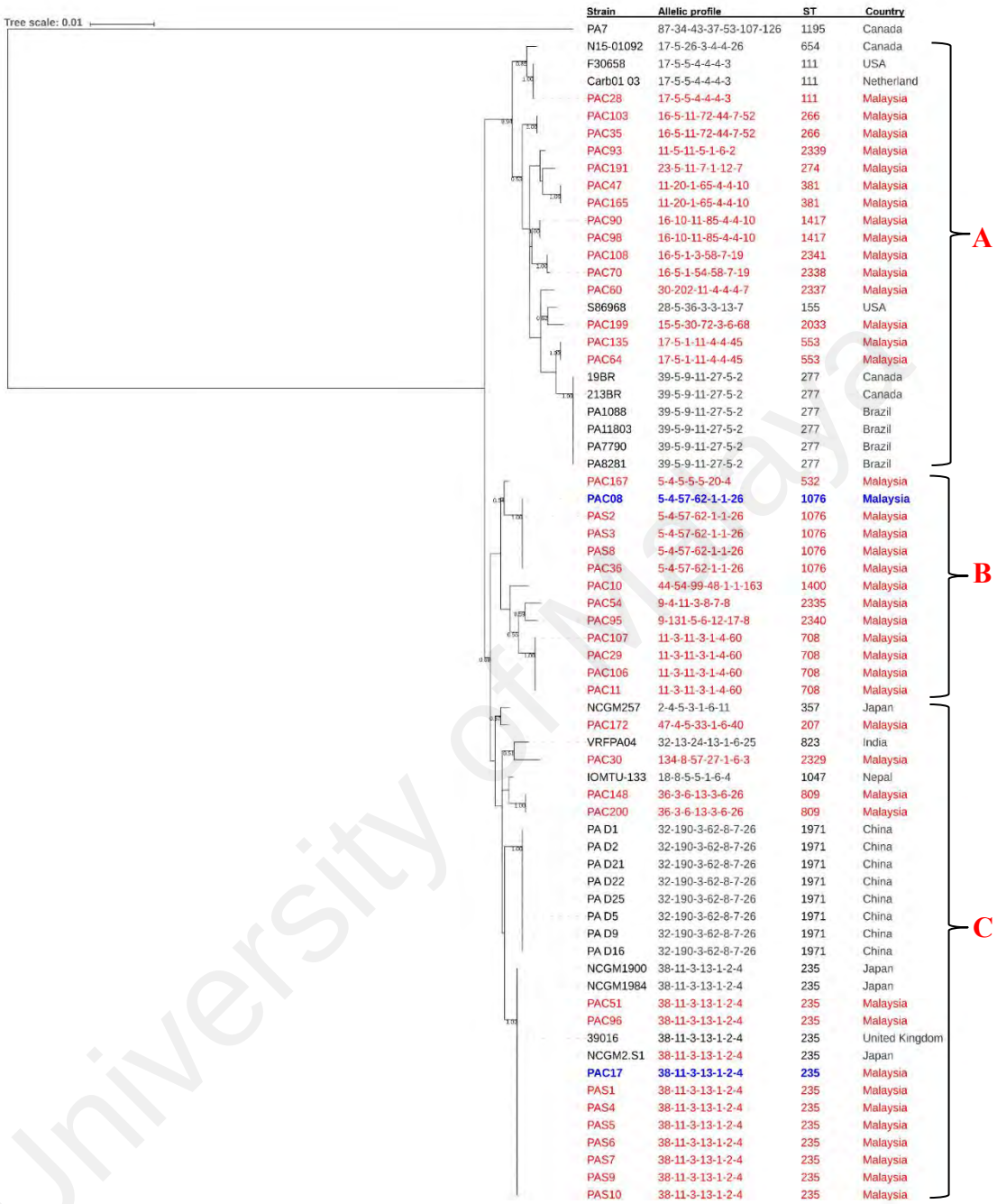


Figure 4.10: Maximum-likehood tree based on MLST showing relationship between PAC08 and PAC17 with clinical strains isolated from other regions. Bootstrap values are given as percentage to indicate the stability of the branching.

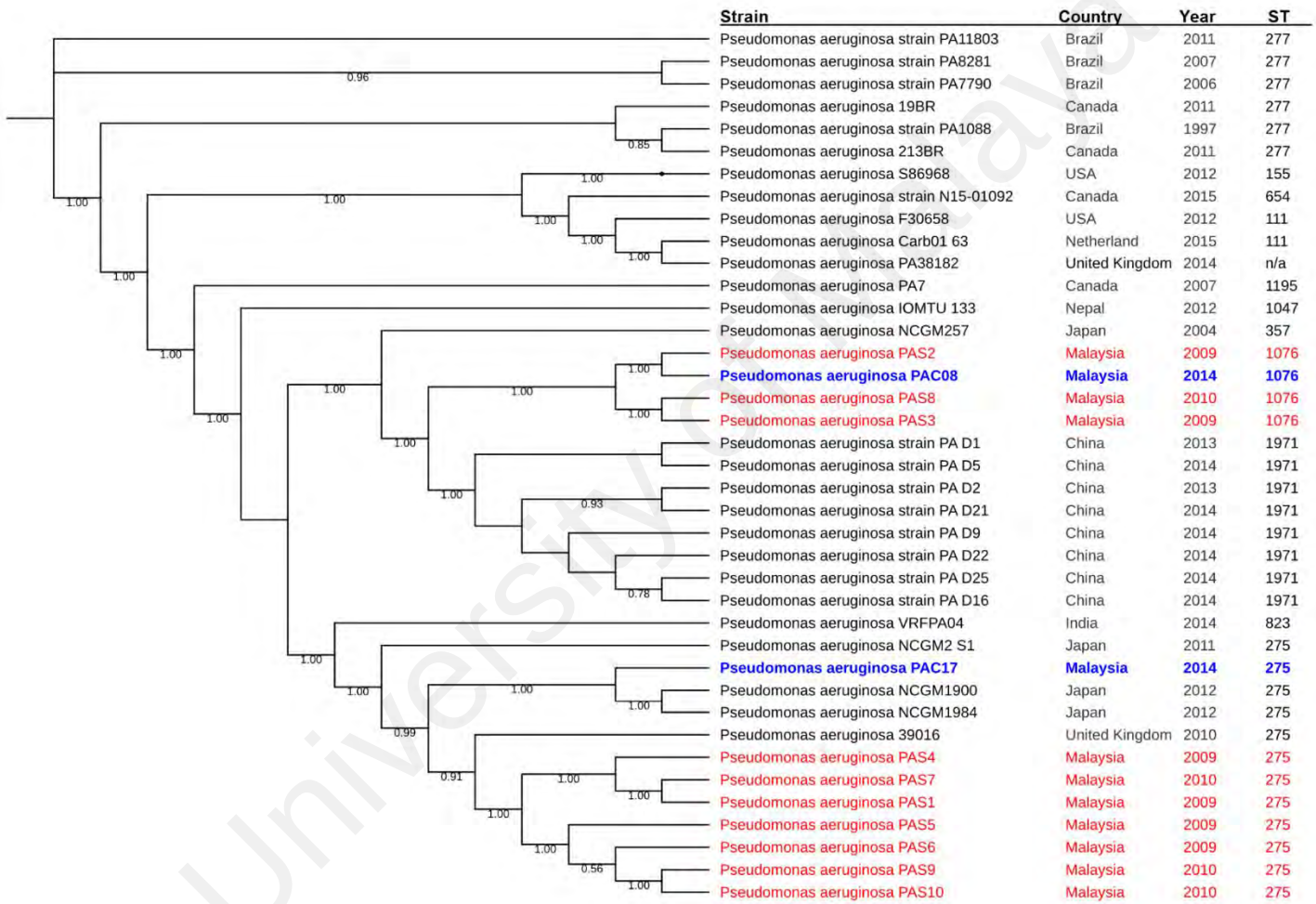


Figure 4.11: Maximum-likehood tree based on full genome SNP show relationship between PAC08 and PAC17 with MDR strains from other regions. Bootstrap values are given as percentage to indicate the stability of the branching

4.7 Virulence Factors in *P. aeruginosa*

In Virulence Factor Database (VFDB), 243 virulence factors were found in *P. aeruginosa* reference strains, which are PAO1, PA7, LESB58 and PA14. These virulence factors of *P. aeruginosa* were classified into several groups which are adherence, antimicrobial activity, anti-phagocytosis, biosurfactant, iron uptake, lipase, protease, quorum sensing system, regulation, secretion system and toxins. All genes in each group were present in PAC08 and PAC17 except gene for LPS O-antigen and pyoverdine. The number of genes related to LPS O-antigen are differed in PAC08 (8 gene) and PAC17 (10 gene). Genes for Type III secretion system (TTSS) translocated effector, exoenzyme *ExoT, ExoU* and *ExoY* were present with 99% identity with reference PAO1 in both strains. However, *ExoS* found in PAC08 and PAC17, was present with 79% identity with reference PAO1.

Table 4.5: Summary of all virulence genes detected in draft genome PAC08 and PAC17

Virulence factors	No of related genes	PAC 08	PAC 17
Adherence			
Flagella	53	53	53
Fimbriae	22	22	22
LPS O-antigen	-[1]	8	10
Type IV pili biosynthesis	24	24	24
Type IV pili twitching motility related proteins	10	10	10
Type IV pili	16	16	16
Antimicrobial activity			
Phenazines biosynthesis	17	17	17
Antiphagocytosis			
Alginate biosynthesis	14	14	14
Alginate regulation	12	12	12
Biosurfactant			
Rhamnolipid biosynthesis	3	3	3
Iron uptake			
Pyochelin	13	13	13
Pyochelin receptor	1	1	1
Pyoverdine	25	20	25
Pyoverdine receptors	1	1	1
Lipase			
Hemolytic phospholipase C	1	1	1
Non-hemolytic phospholipase C	1	1	1
Phospholipase C	1	1	1
Phospholipase D	1	1	1
Protease			
Alkaline protease	1	1	1
Elastase	2	2	2
Protease IV	1	1	1
Hypothetical protein (LepA)	1	1	1
Quorum sensing systems			
Acylhomoserine lactone synthase	1	1	1
N-(3-oxo-dodecanoyl)-L-homoserine lactone QS system	2	2	2
N-(butanoyl)-L-homoserine lactone QS system	2	2	2

Table 4.5, continued.

Virulence factors	No of related genes	PAC 08	PAC 17
Regulation			
GacS/GacA two-component system	2	2	2
Secretion system			
Hcp secretion island-1 encoded type VI secretion system (H-T6SS)	21	21	21
Hcp secretion island-I (HSI-I)	17	17	17
Hcp secretion island-II (HSI-II)	17	17	17
Hcp secretion island-III (HSI-III)	17	17	17
P. aeruginosa TTSS	36	36	36
P. aeruginosa TTSS translocated effectors	4	4	4
Toxin			
Exototoxin-A (ETA)	1	1	1
Hydrogen cyanide production	3	3	3

[1]The number for LPS O-antigen genes are varies

4.8 Antibiotic Resistance

The identification of antibiotic resistance genes through Comprehensive Antibiotic Resistance Database (CARD) detected many genes associated with antimicrobial resistance. CARD had identified complete *P. aeruginosa* multidrug efflux component in both PAC08 and PAC17. As a comparison, reference strains PAO1, VRFPA04 and NCGM2.S1 were included in this study.

4.8.1 Efflux pumps

In both PAC08 and PAC17 strains, several numbers of resistance-nodulation-cell division (RND) efflux pumps which are *MexAB-OprM*, *MexCD-OprJ*, *MexEF-OprN*, *MexGHI-OpmD*, *MexJK-OprM*, *MexMN-OprM*, *MexPQ-OpmE*, *MexXY-OprM, MexVW-OprM* and *TriABC-OpmH* drug efflux system were present. However, *MdtABC-TolC* efflux pump was not complete whereas *MdtA* was not detected in PAC08 and PAC17. In addition, *emrE* which is a small multidrug resistant (SMR) family was present in both *P. aeruginosa* strains as well as reference strains

For major facilitator superfamily (MFS) efflux pumps, 4 components were detected in all strains which are *farAB*, *RosAB, floR* and *tetG*. Efflux pumps *farAB* and *RosAB* were incomplete with the absence of *farA* and *RosA*, respectively. Interestingly, *tetG* (encodes for tetracycline efflux protein) and *floR* (encodes for chloramphenicol exporter) were only found in PAC17. On the other hand, there are 4 efflux pumps from ATP-binding cassette (ABC) family which are *msrE*, *TaeA, msbA,* and *macAB-TolC*. All of these efflux pumps are present in PAC17 while PAC08 lacks of *msrE*. Besides the efflux pump system, the mutated regulatory component for multidrug efflux pump was also detected. A total of 5 mutant regulators were found and these mutants have significant effect for expression of multidrug efflux pump. The presence of multi-drug efflux pump component in PAC08 and PAC17 is summarized in Table 4.6.

Table 4.6: Predicted multi-drug efflux pump from CARD and the presence or absence of genes associated with antibiotic resistance

Predicted Efflux Pump component	Description	PAC08	PAC17	VRFPAO4	PAO1	NCGM2.S1
MexAB-OprM	RND drug efflux	+	+	+	+	+
MexCD-OprJ	RND drug efflux	+	+	+	+	+
MexEF-OprN	RND drug efflux	+	+	+	+	+
MexGHI-OpmD	RND drug efflux	+	+	+	+	+
MexJK-OprM	RND drug efflux	+	+	+	+	+
MexMN-OprM	RND drug efflux	+	+	+	+	+
MexPQ-OpmE	RND drug efflux	+	+	+	+	+
MexVW-OprM	RND drug efflux	+	+	+	+	+
MexXY-OprM	RND drug efflux	+	+	+	+	+
TriABC-OpmH	RND drug efflux	+	+	+	+	+
MdtABC-TolC	RND drug efflux	(+/-)	(+/-)	(+/-)	(+/-)	(+/-)
farAB	MFS drug efflux	(+/-)	(+/-)	(+/-)	(+/-)	(+/-)
RosAB	MFS drug efflux	(+/-)	(+/-)	(+/-)	(+/-)	(+/-)
tetG	MFS drug efflux	-	+	+	-	-
floR	MFS drug efflux	-	+	-	-	-
msrE	ABC drug efflux	-	+	-	-	-
TaeA	ABC drug efflux	+	+	+	+	+
msbA	ABC drug efflux	+	+	+	+	+
macAB-TolC	ABC drug efflux	(+/-)	(+/-)	(+/-)	(+/-)	(+/-)
P. aeruginosa emrE	SMR drug efflux	+	+	+	+	+
evgS	Regulator for *EvgA*	+	+	+	+	+
mexL	Repressor of *mexJK*	+	+	+	+	+
Mutant *adeL*	Regulator of *AdeFGH*	-	-	+	+	+
Mutant *mexS*	Supressor of MexT, Activator of *MexEF-OprN*	+	+	+	+	+
Mutant *nalC* variant S209R	Regulator of *MexAB-OprM*	+	+	+	-	+
Mutant *nalD*	Repressor of *MexAB-OprM*	-	+	-	-	-
Mutant *nfxB*	Repressor of *MexCD-OprJ*	+	+	+	+	+

Note: (+) = refers to present or complete, (-) = refers to absent, (+/-) refers to incomplete. RND=Resistance-Nodulation-Cell Division, SMR= Small Multidrug Resistance, MFS= Major Facilitator Superfamily, ABC= ATP-Binding Cassette

4.8.2 Resistance genes elements

Aminoglycoside genes: Several aminoglycoside resistance-encoding genes were identified in both strains. Both strains and reference strains contain chromosomal-encoded aminoglycoside phosphotransferase (*aph)(3')-IIb*. However, PAC17 contains more aminoglycoside resistance-encoding genes *APH(3'')-Ib* and *APH(6)-Id* which are located in plasmid, integrative conjugative element and transposon. Moreover, PAC17 has aminoglycoside acetyltransferase- *AAC(6')-Ib, ANT(2'')-Ia*, *aadA3* and *aadA6* which appeared to be contained within integrated integron.

β-lactamase genes*:* Both strains and references strains contain *OXA-50* which is mostly found in *P. aeruginosa*. This gene confers susceptibility to ampicillin, ticarcillin, moxalactam and meropenem. PAC08 has *GES-13* which was present in PAC17 and other strains. On the other hand, PAC17 has more genes conferring susceptibility to β-lactamase through the *CARB-3* and *IMP-7*. Two genes encoding extended-spectrum β-lactamase were found in both PAC08 and PAC17 which are *PDC*-7 and PDC-*2*, respectively.

Polymyxin genes: Colistin resistance-encoding genes appeared to form *PmrAB* complex in PAC17 while in PAC08 and reference strains, only *PmrB* is present. Besides that, *PmrC* and *PmrF* are present in all the strains.

Quinolone: PAC08 and PAC17 confer resistance to fluoroquinolone through amino acid changes (T83I) in the protein encoded by DNA gyrase subynit A (*gyraA*). This mutation is commonly found in *P. aeruginosa*.

Chloramphenicol: *catB7* is a gene responsible for Chloramphenicol resistance in *P. aeruginosa* and this gene was detected in all strains. However, extra c*atB2* gene in gene in the integron was found in PAC17.

In addition, there are other genes that encode resistance to fosfomycin, sulphonamide, macrolide, peptide and aminocoumarin antibiotic. Both PAC08 and PAC17 have *fosA* (confers fosfomycin resistance), *mprF* (confers peptide-antibiotic resistance), *sul1* (confers sulphonamide resistance), *alaS* and *cysB* (confer aminocoumarine resistant). For macrolide resistance-encoding gene, only PAC17 has *mphG* gene. A summary of all detectable resistance genesis is in Table 4.7.

The genome annotation through RAST revealed more resistance genes found in PAC08 and PAC17. Both strains possess genes which confer resistance to metals such as zinc, arsenic, copper, cobalt, cadmium and chromium. PAC08 exhibited extra resistance genes such as mercury resistance operon as compared to PAC17.

Table 4.7: Genes associated with antibiotic resistance present in the Malaysian and reference strains of *P. aeruginosa*

Resistance Mechanism	Drug class	Gene	PAC08	PAC17	VRFPAO4	PAO1	NCGM2.S1
Antibiotic inactivation enzyme	Aminoglycoside	*APH(3')-IIb*	+	+	+	+	+
		APH(3')-Ib	-	-	+	-	-
		APH(3')-VIa	-	-	+	-	-
		APH(3'')-Ib	-	+	-	-	-
		APH(6)-Id	-	+	+	-	-
		AAC(6')-Ib	-	+	-	-	-
		ANT(2'')-Ia	-	+	+	-	-
		aadA3	-	+	+	-	+
		aadA6	-	+	-	-	-
	Fosfomycin	*FosA*	+	+	-	+	+
	Macrolide	*mphG*	-	+	-	-	-
	Phenicol	*catB2*	-	+	-	-	-
		catB7	+	+	+	+	+
	β-lactamase	*CARB-3*	-	+	-	-	-
		GES-13	+	-	-	-	-
		OXA-50	+	+	+	+	+
		IMP-7	-	+	-	-	-
		PDC-1	-	-	-	+	-
		PDC-2	-	+	-	-	+
		PDC-7	+	-	+	-	-
Gene altering cell wall charge	Polymyxin	*arnA*	+	+	+	+	+
		PmrA	-	+	-	-	+
		PmrB	+	+	+	+	+
		PmrC	+	+	+	+	+
		PmrF	+	+	+	+	+
Antibiotic target replacement	Sulphonamide	*sul1*	+	+	+	-	+
Antibiotic target modifying enzyme	Peptide	*Brucella suis, mprF*	+	+	+	+	+
Mutation conferring antibiotic resistance	Fluoroquinolone	*Pseudomonas aeruginosa, gyrA* variant T83I	+	+	+	-	+
	Mupirocin	*Bifidobacteria intrinsic, ileS*	+	+	+	+	+
	Aminocoumarin	*aminocoumarin resistant, alaS*	+	+	+	+	+
		aminocoumarin resistant, cysB	+	+	+	+	+

Note: (+) = Presence, (-) = absence.

CHAPTER 5: DISCUSSION

5.1 Genome Assembly and Genomic Features

Generally, genome assembly from sequence data involves multiple steps, starting from pre-process of raw sequence data up to the generation scaffold. Pre-process steps comprised of adapter trimming, quality filtering and error correction. The cleaned raw sequence data was proceeded to creation of contigs, then generation of scaffolds and gaps closure. Usually, these workflows are carried out in multiple software independently in order to generate draft genome or finish genome. In this study, sequence data of PAC08 and PAC17 using the Illumina Miseq platform were comprehensively assembled before proceeded to downstream analysis. Scaffold generated must have good quality in order to have precise comparative analysis. From QUAST assessment, A5 Miseq assembler generated good quality draft genome compared to Velvet, SPADE and IDBA based on the assembly metric included size of genome, N50, contigs number, and length of largest contigs.

In the NCBI database, there are 2668 entries of *P. aeruginosa* genomes consisting of 131 complete genomes and 2538 draft genomes since the first publication of *P. aeruginosa* genome sequence in 2000 (stover-2000). Since then, the number of *P. aeruginosa* genome sequences increases year by year. Up to date, there are no finished sequence of *P. aeruginosa* genome from Malaysia or South East Asia. The term 'finished genome' has been defined by Koren and Phillippy (2015) in which a finished genome should not have any contig gap, error corrected, and confirmation of low coverage region through polymerase chain reaction. However, the process of genome completion can take months to years with excellent quality data and best bioinformatic software. With the latest technology, single molecule real time (SMRT) sequencing such as PacBio, could be an option for producing finished genome (Pacific Biosciences, 2013). It is important

to have finished *P. aeruginosa* genome because functional genomic studies demand high quality and complete genomes sequences as starting point (Fraser et al., 2002).

In this study, based on the genome assembly, the genome sizes of PAC08 (6.89 Mbp) and PAC17 (6.84 Mbp) were not different from the average genome size of *P. aeruginosa* in NCBI genome database which is from 6.0 to 7.5 Mbp. Compared to a selective set (use in this study) of MDR *P. aeruginosa* genome size (as in Appendix C), the genome size for PAC08 and PAC17 were almost similar to most of the genome size which the average genome size is 6.8 Mbp. The contigs generated from *de novo* assembly produced 143 contigs for PAC08 and 125 contigs for PAC17. Both strains have high numbers of contigs generated because of the presence of gap between the contig that could not be close due to lack of overlapping read. Besides that, the lost sequences during the library preparation and repetitive region in *P. aeruginosa* might be reasons for high number of contigs (Ekblom & Wolf, 2014).

Using clustering approach in pan genome analysis, the pan genome of Malaysia *P. aeruginosa* strains contained 7698 genes while core genomes contained 5368 which were highly conserved around 87% of average genomes. The average accessory genes found in the Malaysia strains was 781 genes comprise of 13% from total genes and total predicted accessory genes is 9379. The size of pan genome, core genome and accessory gene of *P. aeruginosa* have been reported by Hilker et al. (2015), Li et al. (2016), Mosquera-Rendón et al. (2016); Ozer et al. (2014) and Valot et al. (2015). The distribution of core genome of selective *P. aeruginosa* by Li et al. (2016), Ozer et al. (2014) and Valot et al. (2015) were almost similar to the *P. aeruginosa* from Malaysia where they estimated core genome for 4805, 5233 and 5316 genes, respectively. For accessory genes, total number of predicted genes demonstrated was approximately similar to study by Hilker et al. (2015), Mosquera-Rendón et al. (2016) and Ozer et al.

(2014). However, Mosquera-Rendón et al. (2016) prediction of core genome contrast with other studies including this study. Mosquera-Rendón et al. (2016) found 2503 genes of core genomes and data analysis indicated that *P. aeruginosa* pan genome is closed.

To date, the study of *P. aeruginosa* pan genomes is debatable because a few studies of *P. aeruginosa* pan genome produced different results. Mosquera-Rendón et al. (2016) study which is most contrary result amongst pan genome studies claimed the data produced represented most accurate result of *P. aeruginosa* pan genomes through analysis of 180 strains of *P. aeruginosa*. According to Rouli et al. (2015), it is difficult to determine number of genomes required for pangenome study. Another aspect affecting pan genome is the quality of genomic data itself whether draft genome or gapless chromosome are used for analysis. The use of gapless chromosomes had been demonstrated by Valot et al. (2015) where 17 of *P. aeruginosa* gapless chromosome from clinical and environment strains. However, the isolation location of strains will affect gene population of selected organism (Rouli et al., 2015).

Multiple bacteriophage sequences found in both the Malaysian *P. aeruginosa* strains were consistent with the other studies which have phage sequences in their *P. aeruginosa* genomes (Li et al., 2016; Murugan et al., 2016; Naughton et al., 2011; Roy et al., 2010). The annotation of these phage sequences showed that the genomes carry genes with unknown function or hypothetical proteins. This result is as expected as stated by Krylov (2014) where the clinical strains of *P. aeruginosa* contain different phages which carry many unknown genes. The presence of *Pseudomonas* phage phiCTX in PAC08 genome is one of factors to the increase of virulence in PAC08. *Pseudomonas* phage phiCTX is moderate phage from *P. aeruginosa* strain which is cytotoxin producer. Through the lysogenization of *Pseudomonas* phage phiCTX, the *P. aeruginosa* strains is converted

into cytotoxin producing strains which is more virulent than non-producing strains (Baltch et al., 1994; Hayashi et al., 1990; Nakayama et al., 1999).

Mobile genetic elements such as Insertion Sequence (IS) elements and integron also play another role in contributing adaptation and survival of *P. aeruginosa* in different environment through genes disruption and genome arrangement. Various IS element studies in *P. aeruginosa* had demonstrated that the transposition of IS element in *P. aeruginosa* genomes play roles in influencing antibiotic susceptibility. To date, ISPa16 (Wolter et al., 2008), ISPa8 (Fowler & Hanson, 2014), ISPa26 (Evans & Segal, 2007), ISPa133 (Ruiz-Martínez et al., 2011), ISPa46 (Diene et al., 2013), IS5 (Wang et al., 2010) and globally-found ISPa1328 (Al-Bayssari et al., 2015; Estepa et al., 2017; Wang et al., 2010) had been found disrupted *oprD* gene through its insertion into gene thus causing carbapenem resistance in *P. aeruginosa*. Another IS element, IS21 has been found to disrupt *MexR* repressor genes, therefore increasing the expression of *MexAB-OprM* efflux system. In this study, IS21 had been found in PAC17 which indicates that these IS elements might disrupt *MexR* gene since *MexR* gene is not detected in antibiotic resistance analysis. Both PAC08 and PAC17 have IS5 element where this IS element was found and might disrupt *oprD* gene, thus reducing susceptibility for imipenem, meropenem and doripenem (Fowler & Hanson, 2014; Wang et al., 2010). In addition, amongst the IS elements found in PAC08 and PA17, IS 3 family is the most abundant in genomes. This finding is in agreement with other study by Al-Nayyef et al. (2015) which found that high copies number of IS 3 family would have affect in genome rearrangement in *P. aeruginosa*.

Another mobile genetic elements which play an important role in dissemination of antibiotic resistance is integron. Multiple gene casssete of integron had been found in PAC08 and PAC17. However, only one class 1 integron in PAC08 is completed with the

presence of integrase and *attC* site. For PAC17, there are 9 incomplete integrons or known as cluster of *attC* sites lacking integron-integrase (CALIN) and the class of these integrons were not determined since the integron integrase was not detected. The finding of gene cassettes in this study was supported by another study conducted by Phoon et al. (2018) where PAC08 has *GES-13* in its class 1 integron and PAC17 has *aadA6* in its class 1 integron. These integrons have been found by Phoon et al. (2018) through analysis of sequencing data of PCR amplicons. This method used by Phoon et al. (2018) is limited as it only targeted genes of interest compared to method used in this study which use whole genome sequence (WGS) to determine integron. The application of using WGS to find integron has offered detail information including gene casette orientation and identity of flanking gene in integron. Nevertheless, the validation of gene casette through PCR amplification are needed so that the genes found are similar to what has been found in WGS analysis.

5.2 Genetic diversity of *P. aeruginosa* based on MLST

In this study, *in-silico* MLST analysis reveal sequence typing of PAC08 and PAC17 as ST1076 and ST275, respectively. These sequence types of PAC08 and PAC17 were validated by the study by Phoon et al. (2018) through traditional MLST method. According to Phoon et al. (2018), PAC08 (ST1076) and PAC17 (ST235) were classified as MDR *P. aeruginosa* due to high value of MIC toward ceftadizime, cefepime, imipenem, and meropenem. Besides of ST1076 and ST235, there are another twenty-three STs reported in the study and six of them are novel allelic profile strains (ST2329, 2335, 2337, 2338, 2340 and 2341). In another Malaysian study, Ramanathan et al. (2017) used MDR *P. aeruginosa* strains ST235 (PAS1, PAS4, PAS5, PAS6, PAS7, PAS9 and PAS10) and ST1076 (PAS2, PAS3 and PAS8), all of which were resistant to ceftadizime, aztreonam, gentamicin and ciprofloxacin. The strains collected from 2009 to 2014

showed that ST235 and ST1076 strains probably circulated and were predominant in the Malaysian hospitals reported in the study. Since then, there was no report regarding the MLST study of *P. aeruginosa* in Malaysia.

Based on *P. aeruginosa* Database in PubMLST, there were 232 entries on ST235 (last accessed 20th January 2018). The data in the database showed that ST235 is present mainly in Europe, North and South America, and a few Asian countries (Japan, India, Philippines, China, Singapore, and Malaysia). A few studies reported that *P. aeruginosa* ST235 is the predominant global ST and another study reported that ST235 and ST175 belong to high-risk epidemic clones associated with multidrug resistant and extensively drug-resistant nosocomial infection (Estepa et al., 2017; Oliver et al., 2015). In another study, ST111 and ST235 are said to be responsible for the spread of carbapenemase gene globally as found in Spain (Viedma et al., 2009), South Korean (Seok et al., 2011) Japan (Miyoshi-Akiyama et al., 2017) and China (Feng et al., 2017; Hong et al., 2016).

According to the allelic profiles of ST235 (38-11-3-13-1-2-4) and ST2613 (172-11-3-13-1-2-4), ST235 might be related to ST2613 because the difference these two STs was only in a single allele which is Acetyl Coenzyme A synthetase (*acs*). Moreover, the ST number assigned to each strain is supposed to be unique and this can be used to discriminate differences in each strain. However, the new allelic profile for *acs*, 172 (from ST2613), had been found in most of the ST235 strains and this makes a particular strain having two ST numbers (ST235 and ST2613). According to Jolley & Maiden (2010), it is possible for a strain to have two copies of *acs* allele with different profile. Even though confirmation of allele through PCR is made, there is high probability that the conserved primer will bind to either copy of *acs* allele, therefore, only one allele will be found. This situation happened in *in silico* MLST through whole genome sequence.

The application of MLST in defining epidemiological or relationship of strains has been widely used. Generally, conventional MLST method requires amplification by PCR using primers of 7 housekeeping genes, then the PCR products are sequenced and the sequenced data are submitted to MLST database for identification. Nowadays, the amplification of genes can be omitted and the sequences of housekeeping gene could be retrieved directly from WGS by simply entering and submitting WGS file into database. The term *in silico* MLST has been given to this method using WGS to search MLST profile. These *in silico* MLST has overcome the conventional MSLT where conventional MLST is time consuming and labour-intensive. However, the drawback of *in-silico* MLST is the cost for WGS is far more expensive compared to conventional MLST.

5.3 Phylogenetic Analysis of Multidrug Resistant *P. aeruginosa*

In this study, multiple phylogenetic trees have been used to confer phylogenetic relationship amongst *P. aeruginosa* strains. In the 16S rRNA phylogenetic tree, PAC08 and PAC17 were clustered together with other *P. aeruginosa* strains except for PA7 which is an outlier of *P. aeruginosa*. In 23S rRNA phylogenetic tree, PAC08 was clustered with PAS3 and PAS8 which are from Malaysia while PAC17 was clustered with MDR *P. aeruginosa* strains from Japan, India, United Kingdom and Malaysia. The use of 16S and 23S gene to confer relationship has been reported by Ali et al. (2012), Gao et al. (2014), Loper et al. (2012), Sankarasubramanian et al. (2014) and others This method is widely used to infer phylogenetic tree because these genes are highly conserved amongst strains belonging to same bacterial species. Li et al. (2016) used 16S gene for153 *P. aeruginosa* genomes from different locations to draw phylogenetic tree. The constructed phylogenetic tree showed only two subgroups. Due to limited nucleotide changes, these genes are not useful for phylogenetic studies between same genus and species even though the bootstrap values presented in phylogenetic tree in this study can

be seen as significant as the percentage is more than 70%. Similarly, in this study, 16S rRNA and 23S rRNA analyses failed to segregate the *P. aeruginosa* from different years and geographical locations.

To overcome this limitation of single gene analysis in phylogenetics, it can be improved by the simultaneous analysis using multiple genes. In this case, the application of Multi Locus Sequence Typing (MLST) using seven housekeeping genes is more reliable in drawing phylogenetic tree as shown in studies conducted by Davies et al. (2004), Jones et al. (2006), and Woerther et al. (2010). Therefore, based on the MLST data in this study, PAC08 and PAC17 were not related to each other. PAC08 was grouped together with other Malaysia strains (Figure 4.10 cluster B) which had ST532, ST1400, ST2335, ST2340 and ST708. According to PubMLST database, ST ST532, ST1400 and ST708 were found in other location such as Australia and Europe. Amongst of these STs, only ST2335 and ST2340 are novel allelic profile strain from Malaysia (Phoon et al., 2018). On the other side, PAC17 was clustered together with isolates from Asia, all of which were MDR strains.

With greater advance in bioinformatic software and the wide application of WGS, a recent study by Tsang et al. (2017) found that phylogeny analysis from MLST cannot fully represent genome phylogeny. MLST too has its limitation as it failed to inferred phylogeny because of the use of seven housekeeping genes. Therefore, Tsang et al. (2017) suggested that phylogenetic based on full genome single nucleotide polymorphism (SNP) and genome are more consistent to each other and might be considered as an alternative in molecular epidemiological study. According to the phylogenetic tree based on full genome SNP, PAC08 was clustered with a strain from Malaysia which has ST1076 allelic profile while PAC17 was grouped together with strains ST235 allelic profile from Japan.

5.4 Resistance Genes

The detection of various genes associated with multidrug resistant phenotypes in PAC08 and PAC17 was supported by a recent study by Phoon et al. (2018). PAC08 and PAC17 are phenotypically resistant to ceftazidime, amikacin, piperacillin-tazobactam, gentamicin, imipenem, ciprofloxacin, cefepime, meropenem and netilmicin (Phoon et al., 2018). However, these strains were found to be sensitive to Polymyxin B.

An increasing rate of resistance of *P. aeruginosa* toward β-lactam drug class in Malaysia has been reported by 2016 National Antibiotic Resistance surveillance report by the Infectious Diseases Research Centre (IDRC) in Institute Medical Research, Malaysia. Generally, the resistance to β-lactam drug class might be triggered by the presence of *OXA-50* gene which can be found in most of *P. aeruginosa* strains, and overexpression of inducible β-lactamase *ampC* (Cabot et al., 2016; Campana et al., 2017; Lambert, 2002; Morita et al., 2013; Oliver et al., 2015). Moreover, both PAC17 and PAC08 strains have different variants of *Pseudomonas*-derived cephalosporinase (*PDC-2* and *PDC-7*) which most probably are associated with reduced susceptibility to ceftazidime and cefepime (Jeukens, et al., 2017; Rodríguez-Martínez et al., 2009). While in PAC17, the source of imipenem- and meropenem resistance might be due to the presence of acquired metallo-β-lactamase (MβL) *IMP-7* gene which is located in gene cassette of class 1 integron (Walsh, et al., 2005). IMP-7 is one of MβL variants which is only found in *P. aeruginosa* from Canada and Malaysia (Ho et al., 2002; Khosravi, Tee Tay & Vadivelu, 2010; Walsh et al., 2005)

PAC17 genome comprises of multiple enzymatic resistant mechanism which are engaged in aminoglycoside resistance. These mechanisms are aminoglycoside phosphoryl-transferases (*APH)*, aminoglycoside adenylyl-transferases/nucleotidyl-transferases (*AAD* or *ANT*) and aminoglycoside acetyl-transferases (*AAC*) (Keith Poole,

2005). Moreover, some of these gene such as *aadA3*, *aadA6*, and *AAC(6')-Ib* are present in an integron cassette which is integrated into chromosome as discussed above. The presence of these acquired aminoglycoside modifying enzymes (AME) probably responsible for the susceptibility of amikacin and netilmicin in PAC17 (Keith Poole, 2005). However, amikacin and netilmicin resistance in PAC08 are not related to these mechanism since PAC08 lacks of these gene except *APH(3')-IIb.* According to Poole (2005), *APH(3')-IIb* is responsible for amikacin and isepamicin resistant in *P. aeruginosa*. Meanwhile, netilmicin resistance in PAC08 might be due to efflux system *MexXY-OprM* (Garneau-Tsodikova & Labby, 2016; Poole, 2005).

Resistance to fluoroquinolone can be seen in both strains where PAC08 and PAC17 are resistant to one of fluoroquinolone agents, ciprofloxacin. The source of fluoroquinolone resistance is probably due to point mutation in gene encoded for *gyrA* (T83I). This point mutation is also present in all reference strains, VRFPAO4 and NCGM2.S1, but not in PAO1 (Murugan et al., 2016). Indeed, in another study, complete genome sequences of *P. aeruginosa* PA7 was shown to have similar point mutation in PAC08 and PAC17 which attribute to fluoroquinolone resistance (Roy et al., 2010).

Besides of these phenotypically resistant of drug classes, there is another drug class that might confer resistance if there is prolong exposure of the strains to this drug class. The presence of *catB7* coding for xenobiotic acetyl-transferase in PAC08, PAC17, and reference strains is responsible for resistance to chloramphenicol drug class (Roy et al., 2010). This *catB7* gene is chromosomal encoded variant of chloramphenicol acetyl-transferase (*cat*) found in *P. aeruginosa* (White et al., 1999). Moreover, another variant of *cat* gene, *catB2* was found in a gene cassette of an integron in PAC17 showing that this gene might be transferred from other bacteria species (Tennstedt et al., 2003; Villa et al., 2002).

Both PAC 17 and PAC 8 were sensitive to Polymycin B (Phoon et al., 2018). The presence of major facilitator superfamily (MFS) type of efflux pump, *PmrAB*, and its component, *arnA, PmrC* and *PmrF*, might be responsible for reducing susceptibility of *P. aeruginosa* to polymycin B. The mechanism of polymyxin resistance is modulated by alteration of lipopolysaccaradie (LPS) by 4-amino-4-deoxy-L-arabinose (L-Ara4N) resulting in net positive charge which reduce its binding to polymyxin and lead to resistance (Abiola et al., 2004). However, due to strong side effect of this drug, polymycin B has become last-resort treatment (Morita et al., 2013).

Multidrug efflux pumps in Gram negative bacteria is well established and identified as a mechanism for resistance to a wide range of antibiotic drug class. Multidrug efflux pumps are classified into five major classes, which are the resistance-nodulation-division (RND) family, the major facilitator superfamily (MFS), the small multidrug resistance (SMR) family, the ATP-binding cassette (ABC) family and multidrug and toxic compound extrusion (MATE) family (Poole, 2004; Webber, 2003). PAC08, PAC17 and reference genomes carry multiple RND family (*MexAB-OprM*, *MexCD-OprJ*, *MexEF-OprN*, *MexGHI-OpmD, MexJK-OprM*, *MexMN-OprM*, *MexPQ-OpmE*, *MexVW-OprM*, *MexXY-OprM*, and *TriABC-OpmH*) which are chromosomally encoded. Each of these efflux systems can export a wide range of antibiotics including tetracycline, chloramphenicol and fluoroquinolone (Piddock, 2006). For example, *MexXY-OprM* is responsible for resistance to aminoglycoside, erythromycin, tetracyclines and glycyl-cycline (Poole, 2005). However, not all of these multidrug efflux pump systems are constitutively expressed (Piddock, 2006).

The presence of component efflux pump in the studied *P. aeruginosa* strains including referenced strains clearly showed their contribution in reduced susceptibility toward antibiotics. The overexpression of these efflux pumps have been recognized to cause cross

resistance and also responsible for removal of tetracyclines, fluoroquinolones, chloramphenicol and some of β-lactams antibiotic (Schwartz et al., 2015; Teixeira et al., 2016). The detected of *TriABC-OprH* operon has been associated with triclosan resistance (Mima et al., 2007). Genes related to small multidrug resistant (SMR) antibiotic efflux pump, *P. aeruginosa EmrE*, play a role in confer resistant to tetraphenylphosphonium, methyl viologen, gentamicin, kanamycin, and neomycin (Li et al., 2003; Nikaido, 2010).

Mutation in drug efflux regulator causes overexpression of drug by efflux pump and this causes resistance to specific antibiotic. Both PAC08 and PAC17 strains were found to have mutation in *mexS* (known as *nfxC*) which is a suppressor for *MexT* functioning as activator for *MexEF-OprN*. A recent study by Richardot et al. (2016) found that *mexS* mutant is rarely found in strains in hospital. The *mexS* mutant causes overexpression of *MexEF-OprN* which resulted in susceptibility to chloramphenicol, trimethoprim and fluoroquinolone (Richardot et al., 2016). On the other hand, mutated repressor *nalD* and regulator *nalC* in PAC17 give great impact to the expression of *MexAB-OprM*. *MexAB-OprM* is major efflux system where contributing to penam, cepham and penicillin resistance Poole, 2004) .

5.1 Pathogenicity Factors

A large number of genes encoding virulence factors were detected in the studied strains showing that the pathogenicity in *P. aeruginosa* is indeed complex. The virulence of *P. aeruginosa* is combination of multiple major groups of virulence factors such as secretion system, adherence and quorum sensing system. Complex secretion systems in *P. aeruginosa* consist of Type VI secretion system (T6SS) and Type III secretion system (T3SS).

In T6SS, there are there clusters of T6SS named Hcp secretion island 1 (HSI-I), Hcp secretion island II (HSI-II) and Hcp secretion island III (HSI-III). Each of these clusters consist of multiple genes and have their own role in pathogenicity. HIS-I is responsible for the export of three toxin proteins to other prokaryotic bacteria, offer fitness advantages for *P. aeruginosa* and take part in biofilm formation (Chen et al., 2015; Hood et al., 2010). HSI-II promotes changes in epithelial cell through P13K-Akt pathway while role of HSI-III was not clear because of limited knowledge of this cluster (Chen et al., 2015; Lesic., et al., 2009; Sana et al., 2012).

Various studies had been performed on identification and prevalence of major virulence genes such as exoenzyme (*exoS*, *exoU*, *exoT* and *exoY*), alkaline protease (*aprA*), exotoxin (*toxA*), elastase (*lasB*), phospholipase (*plcH,* and *plcN*) and alginate (*algD*) in different sources of strains. The detection frequency of *exoS, exoY*, *exoT*, *toxA* and *lasB* gene in urinary tract isolate was more than 70% (Habibi & Honarmand, 2015; Pobiega et al., 2016; Yousefi-Avarvand et al., 2015). For blood isolate, the detection frequency of *exoS*, *plcH*, *lasB*, and *algD* was more than 80% (Mitov et al., 2010). However, the detection of *exoU* gene in PAC08 and PAC17 in this study wasi n contrast to the prevalence of *exoU* gene in a study by Habibi and Honarmand (2015) and Pobiega et al. (2016), in which the frequency detection was less than 30%. *ExoU* gene encodes for cytotoxin capable of killing variety of eukaryotic cells and this gene has greater effect than other T3SS (Yousefi-Avarvand et al., 2015). A study by Mitov et al. (2010) found that *exoU* was more prevalent in multidrug resistant strains and *P. aeruginosa* isolates with *exoU* are most likely resistant to carbapenem, fluoroquinolone, cephalosporin and gentamicin (Garey et al., 2008).

CHAPTER 6: CONCLUSION

In this study, genome sequence of *P. aeruginosa* strain PAC08 (isolated from urine) and PAC17 (isolated from blood) had been successfully assembled and annotated using various bioinformatic softwares. The assembled genome of PAC08 and PAC17 yielded 6891589 and 6841669 base pair, respectively, and both strains have G+C content of approximately 66%. Through the annotation, PAC08 has 6445 CDS and 64 RNA genes while PAC17 has 6336 predicted coding sequence and 66 RNA genes. Even though PAC08 and PAC17 isolated from different location, there was not much difference in term of genomic structures between both strains. However, genomic contents of PAC08 and PAC17 are more diverse with presence of a variety mobile genetic elements such as insertion sequence, bacteriophages, and integrons.

The MLST analysis of PAC08 and PAC17 showed that they belong to ST1076 and ST235, respectively. ST235 is the predominant global sequence type and belongs to high-risk epidemic clones associated with multidrug resistant and extensively drug-resistant nosocomial infection while ST1076 is predominant ST found in urine samples according to the PubMLST database. By using MLST profiles of *P. aeruginosa* from Phoon et al. (2018) and Ramanathan et al. (2017), the phylogenetic relationship of PAC08 and PAC17 was deduced. PAC08 was clustered together with other Malaysian strains which have ST532, ST708, ST1076, ST2335 and ST2340 while PAC17 was clustered with strains from Japan, and United Kingdom, all of which have ST235.

Various resistance genes had been found responsible for multidrug resistance phenotypes in PAC08 and PAC17. The major element contributing to the multidrug resistance in *P. aeruginosa* is multidrug efflux pumps which efficiently efflux out antibiotic. A total of twenty multidrug efflux pump systems had been found in PAC08 and PAC17 which were categorized into Resistance-Nodulation-Cell Divion (RND),

Small Multidrug Resistance (SMR), Major Facilitator Superfamily (MFS) and ATP-Binding Cassette (ABC). Moreover, the mutation in repressor, regulator and suppressor of multidrug efflux disrupted regular function of these component, thus causing overexpression of drugs. Besides that, the presence of resistance gene elements in PAC08 and PAC17 had increased resistance towards aminoglycoside, β-lactamase, Fosfomycin, Macrolide, Phenicol and Sulphonamide.

A total of 243 virulence factors were found in PAC08 and PAC17, as well as in the reference strains, PAO1, PA7, LESB58 and PA14. These virulence factors of *P. aeruginosa* were classified into several groups which are adherence, antimicrobial activity, anti-phagocytosis, biosurfactant, iron uptake, lipase, protease, quorum sensing system, regulation, secretion system and toxins. Each of the virulence factors play roles in pathogenesis and combination of these virulence factor increase chances of invasion of *P. aeruginosa* into host.

Overall, genome characterization of *P. aeruginosa* using next generation sequence was more advantageous compared to traditional method of characterization using PCR. A lot of information could be obtained through whole genome sequence in defining genomic structure of *P. aeruginosa* from Malaysia. It is more beneficial if this study proceeds to transcriptomic study through RNA sequencing where gene expression especially multidrug efflux, can be studied.

The limitation of this study was high analytical skills are needed in interpretation of data obtained from bioinformatics. Moreover, the cost for sequencing especially RNA sequencing in Malaysia is still high. Even though the information obtained through this study surpassed the high cost for sequencing, the genome analysis study mostly require investment in research infrastructures and biocomputing facilities.

REFERENCES

Al-Bayssari, C., Valentini, C., Gomez, C., Reynaud-Gaubert, M., & Rolain, J.-M. (2015). First detection of insertion sequence element ISPa1328 in the oprD porin gene of an imipenem-resistant *Pseudomonas aeruginosa* isolate from an idiopathic pulmonary fibrosis patient in Marseille, France. *New Microbes and New Infections*, *7*, 26–27.

Al-Nayyef, H., Guyeux, C., Petitjean, M., Hocquet, D., & Bahi, J. M. (2015). Relation between insertion sequences and genome rearrangements in *Pseudomonas aeruginosa*. In: Ortuño F., Rojas I. (eds) *Bioinformatics and Biomedical Engineering* (pp. 426–437). Springer, Cham.

Ali, A., Soares, S. C., Santos, A. R., Guimarães, L. C., Barbosa, E., Almeida, S. S., … Azevedo, V. (2012). *Campylobacter fetus* subspecies: Comparative genomics and prediction of potential virulence targets. *Gene*, *508*(2), 145–156.

Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L., & Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics*, *12*(1), 1-10.

Aloush, V., Navon-venezia, S., Seigman-igra, Y., Cabili, S., & Carmeli, Y. (2006). Multidrug-Resistant *Pseudomonas aeruginosa* : Risk Factors and Clinical Impact. *Society*, *50*(1), 43–48.

Baltch, A. L., Smith, R. P., Franke, M., Ritz, W., Michelsen, P., Bopp, L., & Lutz, F. (1994). *Pseudomonas aeruginosa* cytotoxin as a pathogenicity factor in a systemic infection of leukopenic mice. *Toxicon : Official Journal of the International Society on Toxinology*, *32*(1), 27–34.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, *27*(4), 578–579.

Boetzer, M., & Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biology*, *13*(6), 1-9.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Brussow, H., Canchaya, C., & Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews*, *68*(3), 560–602.

Cabot, G., López-Causapé, C., Ocampo-Sosa, A. A., Sommer, L. M., Domínguez, M. Á., Zamorano, L., …Oliver, A. (2016). Deciphering the resistome of the widespread *Pseudomonas aeruginosa* Sequence Type 175 International High-Risk Clone through Whole-Genome Sequencing. *Antimicrobial Agents and Chemotherapy*, *60*(12), 7415–7423.

Campana, E. H., Xavier, D. E., Petrolini, F. V.-B. B., Cordeiro-Moura, J. R., Araujo, M. R. E. de, & Gales, A. C. (2017). Carbapenem-resistant and cephalosporin-susceptible: A worrisome phenotype among *Pseudomonas aeruginosa* clinical isolates in Brazil. *Brazilian Journal of Infectious Diseases*, *21*(1), 57–62.

Carlos Guimaraes, L., Benevides de Jesus, L., Vinicius Canario Viana, M., Silva, A., Thiago Juca Ramos, R., de Castro Soares, S., & Azevedo, V. (2015). Inside the pan-genome - methods and software overview. *Current Genomics*, *16*(4), 245–252.

CDC. (2013). *Pseudomonas aeruginosa in Healthcare Settings*. Retrieved on 1st October 2018 from http://www.cdc.gov/hai/organisms/pseudomonas.html

Chalmers, R., Sewitz, S., Lipkow, K., & Crellin, P. (2000). Complete nucleotide sequence of Tn10. *Journal of Bacteriology*, *182*(10), 2970–2972.

Chaudhari, N. M., Gupta, V. K., & Dutta, C. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Scientific Reports*, *6*, 1-10.

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., & Jin, Q. (2005). VFDB: A reference database for bacterial virulence factors. *Nucleic Acids Research*, *33 (Database Issue)*, D325-D328.

Chen, L., Zou, Y., She, P., & Wu, Y. (2015). Composition, function, and regulation of T6SS in *Pseudomonas aeruginosa*. *Microbiological Research*, *172*, 19–25.

Coil, D., Jospin, G., & Darling, A. E. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*, *31*(4), 587–589.

Cramer, N., Wiehlmann, L., Ciofu, O., Tamm, S., Høiby, N., & Tümmler, B. (2012). Molecular epidemiology of chronic *Pseudomonas aeruginosa* airway infections in cystic fibrosis. *PLoS ONE*, *7*(11), 1-9.

Curran, B., Jonas, D., Grundmann, H., Pitt, T., & Dowson, C. G. (2004). Development of a multilocus sequence typing scheme for the opportunistic pathogen *Pseudomonas aeruginosa*. *Journal of Clinical Microbiology*, *42*(12), 5644–5649.

Cury, J., Jové, T., Touchon, M., Néron, B., & Rocha, E. P. (2016). Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic Acids Research*, *44*(10), 4539–4550.

Darling, A. E., Mau, B., & Perna, N. T. (2010). Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, *5*(6), 1-17.

Davies, H. D., Jones, N., Whittam, T. S., Elsayed, S., Bisharat, N., & Baker, C. J. (2004). Multilocus Sequence Typing of Serotype III Group B Streptococcus and Correlation with Pathogenic Potential. *The Journal of Infectious Diseases*, *189*(6), 1097–1102.

Davison, J. (1999). Genetic exchange between bacteria in the environment. *Plasmid*, *42*(2), 73–91.

Diene, S. M., L'homme, T., Bellulo, S., Stremler, N., Dubus, J.-C., Mely, L., … Rolain, J.-M. (2013). ISPa46, a novel insertion sequence in the oprD porin gene of an imipenem-resistant *Pseudomonas aeruginosa* isolate from a cystic fibrosis patient in Marseille, France. *International Journal of Antimicrobial Agents*, *42*(3), 268–271.

Driscoll, J. A., Brody, S. L., & Kollef, M. H. (2007). The epidemiology, pathogenesis and treatment of *Pseudomonas aeruginosa* infections. *Drugs*, *67* (3), 351-368.

Dubnau, D. (1999). DNA Uptake in Bacteria. *Annual Review of Microbiology*, *53*(1), 217–244.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications, 7*(9), 1026-1042.

Estepa, V., Rojo-Bezares, B., Azcona-Gutiérrez, J. M., Olarte, I., Torres, C., & Sáenz, Y. (2017). Characterisation of carbapenem-resistance mechanisms in clinical *Pseudomonas aeruginosa* isolates recovered in a Spanish hospital. *Enfermedades Infecciosas y Microbiologia Clinica (English Ed.)*, *35*(3), 141–147.

Evans, J. C., & Segal, H. (2007). A novel insertion sequence, ISPA26, in oprD of *Pseudomonas aeruginosa* is associated with carbapenem resistance. *Antimicrobial Agents and Chemotherapy*, *51*(10), 3776–3777.

Feng, W., Sun, F., Wang, Q., Xiong, W., Qiu, X., Dai, X., & Xia, P. (2017). Epidemiology and resistance characteristics of *Pseudomonas aeruginosa* isolates from the respiratory department of a hospital in China. *Journal of Global Antimicrobial Resistance*, *8*, 142–147.

Fowler, R. C., & Hanson, N. D. (2014). Emergence of carbapenem resistance due to the novel insertion sequence ISPa8 in *Pseudomonas aeruginosa*. *PLoS ONE*, *9*(3), 1-7.

Fraser, C. M., Eisen, J. A., Nelson, K. E., Paulsen, I. T., & Salzberg, S. L. (2002). The value of complete microbial genome sequencing (you get what you pay for). *Journal of Bacteriology*, *184*(23), 6403–6405.

Gao, X. Y., Zhi, X. Y., Li, H. W., Klenk, H. P., & Li, W. J. (2014). Comparative genomics of the bacterial genus *Streptococcus* illuminates evolutionary implications of species groups. *PLoS ONE*, *9*(6), 1-12.

Garey, K. W., Vo, Q. P., Larocco, M. T., Gentry, L. O., & Tam, V. H. (2008). Prevalence of type III secretion protein exoenzymes and antimicrobial susceptibility patterns from bloodstream isolates of patients with *Pseudomonas aeruginosa* bacteremia. *Journal of Chemotherapy*, *20*(6), 714–720.

Garneau-Tsodikova, S., & Labby, K. J. (2016). Mechanisms of Resistance to Aminoglycoside Antibiotics: Overview and Perspectives. *MedChemComm*, *7*(1), 11–27.

Gooderham, W. J., & Hancock, R. E. W. (2009). Regulation of virulence and antibiotic resistance by two-component regulatory systems in *Pseudomonas aeruginosa*. *FEMS Microbiology Reviews, 33*(2), 279-294.

Goossens, H. (2003). Susceptibility of multi-drug-resistant *Pseudomonas aeruginosa* in intensive care units: results from the European MYSTIC study group. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *9*(9), 980–983.

Goryshin, I. Y., & Reznikoff, W. S. (1998). Tn5 *in vitro* transposition. *The Journal of Biological Chemistry*, *273*(13), 7367–7374.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075.

Habibi, A., & Honarmand, R. (2015). Profile of virulence factors in the multi-drug resistant *Pseudomonas aeruginosa* strains of human urinary tract infections (UTI). *Iranian Red Crescent Medical Journal*, *17*(12), 1-7.

Hayashi, T., Baba, T., Matsumoto, H., & Terawaki, Y. (1990). Phage-conversion of cytotoxin production in *Pseudomonas aeruginosa*. *Molecular Microbiology*, *4*(10), 1703–1709.

Hilker, R., Munder, A., Klockgether, J., Losada, P. M., Chouvarine, P., Cramer, N., … Tümmler, B. (2015). Interclonal gradient of virulence in the *Pseudomonas aeruginosa* pangenome from disease and environment. *Environmental Microbiology*, *17*(1), 29–46.

Hirsch, E. B., & Tam, V. H. (2010). Impact of multidrug-resistant *Pseudomonas aeruginosa* infection on patient outcomes. *Expert Review of Pharmacoeconomics & Outcomes Research*, *10*(4), 441–451.

Ho, S. E., Subramaniam, G., Palasubramaniam, S., & Navaratnam, P. (2002). Carbapenem-resistant *Pseudomonas aeruginosa* in Malaysia producing IMP-7 beta-lactamase. *Antimicrobial Agents and Chemotherapy*, *46*(10), 3286–3287.

Hong, J. S., Yoon, E.-J., Lee, H., Jeong, S. H., & Lee, K. (2016). Clonal dissemination of *Pseudomonas aeruginosa* Sequence Type 235 isolates carrying blaIMP-6 and emergence of blaGES-24 and blaIMP-10 on novel genomic islands PAGI-15 and -16 in South Korea. *Antimicrobial Agents and Chemotherapy*, *60*(12), 7216–7223.

Hood, R. D., Singh, P., Hsu, F., Güvener, T., Carl, M. A., Trinidad, R. R. S., … Mougous, J. D. (2010). A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host & Microbe*, *7*(1), 25–37.

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*, 1-11.

Inouye, S., Sunshine, M. G., Six, E. W., & Inouye, M. (1991). Retronphage phi R73: An *E. coli* phage that contains a retroelement and integrates into a tRNA gene. *Science (New York, N.Y.)*, *252*(5008), 969–971.

Jander, G., Rahme, L. G., & Ausubel, F. M. (2000). Positive correlation between virulence of *Pseudomonas aeruginosa* mutants in mice and insects. *Journal of Bacteriology*, *182*(13), 3843–3845.

Jeukens, J., Kukavica-Ibrulj, I., Emond-Rheault, J. G., Freschi, L., & Levesque, R. C. (2017). Comparative genomics of a drug-resistant *Pseudomonas aeruginosa* panel and the challenges of antimicrobial resistance prediction from genomes. *FEMS Microbiology Letters*, *364*(18),1-9.

Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., … McArthur, A. G. (2016). CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research*, *45*(D1), D566-D573.

Jolley, K. A., & Maiden, M. C. (2010). BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, *11*(1), 1-11.

Jones, N., Oliver, K. A., Barry, J., Harding, R. M., Bisharat, N., Spratt, B. G., … Crook, D. W. (2006). Enhanced invasiveness of bovine-derived neonatal Sequence Type 17 Group B *Streptococcus* is independent of capsular serotype. *Clinical Infectious Diseases*, *42*(7), 915–924.

Jordan, I. K., Makarova, K. S., Spouge, J. L., Wolf, Y. I., & Koonin, E. V. (2001). Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Research*, *11*(4), 555–565.

Kang, C.-I., & Song, J.-H. (2013). Antimicrobial resistance in Asia: Current epidemiology and clinical implications. *Infection & Chemotherapy*, *45*(1), 22–31.

Khosravi, Y., Tee Tay, S., & Vadivelu, J. (2010). Metallo-β-lactamase-producing imipenem-resistant *Pseudomonas aeruginosa* clinical isolates in a university teaching hospital in Malaysia: detection of IMP-7 and first identification of IMP-4, VIM-2, and VIM-11. *Diagnostic Microbiology and Infectious Disease*, *67*(3), 294–296.

King, E. O., Ward, M. K., & Raney, D. E. (1954). Two simple media for the demonstration of pyocyanin and fluorescin. *The Journal of Laboratory and Clinical Medicine*, *44*(2), 301–307.

Kirisits, M. J., Prost, L., Starkey, M., Parsek, R., & Parsek, M. R. (2005). Characterization of colony morphology variants isolated from *Pseudomonas aeruginosa* biofilms characterization of colony morphology variants isolated from *Pseudomonas aeruginosa* Biofilms. *Applied and Environmental Microbiology*, *71*(8), 4809–4821.

Koh, T. H., Khoo, C. T., Tan, T. T., Arshad, M. A. B. M., Ang, L. P., Lau, L. J., … Ooi, E. E. (2010). Multilocus sequence types of carbapenem-resistant *Pseudomonas aeruginosa* in Singapore carrying metallo-beta-lactamase genes, including the novel bla(IMP-26) gene. *Journal of Clinical Microbiology*, *48*(7), 2563–2564.

Koren, S., & Phillippy, A. M. (2015). One chromosome, one contig: Complete microbial genomes from long-read sequencing and assembly. *Current Opinion in Microbiology*, *23*, 110–120.

Krylov, V. N. (2014). Bacteriophages of *Pseudomonas aeruginosa*: Long-term prospects for use in Phage Therapy. *Advances in Virus Research*, *88*, 227–278.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874.

Kung, V. L., Ozer, E. A., & Hauser, A. R. (2010). The accessory genome of *Pseudomonas aeruginosa*. *Microbiology and Molecular Biology Reviews : MMBR*, *74*(4), 621–641.

Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, *35*(9), 3100–3108.

Lambert, P. A. (2002). Mechanism of antibiotic resistance in *Pseudomonas aeruginosa*. *Journal of Royal Society of Medicine*, *95*(41), 22-26.

Lawrence, J. G., Hendrickson, H., Gottschalk, G., & Schuster, S. C. (2005). Genome evolution in bacteria: Order beneath chaos. *Current Opinion in Microbiology*, *8* (5), 572–578.

Lee, D. G., Urbach, J. M., Wu, G., Liberati, N. T., Feinbaum, R. L., Miyata, S., … Ausubel, F. M. (2006). Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biology*, *7*(10), 1–14.

Lesic, B., Starkey, M., He, J., Hazan, R., & Rahme, L. G. (2009). Quorum sensing differentially regulates *Pseudomonas aeruginosa* type VI secretion locus I and homologous loci II and III, which are required for pathogenesis. *Microbiology*, *155*(9), 2845–2855.

Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, *44*(1), 242-245.

Li, G., Shen, M., Le, S., Tan, Y., Li, M., Zhao, X., … Lu, S. (2016). Genomic analyses of multidrug resistant *Pseudomonas aeruginosa* PA1 resequenced by single-molecule real-time sequencing. *Bioscience Reports*, *36*(6), 1–15.

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*(13), 1658–1659.

Li, X.-Z., Poole, K., & Nikaido, H. (2003). Contributions of MexAB-OprM and an EmrE homolog to intrinsic resistance of *Pseudomonas aeruginosa* to aminoglycosides and dyes. *Antimicrobial Agents and Chemotherapy*, *47*(1), 27–33.

Loper, J. E., Hassan, K. A., Mavrodi, D. V., Davis, E. W., Lim, C. K., Shaffer, B. T., … Paulsen, I. T. (2012). Comparative genomics of plant-associated *Pseudomonas* spp.: Insights into diversity and inheritance of traits involved in multitrophic interactions. *PLoS Genetics*, *8*(7), 1-27.

Mahajan-Miklos, S., Tan, M. W., Rahme, L. G., & Ausubel, F. M. (1999). Molecular mechanisms of bacterial virulence elucidated using a *Pseudomonas aeruginosa-Caenorhabditis elegans* pathogenesis model. *Cell*, *96*(1), 47–56.

Mano, Y., Saga, T., Ishii, Y., Yoshizumi, A., Bonomo, R. A., Yamaguchi, K., & Tateda, K. (2015). Molecular analysis of the integrons of metallo-β-lactamase-producing *Pseudomonas aeruginosa* isolates collected by nationwide surveillance programs across Japan. *BMC Microbiology*, *15*(1),1-8.

Marvig, R. L., Sommer, L. M., Jelsbak, L., Molin, S., & Johansen, H. K. (2015). Evolutionary insight from whole-genome sequencing of *Pseudomonas aeruginosa* from cystic fibrosis patients. *Future Microbiology*, *10*(4), 599–611.

Maurelli, A. T., Fernández, R. E., Bloch, C. A., Rode, C. K., & Fasano, A. (1998). "Black holes" and bacterial pathogenicity: A large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(7), 3943–3948.

Meletis, G., & Bagkeri, M. (2013). *Pseudomonas aeruginosa* : Multi-drug-resistance development and treatment options. *Infection Control*, 33–56.

Mesaros, N., Nordmann, P., Plésiat, P., Roussel-Delvallez, M., Van Eldere, J., Glupczynski, Y., … Van Bambeke, F. (2007). *Pseudomonas aeruginosa*: Resistance and therapeutic options at the turn of the new millennium. *Clinical Microbiology and Infection*, *13*(6), 560–578.

Mima, T., Joshi, S., Gomez-Escalada, M., & Schweizer, H. P. (2007). Identification and characterization of TriABC-OpmH, a triclosan efflux pump of *Pseudomonas aeruginosa* requiring two membrane fusion proteins. *Journal of Bacteriology*, *189*(21), 7600–7609.

Mira, A., Martín-Cuadrado, A. B., D'Auria, G., & Rodríguez-Valera, F. (2010). The bacterial pan-genome: A new paradigm in microbiology. *International Microbiology*, *13*(2), 45–57.

Mitov, I., Strateva, T., & Markova, B. (2010). Prevalence of virulence genes among Bulgarian nosocomial and cystic fibrosis isolates of *Pseudomonas aeruginosa*. *Brazilian Journal of Microbiology*, *41*, 588–595.

Miyoshi-Akiyama, T., Kuwahara, T., Tada, T., Kitao, T., & Kirikae, T. (2011). Complete genome sequence of highly multidrug-resistant *Pseudomonas aeruginosa* NCGM2.S1, a representative strain of a cluster endemic to Japan. *Journal of Bacteriology*, *193*(24), 7010.

Miyoshi-Akiyama, T., Tada, T., Ohmagari, N., Viet Hung, N., Tharavichitkul, P., Pokhrel, B. M., … Kirikae, T. (2017). Emergence and spread of epidemic multidrug-resistant *Pseudomonas aeruginosa*. *Genome Biology and Evolution*, *9*(12), 3238–3245.

Morita, Y., Tomida, J., & Kawamura, Y. (2013). Responses of *Pseudomonas aeruginosa* to antimicrobials. *Frontiers in Microbiology*, *4*(422), 1-5.

Mosquera-Rendón, J., Rada-Bravo, A. M., Cárdenas-Brito, S., Corredor, M., Restrepo-Pineda, E., & Benítez-Páez, A. (2016). Pangenome-wide and molecular evolution analyses of the *Pseudomonas aeruginosa* species. *BMC Genomics*, *17*(45), 1-14.

Murugan, N., Malathi, J., Umashankar, V., & Madhavan, H. N. (2016). Unraveling genomic and phenotypic nature of multidrug-resistant (MDR) *Pseudomonas aeruginosa* VRFPA04 isolated from keratitis patient. *Microbiological Research*, *193*, 140–149.

Nakata, N., Tobe, T., Fukuda, I., Suzuki, T., Komatsu, K., Yoshikawa, M., & Sasakawa, C. (1993). The absence of a surface protease, OmpT, determines the intercellular spreading ability of *Shigella*: The relationship between the ompT and kcpA loci. *Molecular Microbiology*, *9*(3), 459–468.

Nakayama, K., Kanaya, S., Ohnishi, M., Terawaki, Y., & Hayashi, T. (1999). The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: Implications for phage evolution and horizontal gene transfer via bacteriophages. *Molecular Microbiology*, *31*(2), 399–419.

Naughton, S., Parker, D., Seemann, T., Thomas, T., Turnbull, L., Rose, B., … Manos, J. (2011). *Pseudomonas aeruginosa* AES-1 exhibits increased virulence gene expression during chronic infection of cystic fibrosis lung. *PLoS ONE*, *6*(9), 1–8.

Nikaido, H. (2010). Multidrug Resistance in Bacteria. *Annu Rev Biochem.*, *78*(2), 119–146.

Obritsch, M. D., Fish, D. N., MacLaren, R., & Jung, R. (2004). National surveillance of antimicrobial resistance in *Pseudomonas aeruginosa* isolates obtained from intensive care unit patients from 1993 to 2002. *Antimicrobial Agents and Chemotherapy*, *48*(12), 4606–4610.

Ochman, H., Lawrence[2], J. G., & Groisman[3], E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, *405*(18), 299-304.

Oliver, A., Mulet, X., López-Causapé, C., & Juan, C. (2015). The increasing threat of *Pseudomonas aeruginosa* high-risk clones. *Drug Resistance Updates*, *21–22*, 41–59.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., … Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(D1), D206–D214.

Ozer, E. A., Allen, J. P., & Hauser, A. R. (2014). Characterization of the core and accessory genomes of *Pseudomonas aeruginosa* using bioinformatic tools Spine and AGEnt. *BMC Genomics*, *15*(737), 1-17.

Pacific Biosciences. (2013). *The Value of Finished Bacterial Genomes*. Retrieved on 6th June 2019 from https://www.pacb.com/wp-content/uploads/2015/09/microbial_primer.pdf

Parsek, M. R., & Singh, P. K. (2003). Bacterial biofilms: An emerging link to disease pathogenesis. *Annual Review of Microbiology*, *57*, 677–701.

Phoon, H. Y. P., Hussin, H., Hussain, B. M., & Thong, K. L. (2018). Molecular characterization of extended-spectrum Beta Lactamase- and Carbapenemase-producing *Pseudomonas aeruginosa* strains from a Malaysian tertiary hospital. *Microbial Drug Resistance*, *24*(8), 1108-1116.

Piddock, L. J. V. (2006). Multidrug-resistance efflux pumps - not just for resistance. *Nature Reviews Microbiology*, *4*(8), 629–636.

Pobiega, M., Maciag, J., Pomorska-Wesolowska, M., Chmielarczyk, A., Romaniszyn, D., Ziolkowski, G., … Bulanda, M. (2016). Urinary tract infections caused by *Pseudomonas aeruginosa* among children in Southern Poland: Virulence factors and antibiotic resistance. *Journal of Pediatric Urology*, *12*(1), 36.e1-36.e6.

Poole, K. (2004). Efflux-mediated multiresistance in Gram-negative bacteria, *Clinical Microbiology and Infection, 10*(1), 12–26.

Poole, K. (2005). Aminoglycoside Resistance in Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotheraphy*, *49*(2), 479–487.

Rahme, L. G., Stevens, E. J., Wolfort, S. F., Shao, J., Tompkins, R. G., & Ausubel, F. M. (1995). Common virulence factors for bacterial pathogenicity in plants and animals. *Science*, *268*(5219), 1899–1902.

Ramanathan, B., Jindal, H. M., Le, C. F., Gudimella, R., Anwar, A., Razali, R., … Sekaran, S. D. (2017). Next generation sequencing reveals the antibiotic resistant variants in the genome of *Pseudomonas aeruginosa*. *PLOS ONE*, *12*(8), 1-15.

Ricchetti, M., Fairhead, C., & Dujon, B. (1999). Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature*, *402*(6757), 96–100.

Richardot, C., Juarez, P., Jeannot, K., Patry, I., Plésiat, P., & Llanes, C. (2016). Amino acid substitutions account for most MexS alterations in clinical nfxC mutants of *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, *60*(4), 2302–2310.

Rissman, A. I., Mau, B., Biehl, B. S., Darling, A. E., Glasner, J. D., & Perna, N. T. (2009). Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics*, *25*(16), 2071–2073.

Rodríguez-Martínez, J.-M., Poirel, L., & Nordmann, P. (2009). Molecular epidemiology and mechanisms of carbapenem resistance in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*, *53*(11), 4783–4788.

Rouli, L., Merhej, V., Fournier, P.-E., & Raoult, D. (2015). The bacterial pangenome as a new tool for analyzing pathogenic bacteria. *New Microbes New Infect*, *7*, 72–85.

Roy, P. H., Tetu, S. G., Larouche, A., Elbourne, L., Tremblay, S., Ren, Q., … Paulsen, I. T. (2010). Complete genome sequence of the multiresistant taxonomic outlier *Pseudomonas aeruginosa* PA7. *PLoS ONE*, *5*(1), 1-10.

Ruiz-Martínez, L., López-Jiménez, L., d'Ostuni, V., Fusté, E., Vinuesa, T., & Viñas, M. (2011). A mechanism of carbapenem resistance due to a new insertion element (ISPa133) in *Pseudomonas aeruginosa*. *International Microbiology : The Official Journal of the Spanish Society for Microbiology*, *14*(1), 51–58.

Sana, T. G., Hachani, A., Bucior, I., Soscia, C., Garvis, S., Termine, E., … Bleves, S. (2012). The second type VI secretion system of *Pseudomonas aeruginosa* strain PAO1 is regulated by quorum sensing and Fur and modulates internalization in epithelial cells. *The Journal of Biological Chemistry*, *287*(32), 27095–27105.

Sankarasubramanian, J., Vishnu, U. S., Sridhar, J., Gunasekaran, P., & Rajendhran, J. (2014). Pan-Genome of *Brucella* Species. *Indian Journal of Microbiology*, *55*(1), 88–101.

Schwartz, T., Armant, O., Bretschneider, N., Hahn, A., Kirchen, S., Seifert, M., & Dötsch, A. (2015). Whole genome and transcriptome analyses of environmental antibiotic sensitive and multi-resistant *Pseudomonas aeruginosa* isolates exposed to waste water and tap water. *Microbial Biotechnology*, *8*(1), 116–130.

Seok, Y., Bae, I. K., Jeong, S. H., Kim, S. H., Lee, H., & Lee, K. (2011). Dissemination of IMP-6 metallo-B-lactamase-producing *Pseudomonas aeruginosa* sequence type 235 in Korea. *Journal of Antimicrobial Chemotherapy*, *66*(12), 2791–2796.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., & Chandler, M. (2006). ISfinder: The reference centre for bacterial insertion sequences. *Nucleic Acids Research*, *34*(Database issue), D32-D36.

Simpson, J. T., & Pop, M. (2015). The theory and practice of genome sequence assembly. *Annual Review of Genomics and Human Genetics*, *16*(1), 153–172.

Stover, C. K., Pham, X. Q., Erwin, a L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., … Olson, M. V. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature*, *406*(6799), 959–964.

Subedi, D., Vijay, A. K., Kohli, G. S., Rice, S. A., & Willcox, M. (2018). Comparative genomics of clinical strains of *Pseudomonas aeruginosa* strains isolated from different geographic sites. *Scientific Reports, 8*(1), 1-14.

Suraiya Md Noor, S., Hassan, H., Rahman, Z. A., Mohamed, Z., Deris, Z. Z., Hasas, S. A., … Lin, T. K. (2008). Source identification of nosocomial outbreak of *Pseudomonas aeruginosa* by pulsed field gel electrophoresis analysis. *International Medical Journal*, *15*(5), 323-328.

Teixeira, P., Tacão, M., Alves, A., & Henriques, I. (2016). Antibiotic and metal resistance in a ST395 *Pseudomonas aeruginosa* environmental isolate: A genomics approach. *Marine Pollution Bulletin*, *110*(1), 75–81.

Tennstedt, T., Szczepanowski, R., Braun, S., Pühler, A., & Schlüter, A. (2003). Occurrence of integron-associated resistance gene cassettes located on antibiotic resistance plasmids isolated from a wastewater treatment plant. *FEMS Microbiology Ecology*, *45*(3), 239–252.

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., … Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 13950–13955.

Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: The bacterial pan-genome. *Current Opinion in Microbiology*, *11*(5), 472–477.

Tsang, A. K. L., Lee, H. H., Yiu, S.-M., Lau, S. K. P., & Woo, P. C. Y. (2017). Failure of phylogeny inferred from multilocus sequence typing to represent bacterial phylogeny. *Scientific Reports*, *7*(1), 1-12.

Valot, B., Guyeux, C., Rolland, J. Y., Mazouzi, K., Bertrand, X., & Hocquet, D. (2015). What it takes to be a *Pseudomonas aeruginosa*? The core genome of the opportunistic pathogen updated. *PLoS ONE*, *10*(5), 1–15.

Varani, A. M., Siguier, P., Gourbeyre, E., Charneau, V., & Chandler, M. (2011). ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. *Genome Biology*, *12*(3), 1-9.

Viedma, E., Juan, C., Acosta, J., Zamorano, L., Otero, J. R., Sanz, F., … Oliver, A. (2009). Nosocomial spread of colistin-only-sensitive sequence type 235 *Pseudomonas aeruginosa* isolates producing the extended-spectrum beta-lactamases GES-1 and GES-5 in Spain. *Antimicrobial Agents and Chemotherapy*, *53*(11), 4930–4933.

Villa, L., Mammina, C., Miriagou, V., Tzouvelekis, L. S., Tassios, P. T., Nastasi, A., & Carattoli, A. (2002). Multidrug and broad-spectrum cephalosporin resistance among *Salmonella enterica* serotype enteritidis clinical isolates in southern Italy. *Journal of Clinical Microbiology*, *40*(7), 2662–2665.

Walker, T. S., Bais, H. P., Déziel, E., Schweizer, H. P., Rahme, L. G., Fall, R., & Vivanco, J. M. (2004). *Pseudomonas aeruginosa*-plant root interactions. Pathogenicity, biofilm formation, and root exudation. *Plant Physiology*, *134*(1), 320–331.

Walsh, T. R., Toleman, M. A., Poirel, L., & Nordmann, P. (2005). Metallo-β-Lactamases : The quiet before the Metallo-β-Lactamases : the Quiet before the Storm?, *Clinical Microbiology* Review, *18*(2), 306–325.

Wang, J., Zhou, J., Qu, T., Shen, P., Wei, Z., Yu, Y., & Li, L. (2010). Molecular epidemiology and mechanisms of carbapenem resistance in *Pseudomonas aeruginosa* isolates from Chinese hospitals. *International Journal of Antimicrobial Agents*, *35*(5), 486–491.

Webber, M. A. (2003). The importance of efflux pumps in bacterial antibiotic resistance. *Journal of Antimicrobial Chemotherapy*, *51*(1), 9–11.

White, P. A., Stokes, H. W., Bunny, K. L., & Hall, R. M. (1999). Characterisation of a chloramphenicol acetyltransferase determinant found in the chromosome of *Pseudomonas aeruginosa*. *FEMS Microbiology Letters*, *175*(1), 27–35.

Woerther, P., Angebault, C., Lescat, M., Ruppé, E., Skurnik, D., Mniai, A. E., … Andremont, A. (2010). Emergence and dissemination of Extended-Spectrum β-Lactamase–producing *Escherichia coli* in the community: Lessons from the study of a remote and controlled population. *The Journal of Infectious Diseases*, *202*(4), 515–523.

Wolter, D. J., Acquazzino, D., Goering, R. V., Sammut, P., Khalaf, N., & Hanson, N. D. (2008). Emergence of carbapenem resistance in *Pseudomonas aeruginosa* isolates from a patient with cystic fibrosis in the absence of carbapenem therapy. *Clinical Infectious Diseases*, *46*(12), e137–e141.

Wu, W., Jin, Y., Bai, F., & Jin, S. (2015). *Pseudomonas aeruginosa*. In *Molecular Medical Microbiology* (pp. 753–767). Elsevier.

Yousefi-avarvand, A., Khashei, R., Ebrahim-saraie, H. S., Emami, A., Zomorodian, K., Motamedifar, M., … Motamedifar, M. (2015). The frequency of Exotoxin A and Exoenzymes S and U genes among clinical isolates of *Pseudomonas aeruginosa* in Shiraz, Iran. *International Journal of Molecular and Cellular Medicine (IJMCM)*, *4*(1), 167–173.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., … Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, *67*(11), 2640–2644.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: A fast phage search tool. *Nucleic Acids Research*, *39*, W347-W352.