Targeted Sequencing Technology for the Characterisation of Microbes and Antibiotic Resistance Markers from Low-biomass Environmental Samples

by

Qing Yang

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
March 2024

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

The healthcare environment is host to diverse communities of pathogenic bacteria with varying levels of antibiotic resistance. Within these communities, the exchange of genetic material, such as through horizontal gene transfer, further amplifies the spread of antibiotic resistance posing a threat to patients. Surveilling and managing this diversity through genetic characterisation are crucial for effective infection control. By analysing the genetic material of pathogenic bacteria, i.e. their DNA genome, valuable insights in genetic variation within a particular pathogen species can be gained. However, it remains difficult to effectively sequence genes and genomes from trace amounts of target DNA in hospital environmental samples (such as potable water, surfaces, and air) due to low-input or degraded DNA.

In this thesis I implemented a comprehensive approach to detect and analyse pathogenic bacteria and antibiotic resistance genes in low-biomass environmental samples including tap water, p-trap water, wastewater, air dust and surface by combining multiple culture-free targeted sequencing technologies and methodologies. By utilising *16S rRNA* amplicon sequencing and high-throughput-quantitative PCR approaches, the dynamic patterns of pan-pathogenic bacteria, antibiotic resistance genes, and mobile genetic elements in nosocomial environments were sensitively identified, correlated and traced.

Of particular importance, this thesis establishes two novel multiplex-PCR-based panels that enabled rapid and accurate diagnostic typing and investigation of antibiotic resistance in *Legionella pneumophila* and *Pseudomonas aeruginosa*, with a sensitivity down to as few as three bacterial cells using Oxford Nanopore MinION sequencing. The prevalence of sequence type 36 (ST36) and a novel ST of *L. pneumophila* was confirmed from the common water-source during the winter and summer seasons, respectively. Additionally, ST549 of *P. aeruginosa* was found to be associated with various antibiotic resistance traits, including fluoroquinolone, carbapenem, and aminoglycoside resistance, through the occurrence of antibiotic resistance genes and genetic polymorphisms of gyrase and efflux-pump regulator genes linked to antibiotic resistance.

Overall, this thesis demonstrates the utility of novel targeted enrichment technologies for the comprehensive detection and analysis of pathogens and ARGs, particularly in hard-to-sequence

environmental samples containing low starting input. These findings contribute to the understanding of epidemic tracing and outbreak assessment, as well as antibiotic resistance dynamics, thereby facilitating the development of effective surveillance and management strategies.

Acknowledgements

This has been a significant journey, and this period of time holds immense importance in my life.

I would like to thank to myself. Throughout this period of life, I have encountered numerous unexpected and challenging situations. However, I have consistently maintained a strong sense of faith and enthusiasm. I have persevered without stopping or giving up. I firmly believe that I will ultimately achieve my goals.

I would like to express my deepest gratitude to my beloved parents. Their unwavering support and boundless love have been a constant presence in my life. They always selflessly give everything they can to me, and now, they extend that same love and care to my daughter. As time goes by, I see them growing older, but my love for them remains unwavering.

I would like to extend my heartfelt appreciation to my supervisors, Professor Nick Loman, Dr Yu Xia, Dr Jessica Blair, and Dr Joshua Quick. Their expert guidance and supervision have been instrumental in shaping my scientific journey. Their enthusiasm, innovation, persistence, and invaluable contributions to the field of science have continually inspired and motivated me. I am truly grateful for their mentorship and the knowledge they have imparted to me.

I would like to express my deepest gratitude to my husband, Yu. During moments when I found myself trapped in self-doubt and emotional insecurity, he has been an unwavering source of support. His unlimited love and belief in me have been invaluable, especially during my transition back to work after maternity leave. The time we spend together with our daughter, exploring the city and living with mindfulness, is incredibly precious. Together, I cherish what truly matters in life. Our love is enduring and will last for eternity.

I would like to express my heartfelt gratitude to my daughter, Siya, a beautiful and intelligent one-year-old baby. Her infectious smile serves as a constant reminder and motivation for me to persevere. Seeing

her happiness and innocence fills my heart with joy and gives me the strength to keep pushing forward. She is a blessing in my life, and I am grateful for the boundless love and joy she brings to our family.

I would also like to extend my thanks to my other family members and friends. They are my grandparents, my cousins, my uncles and my aunts, as well as best friends Jiahui, Jingting, Qing, Lu, Zhujin, Cheng, Jiwen and Mengxiao. Their friendship and presence in my life have provided me with comfort, encouragement, and a sense of belonging. Their support and understanding have been instrumental in helping me navigate through both personal and professional challenges. I am truly grateful for their friendship and the bond I share.

I would like to express my sincere appreciation to my colleagues in XYLab and Loman's Lab, including Zhanwen, Ziqi, Miao, Bixi, Kaiqiang, Liming, Yuhong, Yuxi, Xiaotong, Cailong, Yi, Baobo, Zhihao, Emily, Pablo, Sam, Rad, Ingra and many others. Their support and assistance have been invaluable throughout my journey, from research studies to my personal life. The time I spent with them, enjoying moments of relaxation and camaraderie, holds a special place in my heart.

Table of contents

Table list

Figure list

List of abbreviations

Abbreviation	Full term
16S rRNA	*16S ribosome RNA*
A. arcticum	*Aquaspirillum arcticum*
A. baumannii	*Acetobacter baumannii*
AA	Amino acid
AAC	Aminoglycoside acetylators
AC	Air conditioner
AF	Aminoglycoside-fluoroquinolone
AME	Aminoglycoside-modifying enzyme
ANOSIM	Analysis of similarities
ANT	Aminoglycoside adenylators
APH	Aminoglycoside phosphorylators
AR	Antibiotic resistance
ARB	Antibiotic-resistant bacteria
ARG	Antibiotic resistance gene
AST	Antimicrobial susceptibility testing
BALF	Bronchoalveolar Lavage Fluid
BCYE	Buffered charcoal yeast extract
bp	Base pair
BSA	Bovine serum albumin
BYEB	Buffered yeast extract broth
CAP	Community-acquired pneumonia
CARD	Comprehensive Antibiotic Resistance Database
CC	Clonal complex

CDC	Center for Disease Control
cDNA	complementary DNA
CF	Cystic fibrosis
CFU	Colony-forming unit
cg	Core gene
cgMLST	Core gene multilocus sequence typing
COG	Clusters of orthologous Genes
Covid-19	Coronavirus disease 2019
CP	Cephalosporin-penam
CsgG	Curlin sigma S-dependent growth subunit G
Ct	Cycle threshold
DB	Database
DHFR	Dihydrofolate reductase
DHPS	Dihydropteroate synthase
DOP-PCR	Degenerate oligonucleotide PCR
E. coli	*Escherichia coli*
E. faecium	*Enterococcus faecium*
ENA	European Nucleotide Archive
ESBLs	Extended-spectrum beta-lactamases
ESGLI	European Society for Clinical Microbiology Study Group on Legionella Infections
ESKAPE	*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa,* and *Enterobacter* spp
FB	Flush Buffer
Gb(p)	Giga base pairs
HCl	Hydrochloric acid
HGT	Horizontal gene transfer
HAC	High-accuracy

HMW	High molecular weight
HT-qPCR	High-throughput qPCR
ICE	Integrative conjugative element
ICUs	Intensive care units
IDT	INTEGRATED DNA TECHNOLOGIES
IS	Insertion sequences
K. pneumoniae	*Klebsiella pneumoniae*
KOH	Potassium hydroxide
L. fermentum	*Lactobacillus fermentum*
L. pneumophila	*Legionella pneumophila*
LB	Luria-Bertani
LD	Legionnaires' disease
LDA	Linear discriminant analysis
LOD	Low limit of detection
LRS	Long-read sequencing
LPS	Lipopolysaccharides
MATE	Multidrug and toxic compound extrusion
Mb(p)	Mega base pair
MCC	Monobactam-carbapenem-cephalosporin
MCE	Mixed cellulose esters
MDA	Multiple displacement amplification
MDR	Multidrug-resistant
MERS	Middle East respiratory syndrome
MGE	Mobile genetic element
MIC	Minimum inhibitory concentration
MLSB	Macrolide/lincosamide/streptogramin B
MLST	Multilocus sequence typing
MPXV	*Monkeypox* virus
MRSA	methicillin-resistant *S. aureus*
MSND	MultiSample NanoDispenser

N. gonorrhoeae	*Neisseria gonorrhoeae*
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
nt	Nucleotide
OD	Optical density
ONT	Oxford Nanopore Technology
OTU	Operational taxonomic unit
P. aeruginosa	*Pseudomonas aeruginosa*
P. koreensis	*Pseudomonas koreensis*
PacBio	Pacific Biosciences
PBP	Penicillin-binding protein
PBS buffer	Phosphate-buffered saline
PCA	Principal component analysis
PCR	Polymerase Chain Reaction
PEP	Primer extension preamplification
PFGE	Pulse field gel electrophoresis
PrimalPCR	PrimalScheme multiplex PCR
PrimalSeq	PrimalPCR-based MinION sequencing
QIIME	Quantitative Insights in Microbiology
qPCR	Quantitative / Real time PCR
QRDR	Quinolone resistant-determining region
R. erythropolis	*Rhodococcus erythropolis*
RF	Robinson-Foulds
RO	Reverse osmosis
RPPs	Ribosomal protection proteins
RT-PCR	Real time PCR
RT-qPCR	Reverse transcription quantitative PCR
S. aureus	*Staphylococcus aureus*
S. maltophilia	*Stenotrophomonas maltophilia*
S. pyogenes	*Streptococcus pyogenes*

SARG	Structured Antibiotic Resistance Gene
SARS-COV-2	Severe acute respiratory syndrome coronavirus 2
SBH	Sequencing by hybridization
SBS	Sequencing by synthesis
SBT	Sequence-based typing
SD	Sequencing depth
SMRT	Single molecule, real time
SNP	Single nucleotide polymorphism
ST	Sequence type
SUP	Super-accuracy
TGS	Third-generation sequencing
Tm	Melting temperature
TPPs	Target protection proteins
WGA	Whole genome amplification
WGS	Whole genome sequencing
WHO	World Health Organization
WWTP	Wastewater treatment plant

Submitted work associated with this thesis

The manuscript associated with Chapter three has been under review in *Applied Microbiology and Biotechnology* Journal. The available manuscript has been announced online in the Preprints.org. Doi: 10.20944/preprints202403.0001.v1. Qing Yang is the first author of the manuscript. The contribution includes the conceptualisation, methodology, investigation, formal analysis and writing. The work was supervised with the collaborating supervisor Dr Yu Xia in the SusTech.

Minor changes were made to the Chapter Three for this thesis, which included individual discussion of seasonal, medium impacts on the compositions of bacterial communities and antibiotic resistance.

Chapter One

Introduction

The thesis consists of six chapters that documented the application of targeted sequencing techonologies to characterise pathogenic bacterial communities and explore antibiotic resistance dynamics using a "One Health" approach. The study specifically analysed low-biomass environmental samples collected from the *Peking University Shenzhen Hospital*, Shenzhen, China. The entire PhD work was conducted and completed separately in the UOB and SusTech over a toal period of 4.5 years. Specifically, in the first year in the UOB, the establishment of two panels described in chapter four and chapter five was carried out. Subsequently, during the period from the second year to the final year in the SusTech, the work described in chapter three and chapter five, as well as the overall writing-up of the thesis, were completed. The narrative explanation of each chapter is as follows.

To start with, the first chapter presented a thorough review that encompassed topics of pathogenic bacteria, antibiotic resistance, detection technologies, and low-input strategies for handling extremely low-biomass samples. Follwing the introductory review, the second chapter detailed the generic materials and methods employed in subsequent three chapters, providing a foundation for the experimental procedures and analytical approaches used in the research. In chapter three, the research integrated gold-standard *16S rRNA* amplicon sequencing and a novel high-throughput quantitative PCR approach to investigate the seasonal dynamic patterns of bacterial communities, antibiotic resistance, and mobile genetic elements in the water and air-conditioner systems of hospital indoor environments. Building upon the comprehensive exploratioin conducted in the studied hospital, the chapter four and chapter five introduced two innovative targeted multipex-PCR panels aiming to enhance the sensitivity, accuracy, and speed of studying pathogenic bacteria and antibiotic resistance from extremely low-biomass samples. These panels used MinION sequencing technology and were specifically developed for prevalent pathogenic bacteria namely *Pseudomonas aeruginosa* and *Legionella pneumophila*, with a detection sensitivity of at least three genome copies. Specifically, chapter four focused on exploring antibiotic resistance acquired by specific *P. aeruginosa* sequence types through nucleotide polymorphisms or horizontal gene transfer. Chapter five described the successuful application of the 51 core-gene panel for *L. pneumophila* with only three cells for the rapid risk assessment of epidemiological investigations. At last, in the sixth chapter, it consisted of a general discussion of the entire PhD work, including a comprehensive overview of the findings, their implications, their limitations, and potential avenues for further investigation and improvement.

1.1 Pathogens

A microorganism, also known as a microbe, is defined as an organism that can only be seen under a microscope, either as a single cell or as a colony of cells. During the period from 1665 to 1683, Robert Hooke and Antoni van Leeuwenhoek, made significant discoveries regarding the existence of microorganisms (Porter 1976; Hooke 1845; Gest 2004). Nearly 200 years later, Robert Koch firstly achieved the successful cultivation of microbe *Bacillus anthracis* from diseased tissues (Koch 1910; Opal 2010). Microbes are present in each of the three domains of life (archaea, bacteria and eukarya) and are are found in diverse environments aincluding soil, water, acidic hot springs, radioactive waste, and even the deep biosphere within Earth's crust (Gerba 2009). The bacteria play a pivotal role in both industry and human health and are essential for human survival (Prakash et al. 2007; Safronova and Tikhonovich 2012). The human body is home to a large and diverse microbiota (Collen 2015). The majority of bacteria that reside in our bodies are harmless or even beneficial commensals, with only a relatively small fraction of bacteria being pathogenic (having the potential to cause disease). The first identified pathogen *Vibrio cholerae*, the bacterium responsible for cholera, was discovered and isolated in the mid-19th century independently by Pacini and Koch (Bulloch 1960; Bentivoglio and Pacini 1995). There are approximately 1,400 known pathogens that can cause diseases in humans (Balloux and van Dorp 2017). To effectively combat and prevent the spread of infectious diseases, it is essential to have a comprehensive understanding of how pathogens are transmitted, and the various strains and types of pathogens involved.

1.1.1 Transmission of pathogenic bacteria

Pathogens can be acquired from an infected host, either human or another animal, or they can be acquired directly from the environment (Gerba 2009). Many human pathogens have a limited survival time outside the host and can only be transmitted through direct or close contact with an infected individual, such as *Neisseria gonorrhoeae* which is transmitted through sexual contact (Unemo et al. 2013). Pathogens that are transmitted through the environment can exhibit varying survival times outside the host, ranging from hours to years. The actual duration of survival is dependent on factors such as the specific organism involved and the environmental conditions it encounters. These factors can greatly influence the viability

and persistence of pathogens in the external environment (Rottier and Ince 2003). Pathogens have the potential to contaminate the environment through various means, such as respiratory secretions from the nose and mouth, as well as being carried on dead skin, faeces, urine, saliva, or sweat (Gerba, 2009; Mitchell 2020). When the contaminated environment is exposed, these organisms can be effectively transmitted to another host, initiating the cycle of infection transmission once again. Certain infectious diseases can be transmitted over great distances through the air, spanning hundreds of meters or even miles. Examples of such diseases include Legionnaires' disease , and hand-foot-mouth disease which is typically caused by various strains of enteroviruses (Gerba 2009).

1.1.2 Evolution and diversity of pathogenic bacteria

The relationship between pathogen diversity and evolution is closely intertwined. Normally, pathogens undergo frequent and rapid evolution, driven by genetic variation (Arber 2014) under selection pressures from the environment (Cui et al., 2013; Alhayek and Hirsch 2023). As a result, pathogen diversity emerges, characterised by the coexistence of multiple strains. Pathogens must to adapt to specific pressures from the host immune system, antimicrobial therapies, and other environmental factors (Diard and Hardt 2017). When pathogens adapt to specific hosts, the acquisition or loss of genes enables bacteria to colonise and thrive in particular host niches (Grote and Earl 2022). For instance, bacteria may acquire genes that enhance their adherence to specific host tissues or help evade host immune responses (Bäumler and Fang 2013). This ongoing evolutionary arms race between pathogenic bacteria and hosts drives continuous evolution and diversification of the pathogens (Adrian et al. 2019).

The key categories of pathogens relevant to this thesis include opportunistic pathogens and environmental pathogens.

Opportunistic pathogens are a group of microbes that typically do not cause infections in healthy individuals. These pathogens may thrive in hosts with weakened immune systems, such as patients in hospital settings or individuals with underlying conditions like cystic fibrosis or AIDS (Alberts 2017). These opportunists can arise from normally harmless commensal symbionts, such as *Streptococcus*

pneumoniae and *Staphylococcus aureus*, or from microbes acquired from the environment, like *Pseudomonas aeruginosa* and *Burkholderia cepacia* (Brown et al. 2012).

Environmental pathogens refer to microorganisms that normally spend substantial parts of their lifecycle outside human hosts, but cause disease with measurable frequency when introduced to humans (Cangelosi et al. 2004). The important difference between environmental pathogens and other human pathogens is their ability to survive and thrive beyond the host (Gerba 2009). The habitat that environmental pathogens live in can be divided into seven types: water, food, soil, air, vectors, living reservoirs, and products of human activity (Cangelosi et al. 2004). Their widespread occurrence in the environment makes them difficult to monitor, assess and control. Their unfavourable effects on human health and development cannot be controlled without first obtaining a comprehensive understanding of their environmental niches, their incidence, and the epidemiology of the diseases they cause. To achieve this, surveillance of the environment to determine the profile of environmental pathogens is required.

One of the most well known environmental pathogens is *Legionella*. In July 1976 after World War II, outbreaks of Legionnaires' disease (LD) occurred among more than 4,000 veterans in Philadelphia, US (Winn 1988). Cases experienced similar severe symptoms including chest pain, fever, lung congestion, and pneumonia which after a long period of investigation was found to be caused by a commonly isolated bacterium, later named *Legionella pneumophila* (Terranova, Cohen, and Fraser 1978; Winn 1988). *L. pneumophila* is typically found natural aquatic systems but can also colonise human-made water systems (e.g. air conditioning cooling towers, hot water systems, and plumbing systems), enabling it to be spread as an aerosol throughout the environment (Cunha et al. 2017). Today, LD continues to strike people every year, usually in the summer and fall months (Gea-Izquierdo et al. 2023). Until 2021, the highest annual rate of LD in the EU/EEA was observed with 2.4 cases per 100,000 population (European Centre for Disease Prevention and Control 2023).

P. aeruginosa is both an opportunistic and environmental pathogen which is responsible for approximately 10% of nosocomial infections (Reynolds and Kollef 2021). Community-acquired infections in CF patients are predominantly attributed to environmental exposure. This is particularly significant as lung infections caused by *P. aeruginosa* are frequently the primary cause of death in these

patients (Ranganathan et al., 2013). In hospitals, the presence of *P. aeruginosa* in the water supply is considered a significant source of infection for patients (Baghal Asghari, Nikaeen, and Mirhendi 2013). A recent study conducted in four large hospitals in the UK demonstrated that 0.9%-16% of all water outlets in augmented care units were contaminated with *P. aeruginosa* (Halstead et al. 2021), demonstrating the potential risk to patients from hospital environmental sources.

1.1.3 Technologies to identify pathogenic bacteria

The identification of pathogenic bacteria is vital for disease diagnosis, epidemiological surveillance, food safety, environmental monitoring, and scientific research (Yamin et al. 2023). Accurate identification enhances the ability to respond effectively to infectious diseases, protect public health, and advance the understanding of the microbial world.

There are several criteria used to classify the methods employed in the identification of bacteria. In general, these techniques can be categorized as either phenotypic methods and genotypic methods (Franco-Duarte et al. 2019). Phenotypic approaches, including microscopy, culture-based methods, and biochemical tests, have long been used for bacterial identification since the early work of Koch and Pasteur (Opal 2010; Prescott et al. 2018). However, molecular techniques, such as nucleic acid-based methods (PCR, DNA sequencing) and proteomic analysis, have greatly enhanced the accuracy, speed, and specificity of bacterial identification.

1.1.3.1 Phenotypic methods

(1) Morphological identification

Bacterial culture was the first method developed to study microbial morphology (Lagier et al. 2015) since culture allows the multiplication of bacterial cells on culture media under controlled laboratory conditions (Singleton and Sainsbury 1981; Parkhill and Wren 2011). Identification of bacteria relies on the morphological features of the cells, which can be visualised under the microscope (Gopinath et al. 2014). Such direct observation of bacteria is the simplest and cheapest way to identify bacteria. Based

on cell wall properties revealed through staining, bacteria can be further classified into Gram-positive, Gram-negative, or acid-fast (Kleanthous and Armitage 2015). Bacteria were initially observed by the Dutch microscopist van Leeuwenhoek using a single-lens microscope (Porter 1976). There are several types of microscopy including optical, electron, and scanning probe microscopy (Wang et al. 2022) which are practical and can identify and obtain morphological information about microorganisms quickly.

(2) Strain typing

Bacterial species can be further subdivided through the process of strain typing. Phage typing is a low-cost phenotypic method used for detecting single strains of bacteria. It has been used to trace the source of outbreaks of infections, especially for *Salmonella* (Baggesen et al. 2010; van der Merwe et al. 2014). The viruses that infect bacteria are called bacteriophages and some of these can only infect a single species of bacteria. These phages are used to identify different strains of bacteria within a single species through a process known as host recognition or host specificity (Stone et al. 2019). For example, it has been used for decades for subtyping of *Salmonella* Typhimurium to determine the epidemiological relation among isolates (van der Merwe et al. 2014).

Serological tests can also be employed to identify closely related species or to classify different strains within the same species. The technique uses antiserum obtained from a known strain, type, or serotype to perform serological testing, specifically targeting and observing serological reactions against the target microorganis (Venbrux, Crauwels, and Rediers 2023).

Historically, phenotypic methods have played a significant role in identifying and characterising bacterial species, primarily relying on isolation and culture techniques. Although these methods still have relevance in specific contexts, they are often labour-intensive, requiring substantial amounts of materials, and may not always provide accurate identification at the species or strain level (Bochner 2009).

1.1.3.2 Molecular methods

The emergence of the "era of molecular age" has brought about a revolution in the ability to detect, identify, characterise, and type bacteria by analysing the genetic characteristics with unprecedented precision and efficiency (Castro-Escarpulli et al. 2015).

On the one hand, genotypic methods have facilitated the characterisation of bacteria that are challenging to culture in the laboratory. On the other hand, genotypic methods have played a crucial role in advancing metagenomics studies (the study of entire genetic material recovered from bulk environmental or clinical samples by a sequencing method), allowing for the exploration of large and diverse unknown bacterial communities present in various environments (Li et al. 2015; Johansson et al. 2023; Markkanen et al. 2023). The majority of molecular methods used for bacterial identification primarily rely on either amplification or sequencing techniques. These methods encompass a spectrum of approaches, starting from individual DNA amplification-based techniques to more advanced integrated methods that involve restriction fragment analysis and genomic sequencing (Franco-Duarte et al. 2019). The advantages and limitations of these approaches can vary significantly. The selection of a particular technology for bacterial identification is influenced by various factors, such as the type of sample under analysis, the desired level of detail and accuracy in the results, available resources and cost considerations, as well as the required turnaround time. These factors must be carefully evaluated to select the most appropriate and feasible technology for a given study or application. Examples of pathogens detections using various molecular technologies in recent years have been discussed in following sections and listed in Table 1.1.

Table1.1 Pathogens detection by molecular technologies in recent years

Technology	Detected pathogens	LOD (Limit of detection)	Primer and assays	Sequencing Platforms	References
Multiplex PCR	*S. aureus, P. aeruginosa, H. influenza, L. pneumophila, K. pneumonia*	-	5 primer pairs	-	(Lapa et al. 2020)
	V. parahaemolyticus, L. monocytogenes and other six foodborn pathogens	-	8 primer pairs	-	(Li et al. 2022)
	Two *S. typhimurium* serotypes	100 CFU/μL	2 primer pairs	-	(Hong et al. 2023)
	DNA methylation	-	8 primer pairs	-	(Wang et al. 2023)
Real time PCR	10 enteric pathogens	2 copies/μL	20 TaqMan assays	-	(Ishii et al. 2013)
	Salmonella	2 copies	7 SYBR Green assay	-	(Barbau-Piednoir et al. 2013)
	S. enterica, Y. enterocolitica and other 10 pathogens	0.161 CFU/μL	12 TaqMan assays	-	(Liu et al. 2019)
RT-qPCR	SARS-CoV-2	6.6 copy/μL	4 primer pairs and probes	-	(Hasan et al. 2020)
	E. coli	-	7 primer pairs	-	(Đermić et al. 2023)
Microarray	123 bacterial strains	10 CFU/ml	26 probes	-	(Cao et al. 2011)
	10 pathogenic bacteria	-	10 probes	-	(Ranjbar et al. 2017)
	15 species	1000 copies/μL	34 probes	-	(Ma et al. 2020)
PFGE	*Y. Enterocolitica* strains	-	-	-	(Thisted Lambertz and Danielsson-Tham, 2005)
	V. parahaemolyticus strains	-	-	-	(Parsons et al. 2007)
WGS	*S. aureus*	-	Complete genome	Sanger	(Holden et al. 2004)
	Vancomycin-resistant *E. faecium,* carbapenem-resistant *E. cloacae*	-	Complete genome	Illumina Miseq	(Reuter et al. 2013)
	R. mucilaginosa, *A. graevenitzii* and other 8 complete bacterial genomes	-	Complete genome	ONT and Illumina Nextseq	(Baker 2022)
	E. coli and *K. pneumonia*	-	Complete genome and plasmids	Flonglle	(Avershina et al. 2022)
	Multidrug-resistant *C. indologenes*	-	Complete genome	Illumina	(Damas et al. 2022)
	SARS-CoV-2	Ct <20	Complete genome	GridION	(Terrazos Miani et al. 2023)
	Gram-negative and gram-positive pathogens	-	Complete genome and plasmids	PacBio	(Korotetskiy et al. 2023)
16S rRNA amplicon sequencing	High-resolution profiling of either microbial mock mixture or *Drosophila* microbiome	-	V1-V3, V3-V5 and V4 region	Sanger	(Fuks et al. 2018)
	Resolution of species and strain level with human stool samples	-	Full length 16S region (V1-V9)	PacBio	(Johnson et al. 2019)
	Human gut microbiota	-	Full length 16S region or V3-V4 region	MinION	(Matsuo et al. 2021)
	Gut bacteria	-	V1-V2 or V3-V4 region	Illumina Miseq	(Kameoka et al. 2021)
	Blood-born bacteria	-	Full length 16S region	MinION	(Huggins et al. 2022)
	Clinical isolates	-	V3-V4 region	ONT, Illumina and Sanger	(Lao et al. 2022)
	Pathogens of patients	-	Full length 16S region	MinION	(Pallerla et al. 2022)

(1) Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) was originally developed in 1983 by Kary Mullis and colleagues, and the first publication describing PCR application in 1985 focused on the analysis of the sickle cell anaemia mutation (Saiki et al. 1985). PCR has emerged as a valuable tool for the rapid detection and identification of bacteria directly from samples. Based on using primers that recognise conserved DNA sequences of bacterial genes, this technique is particularly valuable for identifying bacterial species/strains that may exist in a viable but nonculturable state and are often missed by conventional culture-based methods (Yang et al. 2023). The use of PCR in bacterial identification also helps circumvent situations where phenotypic characteristics are ambiguous and prone to misinterpretation. The speed, accuracy, and reliability of PCR have made it an indispensable tool in the field of infectious disease diagnostics. The primer set can be specifically designed to target particular bacterial species, enabling the detection of the target organism even in the presence of other bacteria. PCR using universal or specific primers have initially been used to amplify the *16S rRNA* genes of bacteria before being sequenced to help in the identification of unknown or novel bacteria species (Miyazaki, Sato, and Tsukuda 2017). Numerous variations of the PCR concept have been subsequently developed, as discussed below.

(i) Multiplex PCR

Multiplex PCR uses multiple primers in a single PCR mixture to detect, identify or differentiate bacteria. Thus, in multiplex PCR reaction, more than one target sequence is amplified in a parallel reaction to produce amplicons for different DNA sequences. While multiplex PCR offers advantages such as cost reduction, sample volume limitation, and rapid detection of multiple bacterial targets, it is crucial to emphasize the significance of primer design in the development of multiplex PCR assays (Sint, Raso, and Traugott 2012). The first multiplex-PCR application was in 1988 as a method to detect deletions in the dystrophin gene (Chamberlain et al. 1988). More applications of multiplex PCR in detecting multiple pathogens, even in the latest application for diagnosing DNA methylation are listed in Table 1.1.

(ii) Real-time (quantitative) PCR

A vital enhancement to PCR for diagnostic applications was the introduction of the concept of monitoring DNA amplification in real time through the monitoring of fluorescence (Holland et al. 1991). In the real-time PCR (also known as qPCR) process, target DNA is amplified and quantified simultaneously within a reaction. There are two strategies for the real-time visualisation of amplified DNA fragments: non-specific fluorescent DNA dyes (e.g. SYBR green) and fluorescently-labelled oligonucleotide probes (e.g. TaqMan probes). These two approaches were developed in parallel (Holland et al. 1991) and are widely used in pathogen detection (Table 1.1). However, probe-based chemistry has gained prominence in multiplex PCR assays due to its higher specificity and lower susceptibility to visualising non-specific PCR products, such as primer dimers (Kralik and Ricchi 2017). The costs associated with equipment and reagents in real-time PCR can be relatively high, especially when conducting assays tailored to target specific bacterial species, such as the commercialised genesig Real-Time PCR detection kits for *L. pneumophila* and *P. aeruginosa*.

(iii) Whole genome amplification

Whole genome amplification (WGA) is a robust method that enables the amplification of an entire genome, starting from small amounts of DNA in the picogram range, to generate exponentially larger amounts of amplified genome products (Wang et al. 2022). WGA has emerged as an invaluable technique for conserving limited valuable samples, especially when employing WGA methods specifically designed to amplify individual cells (Wang et al. 2022). Multiple approaches have been devised for achieving high-fidelity WGA, including multiple displacement amplification (MDA), degenerate oligonucleotide PCR (DOP-PCR) and primer extension preamplification (PEP) (Spits et al. 2006). While DOP-PCR and PEP rely on conventional PCR techniques, MDA can be accomplished through an isothermal reaction setup. In the case of MDA, amplification is facilitated by random hexamer primers and template annealing, utilising the φ29 DNA polymerase enzyme that possesses robust strand replacement activity and exonuclease activity (Liu et al. 2014). By employing a 100 µl reaction system, the amplification product can yield approximately 20-30 µg of DNA with fragment sizes ranging from 10 to 100 kb. This outcome enables the generation of a substantial quantity of consistent and comprehensive whole genome sequences (He et al. 2018). The more details of WGA will be discussed in section *1.3.2.*

(iv) Reverse transcription quantitative PCR

In reverse transcription quantitative PCR (RT-qPCR), RNA is first transcribed into complementary DNA (cDNA) form. Then the cDNA is employed as the template for the qPCR reaction. This technique has been particularly valuable during the SARS-CoV-2 outbreak since 2019 (Hasan et al. 2020). However, due to the inherent instability of RNA, its handling and quantification require meticulous skill in the context of pathogen detection.

Taken together, PCR is a highly sensitive, specific, and rapid technique that can be adapted to various applications, including bacteria identification, genetic testing, disease diagnosis, and forensic analysis. With the small amounts of genetic material, PCR can generate millions to billions of copies of the target sequence. This amplified DNA can be further analysed, sequenced, or used for downstream applications. However, PCR performance can be inhibited by components of enrichment broth, DNA extraction solution, PCR conditions interpreted subsequently, and the PCR contamination.

(2) Pulse field gel electrophoresis

There are several conventional methods continuing to be the implemented for strain identification. One of these methods is pulse field gel electrophoresis (PFGE) developed by Schwartz et al. (1984), employing restriction enzymes and electrophoresis to effectively separate and analyse large DNA fragments. PFGE is often referred to as the "gold standard" for subtyping by the Center for Disease Control (CDC) (Zou et al. 2013). During a seven-year cholera outbreak investigation, PFGE identified a total of 15 distinct pulsotypes of *V. cholerae* associated with the outbreak. Notably, 11 of these pulsotypes were previously unknown, indicating the emergence of new types (Taneja et al. 2012). In a latest study, a combination of PFGE and serum typing was employed to detect and characterise *Vibrio parahaemolyticus* and Enteroaggregative *E. coli* strains associated with a gastroenteritis outbreak (Wu et al. 2024). Nevertheless, PFGE does have significant limitations, such as the requirement for pathogen-specific standardized protocols, susceptibility to the choice of restriction enzymes, a time-consuming and labor-intensive process, and relatively low throughput capabilities (Simar, Hanson, and Arias 2021).

(3) DNA microarrays (gene chip technology)

DNA microarrays have become a practical and effective tool for detecting pathogenic bacteria in both clinical and environmental samples (McLoughlin 2011; Yu et al. 2023). These microbial detection arrays offer a intermediate solution in terms of cost and specificity, bridging the gap between low-cost, specifically focused assays like multiplex PCR and more expensive, broad-spectrum technologies such as high-throughput sequencing (Váradi et al. 2017). Although many array designs typically choose probes from fully sequenced genomes in databases like NCBI, cross-hybridization between probes and similar sequences enables the detection of novel species. This technology has been used in the high-throughput detection of pathogens, such as methicillin-resistant *Staphylococcus* in hospitals (Strommenger et al. 2007), the identification of antibiotic resistance genes (Perreten et al. 2005), and in the automatic analysis of single nucleotide polymorphisms (SNPs) (Zhang et al. 2006). Microarrays are a powerful tool for targeting a large number of genes or regions with high specificity and minimal off-target effects.

(4) DNA sequencing

Literally, DNA sequencing is a technique for determining the precise sequence of nucleotides in a DNA molecule. DNA sequencing is one of the most widely used genotypic methods and plays a pivotal role in deciphering the functionality of genes and other regions within the genome. There are multiple techniques available for DNA sequencing, each possessing unique characteristics. The exploration of alternative sequencing methods is a vibrant and driven field within genomics research with regularly emerging techniques that can help increase throughput with consequent reduction in costs.

(i) Overview of sequencing technologies

Whole-genome sequencing

In recent times, whole-genome sequencing (WGS) has emerged as an increasingly accessible and cost-effective tool for bacterial genotyping. The analysis of a complete bacterial genome not only offers unparalleled insights into into evolutionary origins, but has also brought about a revolutionary shift in our approach to comprehending antibiotic resistance and source attribution in outbreaks. The first

genomes to be sequenced were *MS2* and *ΦX174* which had relatively small genomes (Ng and Kirkness 2010). In the year of 1995, the Institute for Genomic Research sequenced the first complete genome, which was the bacterium *Haemophilus influenzae* (Fleischmann et al. 1995)), using Sanger sequencing taking several years to complete and significant resources. Progress in WGS technologies, e.g. next-generation sequencing, discussed below has resulted in significant enhancements in output generation and analysis speed. Moreover, these advancements have concurrently led to a reduction in overall costs (Quainoo et al. 2017). WGS-based approaches have been evaluated for pathogen identification and antibiotic resistance. For instance, in a study conducted by McGann et al. (2016), WGS was employed to investigate a severe outbreak of vancomycin-resistant *Enterococcus faecium* that affected three patients in an Intensive care unit (ICU). This analysis showcases the ability of WGS to not only identify antibiotic resistance but also infer transmission dynamics and the evolutionary lineage of the outbreak strains, thanks to its comprehensive analytical capabilities. More examples of WGS in pathogen detection are in Table 1.1.

Amplicon-based sequencing

While WGS involves sequencing the entire genome of an organism, providing a comprehensive view of its genetic makeup, amplicon-based sequencing focuses on sequencing specific target regions of interest in the genome. It involves selectively amplifying and sequencing only those regions, typically using PCR with designed primers. This approach is more targeted and focuses on specific genomic regions or genes.

***16S rRNA* PCR-sequencing**

The rapid amplification of DNA targets from relatively low biomass makes PCR one of the most sensitive techniques available for the detection of bacterial targets. The utilisation of PCR-based techniques for identifying bacterial DNA by amplifying and sequencing the hypervariable region (V1-V9) of *16S rRNA* gene has become a standard molecular method, routinely employed in laboratory research and clinical applications (Clarridge 2004; Matsuo et al. 2021). In recent times, the complete sequencing of the full-length *16S rRNA* amplicon has gained momentum, allowing for improved discrimination at the species and strain levels (Johnson et al. 2019; Matsuo et al. 2021). The standard steps involve PCR amplification of the *16S rRNA* gene, followed by sequencing and comparison to known databases for identification. This not only offers a faster alternative to conventional culture-based methods but also facilitates the

identification of bacteria that are challenging to cultivate under laboratory conditions. In a study conducted by Fouad et al. (2002), universal primers targeting the *16S rRNA* gene were designed and utilised to identify bacteria present in the root canals of patients with necrotic pulp tissue. Out of the 24 specimens tested, 22 were found to have detectable bacteria. While the *16S rRNA* gene has gained popularity as a target for PCR-based identification, there are instances where this gene may be identical among closely related species (e.g. *E. coli and Shigella* spp.). In such cases, alternative targets like conserved genes (e.g., *rpoB*, *tuf*, *gyrA*, and *gyrB*) are employed for identification purposes (Coleman and Tsongalis 2016).

Multiplex-PCR based sequencing

The advent of high multiplex PCR has significantly facilitated the simultaneous amplification of numerous amplicons in a single reaction. This capability enables the sensitive and specific sequencing of multiplex PCR-based amplicons, making it an effective approach for assaying a multi-locus sequence typing (MLST) panel consisting of multiple conserved regions. Additionally, this technique can be utilised for various genetic applications, including investigating SNPs at the population level, conducting phylogenetic analysis, and assessing population structure (Brachi, Morris, and Borevitz 2011; Andrews et al. 2016). More details of multiplex-PCR-based sequencing will be discussed in section *1.3.2.*

Amplicon-based sequencing is generally considered a straightforward and culture-independent procedure that has been widely and successfully employed to detect and identify bacteria in various samples (Table 1.1). The choice between WGS and amplicon-based sequencing depends on the research objectives, budget, time constraints, and the level of genomic information required. WGS provides a comprehensive analysis of the entire genome but comes with culture-dependent, higher costs and more data analysis complexity. Amplicon-based sequencing offers a more targeted and cost-effective approach, focusing on specific regions of interest with higher coverage and depth. When applied to clinical settings, several factors come into play that can impact PCR results. Clinical samples commonly contain a limited number of bacteria initially, and additional preprocessing steps are often necessary before PCR can be performed. The purpose of these steps is to eliminate PCR inhibitors, optimise bacterial extraction from the sample, and mitigate the risk of contamination.

Metagenomic sequencing

The targeted identification approaches described above, such as PCR, WGS, and amplicon-based sequencing, indeed enhance the number and proportion of reads related to the specific targets of interest in the sequence data. However, it is important to note that these approaches have a limitation in terms of the breadth of microorganisms that can be identified. They primarily focus on the amplified or targeted regions, potentially excluding other microorganisms present in the sample that do not match the specific targets.

Untargeted metagenomic sequencing involves the sequencing of the total DNA present in a sample without using specific primers targeting particular organisms or genes. This approach captures the entire genetic information within the sample, enabling the detection of microorganisms across various microorganisms, including bacteria, fungi, viruses, and parasites (Yang et al. 2018; Fan and Pedersen 2021). Moreover, metagenomic sequencing has the capability to perform gene functional analyses, including the identification of virulence factors, antibiotic resistance genes, and metabolic networks (Johnson et al. 2019). Currently, with the integration of long-read assembly and short-read error correction techniques, it has become possible to obtain complete bacterial genomes through metagenomic sequencing. In a study by Moss, Maghini, and Bhatt (2020), they demonstrated the successful finishing of bacterial genomes using this approach. Moreover, the advancements in long-read sequencing technologies, such as Oxford Nanopore Technologies sequencing, have further facilitated the assembly of high-quality, long-read-only sequences, allowing for the completion of bacterial genomes up to 5 megabases (Mb) in length (Zhao et al. 2023). The more details of metagenomic sequencing will be discussed in section *1.3.2.*

(5) Profile of novel sequencing platforms

The relatively recent history of DNA sequencing began with the introduction of two fundamental methods in the 20th century, Sanger sequencing (Sanger and Coulson 1975) and Maxam and Gilbert's approach (Maxam and Gilbert 1977). Currently there are numerous sequencing platforms available for DNA sequencing. These platforms utilise different approaches and methodologies to determine the

nucleotide sequence of DNA moleculeds. The novel sequencing platforms that will be discussed subsequently include next-generation sequencing and third-generation sequencing.

(i) Next-generation (second-generation) sequencing

The term "next generation" in the context of DNA sequencing technology signifies a progression in its development after the era of Sanger sequencing (Slatko, Gardner, and Ausubel 2018). Next-generation sequencing (NGS) methods can be categorized into two major types, sequencing by hybridization (SBH) and sequencing by synthesis (SBS). SBH was first developed in the 1980s (Drmanac et al. 1989). This technique used DNA arrays composed of known sequences on filters, which were then hybridized with labeled DNA fragments to be sequenced (Drmanac et al. 2002; Slatko et al. 2018). The majority of SBS technologies utilise a method where individual DNA molecules to be sequenced are either distributed into millions of separate well chambers or anchored to specific locations on a solid substrate, including 454 pyrosequencing (discontinued in 2013), and Illumina technology that is currently dominant technology in the NGS arena.

(ii) Third-generation sequencing (long-read sequencing)

Third-generation sequencing (TGS, known as long-read sequencing (LRS)) is a type of DNA sequencing method currently under active development. In contrast to NGS methods, TGS methods aim to sequence long DNA or RNA molecules. Two companies are currently at the forefront of TGS technology development: Pacific Biosciences (PacBio) and Oxford Nanopore Technology (ONT) which employ different approaches to sequencing single DNA molecules.

PacBio sequencing utilises the concept of single-molecule, real-time (SMRT) sequencing, which captures sequence information during the replication process of the target DNA molecule with the light detection (Travers et al. 2010; Rhoads and Au 2015). PacBio sequencing offers a significant advantage in terms of read length. The PacBio RS II system, equipped with the latest C4 chemistry, provides average read lengths exceeding 10 kb (Rhoads and Au 2015). The newest PacBio sequencing platform, the Revio system, increases throughput to 360 Gb of HiFi reads per day (Manuel et al. 2023).

ONT utilises nanopores as biosensors that are embedded within an electrically resistant membrane. By applying a potential across the membrane, a current is generated that flows exclusively through the nanopore (Deamer et al. 2016). The distinct disruptions in this current can be measured, providing identification about specific nucleotide. Since ONT introduced the first nanopore sequencer, MinION, in 2012, its applications in practical research have experienced significant growth (Mason and Elemento 2012). Ultra-long reads with above 300 kb reads and some close to 4 Mb reads can be sequenced in the ONT (Jain et al. 2018; Sereika et al. 2022). ONT has also developed pocket-sized sequencers are portable and do not require a laboratory setup, making them convenient for use outside of the traditional lab environment. For example, the MinION was used in field laboratories in Africa to sequence Ebola and Lassa virus outbreaks (Quick et al. 2016; Kafetzopoulou et al. 2019).

Oxford Nanopore development and performance

The concept of nanopore sequencing originated in the 1980s and was accomplished through a series of strategies involving both the nanopore itself and the associated motor protein (Wang et al. 2021). Following the introduction of the first nanopore sequencing device, MinION, by ONT in 2012 (Mason and Elemento 2012) and its subsequent commercialization in 2015, ONT has consistently released nine versions of the system to date, including R6 (June 2014), R7 (July 2014), R7.3 (October 2014), R9 (May 2016), R9.4 (October 2016), R9.5 (May 2017), R10 (March 2019), R10.3 (January 2020) and R10.4.1 (July 2022) (Wang et al. 2021).

The performance of ONT can be enhanced by improving the accuracy, extending the length, and boosting the throughput.

In terms of the accuracy, the R9 version of the nanopore system, utilising the Curlin sigma S-dependent growth subunit G (CsgG) nanopore derived from *E. coli*, achieved a significant improvement in sequence accuracy. It demonstrated an accuracy of approximately 87% (Minei, Hoshina, and Ogura 2018), compared to approximately 64% for the R7 version (Ashton et al. 2015). An updated motor protein with this nanopore exhibits a faster translocation rate of approximately 250 bp/s, whereas the R7 version has

a translocation rate of around 70 bp/s (Carter and Hussain 2017). In subsequent developments, a mutant CsgG nanopore and a new motor enzyme were incorporated into the R9.4.1 version, resulting in improved sequencing accuracy, with reported values reaching approximately 98.3% (Gong et al. 2018; Wick, Judd, and Holt 2019). Additionally, these advancements enabled faster sequencing speeds, reaching up to 450 bp/s. ONT introduced the R9.5 version specifically designed to be compatible with the 1D2 sequencing strategy (Seki et al. 2019). This strategy involves measuring a single DNA molecule twice. However, the R9.4 and R9.5 versions of nanopore sequencing technology face challenges when it comes to sequencing very long homopolymer runs (runs of identical bases). The R10 and R10.3 nanopores have been developed with two sensing regions, aiming to achieve higher accuracy specifically when dealing with homopolymer sequences (Huang, Liu, and Shih 2021). Currently the latest R10.4.1 nanopore paired with Kit14 chemistry has increased the accuracy of raw reads to 99% (Q20) with high-accuracy (HAC) basecalling, and has raised up to 99.5% (Q23) with super-accuracy (SUP) basecalling, and has finally reached up to 99.9% (Q30) with duplex basecalling (Ni et al. 2023).

A notable advantage of ONT sequencing is the exceptionally long read lengths made possible by electrical detection. This is because the method relies on the physical process of nucleic acid translocation (Wang et al. 2021). Thus, ONT read lengths are primarily determined by the sizes of the molecules present in the sequencing library. When small fragments outside the desired size distribution are present, they can negatively impact sequencing yield. This is due to the higher efficiencies of both adapter ligation and translocation through nanopores exhibited by these smaller fragments compared to longer fragments (Wang et al. 2021). Several techniques have been employed to extract and purify high-molecular-weight (HMW) DNA. These approaches encompass a range of methods such as spin columns, gravity-flow columns, magnetic beads and phenol-chloroform extraction.

With improvements in ONT and library preparation protocols, the maximum read length has increased from below 800 kb in early 2017 to 2.273 Mb in 2018 (Gong et al. 2019), and over 4 Mb in 2022 (Xu et al. 2022), primarily due to improvements in HMW DNA extraction methods and size selection strategies.

In addition to sequencing length and accuracy, throughput is another important consideration for ONT sequencing applications. The anticipated data output of a flow cell primarily relies on factors such as the quantity of active nanopores, the speed at which DNA/RNA translocates through the nanopore, and the duration of the run. In the early stages of MinION usage, users reported typical yields of several hundred megabases per flow cell (Wang et al. 2021). However, with the application of the Rev D ASIC chip, the current throughput has significantly increased to a maximum of 153 Gb and longer run times from a PromethION flow cell (Nicholls et al. 2019).

1.2 Antibiotic resistance

Antibiotics have transformed the treatment of infectious diseases caused by bacteria. The first antibiotic, salvarsan, was deployed in 1910 to effectively address conditions such as syphilis, relapsing fever, and African trypanosomiasis (Ehrlich and Hata 1911). The discovery of penicillin in 1928 by Alexander Fleming as a product of the fungus *Penicillium notatum* started the golden age of natural product antibiotic discovery (Amábile-Cuevas 1995). Later in 1937, another class of antibiotics, the sulphonamides, was introduced for more theraputic usage (Amábile-Cuevas 1995). These therapeutic agents contributed to historical progress in treating bacterial diseases, ultimately making once deadly infectious diseases now treatable and curable. For some decades after their introduction, antibiotics seemed to have solved the problem of bacterial infectious diseases permanently (Chadwick and Goode 2008).

1.2.1 Occurrence of antibiotic resistance

Soon after the clinical introduction of antibiotics, antibiotic-resistant bacteria (ARB) that can withstand the use of antibiotic therapy started to appear (Berglund 2015). In the early stages, the issue of antibiotic resistance (AR) was initially limited and often overlooked. However, a significant change occurred in the early 1930s when hospitals started observing the emergence of sulphonamide-resistant *Streptococcus pyogenes* strains (Ferretti and Köhler 2016). After that, penicillin-resistant *S. aureus* appeared (Levy and Marshall 2004). Since then, antibiotic resistance has become a matter of great concern (Levy and Marshall 2004). In 2017, ESKAPE (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella*

pneumoniae, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp) pathogens were listed as the “priority status” due to the rising considerable resistance against new antibiotics, such as sulphonamide, trimethoprim, quinolone, tetracycline, vancomycin and more (Health Organization, 2017). AR has been associated with a significant number of deaths worldwide. The World Health Organization (WHO) has estimated that approximately 4.9 million deaths occur annually due to infections caused by ARB while in a study published in 2022, it was found that in 2019 alone, 1.27 million deaths were directly attributed to drug-resistant bacterial infections (Antimicrobial Resistance Collaborators, 2022). Further, it is estimated that the number of deaths attributable to antibiotic resistance will increase to 10 million by 2050 if no actions are taken (O’Neill, Review on Antimicrobial Resistance, and Wellcome Trust (London, England) 2016; Helekal et al. 2023).

The overuse and misuse of antibiotics (even appropriate use), are widely acknowledged as significant factors responsible for the promotion of AR (Jensen et al. 2018). Research has shown that even subinhibitory concentrations of antibiotics, which are approximately 200 times lower than the minimum inhibitory concentration (MIC), can promote the selection of ARB (Andersson and Hughes 2012). In clinical surroundings, antibiotic misuses are ascribed to the prescription of antibiotics without accurate diagnosis of a bacterial infection. Outside of clinical environments, the great success of antibiotics at treating infectious diseases has spurred AR in the aquaculture and agriculture (Manyi-Loh et al. 2018; Hutinel et al. 2022; Markkanen et al. 2023; Gibson et al. 2023), and consequently lead to growing concerns within the "One Health" framework which recognises the interdependence of human health, animal health, and the environment (Food and Agriculture Organization of the United Nations et al. 2023). As a result, certain antibiotics were forbidden to be used as growth promoters in the European Union in the 1990s (Casewell et al. 2003). Currently, although antibiotics still remain generally effective in treating many bacterial infections, there is a escalating worry about the emergence of highly resistant strains that are challenging to treat. As a consequence, the available therapeutic options are becoming limited. There is a legitimate risk that if AR continues to spread, humanity may face clinical conditions reminiscent of the pre-antibiotic era, where bacterial infections were commonly fatal control (Spellberg and Gilbert 2014).

1.2.2 Emergence of antibiotic resistance genes

Antibiotic resistance genes (ARGs) are specific genes found in bacteria that provide them with resistance to antibiotics. ARGs have emerged since before humans started to use antibiotics in therapy, and they have likely existed for as long as antibiotics themselves (Berglund 2015). For example, ARGs that encode resistance to β-lactams, tetracyclines, and glycopeptides in Beringian permafrost dating back 30,000 years (D'Costa et al. 2011) and ARB have been found in a region of a cave in New Mexico, USA, which had been isolated for more than 4 million years (Bhullar et al. 2012).

When ARGs were discovered, they were acknowledged to protect against the effects of antibiotics (Kohanski, Dwyer, and Collins 2010). They often have minimal impact on the fitness of the host cell since many of them function by inactivating the antibiotic without altering the target site (Lai and Cooper 2024). Moreover, the expression of some resistance genes is certainly induced by the antibiotic, resulting in even fewer drawbacks for the host cell in retaining these genes. However, the fitness cost associated with an ARG can vary among different strains of hosts. For instance, Lai and Cooper (2024) discovered that a $bla_{TEM\text{-}116}$ ARG imposed a fitness cost in one host strain but not in another. Moreover, ARGs are also presumed to have evolved to fulfil other purposes than protecting bacteria from antibiotics. For example, an assumption is made that the primary role of ARGs in the environment is to regulate the responses triggered by subinhibitory levels of antibiotics.

1.2.3 Acquisition of antibiotic resistance

Resistance in bacteria can be either intrinsic or acquired. Intrinsic resistance is naturally occurring arising from the biology of the organism of which the mechanism may involve decreasing drug uptake, drug inactivation, and drug efflux (Reygaert 2018), such as the vancomycin resistance in *E. coli* (Hawkey 1998) and biocide triclosan resistance in the genus *Pseudomonas* (Blair et al. 2015) by preventing the drug to penetrate through the bacteria to reach the target sites. As for acquired resistance, it refers to the phenomenon where a bacterium that was once sensitive to antibiotics develops resistance. This resistance can occur through various mechanisms, including changes or mutations in the bacterial DNA or the acquisition of ARGs from other bacterial species through a process known as horizontal gene transfer (HGT).

1.2.3.1 DNA mutations providing resistance to antibiotics

DNA mutations can play a significant role in antibiotic resistance by altering the genetic information of bacteria (Wellington et al. 2013). When considering resistance to particular antibiotics, especially to synthetic agents such as fluoroquinolones and oxazolidinones, DNA mutations have primary clinical importance in certain bacterial species, such as *Mycobacterium tuberculosis, P. aeruginosa* and *Helicobacter pylori* (Huber et al. 2021). For example, fluoroquinolones target DNA gyrase and topoisomerases IV (Ruiz 2003), enzymes essential for bacterial DNA supercoiling. Mutations in the genes encoding subunits of these enzymes, such as *gyrA*, *gyrB*, *parC*, and *parE*, can lead to resistance to fluoroquinolones (Ruiz 2003). The accumulation of amino acid substitutions in these subunits, particularly in the quinolone resistant-determining regions, correlates with increasing resistance. In Gram-negative bacteria, DNA gyrase is usually the primary target, with mutations in *gyrA* associated with low-level resistance, and additional mutations in *parC*, *gyrB*, and *parE* leading to higher MICs (Everett et al. 1996). Resistance can be further enhanced by mutations affecting the expression of efflux pumps (Webber and Piddock 2003). In Gram-positive species, the primary target is often *parC* (Tankovic et al. 1996), although it varies depending on the specific quinolone. Some quinolones such as es-fluoroquinolone garenoxacin target both DNA gyrase and topoisomerase IV equally (Brown 2010), potentially reducing the likelihood of resistance development requiring mutations in both proteins (Ince et al. 2002). This understanding of DNA mutations and their impact on antibiotic resistance is crucial for developing effective strategies to combat resistant bacterial infections.

1.2.3.2 ARGs transmission from cell to cell

ARGs are often located on plasmids or transposons and can be transferred from cell to cell by HGT which includes conjugation, transformation, or transduction. This permits the resistance to rapidly spread among bacteria. This acquisition type of antibiotic resistance causes major problems in treating infectious diseases.

Generally, three pathways of HGT involving capture, accumulation, and dissemination of ARGs are driven by mobile genetic elements (MGEs) (Partridge et al. 2018), which are a collection of genetic material capable of relocating within a genome or being transferred from one species or replicon to another (Frost et al. 2005). MGEs are present in all organisms and serve as significant contributors to the process of evolution (Bigot 2012). The mobilome, consisting of a diverse range of plasmids, transposons, and viruses, represents the ensemble of MGEs present within an organism (Partridge et al. 2018).

(1) Conjugation

During conjugation, genetic material is transferred from donor bacteria to recipient bacteria through direct physical contact between the two cells (Davison 1999). Traditionally, conjugation has been regarded as the main facilitator of ARGs transfer between bacteria which was discovered in the 1950s (Amábile-Cuevas 1995). Conjugation has been seen in many different environments, including soil, seawater, and sewage wastewater (Davison 1999). The most important genetic elements capable of being transferred by conjugation are the plasmids and the integrative conjugative elements (ICEs) (Smillie et al. 2010).

(2) Transformation

Transformation is the uptake of genetic elements from the environment by bacterial cells. This genetic material can include various types of DNA, such as plasmids or fragmented DNA, and is often released into the environment by adjacent lysed bacteria (Pühler and Timmis 2012).

Transformation in the environment is often perceived as a rare occurrence due to the vulnerability of DNA to degradation by nucleases and the dilution effects commonly observed in water environments (Berglund 2015). Nevertheless, DNA can be stabilized through its adherence to particles found in sediment and soil. Dilution effects may also have less significance if transformation takes place within biofilms, as the lysis of newly deceased bacteria can release DNA that neighboring bacteria can take up (Wei and Håvarstein 2012). Natural transformation has been documented in numerous environmental settings, encompassing marine water, groundwater, rivers, and soil (Davison 1999), and it has been

implicated as responsible for the dissemination of penicillin-resistant genes in *Streptococcus* spp. (Johnsborg and Håvarstein 2009). In one study, concentrations of extracellular DNA were compared to intracellular DNA in a river basin in China (Mao et al. 2014). The discovery revealed a higher abundance of extracellular DNA, including ARGs, compared to intracellular DNA. This suggests that extracellular DNA serves as a significant environmental reservoir for genes accessible through transformation.

(3) Transduction

Transduction involves the transfer of foreign DNA or RNA into bacterial or eukaryotic cells using a virus or viral vector as a vehicle (Pühler and Timmis 2012). Typically involved in this context are bacteriophages, which attach to bacterial membranes and injects delivers their genetic material into the cell (Xu and Xiang 2017). Bacteriophage particles are highly effective in facilitating DNA transfer in the environment. Unlike naked DNA, they exhibit greater resistance to environmental degradation, and their compact size facilitates their widespread dissemination (Davison 1999). Moreover, certain bacteriophages possess wide host ranges, with some even capable of infecting bacteria from different classes (Jensen et al. 1998). The characteristics of bacteriophages make them well-suited for gene transfer between bacterial communities that are spatially distant, enabling the transfer of genes from environmental communities to human microbiomes (Muniesa, Colomer-Lluch, and Jofre 2013). Furthermore, evolutionary studies have demonstrated that substantial portions of bacterial genomes have originated from prophages, suggesting the importance of viral alterations to the bacterial chromosome (Brüssow and Hendrix 2002). Through viral metagenome analyses, β-lactamase genes have been detected in activated sludge and urban sewage (Rolain, Canton, and Cornaglia 2012). The *mecA* gene conferring methicillin resistance in methicillin-resistant *S. aureus* (MRSA) has also been found in bacteriophage DNA from a wastewater treatment plant (WWTP) and the receiving water (Colomer-Lluch, Jofre, and Muniesa 2011).

1.2.4 Mechanisms of antibiotic resistance

In addition to investigating the acquisition of AR, gaining a understanding of the specific mechanisms by which bacteria develop resistance to antibiotics is also imperative in the battle against antibiotic

resistance. AR mechanisms can be categorized into four primary groups: restriction of drug uptake, modification, protection or bypass of drug targets, drug inactivation, and increase of drug efflux (Reygaert 2018; Darby et al. 2023). Gram-negative and Gram-positive bacteria indeed exhibit different mechanisms. Gram-negative bacteria commonly employ all resistance mechanisms (Blair, Richmond, and Piddock 2014). In contrast, Gram-positive bacteria are less reliant on limiting drug uptake because they lack an outer membrane with lipopolysaccharides (LPS). They may also have a reduced capacity for certain types of drug efflux mechanisms (Chancey, Zähner, and Stephens 2012). However, Gram-positive bacteria can still develop resistance through other mechanisms such as modifying, protection or bypass drug targets and inactivating drugs.

1.2.4.1 Restrict drug uptake

Bacteria exhibit inherent differences in their ability to restrict the uptake of antibiotics. As mentioned above, the structure and functions of the LPS layer in Gram-negative bacteria provide a barrier to certain types of molecules. This gives those bacteria innate resistance to certain groups of large antimicrobial agents (Blair, Richmond, and Piddock 2014). Gram-positive bacteria do not possess an outer membrane, and restricting drug access is not as prevalent. In addition to the inherent mechanisms, changes in porin channels represent another pathway for limiting drug uptake in bacteria. In bacteria with large outer membranes, substances typically enter the cell through porin channels (Blair, Richmond, and Piddock 2014). There are two primary ways in which porin changes can limit drug uptake, i.e. a decrease in the number of porins present, and mutations that change the selectivity of the porin channel (Kumar and Schweizer 2005). Certain bacteria, such as members of the *Enterobacteriaceae* family, have been observed to develop resistance by reducing the number of porins. This reduction in porin quantity serves as a resistance mechanism against carbapenems (Chow and Shlaes 1991; Cornaglia et al. 1996). Mutations that contribute to alterations within the porin channel have also been identified, such as imipenem-resistant *Enterobacter aerogenes* and β-lactams and tetracycline-resistant *N. gonorrhoeae* (Gill 1998; Thiolas et al. 2004). Furthermore, the formation of a biofilm by a bacterial community also plays an essential role in restricting antibiotic uptake (Mah 2012). Biofilms are dense and adhesive communities of bacteria that produce polysaccharides, proteins, and DNA, creating a protective matrix (Van Acker, Van Dijck, and Coenye 2014). This matrix makes it challenging for antibiotics to penetrate

and reach the bacteria residing within the biofilm. As a result, higher concentrations of antibiotics are often required to achieve effectiveness against biofilm-associated bacteria (Reygaert 2018).

1.2.4.2 Target modification, protection or bypass

Modification of targets has emerged as a clinically significant mechanism of resistance for several important antibiotics. For example, in the case of Gram-positive bacteria, a notable resistance mechanism to β-lactam drugs involves alterations in penicillin-binding proteins (PBPs) (Beceiro, Tomás, and Bou 2013). PBPs are transpeptidases responsible for constructing peptidoglycan in the bacterial cell wall. Changes in the number or structure of PBPs can impact the binding of drugs to their targets (Reygaert 2009). Moreover, in terms of the antibiotic resistance that targets the ribosomal subunits, it can occur through various mechanisms. Ribosomal mutations play a role in the resistance to drugs such as aminoglycosides and oxazolidinones (Roberts 2004). Additionally, resistance can arise through ribosomal subunit methylation, which affects drugs like aminoglycosides, macrolides, oxazolidinones, and streptogramins. This methylation process is commonly associated with *erm* genes (Roberts 2004). Another mechanism of resistance is ribosomal protection, which is observed in the case of tetracyclines (Eliopoulos and Roberts 2003). Furthermore, in the case of drugs that target nucleic acid synthesis, such as fluoroquinolones, resistance commonly occurs through modifications in DNA gyrase or topoisomerase IV (Hawkey 2003). These mutations result in structural changes in gyrase and topoisomerase, leading to a decrease or complete loss of the drug's ability to bind to these components (Redgrave et al. 2014). Last, resistance to drugs that inhibit metabolic pathways, such as sulfonamides and trimethoprim, occurs through mutations in enzymes involved in the folate biosynthesis pathway. Specifically, mutations can affect enzymes like dihydropteroate synthase (DHPS) and dihydrofolate reductase (DHFR) (Huovinen et al. 1995). Additionally, resistance can arise through the overproduction of resistant DHPS and DHFR enzymes (Vedantam et al. 1998).

The protection of drug targets is a significant factor in clinically relevant antibiotic resistance, which affects a wide range of antibacterial drugs and is commonly observed in various bacterial pathogens. This mechanism can be classified into three distinct types (Wilson et al. 2020). The first type involves the action of target protection proteins (TPPs) that actively eliminate the drug from its designated target

site. The second type entails the activity of TPPs that induce conformational changes in the drug target. These changes disrupt the binding of the drug to its intended target site through allosteric mechanisms. The last type is when TPPs induce conformational changes within the target, thereby reinstating its functionality even in the presence of the bound antibiotic (Wilson et al. 2020; Darby et al. 2023). One well-recognised example of antibiotic resistance resulting from target protection is tetracycline resistance. In tetracycline resistance, target protection is mediated by ribosomal protection proteins (RPPs) specific to tetracycline. Currently, there are 13 distinct classes of tetracycline RPPs known (Wilson et al. 2020). Among these classes, Tet(O) and Tet(M) have been extensively studied and are considered the most well-characterised (Connell et al. 2003; Nguyen et al. 2014).

Target bypass is a strategic approach aimed at addressing antibiotic resistance by creating alternative pathways that render the original target unnecessary (Darby et al. 2023; Silva et al. 2023). One of the most well-known examples of target bypass is the development of MRSA. In MRSA, *S. aureus* acquires a variant of PBP known as PBP2a, which shares homology with the original target proteins (PBPs) but has a reduced affinity for β-lactam antibiotics like methicillin (Stapleton and Taylor 2002; Munita and Arias 2016). When methicillin binds to this alternative target site, it fails to effectively inhibit cell wall synthesis because the transpeptidase activity of PBP2a is maintained (Munita and Arias, 2016). This mechanism allows *S. aureus* to circumvent the action of methicillin and ensures its survival even in the presence of the drug. Another example of target bypass is observed in vancomycin-resistant *enterococci*. Normally, vancomycin binds to the terminal D-alanine-D-alanine residues of pentapeptide precursors, thereby preventing the synthesis of the bacterial cell wall (Darby et al. 2023). However, vancomycin resistance in *enterococci* arises from the acquisition of the *van* cluster, with the *vanA* gene cluster being commonly found in clinically vancomycin-resistant strains. The expression of genes within the *vanA* gene cluster can lead to the abnormal synthesis of peptidoglycan precursors (Miller et al. 2014). As a result, instead of binding to the D-alanine-D-alanine residues, vancomycin altenatively binds with reduced affinity to terminal D-alanine-D-lactate or D-alanine-D-serine residues (Miller et al. 2014).

1.2.4.3 Drug inactivation

Bacteria employ two primary methods to inactivate drugs: degradation of the drug itself or transfer of a chemical group to the drug (Reygaert 2018). One prominent group of drug-hydrolyzing enzymes is the β-lactamases, which encompass a wide range of enzymes capable of hydrolyzing β-lactam drugs. Another example of drug inactivation through hydrolysis is seen with tetracycline, which can be hydrolyzed by the *tetX* gene (Blair et al. 2015). As for the drug inactivation caused by the transfer of a chemical group, it frequently involves the transfer of acetyl, phosphoryl, or adenyl groups (Blair et al. 2015). Numerous transferases have been identified that facilitate this mechanism. Acetylation, in particular, is widely employed and has been observed against aminoglycosides (Ramirez and Tolmasky 2010), chloramphenicol (Schwarz et al. 2004), streptogramins (Reygaert 2018), and fluoroquinolones (Robicsek et al. 2006).

1.2.4.4 Increase drug efflux

Bacterial efflux pumps play a significant role in the intrinsic resistance of Gram-negative bacteria to many antibiotics. They actively transport antibiotics out of the bacterial cell and contribute to the resistance of Gram-negative bacteria to drugs that are typically effective against Gram-positive bacterial infections (Reygaert 2018). When these efflux pumps are overexpressed, they can confer high levels of resistance to antibiotics that were previously clinically useful. Bacteria isolated from patients that overexpress efflux pumps have been studied since the 1990s and include *Enterobacteriaceae* (Everett et al. 1996), *P. aeruginosa* (Pumbwe and Piddock 2000) and *S. aureus* (Kosmidis et al. 2012). Mutations in the regulatory network controlling efflux pump expression can lead to the high-level expression of efflux genes observed in multidrug-resistant bacteria. These mutations can occur within a local repressor, a global transcription factor, and intergenic sites that impact the expression of pump genes or their regulators (Blair et al. 2015).

1.2.5 Dissemination of antibiotics and ARGs in the environments

The entry of antibiotics into the environment through diverse pathways is causing concern due to its role in promoting the spread of AR beyond regions with limited regulatory measures. An instance of this is when antibiotics and their metabolites are discharged from hospitals via the urine and feces of patients,

resulting in the presence of these substances in hospital wastewater outlet (Berglund 2015). Likewise, antibiotics enter the wastewater treatment system as a result of individuals consuming antibiotics in their households, following which antibiotics have the potential to be deposited in sludge that is subsequently used as fertilizer on fields. Alternatively, they can be released directly into nearby surface waters or wetlands (Scholz and Lee 2005; Wu et al. 2015). In addition to other routes, antibiotics are commonly used for therapeutic purposes or as growth promoters in livestock and poultry. These types of environments are likely resistance hotspots where ARGs proliferate and new resistant strains are created by HGT. As a result, antibiotics and their metabolites can be disseminated through animal excretion, eventually reaching fields and groundwater (Rahman et al. 2022). A notable aspect is that wherever antibiotics are distributed, it is highly likely that ARB will also be spread through the same pathways (Baquero, Martínez, and Cantón 2008). Consequently, these circumstances lead to environments where antibiotics, ARGs, ARB, and the environmental bacterial flora, which may also contain ARGs and potential ARGs, are intermingled (Berglund 2015). Upon entering the human body, these resistant bacteria can transmit their ARGs to the microbiome of the individual, thereby facilitating the spread of antibiotic resistance (Wellington et al. 2013).

1.2.6 Molecular methods for ARGs detection

To establish the AR and ARGs profile for implementing effective management strategies of their spread, most research uses a combination of phenotypic and genotypic antimicrobial susceptibility testing (AST) strategies (Humphries 2020) although there can be instances where multiple genes or genetic mechanisms can lead to the same phenotype.

Phenotypic methods involve evaluating the growth of microorganisms when they are exposed to an antimicrobial agent (Anjum, Zankari, and Hasman, 2017). These methods rely on in vitro growth in a controlled culture environment, meaning they can only be conducted after the microorganism has been isolated in pure culture. Phenotypic methods have limitations in detecting pathogens present at low levels and can exhibit imprecision, resulting in varying reproducibility of results. Nonetheless, it is important to note that conventional phenotypic methods are generally more cost-effective compared to novel genotypic diagnostics. They offer clear information regarding both resistance and susceptibility of

microorganisms and provide MIC values, which are used by some clinicians to make therapeutic decisions.

Genotypic methods detect the presence of antibiotic resistance genes or SNP mutations that predict AR (Shi et al. 2019). These tests typically yield results within a few hours, and in some cases, within minutes. Furthermore, they can be conducted directly on certain samples without the need for bacterial isolation or culturing (McDermott. et al. 2016; Yang et al. 2022). Genotypic methods such as PCR-only are routinely used by laboratories worldwide both for surveillance and for research and are irreplaceable in terms of cost and throughput compared with other molecular methods. However, methods such as WGS and metagenomic sequencing, which are increasingly being used to provide a more comprehensive picture of all ARGs, could supersede PCR in the future if the cost of WGS and metagenomic sequencing continues to fall and if analytical processes are further simplified. However, it is important to acknowledge that not all isolates can be cultured for WGS. This could be due to specific strain characteristics or limited microbial loads present in the samples. Although metagenomic sequencing allows for direct analysis of bulk samples without culturing process, it is important to note that its sensitivity can still be limited. In such circumstances, amplification-based sequencing is the priority for sketching the ARGs profile, particularly when specific targets are known. In Table 1.2 and following section *1.3.2*, there are more examples and discussions of ARGs detection.

Table 1.2 Identification of antibiotic resistance by molecular approaches in recent years

Technology	Absolute/relative abundance	Throughput	LOD of ARGs	Database or Primer	Host	Samples	References
qPCR	Relative (SYBR)	7 ARGs	1- 100 copies	12 primer pairs	-	Water	(Rocha et al. 2019)
	Relative (TaqMan)	14 ARGs	1 copy/μL	30 primer pairs	-	River	(Paulus et al. 2020)
	Absolute (SYBR Green)	5 ARGs	1.50 gene copies	10 primer pairs	-	Soil and manure	(Tolosi et al. 2021)
	Absolute (SYBR Green)	8 ARGs	1-180 copies	13 primer pairs	-	Manure	(Hembach et al. 2022)
HT-qPCR	Absolute (SYBR Green)	66 ARGs	3 copies/μL	296 primer pairs	-	Wastewater	(Jong et al. 2018)
	Absolute (SYBR Green)	294 ARGs and MGEs	-	296 primer pairs	-	River	(Chen et al. 2019)
	Absolute (SYBR Green)	295 ARGs and MGEs	2.52E-06 copies/16S	296 primer pairs	*Prevotella_9, Bacteroides*, *Thauera*	Sewage	(Ding et al. 2020)
	Absolute (SYBR Green)	375 ARGs, MRGs and MGEs	-	376 primer pairs	103 genera	Leachate	(Yang et al. 2022)
PCR	-	6 ARGs	-	6 primer pairs	*S. aureus*	Meat, milk and their products	(Seedy et al. 2017)
	-	oprD and exoA	-	2 primer pairs	*P. aeruginonsa*	clinical specimens	(Saleh and Motib 2023)
Metagenomic sequencing	Absolute	361 ARGs	0.167 copies/μL	SARG DB	46 species	WWTPs	(Yang et al. 2022)
	Relative (per Gb of reads)	377 ARGs	-	CARD DB	10 phylums	WWTPs	(Dai et al. 2022)
	Relative (ARGs reads in one million reads)	140 ARGs	-	CARD DB	Over 10 genera	Sea	(Su et al. 2022)
	Relative	234 ARGs		ARG-ANNOT DB	26 pathogens	Tracheal aspirate, mini-bronchoalveolar lavage	(Serpa et al. 2022)
WGS	-	65 ARGs	-	2546 ARGs in ResFinder, ARG-ANNOT and CARD DBs	*Salmonella*	Clinical, retail meats	(McDermott et al. 2016)
	-	15 ARGs	-	CARD, and ResFinder DBs	*S. aureus*	Clinical isolates	(Al-Trad et al. 2023)
	-	10 genes SNPs for azole resistance	-	-	*C. albicans*	Strain isolates	(Kumaraswamy et al. 2023)
	-	SNPs for 5 antibiotic resistance	-	-	*B. hyodysenteriae*	Strain isolates	(Vereecke et al. 2023)
Amplicon-based sequencing	-	13 ARGs	-	14 primer pairs	*N. gonorrhoeae*	Urine; urethral swabs	(Zhang et al. 2021)
	-	15 ARGs	-	15 primer pairs	-	Sewage	(Gibson et al. 2023)
DNA microarray	Relative	115 ARGs	20ng/μL	115 probes in NCBI	7 strains	Pure strains and clinical isolates	(Fu et al. 2012)
Machine deep learning	-	50 SNPs for 5 resistance types	-	-	*N. gonorrhoeae*	WGS data	(Shi et al. 2019)

1.3 Low-input strategies

Generally, a significant portion of ongoing microbiome research focuses on areas of the body that harbour high microbial concentrations, such as the gut and mouth (Selway, Eisenhofer, and Weyrich 2020). However, samples used for diagnostic testing often contain fewer microorganisms and are referred to as low microbial biomass samples. Low-biomass environments, which hold particular interest within the scientific community, encompass a wide range of ecosystems and are associated with various research interests. The investigation of these environments can be susceptible to contamination from background DNA, which can potentially overshadow the analysis and interpretation of results. Technical biases, such as PCR over-amplification, present additional challenges in accurately studying these samples. These biases can distort the representation of microbial communities, leading to potential inaccuracies and misinterpretations in the analysis. With the increasing research focus on low microbial biomass body sites, it is crucial to develop and implement new protocols and techniques that maximise the extraction of microbiological information, including pathogenicity and antibiotic resistance from these samples. Additionally, it is important to fully understand the limitations associated with the development of diagnostic tools based on the results obtained from low microbial biomass samples. The technologies I introduced below were based on the molecular techniques I discussed in sections *1.1.3* and *1.2.6*, and more applications for low-input technologies are listed in Table 1.3.

Table 1.3 Sensitive detections of pathogens and antibiotic resistance from various low-biomass samples by novel molecular approaches

Low-input technologies	Detected pathogens	Detected ARGs	LOD	Primers/samples/coverage/platform	References
WGA (MDA)	Microorganisms in five different biotops	-	-	Low biomass (Groundwater, drinking water); high biomass(raw sewage sludge, biofilms, treated biosolids)	(Probst et al. 2015)
	E. coli and *S. aureus*	-	1 bacterial cell	Blood samples	(Anscombe et al. 2018)
	microorganisms in air	-	-	Air samples	(James et al. 2021)
	Salmonella plasmid DNA	-	0.3–1.7 μg	Cover 11 plasmids (37-166kb)	(Yao et al. 2021)
HT-qPCR	-	66 ARGs and five transposase genes	10^{-4} ARGs/16S rRNA gene	Sediment samples of fish farms	(Muziasari et al. 2016)
	-	150 ARGs and five MGE genes	4.3 copies (0.06 copies per bacterial cell)	Phyllosphere and soil samples	(Zhu et al. 2020)
	SARS-CoV-2, and other five pathogenic viruses	-	8 copies/μL	22 primer pairs; Wastewater samples	(Malla et al. 2022)
Metagenomic sequencing	16 species	119 ARGs belonging to 15 antibiotic types (CARD)	-	MinION and Illumina platforms	(Xia et al. 2017)
	Bacterial community in ready to eat food	204 ARG subtypes belonging to 18 ARG types (CARD DB)	2.81x10E-5 ARG copy / 16S rRNA gene	Ready-to eat food; Illumina Novaseq platform	(Li et al. 2020)
	46 predominant putative pathogenic species	361 ARGs (SARG DB)	10cells/μL; 0.167 copies/μL	ONT and Illumina platforms	(Yang et al. 2022)
Multiplex-PCR sequencing	*N. meningitidis (11 housekeeping gene)*	-	-	Samples: 107 reference strains Applied Biosystems Prism 377 automated sequencer	(Maiden et al. 1998)
	Zika virus	-	Ct value: 35.9 (10 genome copies/μL)	Clinical samples; MinION & Illumina platforms	(Quick et al. 2017)
	Seoul virus	-	Ct value: 36.8	Patient serum samples; Illumina platform	(Kim et al. 2018)
	Dengue virus	-	Ct value: 37.9	Patient serum samples; MinION & Illumina platforms	(Stubbs et al. 2020)
	SARS-Cov-2	-	Ct>27	Clinical specimens; Illumina platform	(Kentaro et al.2020)
	SARS-Cov-2	-	Ct value : 38	MinION platform	(Tyson et al. 2020)
	SARS-Cov-2	-	Ct value : 37	MinION & Illumina platforms	(Lin et al. 2021)
	S. enterica (7 housekeeping genes for MLST; 3002 genes for cgMLST)		-	poultry, livestock, food, and human sources; Illumina NovaSeq platform	(Yan et al. 2021)
	X. fastidiosa(7 MLST genes)	-	Ct < 31	Olive leaves; MinION platform platform	(Faino et al. 2021)
	N. gonorrhoeae	13 ARGs	31 copies/reaction	14 primer pairs; MinION & Illumina platforms	(Zhang et al. 2021)
	SARS-Cov-2	-	Ct value : 33.15	36 primer pairs; MinION & Illumina platforms	(Liu et al. 2022)
ONT-ReadUntil	Microbial communities	124 ARGs	At least 8-fold enrichment	Mock communities and BALF clinical samples;	(Cheng et al. 2022)
	41 high-quality metagenome-assembled genomes	-	Double enrichment of rare taxa	MinION platform	(Sun et al. 2023)
	R. ornithinolytica	1,147 ARGs	Double enrichment	MinION and Flongle platforms	(Viehweger et al. 2023)

1.3.1 DNA extraction to enrich target DNA from environmental samples

1.3.1.1 DNA extraction

DNA extraction is a crucial step in enriching target genetic material from low biomass samples. Prior to DNA extraction, it has been observed that preserving samples using ethanol fixation, typically at a concentration of 50%, can be effective (Li et al. 2018), especially when shipping samples at ambient temperature for extended periods. This preservation method helps to maintain sample integrity and minimise degradation of microbial DNA, thereby preserving the maximal differences in microbial composition.

In general, there is a wide range of established DNA extraction protocols and commercially available kits for various sample types. However, it is crucial to carefully evaluate and select the most appropriate method based on the specific characteristics of the sample and the intended downstream application. Efforts to establish appropriate methods of DNA extraction from various environmental samples, e.g. faecal samples (McOrist, Jackson, and Bird 2002), soils (Dineen et al. 2010), and aquatic environments (Urakawa, Martens-Habbena, and Stahl 2010), have highlighted the importance of determining an optimal extraction method. This optimisation is crucial to minimise biases and ensure accurate molecular analyses.

Concerns with extracting DNA from Gram-positive bacteria include their relatively thicker cell wall compared to that of Gram-negative bacteria and the ability to form spores in some bacterial species. Hence, additional treatments such as chemical lysis and hot detergent have been suggested to improve spore lysis (Dineen et al. 2010). In environmental samples primarily composed of microbial cells and their products, such as wastewater inlet and activated sludge (Ye, Peng, and Li 2011), the presence of aggregated cells and associated exopolysaccharides can present a challenge to DNA extraction. These aggregates and exopolysaccharides can create physical barriers that hinder the penetration of cell walls through shearing or chemical reagents typically used in DNA extraction protocols (Davies et al. 1998). Previous studies have demonstrated that mechanical homogenization, particularly bead beating, can

enhance DNA extraction from samples containing complex microbial communities (Guo and Zhang 2013).

With bacterial monocultures, all of the DNA extraction methods yielded relatively pure DNA. When applied to environmental samples, however, DNA purity decreased, which could be attributed to the presence of contaminating products and humic acids in the samples. The low purity of the DNA can also influence the Nanodrop measurements; therefore, DNA yield was also measured with Qubit, whose readings indicated a much lower DNA quantity.

The higher DNA yield from environmental samples sometimes required further purification in order to obtain PCR-amplified products. Lear, Dong, and Lewis (2010) showed that concentrations of DNA amplified by PCR were not influenced by the concentrations of the extracted DNA; therefore, it may be advantageous to use methods such as FastDNA that provided low DNA yield, but that also removed contaminants such as humic acids that inhibit PCR amplification (Lear, Dong, and Lewis 2010).

1.3.1.2 DNA purification and size selection

The DNA obtained using selected commercial kits was typically characterised by short segments, often less than 20 kb. This outcome is likely due to the high shear force applied during the bead-beating processing step. Although such DNA fragments may not be suitable for the construction of nanopore long-read library, they were of sufficient length for the short-read sequencing.

The DNA and PCR amplicons can be purified to remove PCR inhibitors, remaining primers, as well as PCR enzymes and salts. Additionally, purification steps can achieve DNA size selection (Li et al. 2018; Lear et al. 2018). A variety of PCR purification approaches were used. The most commonly adopted PCR purification approaches were the Agencourt AMPure XP system (Pochon et al. 2015; Song et al. 2015), the Qiagen MinElute PCR purification kit (Stoeck et al. 2010; Thomsen et al. 2012), and the Promega Wizard SV Gel and PCR Clean-Up system (Keskin 2014). The AMPure XP method was used frequently in high-throughput sequencing studies and the Promega Wizard system only in Sanger sequencing studies,

whereas the Qiagen MinElute method was used for a range of applications (Li et al. 2018; Lear et al. 2018).

1.3.2 Technologies applied for low-biomass samples or samples containing low-target genes

Building upon the previous discussion on molecular methods for pathogenic bacteria and antibiotic resistance identification, it is important to recognise that certain techniques, such as WGS, may not be well-suited for directly analysing low-biomass samples or samples with low-target genes, especially without prior culturing. However, there are several molecular technologies specifically designed for handling low-biomass samples based on the previously discussed methods. These technologies leverage various approaches, such as target amplification, target sequencing, or single-cell sequencing, to overcome the challenges associated with limited genetic material in precious samples. These advancements in molecular technologies offer more sensitive and accurate analysis, opening up new possibilities for studying bacteria and antibiotic resistance in low-biomass samples.

1.3.2.1 High-throughput qPCR

qPCR is one of the most widely used methods to precisely quantify bacteria in complex ecosystems. The low throughput of conventional qPCR systems can be a limitation, particularly in cases where a large number of samples need to be processed. This limitation can lead to increased labor requirements and higher material costs. The development of high-throughput qPCR (HT-qPCR) has led to a reduction in the workload and material costs (i.e., PCR chemicals) with 100-nL per reaction (5184 reactions in total) using SmartChip Real-Time PCR System developed by TaKaRa. Gradually, HT-qPCR has opened up new fields of application. These include the investigation of synthetic bacterial soil communities (Kleyer, Tecon, and Or 2017), the determination of functional genes in soils (Crane et al. 2018), the quantification of pathogens in spiked feces and environmental water samples (Ishii, Segawa, and Okabe 2013), the quantification of bacteriophages of the species *Lactococcus lactis* and *Leuconostoc* spp. in cheese milk (Muhammed et al. 2017), and the monitoring of SARS-CoV-2 variants and other pathogenic viruses in wastewater (Malla et al. 2022).

Numerous studies have attempted to compare the HT-qPCR technique with other parallel ones. For example, one of the significant advantages of microarray technology over HT-qPCR is the ability to profile a large number of ARGs in a single run. With microarrays, it is possible to profile thousands of ARGs simultaneously, providing a comprehensive view of antibiotic resistance profiles in a sample (Waseem et al. 2019). However, microarrays suffer batch-to-batch variability and are gradually considered as relatively less sensitive and specific. This is further exemplified by the fact that microarray data needs additional validation by qPCR (Jeanty et al. 2010), however, HT-qPCR data does not need additional validations. HT-qPCR, on the other hand, certainly provides better detection limits than by metagenomic sequencing approach with identical yielded data size. For example, HT-qPCR has reportedly detected ARGs to the order of magnitude of 10^{-4} ARGs/*16S rRNA* gene (Muziasari et al. 2016; Manaia et al. 2020). Based on the average bacterial genome size range of approximately 3.67 to 5.56 million bases (diCenzo and Finan 2017), and considering the size of the *16S rRNA* gene as around 1,500 bp (Kao 1998), I can make an estimation. If 10^4 reads are necessary to detect a single copy of the *16S rRNA* gene, achieving the same detection limit for any ARGs as demonstrated by HT-qPCR would need minimally 10^8 reads during metagenomic analysis. Furthermore, in contrast to metagenomic approaches that involve complex bioinformatics tools and pipelines for data analysis, HT-qPCR offers the advantage of a less steep learning curve. The analysis of HT-qPCR data does not typically require extensive familiarity with intricate bioinformatics processes. Additionally, both metagenomic sequencing and microarray hybridization generally require more sample or DNA quantity. HT-qPCR has countered this limitation because it can perform reactions on a nanoliter scale thus consuming only a minute amount of DNA.

The development of HT-qPCR platforms has indeed changed the perception of conventional qPCR and its limitless utilities. Overall, this technique has proved to be very cost-effective as nanoliter scale reactions allow substantial savings in terms of consumables and reagents and also allow more efficient use of the available sample. All these features along with previous strongholds of conventional qPCR (sensitivity and specificity) have made HT-qPCR a technique of better choice in the context of gene detection, e.g. ARGs, especially for those precious samples containing low nucleic-acid quantities.

Nevertheless, HT-qPCR technology has some disadvantages as well. For example, the inability of HT-qPCR to optimise individual assays during a run as all assays would experience the same qPCR cycling conditions. This can be critical because the specific binding of primers often requires different annealing temperatures (Sipos et al. 2007). Also, the reactions at the nanoliter scale make it difficult to recover and sequence amplified products which is easily possible in conventional PCR.

1.3.2.2 MDA

Currently, MDA is the most widely used method for whole genome isothermal amplification. Originally developed for amplifying large circular DNA such as plasmids (Dean et al. 2001; Spits et al. 2006), MDA has now been adapted for amplifying whole genomic DNA (Thurlow et al. 2022). With just a few micrograms of DNA or even a single bacterial cell as the input, it can significantly increase the amount of genetic material available for analysis (Long et al. 2020), and can possibly help overcome challenges encountered in metagenomic approaches when studying low-biomass environments or low-abundance microbiomes, such as air and clean biotopes (Probst et al. 2015; James et al. 2021), ecological micro-niches (Stocker 2012), and single specimens of small organisms (Liu et al. 2011). The overall advantages of MDA have been verified in maintaining the quality of sequencing data and the abundance of species measurements in eight paired metagenomic samples and one titrated mixed control sample (Hammond et al. 2016), and in providing higher high-fidelity DNA yields as well as increasing the number of high fidelity DNA in low biomass samples (Ahsanuddin et al. 2017). However, the disadvantages have also been pointed out which consist of amplification bias (Zong et al. 2012), non-specific amplification (Blagodatskikh et al. 2017), low sensitivity, poor repeatability, a high error rate of operation, and external contamination (Nadal-Ribelles et al. 2019).

1.3.2.3 Metagenomic sequencing

DNA metagenomic characterisation can be employed in bulk samples that contain target genes present at low abundance. As reported, metagenomics integrates with the bioinformatic pipeline using various databases (RefSeq2, SARG, and CARD (Alcock et al. 2020)) could detect below 0.1 ARG copy per *16S rRNA* gene copy depending on different samples (Food and Agriculture Organization of the United

Nations 2013; Li et al. 2020; Qian et al. 2021) or even reached to 1x 10^{-5} ARG copies per *16S rRNA* gene copy in the ready-to-eat food sample (Yiming Li et al. 2020). The unique advantage of metagenomic profiling is its ability to analyse unknown sequences without prior knowledge or dependence on specific target sequences. This sets it apart from other technologies such as HT-qPCR and amplicon sequencing. However, metagenomic sequencing typically necessitates larger sample quantities for library preparation in order to obtain a sufficient number of reads for robust bioinformatic analysis. In particular, LRS may require more than 1000 ng of DNA. This has posed a persistent challenge for researchers, especially when dealing with limited and valuable samples. However, a recent groundbreaking study addressed this issue by utilising lambda DNA as a background carrier DNA during ONT WGS, thereby enhancing the efficiency of library preparation (Terrazos Miani et al. 2023).

1.3.2.4 Multiplex PCR integrated with NGS and TGS

When the cases focus on the specific pathogens in the environmental or clinical samples free of culture, metagenomic sequencing is not potentially the priority for measuring microbial diversity with high specificity, accuracy and sensitivity. For example, when attempting to sequence the Zika virus using a metagenomics approach, Quick et al. (2017) encountered difficulties in recovering an adequate number of viral sequences reads. Despite depleting human rRNA, the low levels of viremia below 1000 copies/μL of RNA posed a challenge in obtaining sufficient viral sequence data. Detection of pathogens through metagenomics is made more complex by specificity issues that arise from misclassification or contamination, nucleic acid stability, and high costs involved in data generation and analysis. To economically generate complete pathogenic genomes from environmental or clinical samples, targeted enrichment is necessary. This can be achieved through two approaches: indirect targeted enrichment via host nucleic acid depletion and direct targeted enrichment using oligonucleotide probes. These methods aim to enrich specific nucleic acid templates during library preparation, thereby enhancing sensitivity for relevant but less abundant target sequences, all while reducing costs.

Ultra-highly sensitive and specific multiplex-PCR-based targeted sequencing approach has been successfully used in enriching viral genomes such as Ebola (Quick et al. 2016), Zika (Faria et al. 2017; Grubaugh et al. 2017) and recently SARS-CoV-2 (Quick 2020). To facilitate the design of primers for

multiplex PCR, several existing tools have been developed, each with a specific focus. For example, Primaclade (Gadberry et al. 2005) focuses on designing minimally degenerate primers for comparative studies of multiple species. PrimerStation (Yamada, Soma, and Morishita 2006) is dedicated to designing human-specific multiplex PCR primers by searching the entire human genome database. Primique (Fredslund and Lange 2007) is designed to generate PCR primers specific to each sequence within a gene family.

Different from the conventional, long-amplicon, single-plex approach commonly in use at present (Kim et al. 2016), the tool called PrimalScheme using greedy algorithms was originally developed for field sequencing of two arboviruses, namely Zika and chikungunya virus (Quick et al. 2017), covering a target gene by amplifying the overlapped regions of the target with multiple short amplicons (400 bp by default). The PrimalScheme takes the nucleotide diversity within homologues into consideration so that the primers generated will be able to sequence variation within species. In addition, with precisely designed primer sets and optimised amplification conditions, the scheme can dramatically reduce the time and effort required for the experiment. More importantly, the highly multiplexed amplification of a target gene by short amplicons would promote the sensitivity of the amplification (Kim et al. 2016; Moezi et al. 2019; Li et al. 2020), making it better suited to low-quality samples. To further shorten the turnaround time, real-time nanopore sequencing was preferred to sequence the amplicons generated by the PrimalScheme multiplex PCR (PrimalPCR). Such novel multiplex PCR strategy has been successfully used in the source tracking of virus outbreaks, including yellow fever virus (Faria et al. 2018), West Nile virus (Hepp et al. 2018) and dengue virus (Hill et al. 2019), and is currently being widely applied to assess the transmission route of SARS-CoV-2 for pandemic management by public health agencies of major countries like UK, USA, Canada and China (Tyson et al. 2020; Lin et al. 2021; Liu et al. 2022). However, it is currently rarely used in bacterial infection tracking, in part due to the difficulty of recovering the larger genomes of bacteria. So far, only Zhang et al. (2021) have used this technology to detect 11 genes associated with antimicrobial resistance in *N. gonorrhoeae* directly from clinical samples, demonstrating the feasibility of the approach for the ST identification.

Noticeably, in multiplex-amplicon integrated sequencing, multilocus sequence typing (MLST) is a commonly employed method for bacterial strain typing during epidemiological assessments. It is a

sequencing-based method that utilises the sequences of internal fragments (typically around 400-500 bp) from a set of commonly used housekeeping genes, usually seven in number (Larsen et al. 2012). The advantages of MLST are that it provides typing data that are unambiguous, portable, more accurate and more discriminatory for most bacteria (Jolley and Maiden 2010), which makes MLST data more suitable for global epidemiological studies. Furthermore, MLST data can be used to investigate evolutionary relationships among bacteria (Floridia-Yapur et al. 2021). Nonetheless, due to the conservation in housekeeping genes, seven-loci-based MLST sometimes lacks the discriminatory power to differentiate bacterial strains (Yan et al. 2021). Consequently, some other typing methods, such as WGS-based mapping (Bosch et al. 2015), SNP-based approach (Reuter et al. 2013), and the core-genome multilocus-sequence typing (cgMLST) schemes with 500 or more genes (Moran-Gilad et al. 2015; David et al. 2016; Qin et al. 2016), have emerged to provide better discrimination among epidemiologically unrelated isolates. However, these schemes require to sequence a great many gene loci, resulting in cost and bioinformatic challenges in large-scale application.

In summary, multiplex-PCR-based sequencing offers numerous advantages, which include the following: (i) high specificity: the majority of sequence reads obtained through this method originate from the pathogen of interest rather than the host, resulting in a significant reduction in sequencing costs; (ii) high sensitivity: this approach enables good coverage of the target pathogen, even at low pathogen loads, enhancing the ability to detect and analyse rare or low-abundance sequences; (iii) simplicity: the design and application of new sets of primers for novel sequences are relatively straightforward, making it a flexible and adaptable method for the detection and characterisation of diverse pathogens and ARGs (Houldcroft, Beale, and Breuer 2017).

1.3.2.5 ONT adaptive sequencing with “ReadUntil”

Nanopore adaptive sampling is a tool for the enrichment of low-abundance species in metagenomic samples (Martin et al. 2022). The inclusion of adaptive sampling as a user-selectable option in the GridION control software has greatly expanded its adoption among users. To use the software, a file containing reference sequences is supplied, and the system can be configured to either deplete or enrich these sequences on a designated set of channels. To accomplish the adaptive sampling, the software

performs basecalling on the initial few hundred bases of each read and compares them with the target reference sequences. Depending on whether the software is configured to enrich or deplete, sequences that match or do not match the references are respectively retained or discarded.

Adaptive sampling presents a promising solution for selectively enriching species of interest within metagenomic samples. It relies on a straightforward library construction method, allowing samples to be processed within an hour without the necessity of amplification (Martin et al. 2022). Martin et al (2022) achieved enrichment up to 13.87-fold for the least abundant species in the longest read length library. Cheng et al. (2022) increased the microbial sequence yield at least 8-fold in all 21 sequenced clinical Bronchoalveolar Lavage Fluid (BALF) samples and precisely detected the ARGs at the species level. Sun et al. (2023) developed the metaRUpore pipeline, which aims to decrease the sequencing coverage of high-abundance populations while moderately increasing the genome coverage of rare taxa by approximately two-fold. This approach has proven effective in facilitating the successful retrieval of near-finished metagenome-assembled genomes of rare species.

1.3.3 Source tracking of epidemiology in the disease outbreak

Low-input strategies have significant applications in epidemiological studies, particularly during outbreak assessments involving low-biomass samples. From the past to the present day, outbreak investigation plays a crucial and challenging role in epidemiology and public health. It serves to identify the source of ongoing epidemics or pandemics and prevent further sporadic cases (Bryson et al. 2021). Conducting epidemiologic and environmental investigations continues to be valuable even after an outbreak has subsided, as it enriches the comprehension of the disease and plays a crucial role in preventing future outbreaks (Dworkin 2010). During an epidemic, infectious diseases have the potential to spread rapidly among a large number of individuals. An example of such an epidemic is the SARS outbreak in 2003, which resulted in the loss of nearly 800 lives worldwide (Wilder-Smith et al. 2020). However, when an outbreak expands to affect multiple countries or continents, it reaches the level of a pandemic. The COVID-19 pandemic, which emerged in 2019, has been particularly devastating, as reported by the WHO, with a death toll of nearly seven million people. This pandemic ranks among the most severe recorded in world history.

1.3.3.1 Employ phylogenetics to track disease outbreaks

At the onset of an outbreak, one of the most critical tasks is to identify the causative pathogen or perform bacterial strain typing (Didelot et al. 2012). These activities help in characterising and confirming the epidemiological linkage within the outbreak, as well as providing valuable insights into the dynamics of the pathogenic population. Once the causative pathogen in the outbreak has been identified, several fundamental questions can be immediately addressed regarding the pathogen itself. These questions include determining whether the pathogen is a novel or previously known infectious agent, and assessing the availability of diagnostics, vaccines, and therapeutics to combat it. At this stage, the generation of pathogen genomics information provides deeper insights into these questions by revealing molecular details that cannot be easily achieved with conventional tools. The molecular strategies, dominated by sequencing techniques, offer enormous support for determining transmission chains and understanding the pathogen at a genetic level.

During outbreak investigations, a key assumption is that minimal genetic differences among pathogen genomes suggest recent transmission or a common source (Didelot et al. 2012). In fact, it is often challenging to provide an exact answer to this question, and thus it is crucial to employ statistical methods of analysis that can accurately quantify uncertainties. Based on this, additional phylogenetics in the outbreak will provide a polishing level of disease details. Phylogenetics is indeed the scientific discipline that studies the evolutionary history and relationships among organisms or groups of organisms by analysing their genomic materials. It has been a powerful tool for microbial epidemiology to track down disease outbreaks globally (Yang and Rannala 2012). A common approach to phylogenetic analysis of the genetic diversity within a microbial population is to construct a phylogenetic tree using sampled microbial genomes, where the branches of the tree are typically interpreted in units of time (Martin 2002). This can provide estimates of the dates for various events, including the date of the last common ancestor at the root of the tree and the dates for each branching event (Grubaugh et al. 2019).

In the preceding sections, I have explored the process of pathogen identification and typing from low-microbial samples by employing a range of molecular tools that are invaluable during outbreak

assessments. When the PCR methods are primarily inclined towards detecting the presence of known pathogens during an outbreak, the sequencing-based approaches are more prevalent in detecting novel pathogens including the initial outbreak of SARS2 (Ksiazek et al. 2003), Middle East respiratory syndrome 3 (MERS3) (Zaki et al. 2012), Lujo virus (Briese et al. 2009), and Ebola virus during the 2013–2016 epidemic in West Africa (Holmes et al. 2016). In the Seoul virus outbreak in South Korea, RT-PCR and multiplex PCR-based NGS were simultaneously applied to establish the difference pattern between *SEOV* S segments and *SEOV* M segments in the phylogenetic tree (LOD of 36.8 Ct value), indicating a genome organization compatible with genetic exchanges in nature (Kim et al. 2018). In the latest large outbreak of mpox virus (MPXV), identified in May 2022 (Kraemer et al. 2022), shotgun metagenomics allowed the rapid reconstruction and phylogenomic characterisation of the first MPXV outbreak genome sequences, showing that this MPXV belongs to clade three and suggesting that the outbreak most likely originated from a single source (Isidro et al. 2022).

1.3.4 Troubleshooting of contamination and biases in low microbial biomass samples

The primary challenge when analysing low biomass samples is the occurrence of contaminating or exogenous DNA, which refers to DNA originating from sources other than the specific sample under investigation. In order to effectively control for contamination and biases, it is crucial to have a comprehensive understanding of when and how these factors occur. Contaminating DNA and biases have the potential to be invited at different stages throughout the process of sample preparation and analysis, with three primary sources being technicians, environments, and reagents or equipment. These sources of contamination typically arise from sampling procedures and the laboratory environment. In recent times, there has been a growing recognition of the presence of contamination and biases in low biomass samples. However, there is still a need for comprehensive inclusion and analysis of controls to be widely implemented and reported. Based on experience, potential solutions to address these issues may include: (i) incorporate controls from the sampling and laboratory environments, as well as for the equipment and reagents used. (ii) minimise the presence of microbial and human contamination being introduced into samples. (iii) maintain consistency and implementing randomization in sample collection. (iv) use quantitative laboratory methods. (v) incorporate bioinformatics approaches to assess and mitigate contamination.

1.4 Objectives of the thesis

This thesis aimed to leverage culture-free targeted sequencing technologies for characterising microbes and antibiotic resistance in low-biomass environmental samples within hospital environments. These technologies involved amplifying target genes from minimal input. Initially, the focus was on combining traditional V4-V5 *16S rRNA* Illumina amplicon sequencing with HT-qPCR to analyse dynamic patterns of bacterial communities, antibiotic resistance, and MGEs influenced by various environmental factors (such as seasonal changes, different mediums, or departmental variations) in the hospital setting. Subsequently, the thesis aimed to explore a novel amplicon sequencing method that integrated tiling multiplex PCR and MinION sequencing, which enabled rapid, sensitive, and efficient retrieval of genomic information from pathogenic bacteria, especially those challenging to culture in a lab setting. The genomic information obtained served multiple purposes, including identifying STs (pathogenicity) of specific pathogenic bacterial species, tracing strain sources, and understanding the development of antibiotic resistance through mutations or horizontal transfer.

By utilising these advanced culture-free targeted sequencing technologies, this thesis strived to greatly improve the comprehension of bacterial communities, antibiotic resistance trends, and the genomic profiles of pathogenic bacteria within low-biomass hospital settings without culturing. The discoveries hold promise for informing the development of effective strategies to manage and mitigate ARB infections in healthcare environments, and may even offer significant value for environmental epidemiological studies during outbreaks.

Chapter Two

Generic Materials and Methods

The entire PhD work employing the methodology outlined in the thesis, was undertaken exclusively by the author (myself).

2.1 Sample collection and pretreatment

All 217 samples used in *Chapter 3* and *Chapter 5* were collected at *Peking University Shenzhen Hospital*, Shenzhen, China (22.55N, 114.10E). The monthly sampling took place between July 2020 and January 2021, specifically on the 15th of each month. During each sampling event, a total of 31 samples were collected from nine distinct locations within the hospital. These locations encompassed two adjacent blood-test rooms, the wastewater treatment plant, and the departments of emergency, gynecology, ophthalmology, otolaryngology, pneumology, and surgery. The samples were categorised into different types, including tap water, p-trap water, air dust, surface swabs, wastewater inlet, and wastewater outlet. To maintain consistency and minimise potential sampling errors, the sampling time was fixed at 9 am for each sampling event and the collection points remained consistent throughout the seven-month period. The specific information for each sample is provided in Table S5.4. After collection, all samples were immediately transported to the laboratory within 30 minutes and were stored at 4°C. The pretreatment process was completed within one day, followed by the DNA extraction process, which was also finished within another one day.

Specifically, the water samples within the hospital indoor environments were collected from taps and plumbing p-traps (a U-shaped bend that is connected to the sink and filters water as it enters a plumbing system). Before collecting the p-trap water, the U-shaped bend of the p-trap was unplugged to allow the stored water in the bend to be discharged. Additionally, the wastewater samples from the inlet and outlet were gathered from the wastewater treatment plant situated 50 meters away from the outpatient building. Then, 500 ml of each type of water was collected in a sterile bottle. Next, bacterial cells were captured and collected by pouring the water through the filtration unit containing the sterile mixed cellulose esters (MCE) membrane with a pore size of 0.22 μm and a diameter of 47 mm. The filtration unit consisted of a funnel, a locking ring, a filter flask, and several rubber tubes connected to a water circulation vacuum pump (-0.098 Mpa). The captured membrane was promptly stored at 4°C for subsequent DNA extraction.

As for the air dust samples, 2 filters (30cm×30cm) of the air conditioner (AC) were collected from each studied department. The filters were washed several times with 1 L of sterile reverse osmosis (RO) water, which was filtered through 0.22 μm MCE membrane to collect microbe cells for subsequent DNA extraction.

In terms of surface samples, sterile cotton swabs were used for the collection. Specifically, three sterile cotton swabs, including one negative control, were used for each surface sample with a swabbing area of approximately 25cm^2 per sample. The swabs were preserved in the sterile phosphate-buffered saline (PBS) solution in a 15mL centrifuge tube at 4°C before subsequent DNA extraction procedure.

2.2 Bacterial culture

2.2.1 Bacterial culture collection

The pure cultures of *L. pneumophila* ATCC 33152 and *P. aeruginosa* ATCC 9027 were obtained from the Guangdong Microbial Culture Collection Center. The *P. aeruginosa* positive control used in *Chapter 4* was originally isolated from a clinical sample and was supplied by MicrobesNG, University of Birmingham, UK.

2.2.2 Activation of frozen culture

For bacterial culture, the appropriate growth medium was prepared and the necessary incubation conditions were created for reviving the strain. As for *L. pneumophila,* the optimal growing condition was at 37 °C for 24-48 hours anaerobically in the presence of 5% CO_2. The required mediums were buffered yeast extract broth (BYEB, Oxoid, USA) for liquid culture and buffered charcoal yeast extract agar (BCYE, Oxoid, USA) for solid culture. For *P. aeruginosa*, the optimal growing condition was at 37 °C for approximately 18 hours aerobically at 150 rpm shaking speed. The necessary mediums were Luria-Bertani (LB) broth (Sigma-Aldrich, USA) and LB agar (Sigma-Aldrich, USA).

To activate the frozen culture in the ampoule approximately 0.5ml of the required broth was firstly transferred into the ampoule. The bacteria were rehydrated for 5-10 minutes and then transferred and subcultured onto appropriate culture media. The solid medium was included as the blank control to detect any contaminants that may have been introduced as the ampoule was opened.

2.2.3 Nutrient broth culture

To make the broth solution, the required dosage (*L. pneumophila*: 21.65 grams of BYEB powder, *P. aeruginosa*: 25 grams of LB powder) of nutrient broth powder was added in 1L of distilled water. For *L. pneumophila*, broth pH was adjusted to 6.9 using a suitable pH adjuster, such as hydrochloric acid (HCl) or potassium hydroxide (KOH). After mixing and dissolving them completely, they were sterilised by autoclaving at 121°C for 15 minutes. After the autoclave, a single colony from the agar plate was selected using a sterile pipette tip, and then the tip was dropped and swirled into approximately 50 mL of liquid broth. Finally, the bacterial culture was incubated at the desired growth conditions for *L. pneumophila* and *P. aeruginosa* as described in section *2.2.2*. A blank control was included for the identification of microbial contamination in the samples as well. The inoculated bacteria could be used for the agar inoculation or for DNA extraction. For long-term storage of the bacteria, it was proceeded with creating a glycerol stock and preserved at -80°C.

2.2.4 Make nutrient agar plates

Nutrient agar and distilled water were measured out and were poured into a clean DURAN bottle for the homogenous mixture. For LB agar, the recipe called for 36 grams of dehydrated agar per 1L of distilled water. In terms of BCYE agar, 28 grams of agar were added to 1L distilled water, and 0.4 grams L-Cysteine solution (Sigma-Aldrich, USA) was then added to 1L prepared BCYE agar. Next, the agar bottle was autoclaved at 121°C for 15 minutes. When waiting for the agar to autoclave, sterile petri dishes were laid out on a heat-resistant counter. The lids were kept on as much as possible to avoid any contamination. Once agar came out from the autoclave, it was carefully poured into the petri dish until it was roughly $^{2}/_{3}$ of the way full. Following that, the agar was solidified at room temperature, which was not more

than 15 minutes. Once the agar solidified and cooled, they were stored upside down in the fridge until used.

2.2.5 Spread and incubate the bacteria in the agar (streak plate)

The desired inoculum volume for spreading, absorption, and calculations was set at 100 microliters (µl). After the application of the inoculum onto the agar surface, the sterile spreader was immediately placed in contact with the inoculum on the surface of the plate and positioned to allow the inoculum to run evenly along the length of the spreader. The goal was to evenly distribute the inoculum and to allow it to be absorbed into the agar. After spreading, it was not allowed to disturb plates for 10 to 20 minutes. Following that, the streak plate was incubated with suggested conditions of *L. pneumophila* and *P. aeruginosa*.

2.2.6 Enumeration

When enumerating colony-forming units (CFUs), plates with between 20 to 150 CFUs were used to calculate the number of CFUs/ml of the original sample. Typically, a dilution series was prepared, often a ten-fold dilution with eight series using PBS. Additionally, the UV spectrophotometer was used to adjust and determine the bacterial CFUs of 1 optical density (OD) value in 1mL broth. For example, for *P. aeruginosa*, 1 OD in 1ml broth equals 8.7×10^8 CFUs.

2.3 DNA extraction, purification, measurement, and storage

2.3.1 DNA extractions with cultured cells and water samples

DNA extraction was conducted using a variety of commercial DNA extraction kits. During the extraction process, a MIULAB horizontal vortex adapter was used, the applied FastPrep homogenisation instrument was JXFSTPRP-64, from Shanghai Jing Xin, China, and the required centrifuge was a Beckman Coulter Microfuge 20R, USA.

To start with the DNA extraction with cultured cells, the harvest cells growing in the broth suspension was firstly centrifuged for 5 minutes at 300× g in a 1.5 ml microcentrifuge tube. Then the supernatant was removed completely and discarded without disturbing the cell pellet. Next, the cell pellet was resuspended in PBS buffer to a final volume of 200 µl, which was ready for the downstream DNA extraction. There were two optional kits for genomic DNA extraction which were QIAamp DNA Blood Mini Kit (QIAGEN, Germany) and Easy-DNA kit (Thermo Fisher Scientific, USA) respectively, of which QIAamp DNA Blood Mini Kit performed better with more yielded DNA. As for the QIAamp DNA Blood Mini Kit, the "Appendix B for Cultured Cells" combining with the protocol "DNA Purification from Blood or Body Fluids" in the provided instructions were referred in this thesis. In terms of the alternative kit of Easy-DNA kit, the extraction process was shown in the provided protocol.

Regarding the DNA extraction from water samples/air-dust samples, the filter membrane was prepared as shown in section *2.1*. As for the surface swab samples preserved in the PBS buffer, they were dried at room temperature for a duration of 15 minutes before DNA extraction. Once the drying process was complete, the swab was carefully inserted into a bead tube/lysing tube provided in the DNA extraction kit. In the selection of DNA extraction kits for the environmental samples, three different kits were utilised for multiplex PCR tests, as described in *Chapter 5* of the thesis. The kits used were the Dneasy PowerSoil Pro Kit (QIAGEN, Germany), FastDNA^{TM} SPIN Kit for Soil (MP Biomedicals, USA), and Dneasy PowerWater Kit (QIAGEN, Germany). After conducting the multiplex PCR tests and evaluating the results based on the presence and quality of PCR bands, it was determined that the Dneasy PowerWater Kit consistently provided better results compared to the other kits (Figure 2.1). Consequently, the Dneasy PowerWater Kit was chosen as the preferred kit for the entire DNA extraction process for all types of collected environmental samples. The extraction procedure was carried out following the provided protocol specific to the Dneasy PowerWater Kit.

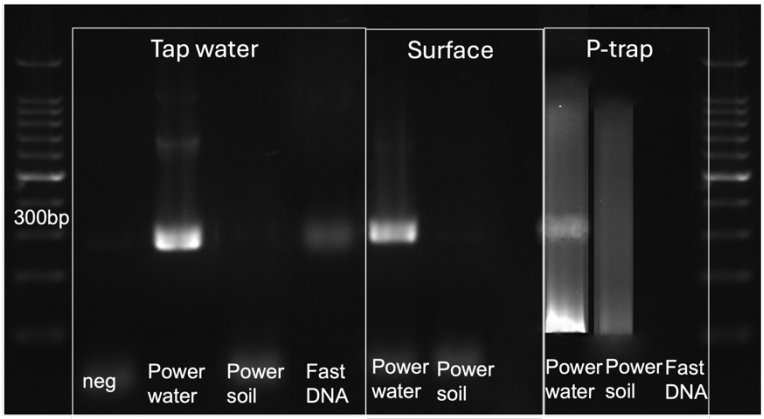


Figure 2.1 Evaluation of three DNA kits (Dneasy PowerWater Kit, PowerSoil Pro Kit and FastDNA™ SPIN Kit for Soil) on tap water, surface, and p-trap water samples. The multiplex-PCR panel in this test was *L. pneumophila* PrimalPCR panel with the target band at 300bp.

2.3.2 Magnetic bead DNA purification and size selection

AMPure XP SPRI beads (Coulter, USA) were used for the DNA purification. Generally, the volume ratio between beads and DNA sample determines the selection of DNA size (Table 2.1). Before starting, the DNA purification beads were warmed to room temperature for 30 minutes, and the fresh 70% ethanol was prepared for the wash step. The purification process was as follows. Firstly, the required volume of beads and DNA were mixed thoroughly by pipetting the entire volume 10 times. Then, the mixed samples were incubated at room temperature for 3-5 minutes for optimal binding. Next, the microfuge tube was placed onto the magnetic separation rack (NEBNext®, England) for 5 minutes to allow the solution to be clear and the beads to be collected on the magnet. After that, the cleared solution in the tube was carefully aspirated and discarded for the entire wash step. Subsequently, to prevent any disturbance to the beads, a volume of 200 µl of 70% ethanol was carefully dispensed to the side of the tube opposite to where the beads were located. The tube was then incubated at room temperature for a duration of 30 seconds. Following that, all the ethanol was aspirated and discarded from the tube. A total of washing round would be repeated for twice. Then, the tube was air-dried for 2-5 minutes to remove residual ethanol until they no longer appeared shiny, but not started to crack. Following the ethanol incubation step, the tube was carefully removed from the magnet. To resuspend the beads, 15-50 µl of the desired elution buffer (such as water, Tris, or Tris-EDTA buffer) was added to the tube. The resuspension process

involved pipetting the entire volume up and down for a total of 10 times. Subsequently, the tube was incubated at room temperature for a duration of 1 minute. After incubation, the tube was placed back onto the magnet, allowing the magnetic beads to separate and adhere to the magnet for a period of 2-5 minutes. Finally, once the beads collected on the magnet, the eluant containing the purified DNA was carefully transferred to a new tube. The purified DNA in the new tube was then quantified using appropriate methods (section *2.3.3*) to determine its concentration or yield.

Table 2.1 Volume ratio between AMpure beads and DNA samples for the DNA size selection

Volume (beads: DNA)	DNA size selection
1.5	Above 100 bp
1.0	Above 200 bp
0.8	Above 300 bp
0.5	Above 1kb
0.4	Above 1.5 kb
0.35	No DNA

2.3.3 DNA measurement by Nanodrop or Qubit instrument

The DNA measurement was conducted by Qubit instrument with Qubit® dsDNA High Sensitivity Assay Kits (Thermo Fisher Scientific, USA), or Nanodrop One Spectrophotometer (Thermo Fisher Scientific, USA) instrument. The measurement steps were shown in the provided protocols of the instruments.

2.3.4 DNA storage

For long-term storage of DNA, it was recommended to store it at temperatures ranging from -90°C to -15°C if the elution buffer does not contain EDTA. However, when the elution buffer contains EDTA, the DNA can be safely stored at frozen temperatures ranging from -80°C to -20°C.

2.4 *16S rRNA* Illumina amplicon sequencing

The whole process of the *16S rRNA* Illumina amplicon sequencing was described in *Chapter 3*. The Illumina PE250 HiSeq sequencing here was conducted by Novogene Co., USA.

The overall bioinformatics commands for PE250 reads are as follows.

(1) Quality control with fastp v0.23.2-3 (Chen et al. 2018)

```
fastp -i forward.fq -o foward.qc.fq -I revers.fq -O revers.qc.fq
```

(2) Merge PE reads with fastq-join v1.3.1-5 (Aronesty 2013)

```
fastq-join -p 2 -m 20 1.fastq 2.fastq -o merge.fastq
```

(3) Extract barcodes under QIIME1 v1.9.1 environment (Caporaso et al. 2010) (QIIME2 has succeeded QIIME1 during the thesis writing)

```
extract_barcodes.py -f merge.fastq -m mapping.txt -o barcode -c barcode_paired_stitched --bc1_len 6 --bc2_len 6 -a --rev_compliment_bc2
```

(4) Split libraries under QIIME1 v1.9.1 environment (Caporaso et al. 2010)

```
split_libraries_fastq.py -i barcode/reads.fastq -b barcode/barcodes.fastq -m mapping.txt -o demult/ -q 20 --max_bad_run_length 3 --min_per_read_length_fraction 0.75 --max_barcode_errors 0 --barcode_type 6
```

(5) Cut primers using cutadapt v1.9 (M. Martin 2011)

```
cutadapt -g forward.primer.fa -a complement_forward_primer.fa -e 0.15 --discard-untrimmed seqs.fna -o amplicon.fa
```

(6) Convert to USEARCH format

```
sed 's/ .*/;/g;s/>.*/&&/g;s/;>/;barcodelabel=/g;s/_[0-9]*;$/;/g' amplicon.fa > seqs_usearch.fa
```

(7) Unique the reads with USEARCH v11 (Edgar 2010)

```
usearch -fastx_uniques seqs_usearch.fa -fastaout seqs_unique.fa -minuniquesize 2 -sizeout
```

(8) Clustering operational taxonomic units (OTUs) with USEARCH v11 (Edgar 2010)

```
usearch -cluster_otus seqs_unique.fa -otus otus.fa -uparseout uparse.txt -relabel Otu
```

(9) Remove chimera with USEARCH v11 (Edgar 2010)

```
usearch -uchime2_ref otus.fa -db rdp_gold.fa -chimeras otus_chimeras.fa -notmatched otus_rdp.fa -uchimeout otus_rdp.uchime -strand plus -mode sensitive -threads 96
```

(10) Filter the sequence with chimeras under QIIME1 v1.9.1 environment (Caporaso et al. 2010)

```
grep '>' otus_chimeras.fa|sed 's/>//g' > otus_chimeras.id
filter_fasta.py -f otus.fa -o otus_non_chimera.fa -s otus_chimeras.id
grep '>' -c otus_non_chimera.fa
```

(11) Rename OTU

```
awk 'BEGIN {n=1}; />/ {print ">OTU_" n; n++} !/>/ {print}' otus_non_chimera.fa > rep_seqs.fa
```

(12) Generate OTU table with USEARCH v11 (Edgar 2010)

```
usearch -usearch_global seqs_usearch.fa -db rep_seqs.fa -otutabout otu_table.txt -biomout otu_table.biom -strand plus -id 0.97 -threads 10
```

(13) Assign taxonomy under QIIME1 v1.9.1 environment (Caporaso et al. 2010)

```
assign_taxonomy.py -i rep_seqs.fa -r 97_otus.fasta -t 97_otu_taxonomy.txt
```

(14) Convert OTU table to biom format under QIIME1 v1.9.1 environment (Caporaso et al. 2010)

```
biom convert -i otu_table.txt -o otu_table.biom --table-type="OTU table" --to-json
```

(15) Add the taxonomy to the last column of OTU table named taxonomy under QIIME1 v1.9.1 environment (Caporaso et al. 2010)

```
biom add-metadata -i otu_table.biom --observation-metadata-fp rep_seqs_tax_assignments.txt -o otu_table_tax.biom --sc-separated taxonomy --observation-header OTUID,taxonomy
```

(16) Convert biom format to txt under QIIME1 v1.9.1 environment (Caporaso et al. 2010)

```
biom convert -i otu_table_tax.biom -o otu_table_tax.txt --to-tsv --header-key taxonomy
```

(17) Summarize the OTU table under QIIME1 v1.9.1 environment (Caporaso et al. 2010)

```
biom summarize-table -i otu_table_tax.biom -o otu_table_tax.sum
```

2.5 HT-qPCR

Here, the instrument to conduct HT-qPCR was the TaKaRa SmartChip Real-Time PCR System (#Cat:64022, Japan) which is composed of the SmartChip MultiSample NanoDispenser (MSND) and the SmartChip Real-Time PCR Cycler.

The qPCR reagent was LightCycler 480 SYBR Green I Master (Roche, Switzerland). It was important to minimise light exposure to the SYBR Master. The blank chips, blotting paper, and cycling film were from SmartChip® MyDesign Kit (430-000244). The 384-well plates were MSND 384-Well Source Plate and Seals (430-000025).

A configuration of "120 assays × 42 Samples" was used for assay and sample assignment, taking into account the availability of 119 target primer sets in the study. The individual primer (forward primer or reverse primer) concentration applied in the laboratory was 10 μM, so the primer pair concentration was 5 μM. In Table 2.2 and Table 2.3, the sample and assay slot assignments in the 384 plates were presented. In cases there was a shortage of 120 assays or 42 samples, it was permissible for the corresponding slots to remain empty.

Prior to the HT-qPCR, the instrument preparation steps such as checking helium tank pressure, checking the water bottle, filling the wash bottle, and running daily warm-up were conducted following the provided instrument protocol. In terms of the whole process, it included: (1) Preparation of 384-well Sample sourceplate and 384-well assay sourceplate according to the slot assignment of "120 assays × 42 Samples". The total volume of the sample PCR reagent mix was 788 μL (493 μL of SYBR Green I Master (2×) and 295 μL of nuclease-free PCR-grade water), of which 15.2 μL was distributed to each well in the 42-sample plate. The DNA volume to each 384-well was 3.8 μL with the optimum DNA concentration from 1-10 ng/μL. Moreover, in terms of the volume of assay PCR reagent mix, it was 1458 μL in total (911 μL of SYBR Green I Master (2×), 547 μL of nuclease-free PCR-grade water and 10.358 mg of sterile bovine serum albumin (BSA, MedChemExpress, USA)), of which 10.9 μL was distributed to each well in the 120-assay plate. Additionally, the volume of 5 μM primer pairs in each well was 2.7 μL. (2) Dispensed sample and assay reagent mixtures into chip and run reactions by MSND. (3) Ran the real-time PCR using the SmartChip Cycler with three steps. The first step was denaturing at 95°C for 10min. The second step had 40 cycles with denaturing at 95°C for 30s and collecting data at 60°C for 30s. The third step was collecting melt-curve data at 97°C.

The output amplification data table generated by the SmartChip qPCR analysis software 2.8.0.65 consisted of 30 columns, encompassing various information. These columns included details such as sample and assay information, Tm (melting temperature) information, specificity information, Ct value (cycle threshold), amplification efficiency, and other relevant data. The calculations of *16S rRNA copy* number and ARGs copy number were described in *Chapter 3*.

Table 2.2 Sample assignment in the 384-well plate for “120 assays × 42 samples” HT-qPCR

	1	2	3
A	S1	S17	S33
B	S2	S18	S34
C	S3	S19	S35
D	S4	S20	S36
E	S5	S21`	S37
F	S6	S22	S38
G	S7	S23	S39
H	S8	S24	S40
I	S9	S25	S41
J	S10	S26	S42
K	S11	S27	
L	S12	S28	
M	S13	S29	
N	S14	S30	
O	S15	S31	
P	S16	S32	

Table 2.3 Assay assignment in the 384-well plate for “120 assays × 42 samples” HT-qPCR

	1	2	3	4	5	6	7	8
A	AY1	AY17	AY33	AY49	AY65	AY81	AY97	AY113
B	AY2	AY18	AY34	AY50	AY66	AY82	AY98	
C	AY3	AY19	AY35	AY51	AY67	AY83	AY99	AY114
D	AY4	AY20	AY36	AY52	AY68	AY84	AY100	
E	AY5	AY21	AY37	AY53	AY69	AY85	AY101	AY115
F	AY6	AY22	AY38	AY54	AY70	AY86	AY102	
G	AY7	AY23	AY39	AY55	AY71	AY87	AY103	AY116
H	AY8	AY24	AY40	AY56	AY72	AY88	AY104	
I	AY9	AY25	AY41	AY57	AY73	AY89	AY105	AY117
J	AY10	AY26	AY42	AY58	AY74	AY90	AY106	
K	AY11	AY27	AY43	AY59	AY75	AY91	AY107	AY118
L	AY12	AY28	AY44	AY60	AY76	AY92	AY108	
M	AY13	AY29	AY45	AY61	AY77	AY93	AY109	AY119
N	AY14	AY30	AY46	AY62	AY78	AY94	AY110	
O	AY15	AY31	AY47	AY63	AY79	AY95	AY111	AY120
P	AY16	AY32	AY48	AY64	AY80	AY96	AY112	

2.6 Bioinformatic tools

In addition to the bioinformatic tools discussed in *Sections 2.4* and *Chapter 4*, this section presents other bioinformatic tools used in the project. The versions of the bioinformatic tools listed in the thesis represent the specific versions utilised during the analysis phase. It is worth noting that there may have been more recent and updated versions available at the time of writing the thesis.

(1) Install Conda and Pip in Linux system

Conda v3-4.0

```
wget http://repo.continuum.io/archive/Anaconda3-4.0.0-Linux-x86_64.sh
bash  Anaconda3-4.0.0-Linux-x86_64.sh
```

Pip for python v3

```
apt install python3-pip
```

(2) Reads Manipulation

Seqkit v2.4.0 (Shen et al. 2016)

Usage for example:

(i) Transform sequences to the reverse strand

```
seqkit seq -r <fasta>
```

(ii) Get subsequences by ID list.

```
seqkit subseq <fasta> <id_list>
```

(iii) Sort sequence by alphabet order of fasta ID.

```
seqkit sort -n -t dna <fasta>
```

bedtools 2.30.0 (Quinlan and Hall 2010)

Compute histograms(default), per-base reports (-d) and BEDGRAPH (-bg) summaries of feature coverage (e.g., aligned sequences) for a given genome.

Usage for example:

```
bedtools genomecov  -ibam <bam>
```

fastp v0.23.2-3 (Chen et al. 2018)

A tool designed to provide fast all-in-one preprocessing for FastQ files.

Usage for example:

Quality filtering for pair-end fastq reads

```
fastp -i in.R1.fq.gz -I in.R2.fq.gz -q 15 -o out.R1.fq.gz -O out.R2.fq.gz
```

(3) Reads mapping/alignment

Minimap2 v2.17 (Li 2018)

Usage For example:

(i) Output alignments in the BAM format:

```
minimap2 -a ref.fa query.fq > alignment.sam
samtools faidx ref.fa
samtools view -bS -o alignment.bam alignment.sam
samtools index alignment.bam
```

(ii) Acquire alignment reads:

```
samtools view -c -F 260 alignment.bam
```

(iii) Map long noisy genomic reads for Oxford Nanopore reads

```
minimap2 -ax map-ont ref.fa ont-reads.fq > aln.sam
```

(iv) Find overlaps between long reads for Oxford Nanopore reads

```
minimap2 -x ava-ont reads.fq reads.fq > ovlp.paf
```

(v) Full genome/assembly alignment

```
minimap2 -ax asm5 ref.fa asm.fa > aln.sam
```

Blast v2.12.0 (Altschul et al. 1990)

Usage for example:

nucleotide to nucleotide BLAST

```
makeblastdb –in ref.fa  -dbtype nucl -parse_seqids
blastn -db ref.fa -query query.fa -outfmt 6 -out out.tab
```

hmmsearch/hmmscan v3.3 (Johnson, Eddy, and Portugaly 2010)

Usage for example

```
hmmsearch/hmmscan  –tbout out.tab –cut_tc –cpu 5  --notextw  ref.hmm
seqfile
```

Bowtie v1.3.1 (Langmead et al. 2009)

An ultrafast, memory-efficient short-read aligner.

Usage for example:

(i) Align short reads to genome

```
bowtie-build  genome.fa  genomeindex
bowtie  genomeindex  short-read.fq
```

(ii) Finding variations with samtools v1.10

```
bowtie -S genomeindex short-read.fq  snp.sam
samtools view -bS -o snp.bam snp.sam
samtools sort snp.bam snp.sorted.bam
samtools pileup -cv -f genomes.fa snp.sorted.bam
```

cd-hit v4.8.1 (Fu et al. 2012)

Usage for example:

Cluster similar DNAs into clusters with 100% similarity threshold

```
cd-hit-est -i query.fa -c 1 -o query.out
```

Mafft v7.487 (Katoh et al. 2002)

```
mafft input.fa > output.fa
```

Roary v3.11.2 (Page et al. 2015)

Usage for example:

Explore core genes and accessary genes from genebank files of pan genome

```
Roary –e –mafft *.gff
```

(4) Reads de novo assembly

Unicycler v0.5.0 (Wick et al. 2017)

Usage for example:

(i) Illumina-only assembly:

```
unicycler -1 short_reads_1.fastq.gz -2 short_reads_2.fastq.gz -o output_dir
```

(ii) Long-read-only assembly:

```
unicycler -l long_reads.fastq.gz -o output_dir
```

(iii) Hybrid assembly:

```
unicycler -1 short_reads_1.fastq.gz -2 short_reads_2.fastq.gz -l long_reads.fastq.gz -o output_dir
```

Canu v2.2 (Koren et al. 2017)

Usage for example:

Assemble nanopore reads:

```
canu -p output -d output-oxford_dir genomeSize=4.8m maxInputCoverage=100 -nanopore input.fastq
```

(5) Reads annotation

Prodigal v2.6.3 (Hyatt et al. 2010)

Fast, reliable protein-coding gene prediction for prokaryotic genomes.

Usage for example:

```
prodigal -i my.genome.fna -o my.genes -a my.proteins.faa
```

Prokka v1.14.5 (Torsten Seemann 2014)

Annotate bacterial, archaeal and viral genomes quickly.

Usage for example:

Have curated genomes I want to use to annotate from:

```
prokka --proteins myannotation.gbk --outdir myoutput --prefix my_contigs.fa
```

(6) ARGs identification

ARGs-OAP v2.3.2 tool (Yang et al. 2016) to predict ARGs with metagenomic data.

```
args_oap stage_one -i inputdir -o outputdir -f fa -t 8
args_oap stage_two -i output -t 8
```

In the input directory, there are fasta-files. For paired-end files, the forward/reverse reads end with _1 and _2 or _R1 and _R2 (followed by .format, see -f, .gz optional), otherwise they will not be considered as a single sample.

After stage_one, a metadata.txt file can be found in output. It summarizes the estimated 16S and cell copy numbers in each sample. After stage_two, the normalised ARGs copies per 16S/cells or hits per million reads will be shown in several *_normalised_*.txt files.

(7) Plasmids identification

Platon v1.5.0 (Schwengers et al. 2020)

Identification and characterisation of bacterial plasmid contigs from short-read draft assemblies.

```
wget https://zenodo.org/record/4066768/files/db.tar.gz
tar -xzf db.tar.gz
platon --db <db-path> --mode {sensitivity,accuracy,specificity}
genome.fasta
```

(8) Sourcetracking

Sourcetracker2 v2.0.1 (Knights et al. 2011)

```
biom convert -i otu_table.txt -o otu_table.biom --to-hdf
sourcetracker2 gibbs -i otu_table.biom -m map.txt -o output/
```

(9) Construct phylogenetic tree

Trim the nucleotide alignment gap using Gblocks v0.91.1 (Castresana 2000)

```
Gblocks <alignment> -t=d -b5=h
```

Fasttree v2.1 (Price, Dehal, and Arkin 2010)

```
fasttree -nt -gtr alignment.fa > tree.output
```

(10) Graphic and visualisation

The graphical user interface (GUI) desktop software for graphics and visualisation used are as follows.

(i) View and edit sequences: Jalview v2.10.5 (Waterhouse et al. 2009), MEGAX v10.2.6 (Huson et al. 2007), Brig v0.95 (Alikhan et al. 2011)
(ii) View and edit assembly: Bandage v0.8.1 (Wick et al. 2015)
(iii) View alignment and SNPs: Tablet v1.19.09.03 (Milne et al. 2013)
(iv) View and edit phylogenetic tree: Figtree v1.4.4 (Rambaut, 2012)
(v) View and edit network: Gephi v0.10.0 (Bastian, Heymann, and Jacomy 2009)

(11) Databases (DBs)

The databases applied in the project included CARD (Alcock et al. 2020), COG (Tatusov et al. 2000), NCBI (Sayers et al. 2022), PubMLST (Jolley et al. 2018) and SILVA (Quast et al. 2013).

2.7 Statistical analysis

The statistical analysis performed in the thesis included Venn analysis, principal component analysis (PCA), analysis of similarities (ANOSIM), Random Forest analysis, Mantel test, Procrustes analysis, linear discriminant analysis (LDA), abundance and diversity analysis, co-occurrence network analysis, Robinson-Foulds distance analysis and SNP distance analysis. The detailed description is in *Chapter 3* and *Chapter 5*. The packages used and scripts created are in Appendix 3.

Chapter Three

Biodiversity and Seasonal Dynamics of the Air-Dust and Water Microbiome in a Large Hospital

The following chapter has been submitted as the manuscript to *Applied Microbiology and Biotechnology* Journal. The available manuscript has been announced online in the Preprints.org. Doi: 10.20944/preprints202403.0001.v1.

Minor changes were made to the chapter for this thesis, which included individual discussion of seasonal, medium impacts on the compositions of bacterial communities and antibiotic resistance, and co-occurrence analysis between bacterial communities and antibiotic resistance.

Abstract

The hospital indoor environment has a crucial impact on the microbial exposures that humans encounter. Resistance to antibiotics is a mechanism used by bacteria to develop resilience in indoor environments, and the widespread use of antibiotics has led to changes in the ecological function of resistance genes and their acquisition by pathogens. By integrating the *16S rRNA* Illumina sequencing and HT-qPCR approaches with water and air dust samples across seven departments in *Peking University Shenzhen Hospital,* China, this chapter yields intriguing findings regarding the dynamic patterns of bacterial communities, antibiotic resistance, and MGEs. The first observation was that while the alpha-diversity of bacterial communities was highly heterogeneous in different hospital indoor environments, the seasonal variation played a vital role in shaping the beta-diversity. Another finding was the presence of widespread mobility of ARGs within the hospital, which was supported by the robust correlation between the abundance of ARGs and MGEs. In the source inferences conducted across seven departments, the primary origins of ARGs and MGEs were hypothesized to stem from the gynecology department and outdoor sources, ultimately accumulating in the blood test room. Further, when examining the association between bacterial communities and ARGs/MGEs, the Spearman correlation analysis reveals a significant correlation within the gynecology department. The results reiterate the importance of surveillance and monitoring of antibiotic resistance, specifically in vancomycin-resistant *Acinetobacter*, MDR-*Pseudomonas* spp. in addition to *Legionella* spp. in man-made water systems, and highlight the significance of understanding mobile ARGs such as *vanB* and *mcr-1*, as well as genetic elements like *Tp614* involved in gene transfer and recombination, and their impact on antimicrobial treatment efficacy.

Keywords

Hospital, *16S rRNA* amplicon sequencing, HT-qPCR, seasonal variation, water and air dust, indoor environment (department), bacterial community, ARGs and MGEs, HGT

3.1. Introduction

Hospital indoor environments are characterised by high infective risk, firstly cause of the compromised immunologic conditions of the patients that make them vulnerable to bacterial, viral, parasitological and fungal opportunistic infections (D'Alessandro and Fara 2017). It has been estimated that around two million patients per year in the United States acquire a nosocomial infection and tragically, at least 90,000 of them succumb to these infections (Pereira et al. 2016). Research has demonstrated that bacteria can persist and accumulate in various locations within the hospital's indoor environment, including white coats (Treakle et al. 2009), stethoscopes (Tang et al. 2011), air conditioners (Li et al. 2021), water faucets (Franco et al. 2020), and water p-traps (Kotay et al. 2020), far longer than previously believed (Kramer, Schwebke, and Kampf 2006). Influenced by the bacterial cell viability and bacterial load (Boyce 2007), the pathogen can be transmitted through contaminated hands or gloves of healthcare workers (Boyce 2007), direct contact with contaminated surfaces (Boyce 2007), splashing of pathogen-contaminated water on sterile goods (Kelsey 2013), and droplets for respiratory pathogens (Hota 2004). The presence of these reservoirs in the hospital environment may heighten the risk of acquiring a nosocomial infection.

Indeed, with extensive usage of antibiotic drugs on patients and routine application of antimicrobial chemicals for sanitation in hospitals, bacteria isolated from hospital environments are frequently resistant to antibiotics. Specifically, in a study conducted by Moges et al. (2014), it was found that a staggering 81.5% of the bacterial isolates from the hospital environment exhibited resistance to multiple antibiotics. Similarly, Phoon et al. (2018) noted that 62.7% of the identified species such as *Staphylococcus aureus*, *Klebsiella pneumoniae*, and *Acinetobacter baumannii* in the tertiary hospital environment were multidrug-resistant (MDR). These pathogens possess the ability to withstand the effects of multiple antibiotics, greatly limiting the available treatment options and significantly increasing the risk of healthcare-associated infection. Of particular concern about the presence of antibiotic resistance is the dissemination of ARGs in the hospital environment. Many of these genes are found on transposons, integrons or plasmids, which can be mobilised and transferred to other bacteria, belonging to the same or different species (Allen et al. 2010). These findings highlight the alarming reality of a "pre-antibiotic era".

Currently, there is a growing focus on studying the diversity, interaction and transmission of microbes and ARGs in various components of the hospital environment, including surfaces (Klassert et al. 2021), air dust (Li et al. 2021, Zhou et al. 2021), water (Sukhum et al. 2022), and during different seasons (Cassone et al. 2021). These studies have shed light on the influence of several factors on the composition of microbes and ARGs in the hospital indoor environment. Factors such as patient/room occupancy (Ramos et al. 2015; ElRakaiby et al. 2019), humidity, temperature (Choe, Smit, and Mermel 2019), air filtration (X. Li et al. 2021) and chemical residues (e.g., antibiotics) (Ben Maamar, Hu, and Hartmann 2020) have been identified as important contributors to microbial and ARG compositions in hospitals. However, despite the wealth of research in this area, there is still a limited number of studies that have thoroughly investigated the comprehensive patterns of the microbiome and antibiotic resistance in a department-specific manner. While numerous studies have examined microbial and antibiotic-resistance profiles in intensive care units (ICUs) (Bokulich, Mills, and Underwood 2013; Oberauner et al. 2013), only a few studies have directed their attention to different ward-room sites (Ramos et al. 2015; ElRakaiby et al. 2019) and various hospital departments (Li et al. 2021). Indeed, the diverse ecological interactions and conditions within different hospital sites have significant clinical implications, yet they have not been extensively explored.

As discussed in the *Introduction* chapter, the detection of bacteria and antibiotic resistance by conventional methods can be time-consuming and labor-intensive. However, molecular techniques provide rapid and sensitive alternatives for these investigations. For example, *16S rRNA* amplicon sequencing uses highly conserved bacterial regions for detecting diverse bacteria, while HT-qPCR is a relatively rapid and convenient method for the simultaneous evaluation of a large number of ARGs with low-quantity DNA samples. The objectives of this chapter were four-fold (workflow in Figure 3.1). Firstly, it aimed to demonstrate the changes in bacterial communities, ARGs, and MGEs in terms of abundance and diversity (alpha- and beta-diversity) influenced by different environmental factors, including seasonal, medium, and department variations within the hospital. Secondly, it denoted to identifying the driving factors shaping the compositions of bacterial communities, ARGs, and MGEs in the hospital indoor environment. Thirdly, it aimed to delve into the mobility of ARGs, and to ascertain whether the proportions of ARGs and MGEs sources could be identified when aggregating towards specific sinks within a particular department. Lastly, it targeted to explore the critical correlations

between bacterial communities and ARGs/MGEs, and to investigate the potential pathogenic bacteria associated with the spread of antibiotic resistance through the horizontal transfer. By utilising this comprehensive approach, the study aimed to provide a deeper understanding of the dynamics of bacterial communities, ARGs, and MGEs in the hospital indoor environment, shedding light on the factors influencing their composition and potential mechanisms of spread.

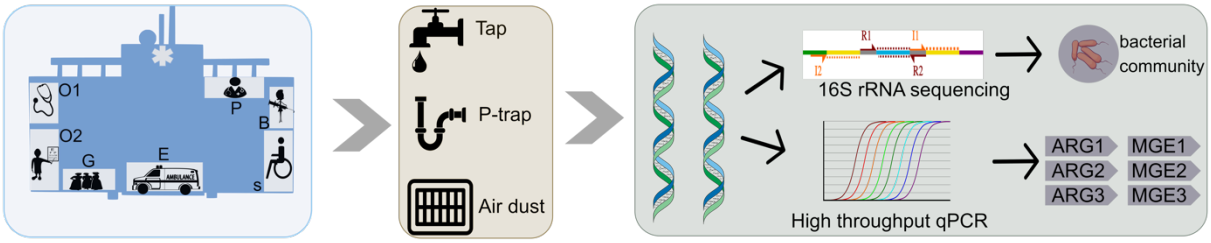


Figure 3.1 Overall workflow of the study. Key locations: O1: otolaryngology department. O2: ophthalmology department. G: gynecology department. E: emergency department. S: surgery department. B: blood-test room. P: pneumology department.

3.2 Methods

3.2.1 Sampling locations and collections

Among 217 collected samples in the hospital environment shown in *Chapter 2*, there were 64 samples (Table S3.1) yielding positive Ct values (Ct<31) in the HT-qPCR experiments (Table S3.3). For the subsequent hospital indoor study, 60 out of the 64 positive samples were utilised, and four samples collected from wastewater plants located outside the hospital were excluded from the analysis. The 60 samples constituted water and air dust samples in seven departments (blood test room, emergency, gynecology, ophthalmology, otolaryngology, pneumology and surgery departments) from early summer (July, August), late summer (September and October) to winter (November, December, January). The samples collection and pretreatment methods have been described in *Chapter 2*.

3.2.2 HT-qPCR

ARGs and MGEs were analysed by HT-qPCR using the Takara (previously Wafergen) SmartChip (#Cat:64022). Referring to the previous references using customed primers for detecting ARGs and MGEs in the hospital (Stedtfeld et al. 2018; Zhu et al. 2020), I selected 109 gene primer sets (Table S3.2) for 11 major classes of antibiotics, six transposase genes, five integrase genes, five plasmid genes and one *16S rRNA* gene primer set. The 11 major classes of antibiotics are aminoglycoside, amphenicol, beta-lactam, fluoroquinolone, multidrug, macrolide/lincosamide/streptogramin B (MLSB), sulfonamide, tetracycline, trimethoprim, vancomycin, and others (peptide, phosphonic acid, and rifamycin). A non-template negative control was used for each primer, and all qPCRs were performed in triplicate. The thermal cycle process amplification was described in *Chapter 2*. In addition, ARGs and MGEs relative copies represented the gene copies to *16S rRNA* copies ratio. We used the following formulas (Zhu et al. 2020):

(1) Gene copies = $10^{\frac{31\text{-Ct}}{10/3}}$

(2) Relative copies = $\frac{\text{Gene copies}}{\textit{16S rRNA}\text{ copies}}$

3.2.3 Bacterial *16S rRNA* sequencing of 64 samples

Bacterial community structures were determined by *16S rRNA* gene amplicon sequencing on a HiSeq platform (Illumina, USA) with PE250 strategy. The V4 to V5 region of bacterial *16S rRNA* gene was amplified with the universal primer set 515F (5′- GTGCCAGCMGCCGCGGTAA-3′) and 907R (5′-CCGTCAATTCMTTTRAGTTT-3′) (Turner et al. 1999), and labelled with unique barcodes (6-nucleotide barcodes) for each sample (Table S3.1). The components of the PCR solution mix and PCR program are shown in Table S3.4 and Table S3.5. Following PCR, the *16S rRNA* amplicon samples were grouped into 3 pools and sent for Illumina pair-end sequencing, each of which contained 2000 ng DNA individually. To ensure quality control of the raw reads obtained from Illumina sequencing, Fastp software (Chen et al. 2018) was used. The post-quality control reads, specifically those related to *16S rRNA* gene amplification, were then imported into the Quantitative Insights in Microbiology (QIIME 1) pipeline (Caporaso et al. 2010). Within this pipeline, several steps were performed, including merging pair-end sequences, extracting barcodes, splitting samples, and removing amplification primers. For

obtaining OTUs, USEARCH V11 (Edgar 2010) was employed with a 97% similarity threshold. Subsequently, taxonomic assignments were achieved by referencing the obtained OTUs against the Silva 138.1 database (Quast et al. 2013).

3.2.4 Statistical analysis

To analyse the changes and similarities in genus complexity among samples, several statistical methods and packages were employed. PCA and ANOSIM were conducted by the R3.5.3 VEGAN package (Dixon 2003) to reveal patterns and differences. The Random Forest test, implemented with the RandomForest package (Genuer and Poggi 2020), was used to identify the most discriminative variables between two sample categories. In ANOSIM, the R value indicates the degree of difference between sample groups, with $R = 1$ representing significant dissimilarity, $R = 0$ indicating no difference, $R > 0.75$ suggesting good separation, $R > 0.5$ indicating differences with some overlap, and $R < 0.25$ representing almost no differences. Moreover, to visualise the common genus and ARGs/MGEs, Venn graphs were generated by the VennDiagram package (Chen and Boutros 2011). For the investigation of the correlation between ARGs and MGEs, the Mantel test was performed with “mantel” function in the Vegan package. Additionally, a pairwise correlation between the abundances of targeting genes and OTUs (genus-level and species-level) was calculated using “corr.test” function, and it was considered statistically robust if the Spearman's correlation coefficient (ρ) was >0.6 and the p-value was <0.05. With the pairwise correlation values, the Gephi platform was used to generate networks (Bastian, Heymann, and Jacomy 2009). Further, to perform the source-tracking analysis, Sourcetracker2 (Knights et al. 2011) was used. Notably, all graphs during the statistical analysis were generated by Rstudio3.5.3 with the ggplot2 package (Wickham 2016). The details scripts for the statistical analysis were in Appendix 3.

3.3 Results and Discussions

3.3.1 Microbial profile

(1) Overall sketch of microbial profile in the hospital

From the *16S rRNA* gene amplicon sequencing, the raw data consisted of 598,1256 reads (2.99 Gbps) for all samples, and the quality-filtered (Q30) data consisted of 591,4227 reads with an average value of 92,410 reads per sample, corresponding to 1439 different bacteria OTUs after filtering and quality control. The mapped level of each taxonomic, i.e. phyla, class, order, family, genus, and species against the Silva 138.1 database (Quast et al. 2013) was 100%, 99.9%, 99.7%, 98.8%, 82.8%, and 9.98%, respectively. The limited percentage of identified bacterial species could be attributed to the utilisation of short reads that may not provide sufficient coverage for the complete identification. However, the rarefaction analysis conducted at the genus level demonstrates that the data obtained was adequate for taxonomic analysis. Despite the limitations of short reads, the analysis was able to provide reliable insights into the composition and diversity of bacterial genera within the sample (Figure S3.1).

At the genus level, there were 684 genera detected in total. Out of the 684 genera detected, a subset of 215 genera were identified as being particularly abundant, accounting for over 0.01% of the total *16S rRNA* gene sequences. These 215 genera were considered representative of the bacterial community and were selected for further analysis in subsequent steps. In terms of the leading genera, *Dechloromonas* (11.0%), *Pseudomonas* (7.59%), *Flavobacterium* (7.56%) and *Acinetobacter* (3.06%) took the leading places (Figure 3.2b). As *Pseudomonas* is the most concerned waterborne pathogen in healthcare facilities responsible for a wide spectrum of infections (such as pneumonia and urinary tract infections) in humans that can be associated with significant morbidity and mortality (Bonadonna, Briancesco, and Coccia 2017), *Dechloromonas* occurs frequently in the soil and wastewater treatment systems associated with nitrogen cycling roles (Petriglieri et al. 2021) and *Flavobacterium* exists more in soil and freshwater that may cause disease in freshwater fish (Bernardet et al. 1996). Regarding pathogenic *Acinetobacter*, it is particularly noteworthy due to the potential implications of its species (e.g. *A. baumannii*, *Acinetobacter nosocomialis* and *Acinetobacter seifertii*) as a source of infections (such as pneumonia, meningitis or bacteremia) in debilitated patients within the hospital (Harding et al. 2018). Other than the dominating concerned genera, several other genera of interest were observed at lower levels in Figure 3.2b. These included *Enterobacter* (0.806%), *Stenotrophomonas* (0.743%), *Bacillus* (0.418%), *Staphylococcus* (0.263%), *Klebsiella* (0.217%), *Mycobacterium* (0.561%), *Streptococcus* (0.116%), *Escherichia-Shigella* (0.0981%), *Legionella* sp. (0.0551%) and *Neisseria* (0.0116%). Notably, *Staphylococcus*, *Klebsiella*, *Acinetobacter*, and *Enterobacter* corresponded to the genus types of ESKAPE pathogens.

Furthermore, species within the *Stenotrophomonas*, *Bacillus*, *Mycobacterium*, *Legionella* and *Neisseria* genera have the potential to lead to infectious diseases such as urinary tract infections caused by *Stenotrophomonas maltophilia*, anthrax induced by *B. anthracis*, tuberculosis contributed by *Mycobacterium tuberculosis*, legionnaire disease infected by *L. pneumophila*, and gonorrhea resulted from *N. gonorrhoeae*. However, it is apparent that not all ESKAPE pathogens were detected in the hospital environment under study. For instance, *Enterococcus* genus was not classified. Although *Enterococci* are commonly associated with hospital acquired infections and are notorious for their resistance to vancomycin (Brinkwirth et al. 2021; Hammerum et al. 2024), the limitations of this study, such as DNA extraction efficiency and the sensitivity of V4-V5 *16S rRNA* amplicon sequencing, may have hindered the identification of *Enterococcus* in the samples.

From species-level respect, the results (Figure 3.2a) indicate the disquieting pathogens in the studied environments. Among the major species, *Pseudomonas* sp. emerged as the predominant species, comprising 45.4% of the samples, with *P. aeruginosa* specifically making up 0.966% of the total. Another alarming pathogen was the presence of *S. maltophilia*, which represented 5.12%. It has been documented that *S. maltophilia* commonly coexists and forms multispecies biofilms with *P. aeruginosa* (Alio et al. 2023). In addition, the study identified the presence of *K. pneumoniae* in 1.80% of the samples. Although the proportion may be relatively lower compared to other pathogens, the detection of *K. pneumoniae* is still significant due to its potential to cause serious infections (e.g. pneumonia, urinary tract and lower biliary tract) and its association with AR (Madebo et al. 2022). In the research conducted by Madebo et al. (2022), *P. aeruginosa* and *K. pneumoniae* were identified as the primary contaminants in the hospital as well. Regarding other troubled pathogens, *Bacillus* sp. (0.0754%) and *Mycobacterium* sp. (0.00233%) were also classified, although the precise species remained unknown. However, prevalent nosocomial pathogens such as *S. aureus*, *Clostridium difficile*, *Candida albicans* and *Neisseria meningitidis* were absent in the study, possibly due to a limited mapping rate at the species level during analysis.

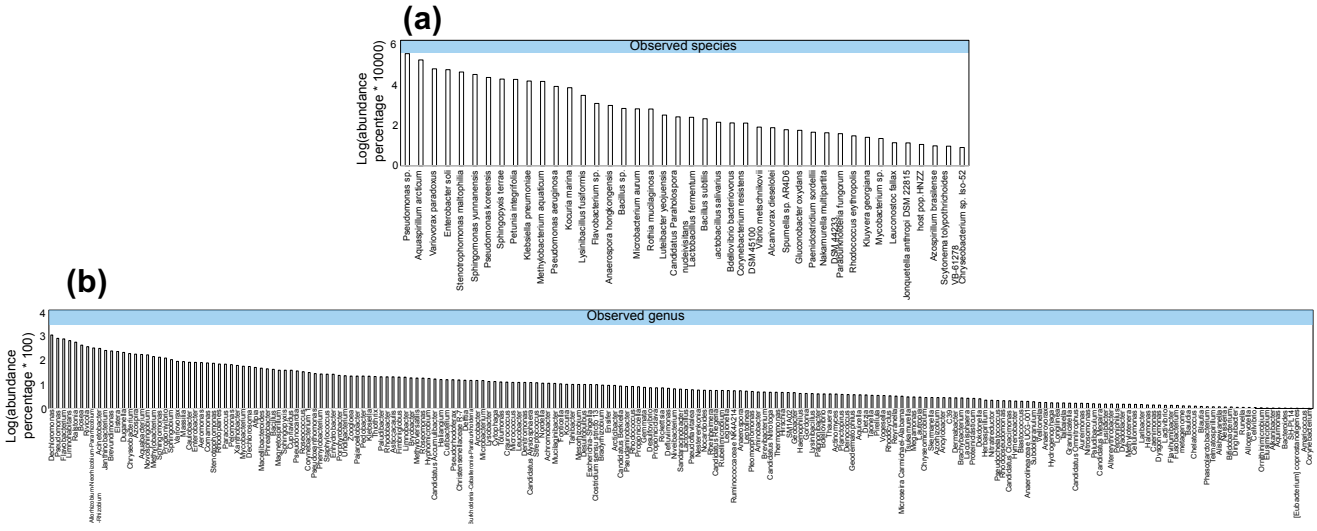


Figure 3.2 Profile of (a) 42 species and (b) 215 genera in the studied hospital environment with 60 samples, where the abundance was represented by the logarithm value of 10^4 percentage of clean sequencing reads mapped to the species level, and by the logarithm value of 10^2 percentage of clean sequencing reads mapped to the genus level, respectively.

(2) Microbial diversity influenced by seasonal patterns

There were notable differences in the dominating microbial contaminants between two seasons. *Limnohabitans* (6.36%) and *Ralstonia* (5.53%) were prominent in summer while *Dechloromonas* (6.48%) and *Flavobacterium* (6.45%) occupied the leading places in winter. As for the alpha diversity, it was observed that bacterial communities exhibited greater diversity during the summer season. This was indicated by a higher average Shannon index of 2.88 (Figure 3.3a). Previous research has reported that temperature and humidity can positively influence microbial diversity (Perencevich et al. 2008; Zhou et al. 2016). The higher temperatures and increased humidity experienced during the summer months create favorable conditions for microbial growth and activity. According to the meteorological information of China Meteorological Administration (China Meteorological Administration 2024), in the specific case of the studied hospital in Shenzhen city, the temperature at the collection points during the summer season, when the air conditioners are in operation, was approximately 32 degrees Celsius. However, it is worth noting that the early summer period may experience even higher temperatures, reaching over 34 degrees

Celsius. During this time, the relative humidity level exceeds 50%. As the summer progresses into the late season, the temperature and relative humidity levels tend to moderate. The temperature during this period ranges around 28 to 30 degrees Celsius, and the relative humidity is over 40%. In contrast, during the winter season when air conditioners are typically not in use at the collection points, the temperature remains approximately 15 degrees Celsius. The relative humidity during this time is above 20%. Given these environmental conditions, especially during summer, which are conducive to the proliferation and diversity of bacterial communities, it is legitimate to anticipate a greater bacterial diversity in the summer samples in comparison to other seasons.

In order to assess the compositional differences between seasonal groups, a PCA analysis was performed. The PCA results reveal the presence of three distinct clusters along the PC1 and PC2 axes, demonstrating clear separation between the summer and winter samples. This indicates that the beta-diversity within the studied environment was influenced by seasonal patterns. This separation was manifested by the ANOSIM result which yielded a p-value of 1e-04 and an R-value of 0.2719 (Table 3.1). The low p-value indicates that the differences between the summer and winter samples were statistically significant. The R-value of 0.2719 suggests a moderate level of dissimilarity between the seasonal groups. The differences in community compositions (genus-level) between summer and winter in the hospital could be attributed to various biological factors. These include changes in human carriage (Koliada et al. 2020), environmental niches (Chawla et al. 2023) and airborne transmission. With regard to the host carriage, the composition of the human microbiome can vary with seasonal factors. During summer, different activity patterns, increased outdoor time, and dietary habits can lead to changes in the human microbiome (Koliada et al. 2020), resulting in observed variations in communities in the hospital environment. Health conditions such as seasonal allergies (e.g. rhinitis) or infections (e.g. influenza) can also impact the microbiome (Choi et al. 2014). In relation to environmental niches, the hospital environment experiences seasonal variations that affect microbial communities. Factors like temperature, humidity, ventilation systems, and cleaning protocols differ between summer and winter, influencing the microbial composition. In addition, airborne transmission also plays a role, influenced by outdoor air quality, ventilation systems (air conditioners), and air circulation patterns. These factors contribute to variations in the types of microorganisms entering and circulating within the hospital environment throughout the seasons.

To gain a deeper understanding of the clusters within the summer season, the summer samples were divided into two subgroups: early summer and late summer. The results (Figure 3.3b) reveal that the two summer clusters were composed of different genera components, with early summer and late summer exhibiting distinct bacterial compositions. This differentiation within the summer season was strongly supported by the ANOSIM result with a p-value of 1e-04 and an R-value of 0.7346 (Table 3.1). To pinpoint the specific genera responsible for driving this separation between early summer and late summer, a RandomForest analysis was performed. The analysis reveals five key genera that contributed significantly to the observed differentiation (Figure 3.3c, 3.3d). These genera included *C39*, *Candidatus Alysiosphaera*, *Dysgonomonas, Roseomonas*, and *Taibaiella*. The presence or abundance of these genera likely played a crucial role in shaping the distinct bacterial compositions between the early summer and late summer clusters. Among these genera, except for *Roseomonas*, the other four were found to be more predominant in the late summer samples. It is worth noting that *Taibaiella* exhibited the most pronounced discrimination, showing a 23-fold higher relative abundance in the late summer samples. Furthermore, it is noteworthy that certain species within the *Roseomonas* genus, e.g. *Roseomonas mucosa* has been known to be opportunistic pathogens for humans (Dé et al. 2004). Their increased relative abundance in early summer samples may suggest a potential health risk during that time. Additionally, *Dysgonomonas* bacteria are recognised as causative agents of gastroenteritis, particularly in immunocompromised individuals (Ryan and Sherris 2013).

Overall, the observed alpha- and beta-diversity patterns underscore that the bacterial community structure was significantly influenced by seasonal variations. However, while diversity is generally considered beneficial for ecosystem health, it can also introduce challenges, particularly in healthcare settings where cross-infection is a concern for immunocompromised patients. As a result, understanding diversity dynamics can provide valuable insights into the ecological processes and environmental factors that shape microbial communities (Brown et al. 2004), contributing to a broader understanding of microbial ecology and ecosystem functioning.

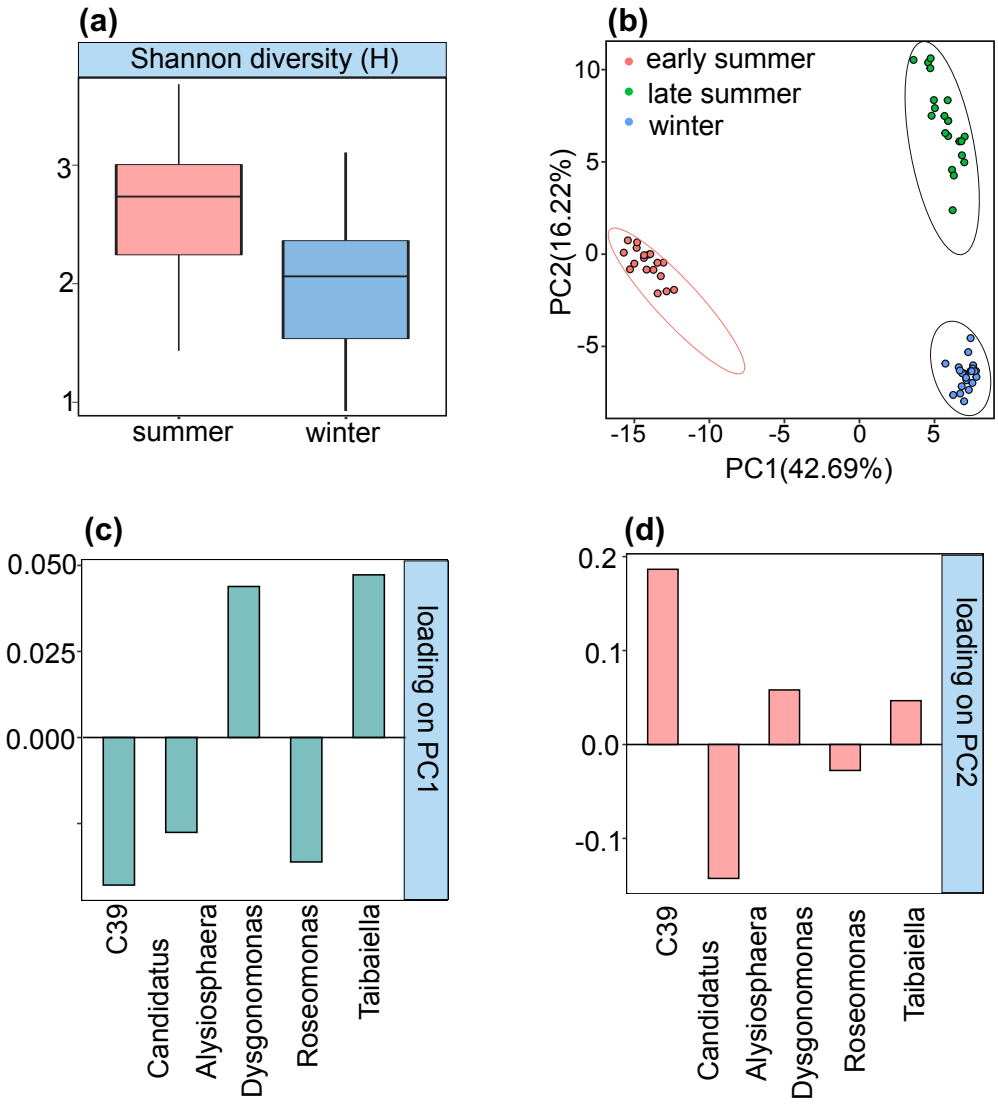


Figure 3.3 (a) Microbial (genus-level) alpha-diversity in two seasons with 60 samples, where the inside line in each bar represents the median value. (b) PCA clusters separated by early summer, late summer and winter with 60 samples, of which the ellipses were drawn in 95% confidence level. (c, d) Five key genera that contributed significantly to the observed compositional differentiation between early summer and late summer.

Table 3.1 ANOSIM of genus composition based on different seasons, mediums and departments

Parameters	Seasons (summer, winter)	Seasons (early- and late- summer)	Mediums	Departments
ANOSIM r	0.2719	0.7346	0.0192	0.0747
ANOSIM p	1e-04	1e-04	0.3138	0.0444

(3) Microbial diversity influenced by different mediums

In the studied water and air dust, the predominant genera differed. In water samples, *Dechloromonas* and *Pseudomonas* were the most abundant, accounting for 10.7% and 7.64% of the total genus *16S rRNA* gene sequences, respectively. In contrast, *Janthinobacterium* and *Brevundimonas* were the prevailing genera in air dust samples, comprising 1.98% and 1.66% of the total genus *16S rRNA* gene sequences, respectively. Indeed, *Janthinobacterium* is generally considered to be a non-pathogenic bacterium and does not pose significant risks to human health. However, it is important to recognise that certain species of *Brevundimonas* can be opportunistic pathogens, particularly in individuals with compromised immune systems (Lee et al. 2011). In addition to the identified leading genera that potentially posed a risk, other pathogenic bacteria known to induce significant opportunistic infections were also detected in both water and air dust samples, such as *Pseudomonas* sp. (45.4%), *Acinetobacter* sp. (3.06%), *Sphingomonas* sp. (1.32%), *Enterobacter* sp. (0.806%), *Aeromonas* sp. (0.792%), *Klebsiella* (0.217%), and *Legionella* sp. (0.0551%). It is well-established that water distributed in healthcare facilities and air linked to outdoor environments are commonly associated with hospital-acquired infections. The transmission routes for waterborne and airborne pathogens involve direct contact, ingestion of water, indirect contact, or inhalation of bioaerosols (Bonadonna, Briancesco, and Coccia 2017). Despite efforts to treat water and disinfect air, it is likely that the water and air in healthcare settings still contain low concentrations of various indigenous microorganisms.

There is an understanding that the microorganisms in aerosols are released through various routes within healthcare facilities (Bonadonna, Briancesco, and Coccia 2017), which is consistent with the findings

that the Shannon diversity of air dust samples was higher than that of water samples, with an index value of 2.88 and 2.52, respectively (Figure 3.4a). Nevertheless, despite the disparity in alpha diversity between water and air dust samples, the overall bacterial structure was not significantly influenced by the medium variations. This lack of correlation was depicted by the PCA result with no distinct clusters separating air dust and water samples (Figure 3.4b) and was further supported by the ANOSIM result with p-value of 0.3138 and an R-value of 0.0192 (Table 3.1). Based on these findings, the implication is that bacterial dissemination occurred frequently in water and air mediums. Explicitly, bacteria present in water sources can become aerosolised and subsequently released into the air. This is particularly notable during the summer season when air conditioners are in operation and are cooled using water from the hospital water supply, which increased the likelihood of microbial exposure and potential transmission to individuals within the healthcare facility (Leung, Tong, and Lee 2019). Remarkably, during the analysis, as air dust samples were lacking in winter, the assessment of genera diversity in water and air dust was carried out without incorporating the water samples from the winter season.

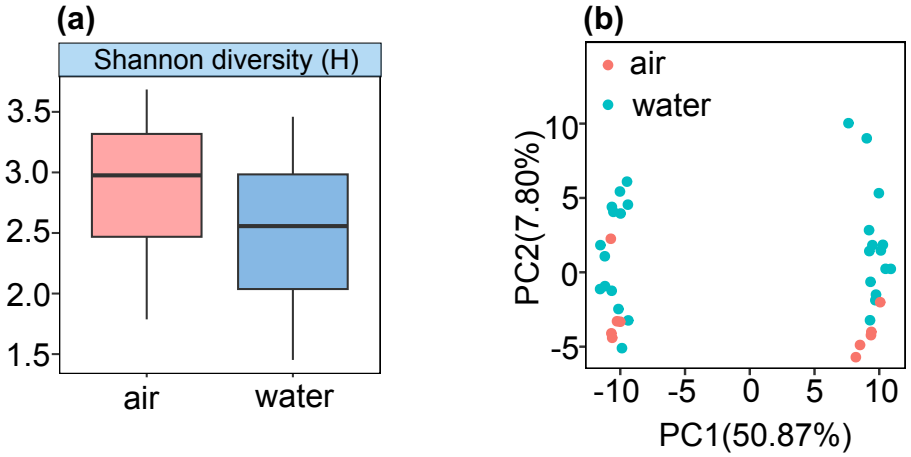


Figure 3.4 (a) Microbial (genus-level) alpha-diversity in two mediums with 38 samples, where the inside line in each bar represents the median value. (b) PCA characteristics of mediums only in summer season with 38 samples.

(4) Microbial diversity in different departments

The analysis of leading genera within different departments reveals interesting patterns that corresponds to the dominant genera observed in different seasons and mediums. These leading genera as mentioned earlier included *Dechloromonas*, *Pseudomonas, Flavobacterium, Limnohabitans, Ralstonia,* and pathogenic *Acineotobactor.* Concerning the genera associated with pathogenesis, their prevalence varied across different departments. For example, in the emergency department, *Pseudomonas*, *Staphylococcus*, *Streptococcus*, and *Neisseria* were the most prevalent. In addition, *Acinetobacter*, *Bacillus*, and *Enterobacter* showed the highest occurrence in the otolaryngology department. Moreover, *Klebsiella* and *Mycobacterium* had the highest abundance in the blood-test room. Furthermore, *Escherichia-Shigella* and *Stenotrophomonas* appeared most frequently in the gynecology and pulmonology departments, respectively. These diverse prevalence patterns raised concerns about the potential for cross-infection across the departments.

When investigating alpha diversity across different departments, significant variations were observed, as depicted in Figure 3.5a. The otolaryngology and surgery departments displayed the highest alpha diversity, with Shannon index values of 2.81 and 2.66, respectively. In contrast, the pneumology department exhibited the lower microbial diversity, recording the lowest Shannon index value of 1.99. This implies a less diverse bacterial structure compared to the other departments. In line with previous observations regarding seasonal patterns, the comparison of alpha-diversity within each department between summer and winter revealed higher Shannon index values in the summer samples than in the winter samples. The most compelling variation between the two seasons was observed in the emergency and pneumology departments, with the Shannon index measuring 2.36 and 2.32 in summer, and contrasting with values of 1.00 and 1.56 in winter, respectively.

Besides the observed fluctuations in alpha diversity, the beta diversity analysis supported by the discrete distribution in the PCA plot (Figure 3.5b) reflects that the bacterial compositions varied independently of the departmental divisions. Indeed, the department-specific variability could be attributed to the fact that bacteria occupy various ecological niches (e.g. air, water, human bodies, and medical equipment), and bacterial responses to different stresses, including antibiotics, can vary based on the specific conditions they encounter (Ben Maamar, Hu, and Hartmann 2020). Additionally, the involvement of other factors, such as different occupancy rates and frequent movement of patients, can also contribute

to different mechanisms of bacterial colonisation, interaction, and evolution in the individual department. These additional factors pose further challenges to elucidating the bacterial diversity among departments within a hospital indoor environment.

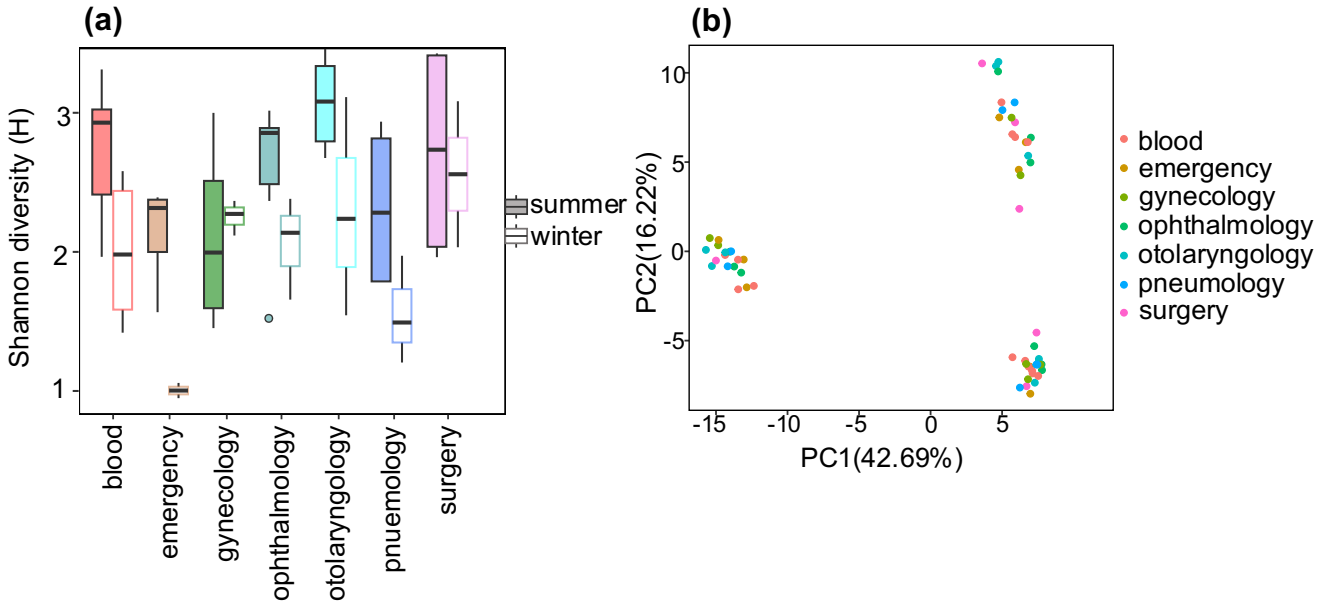


Figure 3.5 (a) Microbial (genus-level) alpha-diversity in seven departments divided into summer and winter with 60 samples, where the inside line in each bar represents the median value, and the outside circles stand for outliers (b) PCA characteristics of departments with 60 samples.

(5) Common bacteria in the hospital indoor environments

Regarding the common genus types present, the Venn analysis reveals several noteworthy findings in the hospital indoor environments. Across seven departments, there were a total of 197 common genus types, indicating the presence of a core microbial community within the hospital environment (Figure 3.6a) In addition to the common community types among the departments, the common bacteria types between water and air dust samples were also observed, with 201 common genera types between these two mediums. To illuminate such phenomenon, in addition to the colonisation of the same genera, common environmental sources like ventilation systems and the activities of healthcare occupants had significant influences on it within healthcare facilities (Hospodsky et al. 2012). Through the interactions

with patients and contact with various surfaces, the occupants could unintentionally facilitate the transfer of bacteria from one patient to another, thereby influencing the spread of healthcare-associated infections.

Assuredly, there is a well-established fact that in the outpatient setting, the blood-test room functions as a focal point for accommodating patients from diverse origins. As a consequence, to further seek the sources of the genera disseminated into the blood-test room, a source-tracking analysis was undertaken. Indeed, the validation that demonstrates the reliability of selecting the "real" sink as the designated sink role in SourceTracker2 analysis has been conducted by Wu et al. (2022). Based on this evaluation, during the analysis, the blood test room was considered the "sink" for the communities (genus-level), while the other six departments were considered the "sources". The result of the analysis (Figure 3.6b) reveals that 46.8% of the genera in the blood test room originated from the other six departments, of which the gynecology and emergency departments contributed the most with 21.6% and 8.81% respectively. However, there were still 53.2% of the sources that could not be identified, indicating a substantial portion of the genera in the blood test room were outdoor origins.

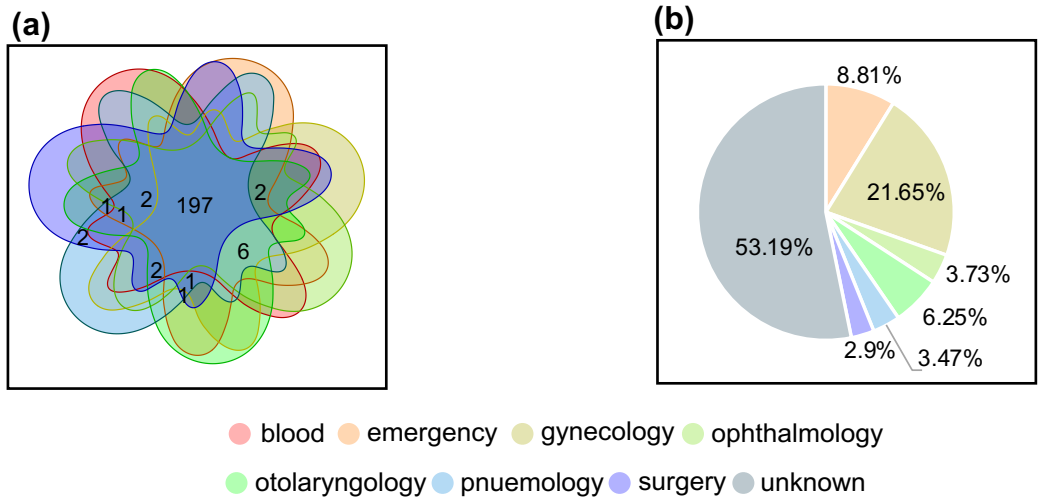


Figure 3.6 (a) Venn analysis of genus types in seven departments with 60 samples. (b) Source-tracking analysis of bacteria in the hospital indoor environments, where the blood department acted as the "sink", and the other six departments served as the "sources".

3.3.2 Broad-spectrum profile of ARGs and MGEs

(1) MDR and MGE genes were the most abundant in the overall abundance profile

The analysis of ARGs and MGEs reveals the detection of 107 target ARGs and MGEs across all samples, In the subsequent discussion, the focus was placed on their relative abundance to mitigate the potential impact of bacterial community size. The total relative copy number per *16S rRNA* gene copy of these ARGs and MGE genes was 74.1, with an average relative copy number of 0.0289. Figure 3.7a and Figure 3.7b illustrate the relative abundance of the 107 ARGs and MGEs belonging to 14 different types in all the samples. It is apparent that the largest proportion of identified genes consisted of MDR and MGE genes, with a combined relative copy number of 24.8 and 20.6, respectively. The MDR genes were further classified into the mobile genes and the genes that were not featured with the mobility with a relative copy number of 5.55 and 19.3, respectively. The MGEs category included integrases, transposases, and plasmids. Integrases had a relative copy number of 11.4, transposases had a relative copy number of 9.14, and plasmids that were represented by *tra* genes had a relative copy number of 0.0600. As acknowledged,

genes conferring antibiotic resistance are often encoded on MGEs, which can be readily shared between bacterial pathogens via HGT (Juhas 2015). To be explicit, integrases play a crucial role in facilitating HGT of ARGs by mediating site-specific recombination and can promote the stabilization of the newly incorporated gene within the recipient DNA, allowing for its stable maintenance and expression (Escudero et al. 2015). Transposases facilitate HGT by a mechanism known as transposition (Wang and Dagan 2024), through which the transposon containing the ARG can be directly cut and pasted to a new location within the recipient DNA. In some cases, transposases can also promote the replication of transposons, leading to the dissemination of ARGs to other bacterial cells (Tokuda and Shintani 2024). With reference to plasmids, they can be transferred horizontally between bacterial cells through a process called conjugation, contributing to the dissemination of antibiotic resistance in bacterial populations (Vrancianu et al. 2020). However, the rates of HGT have not been well quantified in clinical settings (Evans et al. 2020). Certainly, the discovery of MDR and MGE genes was distressing as the prevalence of MDR genes implies that pathogens could be resistant to multiple drugs, and the popularity of MGE genes indicates the potential transfer of ARGs to a wider range of pathogens. Following the MGE genes, ARGs delivering resistance to sulfonamide, MLSB, aminoglycoside and beta-lactam were detected, with 9.28 relative copies, 6.90 relative copies, 4.13 relative and 3.30 relative copies. Considerably, among beta-lactam-resistant genes, there were 57.0% (1.88 relative copies) extended-spectrum beta-lactamases (ESBLs), which has been defined as transmissible β-lactamases that can be exchanged between bacteria (Shaikh et al. 2015). These findings of ARGs and MGEs align with a study conducted by Li et al. (2019) on hospital samples collected from air-conditioning units, which also identified similar patterns of dominating drug resistance.

As for the representative ARGs and MGE genes in studied environments, the dominating 15 representative genes were shown in Figure 3.7c. *merA-marko (*MDR*)*, *int1-a-marko* (integrase), *strB* (sulfonamide resistance), *qacF/H* (MDR-mobile), *tnpA_203* (transposase), and *lnuA* (MLSB resistance) were the high-frequent representative ARGs and MGE genes spread in different environments. *merA-marko* gained the most abundance with 19.2 relative copies, followed by *int1-a-marko* and *strB* with 11.0 and 6.44 relative copies, respectively. However, few representative ARGs were unique in the certain hospital setting, such as bla_{VEB} (beta-lactam resistance) only in the blood test room, *floR* (amphenicol

resistance) exclusively in the emergency department and *bla*$_{ADC}$ (beta-lactam resistance) simply in the ophthalmology department.

In summary, the presence of a diverse range of ARGs and MGEs in the hospital environment highlights the importance of implementing effective strategies to monitor and control the spread of antibiotic resistance. The representative genes provide insights into the specific genetic elements and resistance mechanisms prevalent in different environments. Understanding the abundance and distribution of these genes can inform targeted interventions to mitigate the risk of HGT and combat the spread of antibiotic resistance within the healthcare setting.

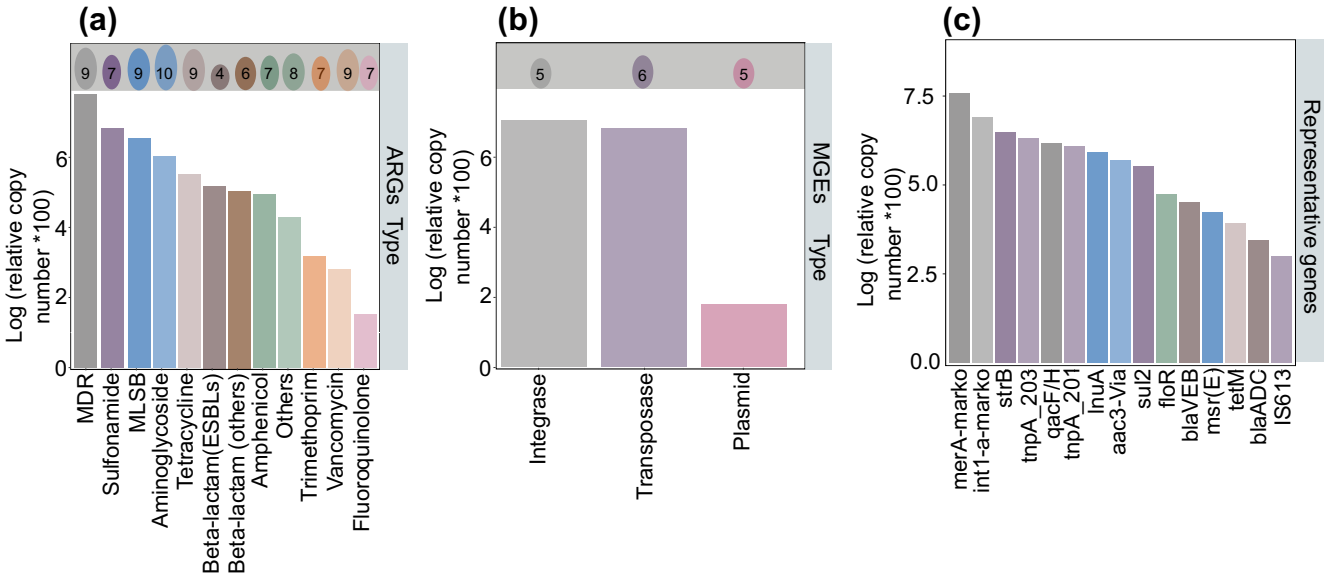


Figure 3.7 (a) (b) Abundance of ARGs and MGEs types in 60 samples, where the abundance represents the logarithm value of 100 times relative copy number per *16S rRNA* gene copy. The grey top area shows the primer number used of each type. (c) Abundance distribution of top fifteen dominating ARGs and MGEs in all studied samples, where the abundance represents the logarithm value of 100 times relative copy number per *16S rRNA* gene copy. The bar colors represent ARGs and MGEs types as shown in figure (a) and figure (b).

(2) ARGs/MGEs abundance and diversity in different seasons

During the summer season, there was an evident increase in the relative abundance and Shannon diversity of ARGs and MGEs as illustrated in Figure 3.8a-3.8d. The distribution of MGE genes exhibited a similar pattern, with summer showing a higher acquisition of copies at 13.0 copies per *16S rRNA* gene, and higher Shannon index value at 1.97. Despite a study conducted by Hashimoto, Hasegawa and Maeda (2019) has reported a higher frequency of gene transfer events in summer, the observed rise in the relative abundance of MGEs in this context might primarily indicate an increase in the presence of MGEs within the bacterial genome. This potentially underscored enhanced gene sharing among microbial communities during the summer season. Conspicuously, plasmid genes showed a unique performance, being exclusively detected in summer samples (as shown in Figure 3.8b). This suggests that plasmids could be more prevalent and active during the warmer months. However, the limited focus on specific plasmid genes may constrain the depth of the plasmid prevalence analysis.

Figure 3.8a depicts a fascinating finding where the prevalence of ARGs and MGEs types in the summer and winter exhibited distinct patterns. Specifically, MLSB-resistant, amphenicol-resistant, and beta-lactam-resistant ARGs predominantly emerged in the summer samples, accounting for 96.7%, 95.0%, and 89.1% of the total, respectively. In contrast, MDR (44.0%), integrase (40.4%), and sulfonamide-resistant (39.7%) genes were frequently found in winter samples. This potentially suggests that the composition of ARGs and MGEs types may be influenced by seasonal patterns. However, despite these observations, the ANOSIM result, as shown in Table 3.2, did not support a significant connection between the components of ARGs/MGE genes and the seasonal variations, indicating while there may be differences in the richness, diversity, and specific types of ARGs/MGEs between seasons, these differences may not be strong enough to be statistically significant. Other factors, such as microbial community dynamics, host factors, or specific environmental conditions, could also contribute to the observed patterns.

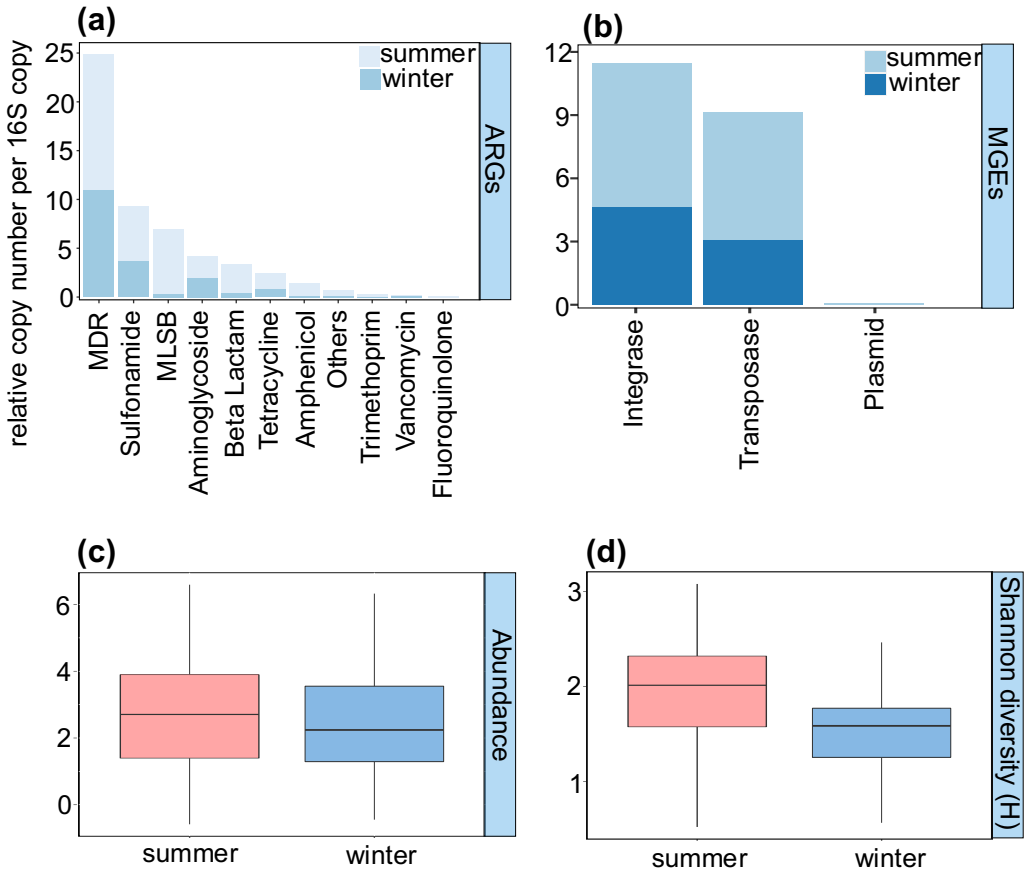


Figure 3.8 Abundance of (a) ARGs and MGEs types (b) only MGEs types in two seasons with 60 samples, where the abundance represents the relative copy number per *16S rRNA* gene copy. (c, d) Boxplot of ARGs/MGEs abundance and Shannon diversity in two seasons, where genes abundance represents the logarithm (10) value of 10e6 relative copy number per *16S rRNA* gene copy, and the inside line in each bar represents the median value.

Table 3.2 ANOSIM of compositions of ARGs and MGEs based on different seasons, mediums and departments

Parameters	Seasons	Mediums	Departments
ANOSIM r	0.006637	0.2487	-0.04852
ANOSIM p	0.4031	0.05954	0.8729

(3) ARGs/MGEs abundance and diversity in different mediums

The overall trend of abundance and diversity of ARGs and MGEs in different mediums reveals that ARGs and MGEs were more abundant in water but were more divergent in the air dust. Water samples exhibited a higher total abundance of ARGs and MGEs, with 28.5 relative copies, compared to air-dust samples, which had 20.0 relative copies. This suggests that water environments may serve as reservoirs for the accumulation of ARGs and MGEs. Likewise, concerning the distribution of MGEs, water samples exhibited higher levels of MGE genes (6.95 relative copies), potentially indicating an elevated exchange of ARGs facilitated by water. Obviously, plasmid genes were unique to air-dust samples (as shown in Figure 3.9b). This set them apart from the majority of other MGE genes, which were found in both water and air-dust samples. Interestingly, despite the higher abundance of ARGs and MGE genes in water samples, the Shannon diversity index indicates a higher genetic diversity of ARGs and MGE genes in air-dust samples (2.41 index) (as shown in Figure 3.9d). This suggests that the air-dust samples in the hospital may harbor a more complex and diverse array of ARGs. The sources of these ARGs in the air-dust samples could include bacterial shedding from patients, contaminated surfaces, or other aerosolisation mechanisms.

In Figure 3.9a, there is an evident discrepancy in the contribution of different types of ARGs and MGEs between two mediums. For example, ARGs and MGEs belonging to MDR, aminoglycoside-resistant and integrase genes were predominantly found in water samples (92.0%, 73.0%, and 68.1%, respectively). On the other hand, ARGs conferring resistance to amphenicol and MLSB antibiotics were more prevalent in air-dust samples (90.6% and 72.1%, respectively). Despite these notable differences, the ANOSIM result (P-value:0.05954; R-value:0.2487) from Table 3.2 conversely indicates that medium variations did not significantly affect the components of ARGs and MGEs.

In summary, the medium observations highlight the distinct characteristics of ARGs and MGEs in different mediums and underscore the importance of considering multiple environmental reservoirs in assessing the spread and diversity of antibiotic resistance in healthcare settings. It's also important that while the ANOSIM result did not show a significant variation, the observed differences in the distribution of ARGs and MGEs in two mediums suggest potential variations in the sources, transmission routes, and

selective pressures associated with antibiotic resistance. These differences could be influenced by factors such as the presence of specific bacteria, physical properties of the mediums, or different exposure routes.

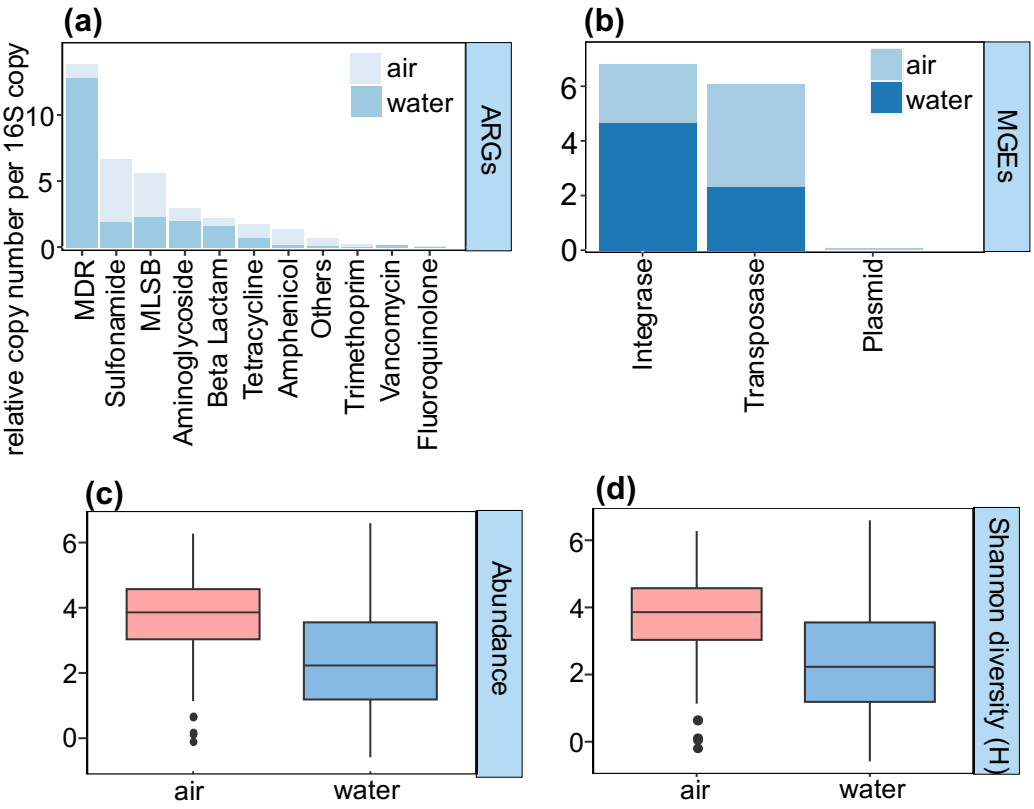


Figure 3.9 Abundance of (a) ARGs and MGEs types (b) only MGEs types in two mediums only in summer with 38 samples, where the abundance represents the relative copy number per *16S rRNA* gene copy. (c, d) Boxplot of ARGs/MGEs abundance and Shannon diversity in two mediums only in summer with 38 samples, where genes abundance represents the logarithm (10) value of 10e6 relative copy number per *16S rRNA* gene copy. The inside line in each bar represents the median value, and the outside circles stand for outliers.

(4) ARGs/MGEs abundance and diversity in different departments

It is shown that the richness of ARGs and MGE genes had a relatively homogeneous distribution across different departments (Figure 3.10a). The blood test room had the highest abundance of ARGs and MGEs,

with 12.9 relative copies, while the otolaryngology department had the lowest gene copies, with 6.18 relative copies. The distribution of MGE genes exhibited a different pattern (Figure 3.10c), with the emergency department showing the highest abundance of MGE genes at 4.00 relative copies, implying a higher potential for the ARGs sharing in that department. Plasmid genes consistently exhibited a distinct distribution and were only found in the emergency and otolaryngology departments. This suggests that plasmid-mediated gene transfer may be more prevalent in these two departments compared to other departments.

In contrast to the relatively stable distribution of ARGs and MGEs abundance, the diversity of ARGs and MGEs showed fluctuations across departments. The surgery samples exhibited a notable increase in diversity, with an average Shannon index of 2.26 (Figure 3.10b). This could be attributed to factors such as the presence of diverse microbial sources from wounds and injuries, the use of antimicrobial chemicals in indoor environments (Hartmann et al. 2016), and the contribution of human microbes to the overall richness and diversity of ARGs in the surgery department (Prussin and Marr 2015). Similarly, the ANOSIM analysis indicates that department variations did not significantly affect the components of ARGs and MGEs as well (p-value: 0.8729; R-value: -0.0485).

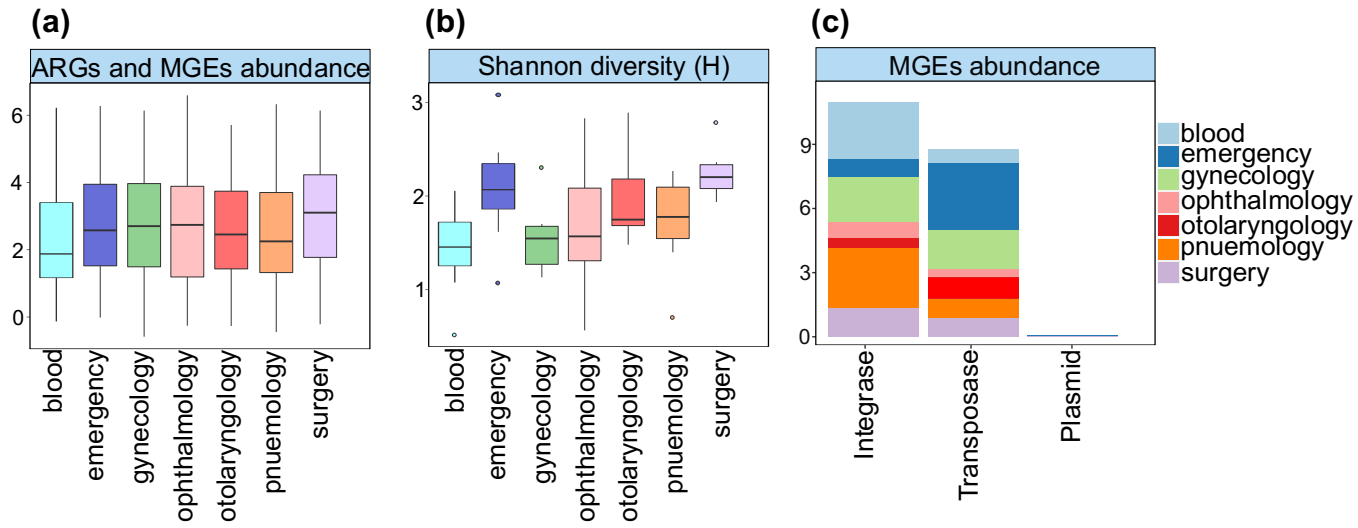


Figure 3.10 (a) (b) Boxplot of ARGs/MGEs abundance and Shannon diversity in seven departments with 60 samples, where genes abundance was represented by the logarithm(10) value of 10e6 relative copy

number per *16S rRNA* gene copy. The inside line in each bar represents the median value, and the outside circles stand for outliers. (c) Abundance of MGEs types in seven departments with 60 samples, where the abundance represents the relative copy number per *16S rRNA* gene copy.

(5) Co-occurrence between ARGs and MGEs

The preceding discussion has stated that MGEs can contribute to the HGT of ARGs among different microbes, and the host range of the MGEs carrying ARGs is important for determining how far it will spread (Ben Maamar, Hu, and Hartmann 2020). Consequently, to further prove the mobility of ARGs, the Mantel test between ARGs and MGEs was conducted. The results unveil a robust correlation between the abundance of ARGs and MGE genes in the hospital indoor environment. Fascinatingly, with respect to each department, such significant correlation also appeared in the emergency, surgery, pneumology and otolaryngology departments (Table 3.3). This hints that the abundance of ARGs likely to be associated with the presence of MGEs, bolstering the concept of horizontal transfer of ARGs among bacteria within the hospital setting. Similar correlations between ARGs and MGE genes have been frequently documented in hospital aerosols and wastewater as well (Wu et al. 2022; Jiao et al. 2023; Markkanen et al. 2023).

With regard to the phenomenon of common ARGs/MGEs subtypes in the hospital indoor environments, it is manifested in Figure 3.11a. Pointedly, the indoor hospital exhibited a core set of 38 ARGs and 8 MGEs across seven departments. On the one hand, this may suggest the presence of a common pool of genes circulating within the hospital. On the other hand, this may be caused by the common communities across the departments as discuss. Significantly, akin to the unique trends observed in representative ARGs, the singular presence of specific ARGs was once again apparent. For example, the ophthalmology department had vancomycin-resistant genes *vanSB* and *vanWB*. The pneumology department had peptide-resistant gene *mcr-2* and MLSB-resistant gene *ermY*. Additionally, the otolaryngology department had the aminoglycoside-resistant gene *armA*. Among those unique ARGs, *vanSB*, *vanWB* and *ermY* have been acknowledged as chromosome mediated ARGs (Leigh et al. 2022). Consequently, their unique presence in specific departments could be attributed to their inherent mobility. With respect to *mcr-2* and *armA* gens, they have been known as mobilised resistance genes mediated by plasmids that

horizontally transferred to multiple organisms (Liassine et al. 2016; Saadatian et al. 2018). Therefore, the exclusive presence of these genes in particular departments could be attributed to their restricted transferability mediated by specific MGEs.

Once more, a source tracking analysis was performed to probe the origins' proportions of ARGs and MGEs found in the blood test room. To the same token, the analysis maintained the postulation that the blood test room functioned as a 'sink' for the accumulation of ARGs/MGEs, with the other six departments acting as the 'sources'. According to the source tracking results (Figure 3.11b), it was found that 50.8% of ARGs and MGEs present in the blood test room originated from the other six departments. Among these, the gynecology department was identified as a significant contributor, accounting for 31.5% of the ARGs and MGEs, followed by the pneumology and surgery departments contributing 7.90% and 5.45% respectively. Additionally, given the gynecology department as one of primary sources that contributed 21.6% of the genera to the blood-test room, it underscores the substantial impact of this department on the distribution of bacteria, ARGs, and MGEs. However, it is crucial to recognise that a substantial portion (49.2%) of the source of ARGs/MGEs could not be identified, implying the possibility of outdoor sources. A similar source-tracking study was also conducted by Li et al. (2021) in the same hospital, which pointed out that in the air dust, the outpatient hall was one of the main ARG transmission sources to the ophthalmology and pediatrics departments.

Undeniably, although the Venn and Sourcetracker2 analyses were effective in identifying common bacterial community types and ARGs/MGEs subtypes, as well as tracing the sources proportions based on logical assumptions, a comprehensive understanding of the extensive dissemination of bacteria, ARGs and MGEs remained elusive. To further elucidate the transmission dynamics, it is imperative to identify the genome sequences of the involved bacteria. This will enable a detailed comparison of genetic relatedness, aiding in explaining the sharing phenomenon and facilitating precise source tracking (further discussion is in *section 6.6*).

Table 3.3 Mantel test of ARGs and MGEs abundance

Mantal test	Whole hospital	Emergency	Gynecology	Ophtha-lmology	Otolary-ngology	Pneu-mology	Surgery	Blood
p value	0.001	0.033	0.242	0.154	0.002	0.014	0.014	0.14
r value	0.2100	0.4543	0.1831	0.1876	0.6183	0.5351	0.6	0.1387

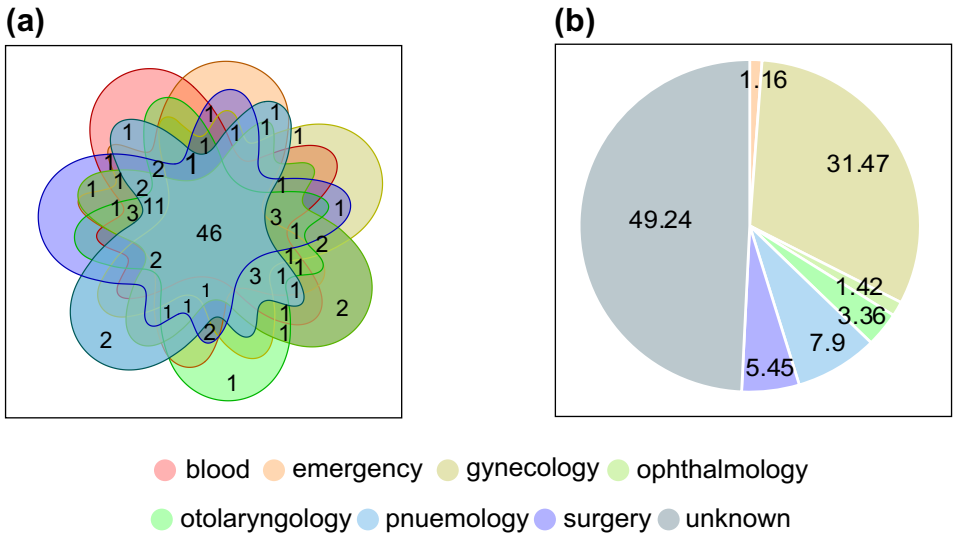


Figure 3.11 (a) Venn diagram of sharing ARGs/MGEs types in seven departments with 60 samples. (b) Source-tracking analysis of ARGs and MGEs across seven departments, where the blood department acted as "sink", and other six departments served as "sources".

3.3.3 Correlation between bacterial communities and ARGs/MGEs

(1) Insignificant correlation between bacterial communities and ARGs/MGEs in the overall studied hospital environment

ARGs were able be found in all studied environments. Indeed, the use of antibiotics by humans has been found to stimulate the acquisition of ARGs by pathogenic bacteria (Martinez et al. 2009). To investigate whether the bacterial community (genus-level) was correlated with ARGs and MGEs composition, I used the Procrustes analysis and the Mantel test to correlate profiles with 60 samples. The Procrustes test shows that there was a lack of goodness-of-fit (sum of squares M^2 = 0.6583, p =1e-06, 10000 permutations) on the basis of Bray−Curtis dissimilarity metrics, reflecting the potential inconsistency between the bacterial community composition (genus-level) and the composition of ARGs and MGEs. The mantel test supported this finding as well (r=0.03715, p=0.113).

To further dig into the correlation between specific subtypes of ARGs/MGEs and genera, pairwise Spearman's rank correlation was calculated. Among the numerous correlations tested (23,221 pairs), only nine genera showed a significant correlation with the common tetracycline-resistant *tet(36)* in pair ($p<0.05$, $\rho>0.6$). Previous research has argued that *tet(36)* was first discovered in *Bacteroides* sp. strain and the HGT of *tet(36)* was claimed to occur frequently between divergent phylogenetic groups in the environment (Whittle et al. 2003). Furthermore, upon investigating the association between ESKAPE-related genera (*Pseudomonas*, *Enterobacter, Staphylococcus*, *Klebsiella*, and *Acinetobacter*) and 107 ARGs/MGEs (refer to Table 3.3), none of the 540 pairs analysed exhibited a significant correlation. This lack of significant correlation implies a lack of relevance of the potential ARGs carried by ESKAPE organisms.

To delve into the correlations at a higher resolution on the species level (42 species), the analyses were replicated, resulting in similar insignificant outcomes as shown in Table 3.3. Only one pair of species and ARG demonstrated a Spearman's correlation with $\rho > 0.8$ ($p < 0.01$), specifically *Rhodococcus erythropolis* with *armA*. In a study by Li et al. (2015), it was proposed that non-random co-occurrence patterns between ARGs and microbial taxa might suggest potential host information of ARGs if they shared significantly similar abundance trends across different environments (Spearman's $\rho > 0.8$, p-value < 0.01). Consistent with this hypothesis, *R. erythropolis* was deemed a potential host carrying *armA* and conferring resistance to aminoglycosides, and the minimal pathogenic potential of *R. erythropolis* in lymphocytic leukaemia (Park et al. 2011) exacerbated this resistance phenomenon. Furthermore, two other non-pathogenic species exhibited a pairwise coefficient $\rho > 0.6$ ($p < 0.05$) with two ARGs, indicating a strong correlation but lacking a distinct host association. These species and corresponding ARGs were *Leuconostoc fallax* with *aac(6')-iic* (aminoglycoside), and *Lactobacillus fermentum* with *fosb* (phosphonic acid). For *L. fermentum to* be regarded as a potential probiotic (Zheng et al. 2020), the horizontal transferability of *fosB* (Wang et al. 2021) could probably lead to the establishment of reservoirs of resistance genes in the gut microbiota, impacting long-term health and treatment outcomes.

Taken together, these results illustrate that in the studied hospital environment, the composition of bacterial communities (genus and species level) did not strongly influence the presence or abundance of ARGs and MGEs. The factors such as selective pressure from antibiotic usage, HGT, and other

environmental factors may play a more significant role in shaping the prevalence and distribution of antibiotic resistance. However, while there might not be a significant overall correlation between bacterial composition and ARGs/MGEs profiles, certain scenarios can unveil the potential hosts of specific ARGs, like the case of pathogenic *R. erythropolis* carrying *armA*.

Table 3.4 Correlation between OTUs and ARGs/MGEs

	Bray-curties M^2	Bray-curties p	Mantel r	Mantel p
Hospital (genus-level)	0.6583	1e-06	0.03715	0.113
Hospital (genus-level, *ESKAPE*)	0.7946	1e-04	r=0.498	1e-04
Hospital (species-level)	0.7363	1e-05	-0.03455	0.8181
Gynecology department (genus-level)	0.1007	0.0025	0.7571	0.018
Gynecology department (species-level)	0.2240	0.003770	-0.04057	0.47758

(2) Co-occurrence between bacterial communities and ARGs/MGEs in the gynecology department

Absorbingly, although there was no correlation observed between the bacterial community composition and ARGs/MGEs profile in the overall studied environment, a strong correlation between them (genus-level) was discovered specifically in the gynecology department which was supported by the Procrustes'residuals value. In general, the residuals value represents the closeness between the ARGs/MGEs and genus composition. A lower residuals value indicates a closer relationship. In the gynecology department, I found that the average residuals value of gynecological samples was significantly low, suggesting a strong association between ARGs/MGEs and bacterial composition in that specific department. To further evaluate this strong correlation, Procrustes and Mantel analyses were conducted exclusively targeting the ARGs/MGEs and bacterial composition (genus-level) in the gynecology department. Both analyses gave compatible results, showing a strong correlation between ARGs/MGEs and genera (Table 3.3).

Based on these findings, I proceeded with pairwise Spearman's correlation analysis between nine gynecology samples comprising 43 ARGs/MGEs and 181 genera. The results reveal a significant relationship between all 43 target genes and 181 genera, with p-values below 0.05 and ρ values exceeding 0.6. Furthermore, 32 ARGs/MGEs and 117 genera exhibited ρ values surpassing 0.8, with p-values below 0.01, leading to their inclusion in the undirected network analysis (Figure 3.12a, 3.12b). Notably, *Afipia* and *floR* (amphenicol) demonstrated the strongest correlation, with ρ values peaking at 0.97.

By the same token, at the species level, the pairwise Spearman's correlation analysis was conducted across seven gynecology samples containing 43 ARGs/MGEs and 29 species. The results indicate a significant relationship between only 18 ARGs (no MGEs) and 10 species, with p-values below 0.05 and ρ values above 0.6. Additionally, 10 ARGs and 8 species exhibited ρ values exceeding 0.8, with p-values below 0.01. These pairs were further subjected to another undirected network analysis (Figure 3.12c). Noteworthy pairs such as *Luteibacter yeojuensis* and *floR* (amphenicol), along with *Aquaspirillum arcticum* and *vga(A)LC* (MLSB), displayed the strongest correlation, reaching ρ values of 1.

(3) Network analysis between bacterial communities (genus-level) and ARGs/MGEs in the gynecology department

The created network of the gynecology department consisted of 185 nodes (each node represents a subtype of ARGs, MGEs, or microbial taxa) and 366 edges (Figure 3.12). The modularity index was 0.642 (value > 0.4), suggesting that the network had a modular structure. Based on the modularity class, the entire network was clearly separated into three major modules. Compared with a random association, clusters of nodes in the same module contained more interactions among themselves than with other nodes. Module I was the largest module comprising 82 nodes, followed by Modules II and III including 23 and 20 nodes, respectively.

From the network, it was witnessed that transposase *Tp614* and aminoglycoside-resistant *sat4* in Module I had the most positive correlations with 151 types of genera individually. Besides *Tp614,* MGE genes including integrase *int1-a-marko* and transposase *tnpA 201* in Module III had more positive correlations

with 18 and 17 types of genera as well. This substantial number of correlations involving MGEs in the network indicates the dissemination of ARGs and the potential acquisition of these ARGs by other microbes within the gynecology department. This phenomenon underscores the impact of high antibiotic selective pressure in the hospital environment. Other ARGs, such as *bacA*, *ampC* in Module III and *tetX* in Module V, which conferred resistance to bacitracin, beta-lactamase, and tetracycline, respectively, were associated with multiple candidate genera (20, 15, and 12 genera, respectively), suggesting a higher likelihood of these ARGs being carried by a diverse range of microbial hosts. However, it is crucial to note that while numerous ARGs were associated with multiple genera, only a few ARGs were found to be potentially carried by a single host. For example, MLSB-resistant *msr(E)* was exclusively correlated by *Chelatococcus,* and *Paludibacter* was the only correlation with MLSB-resistant *lnuA*. These observations indicate that certain ARGs may have a more restricted distribution and are less likely to be transferred between different microbial taxa.

On the other hand, from the community perspective, it is essential to highlight genera that exhibited correlations with multiple resistances. For instance, *Anaerovorax* demonstrated the highest number of associations with six ARGs/MGEs, including *ampC* (beta-lactam resistant)*, int1-a-marko* (integrase), *tetX* (tetracycline resistant), *tnpA 201* (transposase), and *Tp614* (transposase). Similarly, genera like *Aureimonas*, *Laribacter*, *Cellvibrio*, *SH-PL14*, *Flavobacterium*, *Macellibacteroides*, *Ralstonia*, *Hyphomicrobium*, and *Arcobacter* were also correlated with six ARGs but not with any MGEs, including *aac(6)-iic* (aminoglycoside resistant), *ampC* (beta-lactam resistant), *KPC* (beta-lactam resistant), *lnuA* (MLSB resistant), *mcr-1* (peptide resistant), *sul2* (sulfonamide resistant), *tetO* (tetracycline resistant), *vanSB* (vancomycin resistant) and *vanYD* (vancomycin resistant). Despite those genera not being closely associated with pathogenesis, their significant correlations with transferable ARGs like *mcr-1*, and MGEs such as *tnpA 201* and *Tp614*, point out a potential for broader dissemination to additional pathogens.

Given the emphasis on the concerned clinical genera in the network, the significant correlations linked to ARGs appeared as well. In the case of *Legionella*, it was found to be strongly correlated with ARGs of beta-lactam-resistant $blaOXA_{10}$ and MLSB-resistant *va0*, which has not been reported by other studies. Currently, beta-lactam, fluoroquinolones, macrolides, and rifampicin are reported as the active antibiotics to which *Legionella* spp. are susceptible (Nimmo and Bull 1995; Sharaby et al. 2019; Pappa et al. 2020).

Implicitly, the resistant delivery to beta-lactam in this study might raise the concern for the inactive therapy of the antimicrobials such as amoxicillin that belongs to beta-lactam to legionellosis patients. More seriously, as the *blaOXA*$_{10}$ is a frequently encountered ARG capable of HGT in hospitals (Golshani and Sharifzadeh 2013), the presence of beta-lactam antibiotics in the environment may promote the evolution of microbial resistance mechanisms (D'Costa et al. 2006;Almahmoud et al. 2009). Another notable discovery affirmed a correlation between *Enterobacter* (ESKAPE-related) and the MDR *cefa_qacelta*, pinpointing the association between this genus and multidrug resistance. What is more, *Acinetobacter* (ESKAPE-related) was acknowledged to be co-occurring with vancomycin-resistant *vanB*, of which the presence has been known to signify a heightened level of resistance to vancomycin (Luqman et al. 2024). As suggested by Chang et al. (2003), this resistance may have been acquired through the acquisition of *vanB* from vancomycin-resistant *enterococci*. Worse, while vancomycin-containing regimens have been reported to offer therapeutic benefits against infections caused by colistin (peptide)-resistant *A. baumannii*, the emergence of this co-occurrence between *Acinetobacter* and *vanB* implies hidden treatment failures in corresponding infections in the future.

Moreover, of particular concern were the correlations observed between *sat4*/*Tp614* and multiple alarming genera, including *Bacillus*, *Escherichia-Shigella*, *Klebsiella* (ESKAPE-related), *Neisseria*, *Staphylococcus* (ESKAPE-related), *Stenotrophomonas* and *Streptococcus*. In point of fact, *Tp614* represents the signatures of various genetic elements involved in gene transfer and recombination. As a result, it is reasonable to estimate that *Tp614* would carry *sat4* transferred in these pathogenesis-related genera which potentially impeded the antimicrobial treatment to the related infections. A similar pattern was shown by Enany and Alexander (2017) that transposase *Tn5404* carried genes of *sat4* in the *S. aureus* that was spread in the bacterial populations. Actually, *sat4* coding for aminoglycoside resistance is frequently found in clinical and urban wastewater carried by clinical strains (Zaheer et al. 2020), and the risk always arises when *sat4* forms the cluster of *aadE/ant(6)-la - sat4 - aph(3')-IIIa*, which is commonly associated with insertion elements from *Tn5405* transposons (Zaheer et al. 2020). Noticeably, the network presents that *Sandaracinobacter* and *Azorhizobium* had a close loop cycle with *bla*$_{PSE}$. This implicates that *bla*$_{PSE}$ was the low-risk ARG that might not transfer in the gynecology department.

(4) Network analysis between bacterial communities (species-level) and ARGs in the gynecology department

The network created at the species level using seven samples comprised 18 nodes, connected by 19 edges (Figure 3.12c), with a modularity index of 0.8. Within this network, eight species were identified as potential hosts carrying ARGs, i.e. *Anaerospora hongkongensis*, *A. arcticum*, *Enterobacter soli*, *L. yeojuensis*, *Pseudomonas koreensis*, *Pseudomonas* sp., *Sphingomonas yunnanensis*, and *Variovorax paradoxus*. Fascinatingly, in the precedingly discussed network at the genus level, no *Pseudomonas* genus displayed correlations with any ARGs/MGEs. However, in the current network at the species level, notable correlations of *Pseudomonas* spp. (including *P. koreensis* and *Pseudomonas* sp.) had surfaced with six ARGs in total, including beta-lactam-resistant *ampC*, *bla-PSE* and *bla-ACT*, vancomycin-reisistant *vanSB*, peptide-resistant *mcr-1*, and MLSB-resistant *lnuA*. Regarding *mcr-1*, despite being a mobilised polymyxin resistance enzyme mediated by plasmids (Tahmasebi et al. 2020), it was only found to co-occur with *P. koreensis* and *Pseudomonas* sp. rather than all examined species. The co-harboring of *mcr-1* with multiple other ARGs resistant to beta-lactam, vancomycin, and MLSB paints a grim picture of highly resistant *Pseudomonas* spp.. This situation presents an opportunity for horizontal transfer into MDR pathogens, potentially leading to a substantial impact on the efficacy of these critical last-resort antibiotics (Tahmasebi et al. 2020).

In summary, the network analysis reveals the modular structure of the gynecology department's microbial community, with specific ARGs, MGE genes, and microbial taxa (genus- and species- level) forming distinct modules. The network provides the evidence that MGEs played a prominent role in facilitating the dissemination of ARGs, indicating the potential for HGT among different microbial organisms. While majority of ARGs showed correlations with a wide range of communities, others such as bla_{PSE} were exclusively associated with specific genus, highlighting the immobility and conservation of certain ARGs. Taken together, this finding underscores the need for careful surveillance and monitoring of antibiotic resistance, particularly in vancomycin-resistant *Acinetobacter*, MDR- *Legionella* spp. and *Pseudomonas* spp. in man-made water systems. The findings are also essential for monitoring and understanding the transferrable ARGs such as *mcr-1* along with *vanB*, and genetic elements such as *Tp614* involved in gene transfer and recombination, as well as their impact on antimicrobial treatment efficacy.

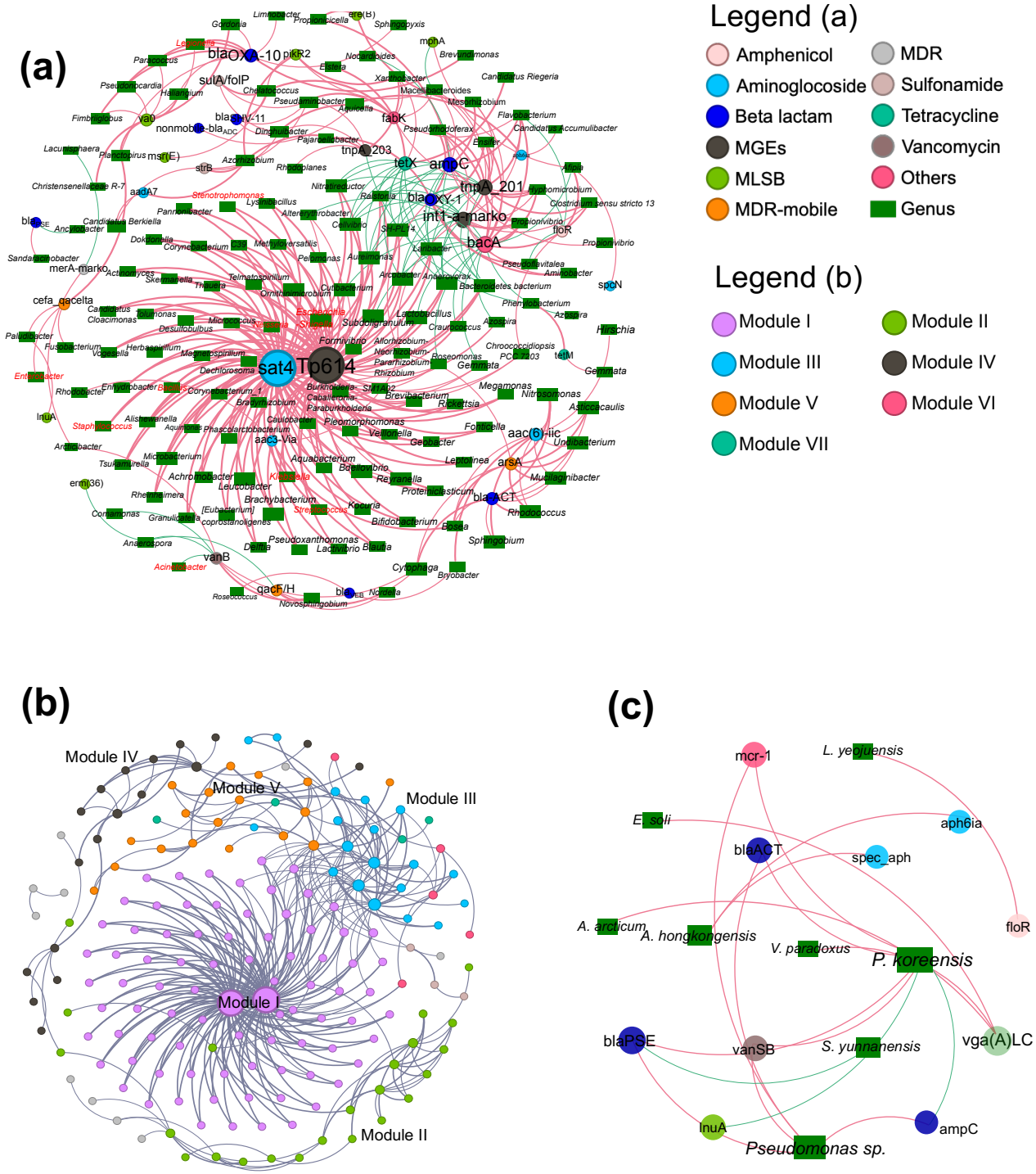


Figure 3.12 (a) Co-occurrence analysis between bacterial communities (genus-level) and ARGs/MGEs in gynecology department with nine samples. Circle colors represent ARGs and MGEs types, and dark-green boxes represent the genera. Line colors mean positive (red) or negative (blue) spearman correlation.

Font/node size represents the degree of connections. Font color with red represents the concerned genera. Line width represents the absolute value of the correlation index, i.e. the wider the line shows, the bigger the absolute value is. (b) the same network as Figure(a), while the node colors here represent the Module classes. (c) Co-occurrence analysis between bacterial communities (species-level) and ARGs in gynecology department with seven samples. Line colors mean positive (red) or negative (green) spearman correlation. The other legends of circles and boxes are same as Figure(a).

3.4 Conclusions

Hospital buildings are indeed dynamic environments that harbor a diverse range of bacteria derived from both the surrounding environment and the individuals present. These factors play a critical role in determining the infectious risk faced by patients within these facilities. One important factor influencing the bacterial composition in hospitals is the usage of antibiotics. Antibiotic administration creates selective pressure, favoring bacteria that possess resistance mechanisms against the specific antibiotics used. As a result, antibiotic-resistant strains can emerge and proliferate within the hospital environment.

The study conducted at *Peking University Shenzhen Hospital* has furnished valuable insights into the variations, correlation and source tracking of bacterial communities, ARGs and MGEs, providing the guidance for infection control within the hospital indoor environment. The study specifically investigated the impact of different seasons, mediums, and departments on these elements. By examining alpha- and beta-diversity, the bacterial community in the air dust was observed to be highly heterogeneous, and the seasonal variation was confirmed as a significant factor that shaped the bacterial composition. In terms of ARGs and MGEs, the significant correlation between ARGs and MGE genes raised concerns about the potential for horizontal transfer of ARGs through the hospital environment. However, the primary factor shaping the composition of ARGs and MGE genes was still unclear. Notably, in the gynecology department which served as the major source of bacterial communities and ARGs/MGEs, the significant associations between bacterial communities and ARGs/MGEs were found. The network profile highlights the ARGs linked to bacteria such as *Acinetobacter*, *Pseudomonas* spp. in addition to *Legionella* spp., and potential horizontal transfer carried by MGEs like transposase *Tp614* and *tnpA 201*.

These findings emphasize the importance of comprehending antibiotic resistance dynamics in hospital environments and address the need for effective strategies to mitigate the spread of resistance genes.

3.5 Limitations

With DNA-extraction-based molecular methods, variations in cell membrane characteristics, as observed in Gram-positive and Gram-negative bacteria, could lead to differing DNA extraction efficiencies. Furthermore, the *16S rRNA* amplification process, merely encompassing two hypervariable regions, provided limited resolution for species-level identification. While genus-level analysis was effective and commonly applied in studying bacterial communities, a finer resolution at the species level with full *16S rRNA* involved or even whole genome analysis could offer a more detailed insight into shared communities, bacterial dissemination and even the co-occurrence between communities and ARGs within the environment.

Moreover, in HT-qPCR, the profiling of ARGs might be skewed due to uneven amplification efficiencies with over hundred assays. Additionally, inconsistencies in *16S rRNA* gene copy numbers among microorganisms could introduce bias when normalising against the *16S rRNA* gene. Further, relying solely on relative quantification without exploring the genetic sequences of ARGs, HT-qPCR was not able to reveal detailed insights into horizontal transfer or other mechanisms associated with AR acquisition.

3.6 Paradoxical results with wastewater samples

There were also four wastewater samples (Table S3.1) involved in the *16S rRNA* Illumina amplicon sequencing and HT-qPCR, of which two were inlet samples and two were outlet samples collected in August 2020 (early summer) and October 2020 (late summer). Drawing from the earlier conversation, the Shannon index value of bacterial communities for air-dust samples was recorded at 2.88. Surprisingly, the alpha-diversity at the genus level in both inlet and outlet samples was even lower than that of air dust samples, registering a Shannon index value of 2.82 and 2.54, respectively (Figure S3.2). This outcome appeared peculiar as the bacterial diversity in the wastewater samples was anticipated to be higher than

that in the air. Upon visually inspecting these samples, it was also evident that the wastewater samples exhibited more complex communities compared to the air dust samples. To the same token, the Shannon diversity of ARGs and MGEs followed a similar peculiar pattern. The Shannon index for the inlet and outlet samples was 2.85 and 2.32, respectively, while the air dust sample had a value of 2.41, which was also higher than that of the wastewater outlet samples.

Chapter Four

Development of a Multiplex-PCR Panel for Sequence- and Resistance-Typing of *Pseudomonas aeruginosa* Using Nanopore Sequencing

This thesis comprises two multiplex-PCR-based MinION sequencing panels of *L. pneumophila* (Chapter 5) and *P. aeruginosa* (this chapter).

4.1 Motivation for developing the *P. aeruginosa* panel

The emergence of a MDR bacterial population has led to the concern of entering a "post-antibiotic era" for infectious diseases. *P. aeruginosa* is an important pathogen responsible for nosocomial infections in ICUs and has developed resistance to many commonly used antibiotics, including aminoglycosides, quinolones and β-lactams (Pang et al. 2019). The spread of these MDR strains globally has led to the identification of "high-risk clones" such as STs 235, 111, 233, 244, 357, 308, 175, 277, 654, and 298 (Del Barrio-Tofiño, López-Causapé, and Oliver 2020). For example, ST-175 is widely distributed worldwide (Cholley et al. 2011), which is a coloniser of respiratory secretions in CF patients, and has been associated with the multi-resistant isolates in the hospital (García-Castillo et al. 2011; Gomila et al. 2013).

While the import of resistance through MGEs remains a concern, the most formidable challenge posed by *P. aeruginosa* is its remarkable ability to rapidly develop resistance during the treatment of infections. This includes mechanisms including over-expression of efflux pumps, impermeability resulting from porin modification or loss, target modification, and enzyme-mediated antimicrobial inactivation. These resistance mechanisms are significantly induced by nucleotide mutations in *P. aeruginosa* (Huber et al. 2021). Understanding the contribution of these mutations to antibiotic resistance is crucial for identifying novel emerging resistance patterns and developing effective treatment strategies.

To address the need for sensitive, specific, and rapid identification of *P. aeruginosa* STs and the potential antibiotic resistance that threatens effective therapy in the hospital environment, a targeted multiplex PCR sequencing panel was developed.

4.2 Panel development

4.2.1 Seven MLST genes and three housekeeping genes of *P. aeruginosa*

To characterise *P. aeruginosa* and relevant STs in the environment, three housekeeping genes (*16S rRNA1*, *16S rRNA2* and *gyrA*) and seven MLST genes (*acsA, aroE, guaA, mutL, nuoD, ppsA and trpE*) were retrieved from whole genome sequences of *P. aeruginosa* strains deposited in NCBI database. The length of MLST genes in this panel and the region to determine the ST from the PubMLST database (Jolley et al. 2018) is shown in Table 4.1.

Table 4.1 Length of seven MLST genes in the *P. aeruginosa* panel and PubMLST database

MLST genes	Nucleotide length in panel (bp)	Region in PubMLST	Nucleotide length in PubMLST (bp)
ppsA	2376	977-1347	370
aroE	825	172-670	498
nuoD	1782	627-993	366
trpE	1479	851-1294	443
acsA	1956	954-1344	390
guaA	1578	690-1063	373
mutL	1902	522-964	442

4.2.2 Antibiotic resistance delivered by *P. aeruginosa*

Here, the ARGs of *P. aeruginosa* were sourced from the CARD database (Alcock et al. 2020). The comprehensive selection process, encompassing AR mechanisms, AR families, and potential hosts of ARGs, relied entirely on the data available in the CARD database as well. Five aspects were considered to determine the target genes, namely:

(1) Frequently used antibiotics: to reflect the prevalent ARGs in the hospital setting, it is important to consider the commonly used drugs, such as aminoglycosides, carbapenems, cephalosporins, glycopeptides, macrolides, MLS, phenicols, sulfonamides and trimethoprim. The selection of popularly applied drugs can help identify the ARGs that are most relevant and are likely to be encountered in the hospital environment.

(2) Exclusion of intrinsic mechanisms of AR: intrinsic mechanisms such as efflux pumps and alterations in membrane permeability are naturally occurring defense mechanisms in bacteria. These mechanisms

can contribute to intrinsic resistance and are not necessarily indicative of acquired resistance developed by pathogens. In the context of targeting genes for analysis, it may be reasonable to exclude resistance genes associated with these intrinsic mechanisms if they have less significance in reflecting the development of acquired resistance by pathogens.

(3) Horizontal gene transfer: HGT is a crucial mechanism for the transmission and acquisition of ARGs. Through HGT, bacteria can acquire novel ARGs from other bacteria, leading to the spread of resistance traits within microbial populations. Conjugative MGEs such as plasmids, transposons, and ICEs play a significant role in facilitating HGT. These elements carry ARGs and can transfer them between bacterial strains or species. Therefore, genes associated with resistance to widely used antibiotics like beta-lactams, aminoglycosides, and vancomycin, which are frequently mediated by MGEs, are of great importance.

(4) Mutation (Horcajada et al., 2019): the mutations of transcriptional regulator genes, such as *mexT, mexS, mexT, mexR, mexZ, nfxB, nalC and nalD*, and outer membrane porin precursor genes, such as *oprD*, have been identified as key mechanisms contributing to antibiotic resistance. Mutations in these genes can lead to overexpression of efflux pumps and reduced expression of porins on the outer membrane, resulting in increased resistance to new drugs. Therefore, in order to detect novel resistance due to overproductions, those genes were regarded as the target genes.

4.2.3 Sequence database of target genes of *P. aeruginosa*

The database here included clustering, alignment, and trimming of target genes described in section *2.4.2*.

To analyse the homologous genes acquired from *section 4.2.2,* sequences with a similarity greater than 95% were clustered using the cd-hit algorithm (Fu et al. 2012) (`cd-hit-est -i input.fa -aL 0.95 -aS 0.95 -c 0.95 -o output.cluster -d 0 -g 1; make_multi_seq.pl input.fa output.clstr output.fa n`). In addition, sequences with a high identity of 99-100% to the representative sequence within the cluster were excluded from the analysis. This step was taken to avoid bias in primer selection, as highly similar sequences could lead to non-specific amplification. Next, the homologous genes in the cluster were subjected to MAFFT (Katoh et al. 2002) to generate a multi-

sequence alignment (`mafft input.fa > output.fa`), after which the JalView (Waterhouse et al. 2009) was used for viewing and trimming the alignment. By using JalView, the sequences within the cluster were adjusted to approximately the same length, ensuring consistency in subsequent analyses. Notably, as the target amplicon length was defined as 400 bp±20%, any genes below 420 bp in length were excluded from the analysis.

4.2.4 Primer design of target genes

In terms of principles for designing primers for the tiling multiplex PCR, it consisted of four elements. (1) Primer length: usually, primers of short length, in the range of 18-25 bases were used. (2) Melting temperature: primers with similar Tm, preferably between 55°C-60°C were used. For sequences with high GC content, primers with a higher Tm (preferably 75°C-80°C) were recommended. A Tm variation of between 3°-5° C was acceptable for primers used in a pool. (3) Specificity: in a multiplex assay, where multiple target sequences are present in a single reaction vessel, competition can occur. To ensure accurate and reliable results, the primers should be designed to specifically amplify their intended target sequences without cross-reactivity or interference from other non-target sequences. (4) Avoid heterodimers formation: the designed primers were checked for the formation of heterodimers with all the primers present in the reaction mixture using primer3-py (select top scoring, non-interacting pair).

To generate primers, PrimalScheme (Quick et al. 2017) was used (`pip install primalscheme; primalscheme multiplex input.fa`). Notably, the target genes of *P. aeruginosa* exhibited relatively high GC contents ranging from 26.3 % to 71.2%. Regarding the PrimalScheme, it allows for customised parameter settings within the script of *config.py*. For the *P. aeruginosa* scheme, an attempt was made to optimise the primer design by setting specific criteria. These criteria included: (1) Amplicon Length: the desired amplicon length was set between 380 bp and 420 bp. (2) GC Content: the minimum and maximum GC content of the primers was set to 40% and 65%. (3) Tm: the minimum, optimum, and maximum Tm of the primers were set to 64.5°C, 66°C, and 67.5°C, respectively. (4) Primer Size: the minimum, optimum, and maximum size of the primers were set to 18 bp, 20 bp, and 34 bp, respectively. The evaluation of performance of customised parameters was based on two criteria: the number of genes for which primers could be successfully generated and the amplicon coverage across the full nucleotide

length. After the evaluation, all designed primers were ordered from Integrated DNA Technologies (IDT), and 100 μM primer stock solution was prepared for the downstream usage.

4.2.5 High-GC PCR reactions

The PCR reagents used here were dNTP mix 25mM each (Thermo Fisher Scientific, USA), and Q5® High-Fidelity DNA Polymerase pack (NEBNext®, England) containing Q5 High-Fidelity DNA Polymerase, Q5 Reaction Buffer (5x) and Q5® High GC Enhancer. With the high GC contents and melting temperatures of primers, the reagent of GC enhancer was added to the PCR reaction solution to help break the stable structures at GC-rich regions.

(1) High-GC single-plex PCR

Prior to conducting the multiplex PCR with the addition of a GC enhancer, the single-plex PCR was performed firstly to investigate the efficiency of the GC enhancer. The total volume of mastermix was 25 μL. The single-plex PCR reactions were performed with three primer pairs along the DNA samples (around 10ng/μL) extracted from the pure *P. aeruginosa* strain. The primers' and PCR components' information, as well as the single-plex PCR programme are shown in Tables S4.1-S4.3.

(2) High-GC-two-step-touchdown-multiplex PCR

The DNA samples used were extracted from pure strains of *P. aeruginosa*, as well as two tap water samples as described in *section 4.3*. The total volume of mastermix was 25 μL. The components' information and the touch-down programme condition are shown in Table S4.4 and Table S4.5. To assess the LOD value of the multiplex PCR, a three-fold serial dilution of DNA samples extracted from *P. aeruginosa* strain cells was performed. The dilutions ranged from 300 cells down to 3 cells per reaction. The concentration of genome copies present in these DNA samples was quantified using qPCR, as outlined in *section 4.2.6*. To ensure the robustness and reliability of the assay, the multiplex PCR was optimised by repeatedly conducting the assay with the serially diluted DNA samples. This optimisation process was carried out independently at least five times in both UOB and SusTech.

(3) Electrophoresis

Initially, the 2% agarose gel (2g agarose per 100 ml TBE buffer) suited for 400 bp amplicons was created and boiled. Then, the SYBR Safe DNA Gel Staing was added and mixed to the gel when the gel cooling to approximately 50°C. Next, the stained gel was poured into a mold with a comb being placed in the mold. After the gel was solidified into a matrix, the comb was removed. Then 9 μL of DNA sample in addition to 1μL of loading buffer (Thermo Fisher Scientific, USA) were mixed, of which 9 μL mixture was loaded to each well of the solid gel. Last, ran the gel for 40 minutes (short DNA) with a voltage of 120 volts and a current value of 0.01A. After the DNA electrophoresis finished, the amplicon bands were visualised via the UV transilluminator (BIO-RAD, USA).

4.2.6 Real-Time PCR

The instrument for the RT-PCR was Applied Biosystems™ QuantStudio 6 and 7 Flex Real-Time PCR Systems (USA). To quantify *P. aeruginosa,* the kit used here was Primerdesign Ltd ™ *Pseudomonas aeruginosa* genesig Standard Kit. The RT-PCR process was conducted following the provided protocol in the kit.

4.2.7 MinION amplicon sequencing

(1) ONT kits and other reagents

In this chapter, DNA sequencing was performed using ONT kits. It should be noted that some of the ONT kits used in the experiments were discontinued at the time of writing. However, the reagents included in the latest available kits were found to be highly comparable to those used in the previous kits. Although specific kit versions have been discontinued, the fundamental components and methodologies remain largely consistent in the newer kits. As a result, the data generated from the latest kits can be considered representative of the sequencing results obtained using the previous kits.

The kits here are the latest version updated by ONT, which include Native Barcoding Kit 96 V14 (SQK-NBD114.96) instead of old SQK-LSK109, SQK-LSK110 and EXP-NBD196. Sequencing Auxiliary Vials V14 (EXP-AUX003), Native Barcoding Expansion V14 (EXP-NBA114), Flow Cell Priming Kit (EXP-FLP004), and Flow Cell Wash Kit (EXP-WSH004).

Other reagents from NEW ENGLAND BioLabs *InC.* (NEBNext®, England) used were NEB Blunt/TA Ligase Master Mix, NEBNext FFPE Repair Mix, NEBNext Ultra II End repair/dA-tailing Module, and NEBNext Quick Ligation Module.

(2) Library preparation

Sequencing libraries were prepared using the protocol developed by Quick et al. (2017). Changes included in my procedures were:

(1) In the sample normalisation step (step 12A (ii)), the normalisation DNA input was 30ng per sample (three samples in total) if the sample only had specific 400bp bands.

(2) Following the end-prep incubation step (step 12A (iii)), the cleanup with SPRI beads was conducted for each sample. The DNA was eluted in 5µl EB.

(3) In the 'Barcode Ligation' step (step 12A (iv)), the volume of each component decreased to: NBXX barcode of 1.25 µl, Blunt/TA Ligase Master Mix DNA of 5 µl, DNA of 4.55 µl, making a total of 10.8 µl.

(3) MinION sequencing

The prepared DNA library with 10 ng-20 ng was loaded on the new flow cell or the washed flow cell following the instructions of the above Baseline protocol. In this study, I used MinION flowcell R9.4.1, which has been discontinued and replaced by the flowcell R10.4.1.

(4) Use flowcell maximally

(i) Washing the flowcell

To reuse a flow cell for multiple sequencing runs, it is essential to perform a thorough washing process. This step ensured the removal of any residual contaminants or impurities that might have accumulated on the flow cell during previous sequencing runs. Washing could also help minimise the carryover of nucleotide molecules from previous runs, reducing the risk of cross-contamination between samples. Additionally, it played a vital role in maintaining the cleanliness and optimal functionality of the active pores on the flow cell, which were responsible for accurate base calling during sequencing. By performing regular and effective washing, the flow cell could be prepared for subsequent runs, ensuring reliable and high-quality sequencing data while prolonging the lifespan of the flow cell.

There were two scenarios when using the Flow Cell Wash Kit (EXP-WSH004). One was “wash and store”, and another one was “wash and reuse” the flowcell straight away. After the washing step, it was generally not recommended to reload different DNA libraries with the same barcode, which could lead to cross-contamination and compromise the accuracy and integrity of the sequencing data.

(ii) Refueling the flowcell

Refueling is the replenishment of motor fuel in the sequencing experiment through the addition of Flush Buffer (FB) from the Flow Cell Priming Kit. Normally, in situations where the "translocation speed" during a sequencing run dropped significantly below the average level, it was common practice to perform a refueling action to ensure optimal sequencing performance. To refuel the flowcell during the sequencing run, the following steps were implemented: (i) "Pause run": The sequencing run was temporarily halted to allow for the refueling process. (ii) FB flush: A FB flush was performed to remove any potential contaminants or obstacles that might be affecting the translocation speed. This step helped maintain the efficiency and accuracy of the sequencing process. (iii) Resuming the run: After the FB flush, the sequencing run was resumed to continue the data acquisition process.

4.2.8 Bioinformatic analysis of amplicon reads

(1) Consensus reads

The raw FAST5 files generated during the sequencing process were subjected to basecalling using Guppy v4.2.2, which is a software tool provided by ONT. Basecalling is the process of converting the raw electrical signals obtained from the nanopore sequencing into nucleotide sequences. During basecalling, reads with a quality score (q-score) below 7 were considered to have lower confidence and were subsequently discarded. Quality scores provide an estimation of the accuracy of each base in a read, with higher scores indicating higher confidence in the basecall. After the basecalling step (`guppy_basecaller –i fast5_dir –s out_dir --flowcell <flowcell> --kit SQK109 –c dna_r9.4.1_450bps_hac.cfg \ --num_callers 1 --cpu_threads_per_caller 14 -x "cuda:0"`), the demultiplexing process was performed using Guppy_barcoder (`guppy_barcoder -i fastq_dir -s out_dir --barcode_kits EXP-NBD196 -x 'cuda:0' --trim_barcode --require_barcodes_both_ends --trim_adapters --recursive --compress_fastq`). Demultiplexing involves assigning each sequence read to its respective sample based on the barcode information. The demultiplexed FASTQ files obtained from Guppy_barcoder were then subjected to further analysis using the Artic bioinformatics pipeline (`artic minion <gene_scheme> <sample.fastq>;` https://github.com/artic-network/fieldbioinformatics). The pipeline consisted of three key steps. Firstly, the demultiplexed FASTQ files were aligned to a reference sequence. The reference sequence used in this case was a pseudomolecule that encompassed 103 target genes, which were also the representative genes for the primer design. Secondly, the resulting alignment files were processed in BAM format. This involved converting them to BAM format, indexing, sorting, and trimming using samtools software (Danecek et al., 2021). These processes optimised the alignment files for subsequent analysis. Lastly, the consensus sequence generation step was vital. The bcftools software (Danecek et al., 2021) was employed to generate the consensus sequence for each gene of interest. The consensus sequence represented the most likely nucleotide at each position, accounting for the observed variations in the aligned reads. To ensure reliable variant calling, regions with sequence depths below 20× were masked by replacing the corresponding bases with "N". This masking helped filter out regions with low coverage, reducing potential errors or biases in the variant calling process. By

following these steps in the Arctic pipeline, reliable and accurate consensus sequences were obtained for each gene, enabling further analysis and interpretation of *P. aeruginosa* data.

(2) Performance evaluation of amplicon sequencing

The performance of amplicon sequencing was evaluated by five elements, which were yielded basecalled reads, alignment ratio of tested samples, sequencing depth, base coverage and non-mask region of each target gene. Specifically, the basecalled reads referred to the number of reads obtained after basecalling. The alignment ratio represented the proportion of Fastq reads that were successfully aligned to the target genes. The sequencing depth quantified the average number of times a specific base in the target gene region was sequenced, which was calculated via the bedtools described in *section 2.6*. The base coverage represented the regions of the gene that were successfully sequenced. The non-mask region was the portion of the target gene where the sequencing depth exceeded a specified threshold, typically set at 20×. It indicates the percentage of the gene region that had sufficient sequencing depth for reliable analysis and interpretation.

(3) STs determination of *P. aeruginosa*

To determine the ST of *P. aeruginosa* isolates, there were two options. One option was from the PubMLST Website (Jolley et al. 2018), which provides a database of *P. aeruginosa* ST strains. The database currently contains 4,528 updated *P. aeruginosa* ST strains. On the website, the consensus read of each MLST gene was input, and then the allele and ST assignments were performed using the tools available. Another alternative was using the mlst tool developed by Seemann (https://github.com/tseemann/mlst). This tool allowed us to scan contig files against PubMLST typing schemes, including the *P. aeruginosa* scheme. By providing contig files as input, the mlst tool could determine the ST of *P. aeruginosa* isolates (`mlst consensus.mlst.fa`).

(4) SNP-related analysis of antibiotic resistance

The drug resistance encompassed by *P. aeruginosa* were identified with my customised ARGs reference in conjugation with the Artic pipeline. Besides, by detecting confident SNPs with a sequencing depth exceeding 20× and a variant-calling quality over 20 using Medaka developed by ONT (`medaka variant reference.fasta input.hdf input.vcf`), the antibiotic resistance associated with gene mutations was able to be predicted (Take six related genes for instance as follows). During the variant calling process with Medaka, both major SNPs and minor SNPs can be detected.

(i) *gyrA* mutation to fluoroquinolone resistance: In Gram-negative bacteria including *P. aeruginosa*, amino acid alterations found in *gyrA* are associated with high-level fluoroquinolone resistance (Park et al. 2020). *P. aeruginosa* becomes resistant to fluoroquinolone through mutation of the quinolone resistant-determining region (QRDR), which is 67–106 amino-acid (AA) motif in the GyrA protein (Takenouchi, Sakagawa, and Sugawara 1999; Cabot et al. 2016; Farahi, Ali, and Gharavi 2018; Park et al. 2020).

(ii) *oprD* mutation to carbapenem resistance: The main porin for uptake of carbapenems in *P. aeruginosa* is the outer membrane protein OprD (C. H. Kim et al. 2016). Mutations in *oprD* caused by nucleotide alterations in the *oprD* structural gene, such as SNPs between nucleotide positions 62 and 72, have been found to be the major mechanisms leading to the inactivation of OprD with concomitant loss of the porin from *P. aeruginosa* outer membranes and increases of the MICs of carbapenems (Kao et al. 2016). However, most of the amino acid substitutions observed in OprD are conserved mutations consisting of changes from one hydrophobic residue to another that is also hydrophobic (Ocampo-Sosa et al. 2012).

(iii) *mexT* and *mexS* mutations: *mexT* is a regulator gene that positively regulates the MexEF-OprN multidrug efflux system to efflux fluoroquinolones, trimethoprim, and chloramphenicol (Kumar and Schweizer 2011). The nucleotide mutations of *mexT* can lead to alterations in the regulatory mechanisms of antibiotic resistance of *P. aeruginosa*. This can result in increased expression of efflux pumps, thereby decreasing the drug susceptibility of *P. aeruginosa* (Köhler et al. 1999). Additionally, mutations in *mexS*, a gene immediately upstream of *mexT*, have been shown to enhance mexEF-oprN expression and decrease OprD production, providing a multidrug resistance profile indistinguishable from that of previously described mutants (Sobel, Neshat, and Poole 2005).

(iv) *mexZ* mutation: The occurrence of mutational inactivation in the *mexZ* repressor genes can lead to *P. aeruginosa* frequently becoming resistant to aminoglycosides through the overexpression of the MexXY-OprM efflux pump (Aires et al. 1999).

(v) *nalD* mutation: NalD is a transcriptional repressor of the efflux system MexAB-OprM and mutations in NalD have been linked to aztreonam (beta-lactam) resistance (Sobel et al. 2005). For example, a point mutation F198L found in NalD are consistent with the finding that NalD mutation can increase aztreonam MIC on average by over 2 folds (Yan et al. 2019).

4.3 Validation of the *P. aeruginosa* panel

In this section, I will present the validation of the *P. aeruginosa* panel involving 103 target genes. For the validation in UOB, three samples were used, which included one positive control and two tap water samples. The positive control strain, PAO1, was isolated from a clinical sample and provided by *MicrobesNG* in Birmingham, UK. The tap water samples were collected in September 2019 from the emergency department at *Queen Elizabeth Hospital Birmingham*, UK. In the validation conducted at SusTech, the positive control strain used was *P. aeruginosa* ATCC 9027 (PAO1). The overall validation work with three tested samples could be finished within two days. The entire bioinformatic analysis performed in this chapter relied on the MinION sequencing data generated at UOB.

4.3.1 Generated primers

Referring to the CARD database, a total of 242 annotated ARGs were identified in *P. aeruginosa*. After applying the selection criteria, a total of 209 potential target ARGs from 42 gene families were chosen for further analysis. These 209 candidate ARGs were categorized into 93 ARG clusters. Among these clusters, eight genes were identified as responsible for enhancing resistance levels and playing a role in shaping the mutational resistome of *P. aeruginosa*. Additionally, 67 ARGs were associated with HGT. Figure 4.1a and Figure 4.1b present the types and mechanisms of the 93 target ARGs. These ARGs, along with an additional 10 housekeeping genes, were compiled into a total of 103 input FASTA files (Table

S4.6). By utilising customised parameters in the PrimalScheme software, a total of 536 primers were generated, successfully recovering 103 of the desired target genes. The average amplicon coverage achieved was 97.1%, indicating a high level of specificity and efficiency in capturing the intended genetic regions. For these 536 primers, they were divided into two pools to prevent the formation of chimeric products between neighboring amplicons. Pool 1 consisted of 268 primers, while pool 2 also contained 268 primers. The specific primer sequences and their respective assignments in the two pools can be found in Table S4.7.

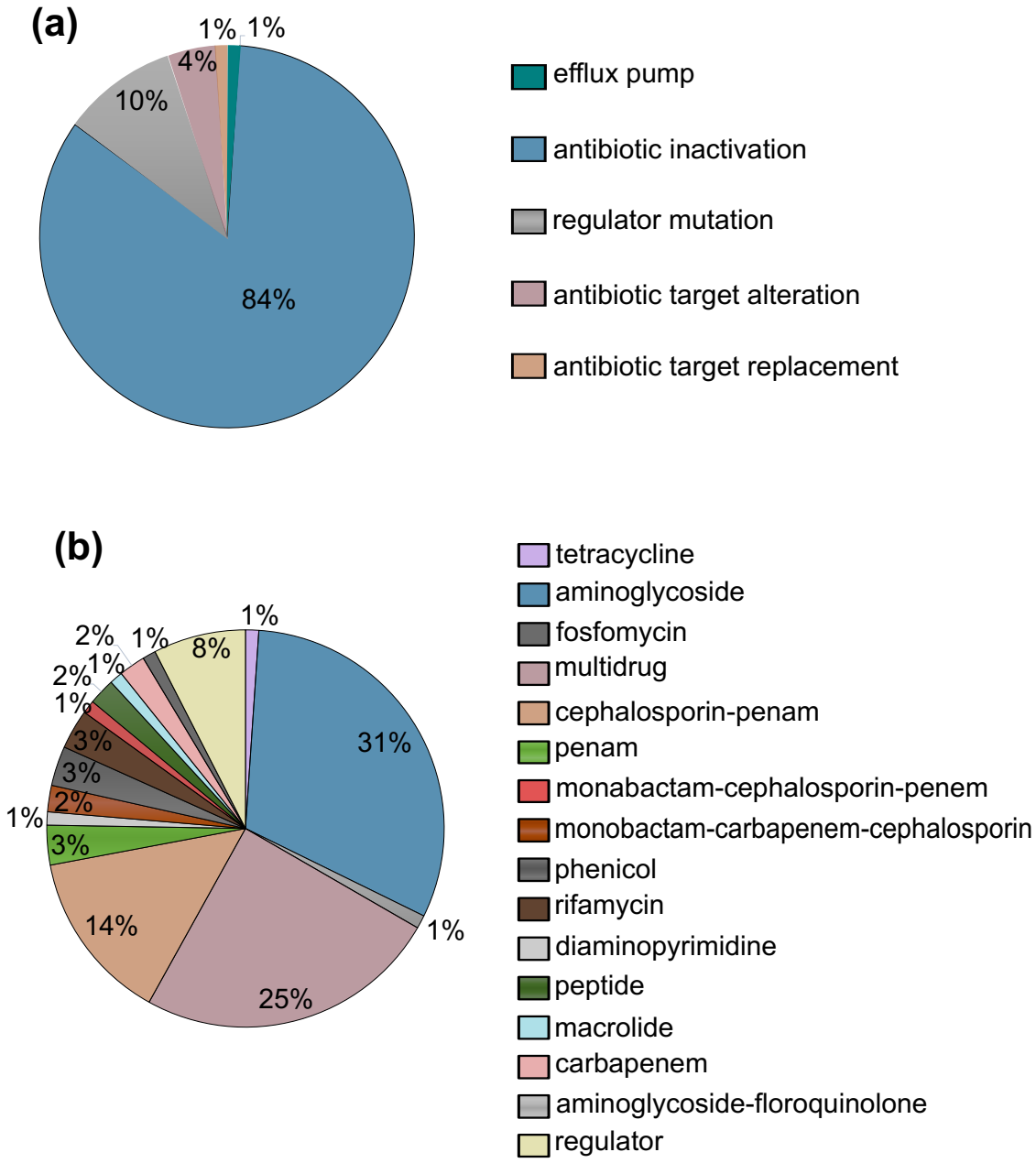


Figure 4.1 (a) Types of 93 target ARGs; (b) Resistance mechanisms of 93 target ARGs.

4.3.2 Sensitivity of high-GC two-step touchdown multiplex PCR

The GC enhancer was necessary for successful amplification in cases where the target regions had a high GC content and required a high annealing temperature. The tiling multiplex PCR results demonstrate that the optimised PrimalPCR process could obtain target amplicons from the sample containing as low as three *P. aeruginosa* cells per reaction. The optimised amplicon bands could be observed with 300 and 30 copies per reaction (Figure 4.2).

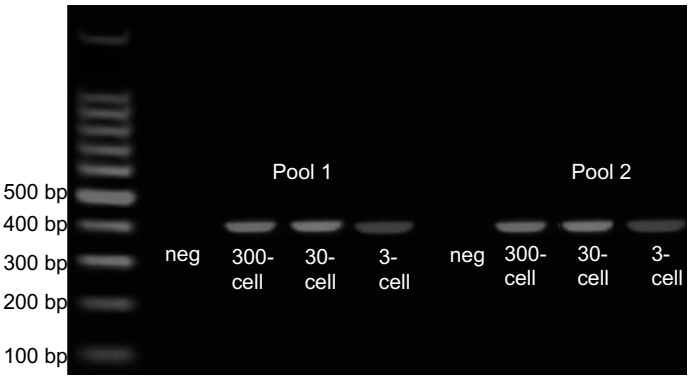


Figure 4.2 Optimisation of PrimalPCR with DNA samples extracted from isolated *P. aeruginosa* strain. DNA marker showed here is 100 bp ladder. DNA templates applied in pool1 and pool2 are (from left to right): nuclease-free water, DNA templates containing 300 cells, 30 cells, and 3 cells respectively.

4.3.3 Performance evaluation of PrimalPCR-based MinION Sequencing

During the targeted MinION sequencing, a total of 432,151 Fastq basecalling reads were generated from three tested samples with three *P. aeruginosa* cells. The analysis reveals variations in the number of Fastq reads among the samples. Specifically, the positive sample yielded 100.827K reads, the first tap water sample had 200.413K reads, and the second tap water sample obtained 130.911K reads. In contrast, the negative sample obtained 20 Fastq reads that were found to be unaligned with the reference sequences of the target genes. Furthermore, the alignment rates of the samples provided additional insights. The positive sample demonstrated a 100% alignment rate, indicating that all the reads successfully aligned with the reference genes. The first tap water sample exhibited a slightly lower alignment rate of 97.0%,

while the tap second water sample had an alignment rate of 97.2%. Among 103 target genes of *P. aeruginosa*, targeted sequencing enabled the recovery of 28 target genes in total (Table 4.2). All ten housekeeping genes and eight regulator genes were detected in all three samples, indicating their presence and successful recovery. Regarding other ARGs, the positive sample contained seven ARGs, the first tap water sample had eight ARGs, and the second tap water sample harboured ten ARGs.

Table 4.2 Number of recovered genes in three tested samples by *P. aeruginosa* panel

Samples	Recovered genes in total	Recovered housekeeping genes	Recovered regulator genes for antibiotic resistance	Recovered other ARGs
Postive	25	10	8	7
W1	27	10	8	8
W2	28	10	8	10

The average sequencing depth (SD) of recovered genes in three samples was 1523×. However, there were variations among the positive and water samples (Figure 4.3a). The positive sample had an average SD of 867×, the first tap water sample had 2457×, and the second tap water sample had 1245×. Regarding the base coverage depicted in Figure 4.3b, the positive sample and the two water samples achieved 96.2%, 95.8%, and 95.9% base coverage, respectively. The non-mask region percentage across the detected genes exhibited a similar distribution (Figure 4.3c), with the positive sample and the two water samples covering 96.2%, 95.6%, and 95.4% of the gene region that had sufficient sequencing depths (over 20×) for reliable analysis and interpretation. Noticeably, even though the non-mask region of MLST genes such as *acsA*, *aroE*, *mutL*, and *ppsA* genes did not reach 100% coverage, the coding region necessary for determining the sequence types was 100% recovered. This indicates that the sequence types can be accurately determined using the consensus reads generated by the *P. aeruginosa* panel. Despite the observed fluctuations in the SD, base coverage, and non-mask region percentage, the overall performance evaluation demonstrates that the PrimalPCR-based MinION sequencing (PrimalSeq) panel for *P. aeruginosa* can efficiently and specifically recover target genes within a 48-hour timeframe for the downstream analysis such as the STs determination and antibiotic resistance identification.

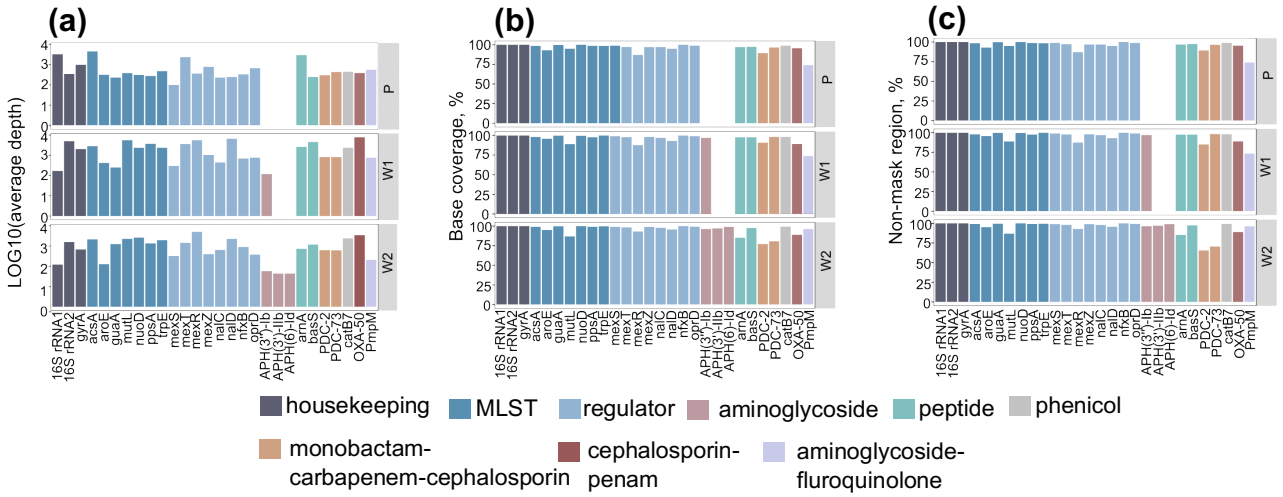


Figure 4.3 Performance evaluation of MinION sequencing across recovered genes in three tested samples. (a) Distribution of logarithm (10) value of average SDs; (b) Distribution of base coverage (%); (c) Distribution of non-mask region (%).

4.3.4 STs determination

Based on the MLST analysis using the seven-loci scheme, it was determined that the *P. aeruginosa* strains present in the three samples belonged to ST549, which was identical to the sequence type of the PAO1 strain.

The connection between high-risk clones and horizontally acquired resistance mechanisms *in P. aeruginosa* is well-established (Del Barrio-Tofiño, López-Causapé, and Oliver 2020). Among extended-spectrum beta-lactamases- or metallo-β-Lactamase-producing *P. aeruginosa* isolates, a majority of them are found within specific clones. The most prevalent clone is ST235, followed by ST111 (Oliver et al. 2015). These clones often exhibit mutations in the QRDR, leading to fluoroquinolone resistance. Additionally, mutations in the oprD gene, associated with carbapenem resistance, are frequently observed in these clones (Jaillard et al. 2017; Del Barrio-Tofiño et al. 2017). A recent genomic analysis has suggested that the presence of DprA, a determinant involved in homologous recombination and present in transformable species, specifically in ST235, may enhance the ability of this high-risk clone to acquire

and maintain foreign resistance elements at a higher rate compared to other *P. aeruginosa* clones (Treepong et al. 2018).

4.3.5 Antibiotic resistance in tested three samples

The relatively expansive genome size of *P. aeruginosa*, coupled with its genetic adaptability, not only enables its versatile metabolic responses to various environments, but also facilitates the acquisition of AR (Lee et al. 2023). Acquired resistance mechanisms encompass the horizontal transfer of ARGs and mutational resistance. It has been proposed that the acquisition of ARGs is contingent upon the strain's existing resistance to antibiotics; essentially, lower resistance to antibiotics correlates with a decreased acquisition of ARGs (Ahmed 2022). While both acquired resistance mechanisms contribute to MDR/extensively drug-resistant *P. aeruginosa*, the emergence of strains harboring horizontally acquired ARGs is increasingly recognised as a growing threat in clinical settings (Del Barrio-Tofiño et al. 2020).

(1) Intrinsic and horizontally acquired antibiotic resistance

Regarding ARGs, the positive sample contained seven ARGs, the first tap water sample had eight ARGs, and the second tap water sample harbored ten ARGs. Notably, all samples had seven common ARGs *arnA*, *basS*, *PDC-2*, *PDC-73*, *catB7*, *OXA-50* and *PmpM*, which conferred resistance to the peptide, monobactam-carbapenem-cephalosporin (MCC), phenicol, cephalosporin-penam (CP), and aminoglycoside-fluoroquinolone (AF) classes, respectively. The blast results for these seven ARGs, conducted through the web interface (Johnson et al. 2008), exhibited 100%-similarity alignment solely with *P. aeruginosa* strains. Apart from the conserved genes *arnA* and *basS*, the remaining five ARGs are intrinsic to the AR profile of *P. aeruginosa*. To be detail, in addition to conferring resistance to peptides by modifying the peptide target, *arnA* is also involved in amino sugar and nucleotide sugar metabolism (Gatzeva-Topalova et al. 2005), and *basS*, a subset of the *pmrB* gene, plays a role in peptide efflux mechanisms within the *P. aeruginosa* (McPhee et al. 2003).

Regarding the intrinsic ARGs of *P. aeruginosa*, *PmpM* represents the first documented instance of an H^+-coupled multidrug efflux pump within the multidrug and toxic compound extrusion (MATE) family,

notably effective against AF class antibiotics (He et al. 2004). *PDC-2* and *PDC-73* are chromosomally encoded class C β-lactamases specific to *P. aeruginosa* (Colque et al. 2022), with the PDC class recognised for its intrinsic antibiotic resistance properties (Pang et al. 2019). However, the inducible expression of PDC β-lactamases to provide the AR is in response to the persistent presence of antibiotics or other environmental stimuli, and this characterisation is called adaptive resistance (Torrens et al. 2019). *OXA-50* has also been recognised for the intrinsic AR (Pang et al. 2019), although other *OXA*-genes like *OXA-10* has been reported to reside in integrons and function as mobile ARGs (Lee et al. 2023). Additionally, the research conducted by Ahmed (2022) has also indicated that both susceptible and resistant *P. aeruginosa* strains harbor *OXA-50* and *catB7*, suggesting the potential for natural transformation in *P. aeruginosa* to acquire these genes (Nolan et al. 2020).

An interesting observation was made regarding aminoglycoside resistance. In the positive isolate sample, no resistance to aminoglycoside was detected. Nonetheless, the first tap water sample was found to contain aminoglycoside-resistant gene *APH(3")-Ib.* In the second tap water sample, aminoglycoside resistance was identified with the presence of multiple aminoglycoside-resistant genes, including *APH(3")-Ib*, *APH(3')-IIb* and *APH(6)-Id*. There is a concept positing that aminoglycoside-modifying enzymes (AMEs) originate from antibiotic producers is rooted in the HGT events occurring between soil microorganisms and pathogenic bacteria (Yoon et al. 2017). Nevertheless, unlike *APH(3")-Ib* and *APH(6)-Id*, *APH(3')-IIb* is situated in the chromosome of *P. aeruginosa* (Zeng and Jin 2003), with limited evidence suggesting its mobility. The blast analysis of consensus sequence of *APH(3')-IIb* also demonstrated exclusive alignment to *P. aeruginosa* with 100% similarity. In contrast, substantial evidence supports the horizontal transfer nature of *APH(3")-Ib* and *APH(6)-Id* genes (Tyson et al. 2015; Genthe et al. 2016; Leekitcharoenphon et al. 2021; Kyung et al. 2023). For example, *APH(3")-Ib* is recognised as a plasmid-encoded ARG in multiple hosts (Scholz et al. 1989) and is also present on ICEs in *Histophilus somni* (Farghaly et al. 2023), suggesting a HGT pathway via the conjugation. Moreover, *APH(6)-Id* has been screened within integrons in *A. baumannii* (Sezmis et al. 2023) and in the transposon *Tn5393* of *Pseudomonas syringae* (Collins et al. 2007). Further, the blast results showing a 100% similarity of *APH(3")-Ib* and *APH(6)-Id* also suggest that beyond being present in the genome of *P. aeruginosa*, these genes were also found in the genomes of other species, such as *E. coli* and *K. pneumophila*.

In consequence, concerning two water samples, the PrimalSeq panel may not conclusively establish that horizontally-related ARGs like *APH(3'')-Ib* and *APH(6)-Id* can serve as direct indicators of aminoglycoside resistance in *P. aeruginosa*. Further tests utilising MIC testing and complete genome analysis are essential to pinpoint the primary hosts responsible for harboring these ARGs.

(2) Nucleotide polymorphisms related to antibiotic resistance

P. aeruginosa can develop further antibiotic resistance via the acquisition of chromosomal mutations. Table 4.3 presents a summary of the principal genes known to enhance resistance levels and shape the mutational resistome of *P. aeruginosa*. The antibiotics related to such resistance mechanisms are aminoglycoside, cefepime, ceftazidime-avibactam, ceftolozane-tazobactam, fluoroquinolone, fosfomycin, imipenem, meropenem, piperacillin-tazobactam and polymyxins, among which ceftazidime belongs to cephalosporin, piperacillin belongs to penicillin beta-lactam, and avibactam in addition to tazobactam belong to non-β-lactam β-lactamase inhibitors.

Table 4.3 Main genes increasing antibiotic resistance levels of *P. aeruginosa* (adapted from Horcajada et al., 2019)

Genes	Resistance mechanism	Antibiotics resistance
gyrA	Quinolone target modification	Fluoroquinolones
gyrB	Quinolone target modification	Fluoroquinolones
parC	Quinolone target modification	Fluoroquinolones
parE	Quinolone target modification	Fluoroquinolones
pmrA, *pmrB*, *phoQ*, *cprS*, *colR*, *colS*	Lipopolysaccharide modification	Polymyxins
parR	Lipopolysaccharide modification	Polymyxins
	OprD downregulation	Imipenem, meropenem
	MexEF-OprN hyperproduction	Fluoroquinolones
	MexXY hyperproduction	Fluoroquinolones, aminoglycosides, cefepime
parS	Lipopolysaccharide modification	Polymyxins
	OprD downregulation	Imipenem, meropenem
	MexEF-OprN hyperproduction	Fluoroquinolones
	MexXY hyperproduction	Fluoroquinolones, aminoglycosides, cefepime
mexR, *nalC*, *nalD*	MexAB-OprM hyperproduction	Fluoroquinolones, ceftazidime, cefepime, piperacillin-tazobactam, meropenem, ceftazidime-avibactam
nfxB	MexCD-OprJ hyperproduction	Fluoroquinolones, cefepime
mexS	MexEF-OprN hyperproduction	Fluoroquinolones
	OprD downregulation	Imipenem, meropenem
mexT	MexEF-OprN hyperproduction	Fluoroquinolones
	OprD downregulation	Imipenem, meropenem
cmrA, *mvaT*, *PA3271*	MexEF-OprN hyperproduction	Fluoroquinolones
mexZ, *PA5471.1*, *amgS*	MexXY hyperproduction	Fluoroquinolones, aminoglycosides, cefepime
oprD	OprD porin inactivation	Imipenem, meropenem
ampC	AmpC structural modification	Ceftolozane-tazobactam, ceftazidime-avibactam
ampD, *ampDh2*, *ampDh3*, *ampR*, *dacB*, *mpl*	AmpC hyperproduction	Ceftazidime, cefepime, piperacillin-tazobactam
ftsI	β-Lactam target modification	Ceftazidime, cefepime, piperacillin-tazobactam, ceftolozane-tazobactam, ceftazidime-avibactam, meropenem
fusA1	Aminoglycoside target modification	Aminoglycosides
glpT	Inactivation of transporter protein	Fosfomycin
rpoB	Rifampin target modification	Rifampin

(i) Fluoroquinolone resistance

It has been reported that fluoroquinolone resistance in *P. aeruginosa* is primarily mediated through mutations in the *gyrA* gene, with mutations in the *parC* gene and efflux regulatory genes being secondary factors (Higgins et al. 2003). When the *gyrA* gene undergoes mutation, *P. aeruginosa* exhibits resistance to fluoroquinolones by altering the QRDR (López-Causapé et al. 2018). Previous studies, utilising WGS and targeted *gyrA* sequencing, have identified various mutations in QRDR motifs (Table 4.4), with 67-106 AA motifs of the GyrA protein playing a crucial role in quinolone resistance development in *P. aeruginosa*. Specifically, residues T83 (ACC-ATC) and D87 (GAC-GGC) have been consistently associated with the QRDR motif and are considered essential for the development of fluoroquinolone resistance in *P. aeruginosa* in clinical environments (Takenouchi et al. 1999; Cabot et al. 2016; Farahi et al. 2018; Park et al. 2020).

In the analysis of the *gyrA* gene in the panel, six SNPs were identified in the positive sample, 12 SNPs in the first tap water sample, and 10 SNPs in the second tap water sample (as shown in Table 4.5). Notably, the study reveals consistent mutations of T83 and D87 in the QRDR motif across all samples, which are known to be significant in conferring fluoroquinolone resistance. In addition to these two important mutations in the *gyrA* gene, other mutations were also detected, all of which were synonymous (silent) mutations that did not result in AA alterations.

In addition to *gyrA* mutations, overexpression of specific efflux pumps in *P. aeruginosa*, such as MexCD-OprJ (*nfxB*), MexEF-OprN (*mexS*, *mexT*), MexAB-OprM (*mexR*, *nalC*, *nalD*), or MexXY-OprM (*mexZ*), can also impact the susceptibility of the bacterium to fluoroquinolones (Bruchmann et al. 2013). Previous studies have reported mutations in these genes that contribute to fluoroquinolone resistance, and these findings are listed in Table 4.4. However, it is worth noting that there is limited research specifically focusing on mutations in the *mexT* or *mexS* genes and their contribution to fluoroquinolone resistance. Among those genes in my panel analysis, successful SNP calls were made for the *mexT*, *nfxB*, and *mexR* genes, and SNPs associated with fluoroquinolone resistance were only detected in *nfxB* and *mexR*. Specifically, for the *nfxB* gene, the SNP was identified in the 263-nt motif, resulting in a change from CTC to CCC, leading to an AA alteration from leucine to proline. In the case of *mexR* gene, the SNP was

observed in the 157-nt motif, causing an AA change from asparagine to aspartic acid. These SNPs align with the findings of previous studies mentioned in Table 4.4, indicating their relevance to fluoroquinolone resistance in *P. aeruginosa*.

As a conclusion, while the occurrence of *PmpM* signified inherent resistance to fluoroquinolones, the identification of missense SNPs in the *gyrA*, *nfxB* and *mexR* genes presented the compelling evidence for the diminished susceptibility of *P. aeruginosa* to the fluoroquinolone treatment.

Table 4.4 Genes mutations contributing to fluroquinolone resistance in the recent studies

Original codon	New codon	AA alteration	AA position	Genes	Resistance	References
ACC	ATC	Threonine - Isoleucine	83	*gyrA*	Fluroquinolone	(Park et al. 2020)
ACC	GCC	Threonine - Alanine	83			
GAC	GGC	Aspartic acid - Glycine	87			
CTG	CAG	Leucine - Glutamine	55			
GAC	AAC	Aspartic acid - Asparagine	82;87			
GAC	TAC	Aspartic acid - Tyrosine	87			
ACC	ATC	Threonine - Isoleucine	83			(Farahi et al. 2018)
ACC	ATC	Threonine - Isoleucine	83			(Takenouchi et al. 1999)
GAC	GGC	Aspartic acid - Glycine	87			
GAC	AAC	Aspartic acid - Asparagine	87			
ACC	ATC	Threonine - Isoleucine	83			(Cabot et al. 2016)
GAC	AAC	Aspartic acid - Asparagine	87			
GCC	GAA	Alanine - Glutamic acid	124	*nfxB*		(Higgins et al. 2003)
AGC	CGC	Serine - Arginine	36			(Morero et al. 2011)
CAC	CCC	Histidine - Proline	57			
ACC	CCC	Threonine - Proline	145			
CAG	CCG	Glutamine - Proline	175			
CAT	CCT	Histidine - Proline	179			
CTG	CCG	Leucine - Proline	26			
CAC	CGC	Histidine - Arginine	87			
CTC	CCC	Leucine - Proline	88			
CAT	CCT	Histidine - Proline	107	*mexR*		(Higgins et al. 2003)
CGG	GGG	Arginine - Glycine	21			
AGC	GGC	Serine -Glycine	26			
AAC	GGC	Asparagine - Glycine	79			
AAC	GAC	Asparagine - Aspartic acid	53			
GTG	-	Valine fs	151	*nalD*		(López et al. 2022)
CGC	-	Arginine*	164	*nalD*		
GAG	-	Glutamic acid - fs	98	*mexZ*		
ACCCGACGAA	-	Δnt216-225	-	*mexZ*		
GTT	-	Valine *	43	*mexZ*		

Table 4.5 Mutations contributing to fluroquinolone resistance in three samples in this study

position	AA position	Original codon	Original AA	New codon	New AA	Gene	Samples
248	83	ACC	Threonine	ATC	Isoleucine	*gyrA*	P
260	87	GAC	Aspartic acid	GGC	Glycine		
1800	600	GTC	Valine	GTG	Valine		
1881	627	GAT	Aspartic acid	GAC	Aspartic acid		
1941	647	GGT	Glycine	GGC	Glycine		
2430	810	CGC	Arginine	CGT	Arginine		
157	53	AAC	Asparagine	GAC	Aspartic acid	*mexR*	
263	88	CTC	Leucine	CCC	Proline	*nfxB*	
248	83	ACC	Threonine	ATC	Isoleucine	*gyrA*	W1
260	87	GAC	Aspartic acid	GGC	Glycine		
396	132	CAC	Histidine	CAT	Histidine		
531	177	GGC	Glycine	GGT	Glycine		
780	260	ATT	Isoleucine	ATC	Isoleucine		
891	297	TCT	Serine	TCC	Serine		
966	322	TAT	Tyrosine	TAC	Tyrosine		
1800	600	GTC	Valine	GTG	Valine		
1881	627	GAT	Aspartic acid	GAC	Aspartic acid		
1941	647	GGT	Glycine	GGC	Glycine		
2157	719	AAG	Lysine	AAA	Lysine		
2430	810	CGC	Arginine	CGT	Arginine		
157	53	AAC	Asparagine	GAC	Aspartic acid	*mexR*	
263	88	CTC	Leucine	CCC	Proline	*nfxB*	
248	83	ACC	Threonine	ATC	Isoleucine	*gyrA*	W2
260	87	GAC	Aspartic acid	GGC	Glycine		
780	260	ATT	Isoleucine	ATC	Isoleucine		
891	297	TCT	Serine	TCC	Serine		
966	322	TAT	Tyrosine	TAC	Tyrosine		
1800	600	GTC	Valine	GTG	Valine		
1881	627	GAT	Aspartic acid	GAC	Aspartic acid		
1941	647	GGT	Glycine	GGC	Glycine		
2430	810	CGC	Arginine	CGT	Arginine		
157	53	AAC	Asparagine	GAC	Aspartic acid	*mexR*	
263	88	CTC	Leucine	CCC	Proline	*nfxB*	

(ii) Carbapenem resistance

In the study, two specific ARGs, *PDC-2* and *PDC-73*, were identified as resistance genes for MCC antibiotics. Generally, carbapenems are effective antibiotics against *P. aeruginosa* because they can readily penetrate the bacterium's outer membrane due to their small hydrophilic nature. They enter the bacterial cell through porin proteins, which serve as aqueous channels. The primary porin responsible for carbapenem uptake is the outer membrane protein OprD. Resistance to carbapenems in *P. aeruginosa* often arises from the loss of OprD or downregulation of *oprD* mRNA expression in the outer membrane. This resistance mechanism can be the result of inactivating mutations or insertion sequences in the *oprD* gene. Alternatively, remote mutations may occur that upregulate the efflux system MexEF-OprN (*mexS*, *mexT*), simultaneously downregulating *oprD* expression (El Amin et al. 2005), although the mutations of *mexT* or *mexS* also confer resistance to fluoroquinolone described earlier.

The specific mutations in the *oprD* gene listed in Table 4.6 correspond to examples found in strains with intermediate resistance or resistance to carbapenems such as imipenem and meropenem. These mutations can have varying effects on the level of resistance to different carbapenems. Regarding the mutation of *oprD* gene in this study (Table 4.7), no SNP was identified in the positive sample. However, a total of 21 SNPs were detected in the first tap water sample and 9 SNPs were observed in the second tap water sample. Among these SNPs, three SNPs were associated with carbapenem resistance. The first SNP occurred in the 127-nt motif, resulting in the replacement of GAC (aspartic acid) with AAC (asparagine). The second SNP was found in the 628-nt motif, leading to the replacement of ATT (isoleucine) with GCT (alanine). The third SNP occurred in the 719-nt motif, causing the replacement of AGC (serine) with ACC (threonine). These alterations may impact the pore specificity and conformation of OprD, as they involve the replacement of a positively charged AA with a neutral-polarity AA. Similar AA changes have been observed in clinical strains with reduced susceptibility to imipenem (El Amin et al. 2005) or resistance to both carbapenems (Sanbongi et al. 2009). In terms of the other SNPs of the *oprD* gene detected, the majority of them were synonymous mutations, meaning they did not result in changes to the AA sequence of OprD. Nevertheless, there was one non-synonymous SNP identified in the 688-nt motif of the W1 sample, where a substitution from GAA (glutamic acid) to AAA (lysine) occurred. Both glutamic acid and lysine are hydrophilic residues, and it is expected that such substitutions would not significantly impact the integrity of the porin.

As for *mexT* mutations to carbapenem, it has been reported by Ocampo-Sosa et al. (2012). The study points out that all of the MexT variants identified in carbapenem-resistant isolates retained a conserved domain spanning residues 144 to 340. Table 4.6 provides a list of mutations found in *mexT* in carbapenem-resistant strains. In this study, two *mexT* SNPs were detected in all three tested samples (Table 4.7). One SNP was found in the 514-nt motif, resulting in a missense alteration from phenylalanine to isoleucine. This particular alteration has been associated with the overexpression of the MexF efflux pump. The second SNP was identified in the 936-nt motif, causing a silent alteration of the amino acid arginine. These findings are consistent with the previous report and suggest that *mexT* mutations, including those observed in this study, may contribute to carbapenem resistance in *P. aeruginosa*.

In summary, based on the SNP analysis conducted in this study, I found that SNPs in the *oprD* and *mexT* genes were associated with carbapenem resistance in two water samples, while the *P. aeruginosa* in the positive sample potentially gained the carbapenem resistance solely by the alteration of MexT. These findings align with existing knowledge and literature that have also identified these genes as being associated with carbapenem resistance in *P. aeruginosa*.

Table 4.6 Genes mutations contributing to carbapenem resistance in the recent studies

Original codon	New codon	AA alteration	AA position	Genes	Resistance	References
GAC	AAC	Aspartic acid - Asparagine	42	*oprD*	Carbapenem	(El Amin et al. 2005; Kao et al.2016)
AGC	AGA	Serine - Arginine	59			
GAC	GCC	Aspartic acid - Alanine	61			
ACC	AGC	Threonine - Serine	103			
AAG	ACG	Lysine - Threonine	115			
TTC	CTC	Phenylalanine - Leucine	170			
GAG	CAG	Glutamic acid - Glutamine	185			
GTC	ACC	Valine -Threonine	189			
ATT	GCT	Isoleucine - Alanine	210			
AGC	ACC	Serine -Threonine	240			
AAC	ACC	Asparagine -Threonine	262			
GCC	TCC	Alanine - Serine	267			
TTC	ATC	Phenylalanine - Isoleucine	172	*mexT*		(Ocampo-Sosa et al. 2012)
CGG	CGC	Arginine - Arginine	312			
GGC	AGC	Glycine - Serine	300			

Table 4.7 Mutations contributing to carbapenem resistance in three samples in this study

Position	AA position	Original codon	Original AA	New codon	New AA	Genes	Samples
514	172	TTC	Phenylalanine	ATC	Isoleucine	*mexT*	P
936	312	CGG	Arginine	CGC	Arginine		
126,127	42,43	CTC,GAC	Leucine, Aspartic acid	CTG,AAC	Leucine, Asparagine	*oprD*	W1
183	61	GAC	Aspartic acid	GAT	Aspartic acid		
192	64	GAC	Aspartic acid	GAT	Aspartic acid		
576	192	CGT	Arginine	CGC	Arginine		
579	193	GGC	Glycine	GGT	Glycine		
582	194	GAA	Glutamic acid	GAG	Glutamic acid		
588	196	TAT	Tyrosine	TAC	Tyrosine		
600	200	GCA	Alanine	GCC	Alanine		
604	202	GAG	Glutamic acid	CAG	Glutamine		
609	203	ACC	Threonine	ACA	Threonine		
624	208	GAT	Aspartic acid	GAC	Aspartic acid		
628	210	ATT	Isoleucine	GCT	Alanine		
645	215	GCA	Alanine	GCG	Alanine		
654	218	GAT	Aspartic acid	GAC	Aspartic acid		
672	224	CTG	Leucine	CTC	Leucine		
675	225	TAC	Tyrosine	TAT	Tyrosine		
681	227	GCC	Alanine	GCA	Alanine		
684	228	GAA	Glutamic acid	GAG	Glutamic acid		
687,688	229,230	CTC,GAA	Leucine, Glutamic acid	CTG,AAA	Leucine, Lysine		
708	236	TAT	Tyrosine	TAC	Tyrosine		
719	240	AGC	Serine	ACC	Threonine		
514	172	TTC	Phenylalanine	ATC	Isoleucine	*mexT*	
936	312	CGG	Arginine	CGC	Arginine		
126,127	42,43	CTC,GAC	Leucine, Aspartic acid	CTG,AAC	Leucine, Asparagine	*oprD*	W2
183	61	GAC	Aspartic acid	GAT	Aspartic acid		
192	64	GAC	Aspartic acid	GAT	Aspartic acid		
201	67	CAA	Glutamine	CAG	Glutamine		
219	73	TAT	Tyrosine	TAC	Tyrosine		
234	78	ACC	Threonine	ACT	Threonine		
243	81	ACT	Threonine	ACC	Threonine		
628	210	ATT	Isoleucine	GCT	Alanine		
719	240	AGC	Serine	ACC	Threonine		
514	172	TTC	Phenylalanine	ATC	Isoleucine	*mexT*	
936	312	CGG	Arginine	CGC	Arginine		

(iii) Aminoglycoside resistance

P. aeruginosa commonly develops resistance to aminoglycosides through the acquisition of AME genes. These AME genes can be categorized into three primary families: phosphorylators (APH), acetylators (AAC), and adenylators (ANT) (Atassi et al. 2023). HGT is a common mechanism by which *P. aeruginosa* strains acquire AME genes. However, in addition to AME-genes' acquisition, resistance to aminoglycosides can also occur through mutations in the bacterial genome. One well-studied example involves mutations in the *mexZ* gene, which can lead to nonsynonymous mutations, frameshift mutations, or premature stop codons (Atassi et al. 2023), such as the 195-AA alteration from glycine to glutamic acid reported by Cabot et al. (2014). These mutations result in the overexpression of the MexXY-OprM efflux pump (Cabot et al. 2014; Abbara et al. 2019; Atassi et al. 2023).

In this study, no mutations were detected in the *mexZ* gene. However, the acquisition of APH genes, namely *APH(3'')-Ib*, *APH(3')-IIb*, and *APH(6)-Id*, was observed. On the one hand, these findings potentially suggest that in my studied *P. aeruginosa* strains, aminoglycoside resistance was primarily mediated by the acquisition of AME genes rather than mutations in the *mexZ* gene. However, the role of *APH(3')-IIb* in delivering aminoglycoside resistance is controversial, as *APH(3')-IIb* was found to be prevalent in the clinical strains that were sensitive to aminoglycoside drugs (Atassi et al. 2023). On the other hand, there is an explanation that in addition to *P. aeruginosa*, more species was likely the potential hosts carrying those *APH* genes, such as *E. coli* and *K. pneumoniae* mentioned above.

4.3.6 Conclusions

In this Primalseq panel of *P. aeruginosa*, the approach allows for the identification of specific STs by targeting the seven MLST genes, which provides valuable information regarding the genetic relatedness and epidemiology of *P. aeruginosa* strains. Additionally, beyond MLST, by targeting 93 ARGs with 16 categories, the use of MinION amplicon sequencing technology allows for sensitive recovery of target antibiotic resistance, and accurate detection of genetic variations, e.g. SNPs, which can promote the assessment of potential antibiotic resistance risks and enabling a better understanding of the resistance profile and the selection of appropriate treatment options. Remarkably, the ability to perform this panel with as few as three *P. aeruginosa* cells is particularly advantageous in scenarios where the availability of bacterial isolates or the concentration of bacterial samples may be limited.

4.3.7 Limitations and improvements

During the validation process, only three samples were tested. Therefore, for practical implementation, a larger scale of environmental samples should be included to assess the efficacy of this panel thoroughly. In addition, in terms of the identification of STs in *P. aeruginosa,* the panel relies on distinguishing various allele numbers of MLST genes by utilising the sequences of seven MLST genes from the PAO1 strain as a reference. Genetic mutations identified are then used to rectify nucleotides at corresponding base positions within the MLST genes, resulting in corrected MLST sequences. However, the presence

of multiple *P. aeruginosa* strains in a sample can complicate the process. Multiple mutations within the same base position of a specific MLST gene can obscure the determination of the correct combinations of allele numbers, ultimately affecting the accurate identification of STs. Consequently, to circumvent such limitation, the amplification-based sequencing which relies on alignment and variant-calling results to determine STs can be replaced with WGS or the metagenomic sequencing through which STs can be ascertained via *de novo* assembly.

Moreover, the identification of ARGs (particularly the intrinsic ARGs) only functions as a genotypic indicator of resistance, rather than directly representing the phenotypic expression of resistance. In the validation work, the focus was solely on genetic analysis without conducting phenotypic MIC testing, which could introduce uncertainty when determining antibiotic resistance. For example, in terms of carbapenem resistance, only the MCC-resistant *PDC-2* and *PDC-73* genes were identified. However, the targeted ARGs that were simply resistant to carbapenem (e.g *HMB-1*) were not detected. It is important to note that although the presence of MDR-genes generally provides a greater advantage for bacterial evolution compared to a single resistance gene, the carbapenem resistance potentially associated with SNPs of *oprD* and *mexT* genes still need to be validated through phenotypic resistance testing. Additionally, it is conspicuous that certain identified ARGs, such as horizontally-originated ARGs of *APH(3'')-Ib* and *APH(6)-Id*, were not exclusively found in *P. aeruginosa* strains. Therefore, this panel of ARGs may be more suitable for scenarios where a specific *P. aeruginosa* strain is the predominant contaminant, especially when used in conjunction with MIC tests. Otherwise, in scenarios where *P. aeruginosa* is not the dominant organism or there are multiple *P. aeruginosa* strains, whole-genome analysis is essential to offer resolution regarding the hosts of these ARGs.

Chapter Five

Amplicon Sequencing of Core Genes in *Legionella Pneumophila* Reveals Persistent Colonisation of the Hospital Plumbing and Evidence of Onward Transmission to Patients

In this chapter, apart from selecting the 51 core genes, all other aspects of the research were completed by the author (myself).

5.1 Background

L. pneumophila is the most or the second most frequent community-acquired pneumonia (CAP) that requires admission in ICU. Since LD is mainly transmitted via inhalation of infectious aerosols (Cunha, Burillo, and Bouza 2016), it is of great epidemiological significance to perform molecular epidemiology analysis of *L. pneumophila* isolates and to trace the source of infection.

As for the epidemiological typing of *L. pneumophila*, the current "gold standard" method is sequence-based typing (SBT) (Gaia et al. 2003; Ratzow et al. 2007) developed by the European Society for Clinical Microbiology Study Group on Legionella Infections (ESGLI). The scheme uses a combination of seven housekeeping and virulence genes (*flaA, pilE, mip, mompS,* and *proA, asd, neuA*) to determine STs among *L. pneumophila* strains. Over 2,000 STs have now been reported. However, there is a trade-off between discriminatory power and epidemiological concordance. This method based on only seven genes cannot differentiate strains within ST1 which has caused 85% of LD cases, leaving large number of source-tracking investigations remain unresolved (Borchardt, Helbig, and Lück 2008; Harrison et al. 2009; Tijet et al. 2010). David et al. (2016) has revealed that *L. pneumophila* strains could be most usefully typed using a cgMLST scheme with approximately 50 core genes (cg).

In this chapter, I present the first PrimalPCR integrating MinION sequencing protocol for *L. pneumophila* typing based on the 51-core-gene scheme (Figure S5.1), of which the determination of 51 core genes was pre-selected by others. With the optimised touch-down PCR condition, samples with extremely low cell number of *L. pneumophila* could be sequenced for accurate ST determination and for revealing the *L. pneumophila* transmission pattern with a hospital environment in different seasons in eight departments with six medium types, including tap/trap water, surface swab, air condition dust and wastewater. This study provides important evidence into the application of MinION sequencing coupling with the multiplex-PCR on the bacteria tracing. Meanwhile, the study offers a deeper insight into *L. pneumophila*

transmission behavior, paving the way for core-gene-based epidemiology of the outbreak assessment directly from environmental samples with the extremely low biomass in the field.

5.2 Methods

5.2.1 Activation, growth and culture of *L. pneumophila*

The methods for activating, growing and culturing of *L. pneumophila* have been described in *Chapter 2*.

5.2.2 Hospital sample information and treatment

217 hospital environment samples were collected from July 2020 to January 2021 monthly in *Peking University Shenzhen Hospital*, Shenzhen, China. Samples of six different medium categories (tap, p-trap, surface, air dust, sewage inlet, sewage outlet) were collected from eight sites (blood test room, emergency, gynecology, ophthalmology, otolaryngology, pneumology and surgery departments as well as wastewater treatment plant) of the studied hospital (Table S5.4). The collection methods and pretreatment approaches of each type of sample have been described in *Chapter 2*. Following the collection and pretreatment, samples were stored at -20°C temporarily and were subject to DNA extraction within one day.

5.2.3 DNA Extraction of *L. pneumophila* pure strains and environmental samples

The pure strains of *L. pneumophila* were collected from cultured BYEB broth. 100 µl broth containing live cells was used for the DNA extraction by QIAamp DNA Blood Mini Kit (QIAGEN, Germany) following the default protocol. DNA of environmental samples was extracted with the Dneasy PowerWater Kit (QIAGEN, Germany). The final DNA was eluted into 100 µl EB for the downstream use. More details can be referred to *Chapter 2*.

5.2.4 Primer design

The references of 51 core genes of *L. pneumophila* were designated for generating primers using the Primal Scheme (https://github.com/aresti/primalscheme) by setting the amplicon length of 300 bp, min-overlap of 50 bp and melting temperature at 65°C. As a result, 448 primers were generated, of which 252 primers were assigned in pool 1, and 196 primers were assigned in pool 2.

5.2.5 Process of PrimalPCR

The two-step touchdown PCR program and the components' information of PCR is shown in Table S5.5 and Table S5.6 respectively. It is noteworthy that as the PrimalSeq panel is extremely sensitive, the false positive would appear if there is trace contamination in the PCR environment. Therefore, conducting the lab in the PCR hood is strongly recommended, with all PCR water and buffer sterilised under the UV light for an hour prior to the use.

5.2.6 Optimisation of PrimalPCR

In order to optimise the multiplex-PCR for this *L. pneumophila* scheme, four different conditions were tested, as described in Table S5.7. The variations among these conditions included the choice of DNA polymerase and the implementation of a touchdown temperature. The DNA samples used in this stage were extracted from the *L. pneumophila* ATCC 33152 strain as described in *section 2.2.* To facilitate the optimisation process, the DNA samples were subjected to 10-fold serial dilution (starting from approximately 10^8 copies) using nuclease-free water. There were ten different genome copies ranging from 1400 copies to 0.3 copies for the optimisation of PrimalPCR, as outlined in Table S5.8.

5.2.7 RT-PCR of *L. pneumophila*

The RT-PCR procedure was instructed by the protocol of *Legionella pneumophila* v1.1 genesig standard kit (UK).

5.2.8 MinION Sequencing

The library preparation with LSK109 and EXP-NBD 196 kits for the MinION sequencing was described specifically in *Chapter 4.2.7*. The prepared DNA library with 10ng-20ng was loaded on the new flow cell R9.4.1 or the washed flow cell R9.4.1.

5.2.9 Bioinformatic analysis

(1) Consensus reads generation of 51-core-gene MinION sequencing

The overall process to generate consensus reads was described comprehensively in the *section 4.2.8*, of which a pseudomolecule comprising 51 target genes was used as the reference sequence for mapping the basecalled reads during the alignment process. With generated consensus reads of 51 core genes of each sample, they were sorted in the same order and catenated to generate a pseudo-genome. The pseudo-genomes of all samples were aligned using mafft (Katoh et al. 2002). The gaps of aligned pseudo-genomes of all samples were trimmed by Gbblock (Castresana 2000). The percentage of "N" in each aligned pseudo-genome was calculated as a masked ratio. At last, the phylogenetic tree was built with the Fasttree (Price, Dehal, and Arkin 2010) and visualised by Figtree (Rambaut, 2012). Normally, the entire process, starting from samples collection to the final tree visualisation, was completed within a timeframe of three days. This timeline applied when working with a maximum of 24 samples by one person.

To insert the whole-genome references of *L. pneumophila* STs into the pseudo-genome tree of studied environmental samples, whole-genome sequences were randomly sliced into 300,000 pieces of 350 bp reads to simulate the amplicons delivered by nanopore sequencing.

As for Illumina pair-end reads downloaded from ENA, snippy (Seemann, 2018) was used to call the SNPs and to generate the consensus read for the tree construction.

(2) Validation of 51 core genes

(i) 51-core-gene tree building based on *L. pneumophila* reference genomes

L. pneumophila references of 76 finished whole genomes were downloaded from the NCBI genome database. The circular interpretation of *L. pneumophila* genomes was displayed by the BRIG (Alikhan et al. 2011). Additional shotgun sequencing raw reads and assembly reads of eleven environmental *L. pneumophila* samples were obtained from European Nucleotide Archive (ENA) (Table S5.3). The 51 core genes were extracted from the whole genome reference with BLASTN (Camacho et al. 2009). Sequence alignment, trimming, and tree building were conducted following the identical procedure as described above.

(ii) Whole-genome tree building based on *L. pneumophila* reference genomes

Roary tool (Page et al. 2015) was used in this process to obtain the genome alignment within all 91 *L. pneumophila* whole-genome references. Prokka (Torsten Seemann 2014) was used to produce the annotation file for Roary analysis. The resulting alignment file of 1,226 genes was used to build the whole-genome tree by Fasttree (Price, Dehal, and Arkin 2010).

(iii) Tree comparison and calculation of pairwise SNP distance

51-core-gene tree and whole genome (wg)-tree were compared using the generalised Robinson-Foulds (RF) distance (Smith 2020) with the assistance of the "TreeDist" package available at https://ms609.github.io/TreeDist/. Within the "TreeDist" R package, the functions utilised for tree distance calculation, tree similarity assessment, and tree visualisation were 'TreeDistance', 'SharedPhylogeneticInfo', and 'VisualiseMatching', respectively. Normally, when the RF distance is 1, the similarity between the two trees is 0. The overall pipeline for the tree comparison was demonstrated in the R markdown script which is shown in Appendix 3.

In order to analyse the genetic relatedness within the clade on the phylogenetic tree, the pairwise SNP distances were calculated. This was achieved by using the "dist.alignment" function from the "seqnir" R package (Fitch 1966) with the aligned fasta file of the clade (Appendix 3).

(iv) STs identification

NCBI is a straightforward but limited way for acquiring the STs information. Besides, Legsta tool (Gaia et al. 2005; Ratzow et al. 2007) can be used for the identification of STs, while the phylogenetic relationship between STs within clonal complexes was implemented by using the goeBURST (Francisco et al. 2009) in the PHYLOViZ program (Ribeiro-Gonçalves et al. 2016).

5.3 Results and Discussion

5.3.1 Locus of 51 core genes

Comparative genomics of *L. pneumophila* strains have revealed that around 300 genes (~10%) are specific for each strain showing a substantial high level of genome diversity of *L. pneumophila* species (Gomez-Valero and Buchrieser 2013). To ensure the representativeness, 51 core genes which remain essentially conserved throughout *L. pneumophila* evolution were selected as the amplification targets for the scheme (thereafter referred as 51 cgMLST scheme). In addition to 6 MLST genes, the set of 51 core genes also comprised 21 housekeeping genes and 24 conserved regulator genes, which were further categorized into 16 clusters of orthologous genes (COG) types associated with metabolism, biogenesis, and cell structure (Tatusov et al. 2000). Notably, the 6 MLST genes are also part of the previous ESGLI scheme. In Figure 5.1a, the locus tags corresponding to these 51 core genes are displayed. To facilitate the amplification of these genes, a total of 448 primers were designed for multiplex PCR. These primers targeted adjacent regions of the 51 genes and were assigned to two separate primer pools, as depicted in Figure 5.1b. By amplifying neighboring amplicons in separate pools, the generation of chimera products was minimised.

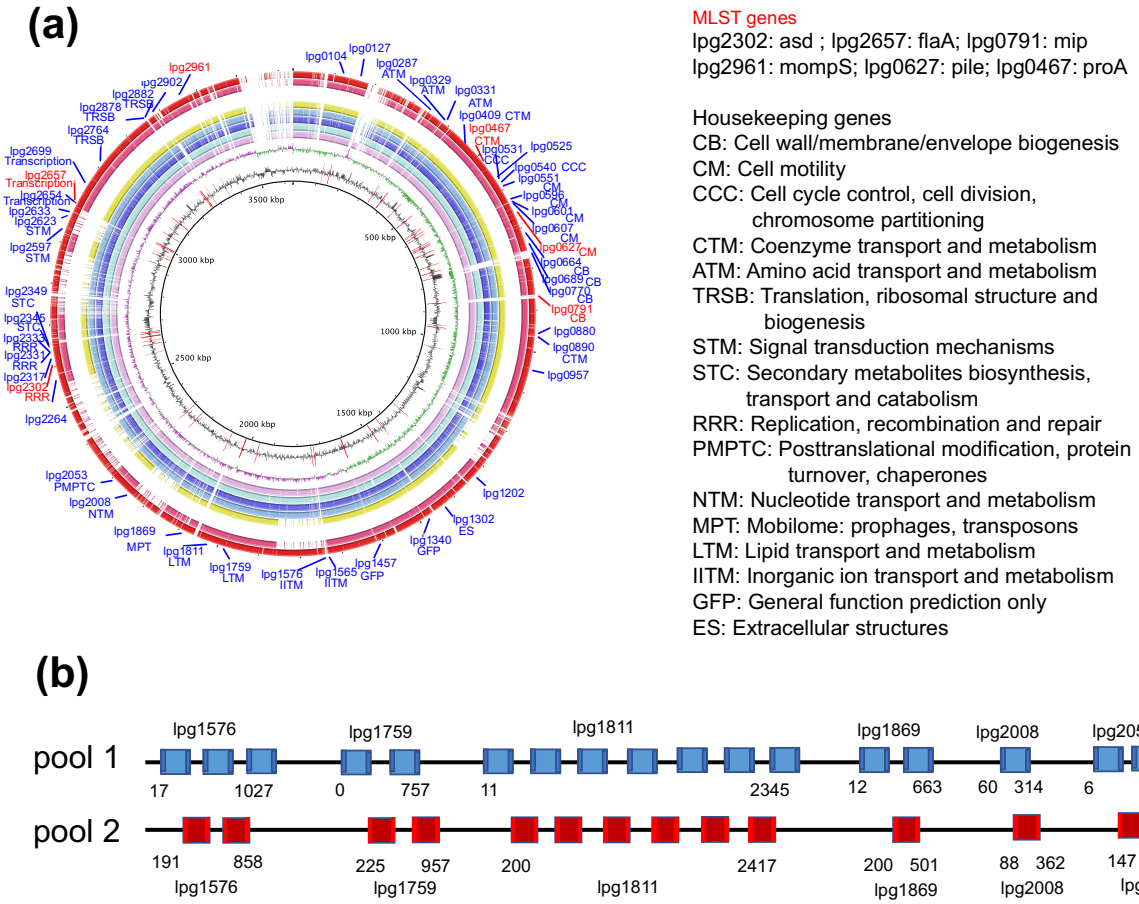


Figure 5.1 (a) Locus tag of 51 cgMLST scheme developed in this study. The red font represents MLST genes designed by ESGLI, and blue font represents housekeeping genes assigned to 16 COG types. (b) Schematic diagram showing the genomic location of designed primers PrimalPCR in pool1 and pool2.

5.3.2 Reliability of 51 cgMLST scheme by MinION sequencing

To assess the reliability of the MLST scheme based on the 51 core genes, a total of 91 *L. pneumophila* genome references were collected from NCBI and ENA databases. Then STs of these 91 references determined by 51 cgMLST scheme were compared to those STs determined by WG-based scheme which contained 1,226 core genes. In this result, a total of 75 reference genomes were assigned to 22 STs, while the STs of the remaining 16 genomes were still uncertain. We observed a highly congruent clustering

pattern among the STs in two trees, where 91 genomes showed consistency between the trees (refer to Figure 5.2).

To statistically compare the two trees, I employed the generalised RF algorithm to assess their similarity. The 51-cg tree consisted of 67 bipartition splits, dividing the 91 genome tips. On the other hand, the WG tree had 88 bipartition splits among the same 91 genomes. By applying the RF algorithm, I identified 67 matching splits between the two trees. The overall distance and similarity values were calculated as 0.277 and 0.709, respectively. Among these matching splits, I annotated eight pairs in Figure 5.2, labelled as split A, B, C, D, E, F, G, and H. Split D emerged as the largest branch in both trees, encompassing the highest proportion of genomes (similarity score of 86.1%). This split formed a clonal complex (CC) centred around ST36, comprising 31 genomes. Additionally, split D included three genomes of ST187, two genomes of ST40, one genome of ST37, and one genome of ST211. Following split D, split C represented the second-largest branch, consisting of 22 genomes. Out of these, 15 genomes belonged to ST1, three genomes to ST62, and the remaining genomes were assigned to ST222, ST23, and ST2439. The third-largest branch comprised split A and split E. Split E predominantly contained ST84 (two genomes) and ST15 (two genomes). However, the specific STs within split A are currently unknown. Moreover, split B, split F, split G, and split H constituted the remaining structure of the tree, incorporating ST731 (five genomes), ST40 (two genomes), and several other STs (seven genomes). Apparently, these splits showed strong bootstrap support, indicating a robust differentiation of various *L. pneumophila* lineages using the 51-cgMLST scheme. Overall, the findings demonstrate the competitive performance of the developed 51-cgMLST scheme compared to the WG-based phylogenetic inference. This underscores the reliability of the scheme for epidemiological typing in *L. pneumophila*.

Besides the validity, the rapid ST determination is inevitably important for source tracking in outbreak incidents. Therefore, MinION was used to sequence amplicons derived from PrimalPCR. Despite the real-time fast sequencing of MinION, the feasibility to obtain reliable *L. pneumophila* typing based on error-prone nanopore reads was tested with seven *L. pneumophila* strains belonging to ST181, ST9, ST77, ST1, ST128 and ST6. By comparing the phylogenies between cgMLST-scheme recovered by nanopore sequencing and WG-scheme obtained by Illumina sequencing, it is found that the nanopore-generated

consensus sequences all formed monophyletic clusters with their Illumina counterparts (Figure 5.3b), highlighting reliability of *L. pneumophila* typing determined based on nanopore sequencing.

As a conclusion, the cgMLST scheme surpasses the MLST scheme in *L. pneumophila* typing due to its superior strain/ST identification discriminative power. It excels in scenarios where direct WGS from environmental samples is challenging due to issues like bacterial culturing difficulties or failed WGA processes. Additionally, the streamlined nature of cgMLST, compared to extensive WGS, offers time and computational efficiency advantages during outbreak assessments for studying evolutionary events and tracing outbreak sources. However, it is evident that for analysing recombination events in *L. pneumophila* strains, MLST remains a preferred and more straightforward approach compared to phylogenetic tree analysis.

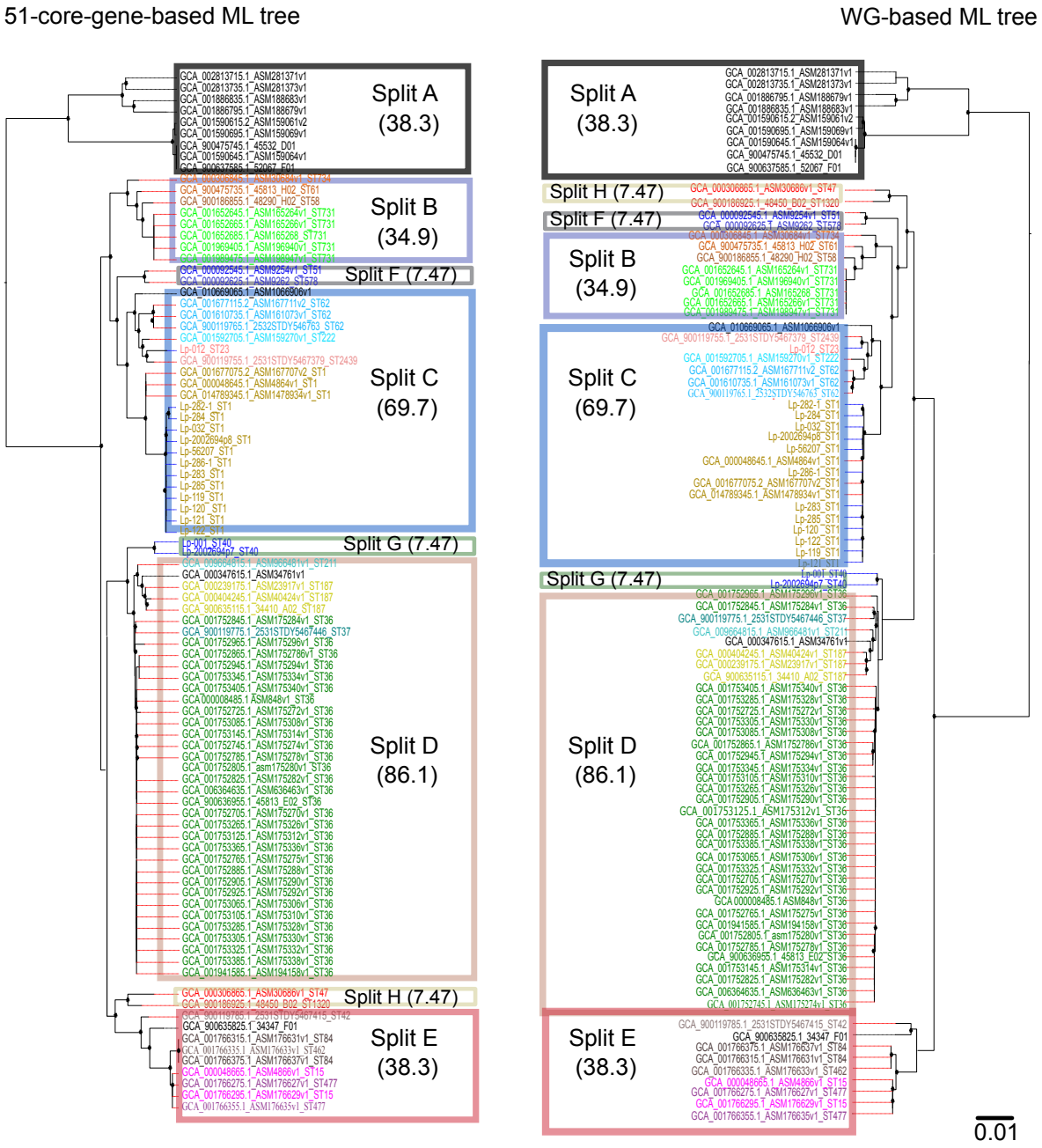


Figure 5.2 Evaluation of the reliability of 51 cgMLST scheme. Left phylogeny was constructed with the 51 core genes extracted from 91 *L. pneumophila* references. Right phylogeny was constructed with the whole genomes of 91 *L. pneumophila* references by Roary tool. Branches with boost strap value over 0.9 were shown in black dots.

5.3.3 Sensitivity of the 51 cgMLST scheme by PrimalPCR

The high-sensitivity detection is always challenging but crucial in the culture-free outbreak assessment (Wikramaratna et al. 2020), especially when the studied environment contains low target microbial abundance. Indeed, multiple studies have observed false negative results on at least one occasion for infected individuals, such as the SARS-CoV-2 investigation using the real-time PCR technology (Xie et al. 2020; Fang et al. 2020; Wölfel et al. 2020). Such false-negative results have confined the analysis of correct diagnosis (Hao, Wu, and Wang 2020) and subsequent community transmission (Cao et al. 2020). Currently, the major obstacle for improving the sensitivity is the sacrifice of the specificity concurrently. To overcome this, I improved and optimised 51 cgMLST-PrimalPCR scheme. An important aspect to achieve the high sensitivity and specificity of the scheme is the multiple primers designed for short amplicons (300bp), which has been proved in the previous studies. Quick et al. (2017), who have developed the PrimalSeq scheme, enriched viral genomes from samples containing as few as 50 genome copies per reaction. Stubbs et al. (2020) also successfully applied this scheme to DENV-infected clinical samples with at least 37.9 Ct value. Another equally critical respect for the high sensitivity relies on the touchdown PCR program that I have improved in this study, making it more robust suited to low quality and load samples (Korbie and Mattick 2008).

The LOD determination results demonstrate that the PrimalPCR could obtain target amplicons at lowest *L. pneumophila* strain concentration of 3 cells per reaction in both pools. The optimised amplicon band could be observed with 14 *L. pneumophila* copies per reaction (Figure 5.3a). With the amplicons of 3 copies of *L. pneumophila* cells, the MinION sequencing was performed on one MinION flow cell with 754 pores for 24 hours. The sequencing run generated a total of 582.2K Fastq reads, with average length of 356 bp. Impressively, 99.61% of these reads were successfully aligned to the *L. pneumophila* reference genes. In terms of a base coverage, more than 90% bases of the 51 core genes were recovered from the nanopore reads. On average, these genes were sequenced at a depth of 4841×. As illustrated in Figure 5.3c, the distribution of the base coverage among the 51 core genes was found to be even. This observation suggests the highly successful concurrent amplification of different primers within a single pool. However, the apparent fluctuation in the SD across 51 genes (Figure 5.3c) indicates an inevitable

imbalance of primers' amplification efficiency within the scheme. Such primer bias is prone to occur in the short-amplicon multiplex PCR panel, and along with the increasing number of primers in one pool, the PCR efficiency caused by the primer bias is more likely to drop out (Liu et al. 2022). Worth to note is that high SD was not necessarily associated with high degree of base coverage. For example, the maximum depth reached 13,938× in the *lpg2623* locus, but this locus acquired the lowest base coverage of 58.3%. Moreover, *lpg2333* had minimum depth of 119×, despite 94.7% bases of this locus were sequenced. As the primers developed by PrimalScheme exhibited a theoretical coverage of over 90% for all the core genes, the uncovered regions at high SD reflect the occurrence of primer deficiency during multiplex-PCR. In order to ensure the reliability for STs determination, regions with below 20× depth on the core genes were masked with "N" before generating consensus sequences. Figure 5.3d prominently displays the base positions with over 20× depth for each gene. Genes with low base coverage consistently exhibit missing regions that were not amplified. For example, in *lpg2331*, regions encompassing nearly the initial 10% and from approximately 52% to 55% were not covered, despite the primer design being capable of covering these areas. Presumably, this absence could be attributed to competition during multiplex-PCR or possibly to unexpected mutations occurring within these regions. As calculated, the final pseudo-genome generated by combining the consensus sequences of 51 core genes covered 82% of the core gene length. This led to an effective alignment length of 34,482 bp for *L. pneumophila* typing, which was approximately 14 times longer than that of the ESGLI scheme using seven MLST genes. Therefore, given the apparent importance of sufficient alignment length for discriminating closely related strains during typing, the large pseudo-genome size obtained by 51 cgMLST scheme will enable a more precise epidemiological source tracking through differentiation of previously unresolved *L. pneumophila* STs.

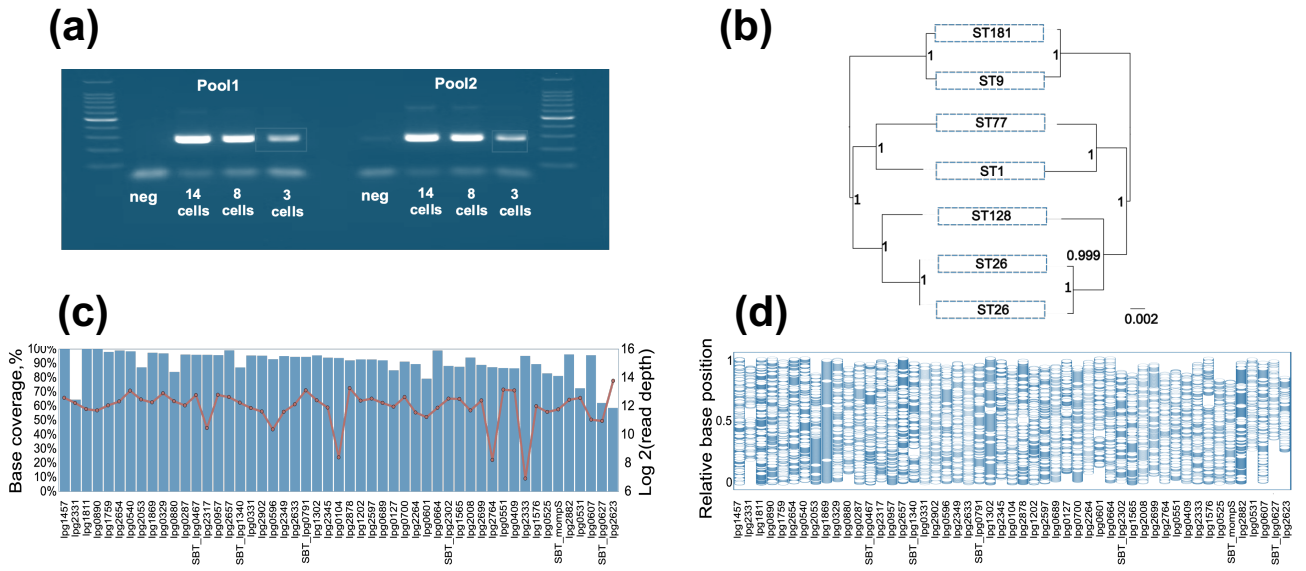


Figure 5.3 (a) Optimisation of PrimalPCR with DNA samples extracted from *L. pneumophila* strain ATCC 33152. DNA marker showed here is 100 bp ladder. DNA templates applied in pool1 and pool2 were (from left to right): nuclease-free water, DNA template containing 14 cells, 8 cells, and 3 cells respectively. (b) Evaluation of the accuracy of PrimalPCR-based MinION sequencing with seven *L. pneumophila* isolates. The left part was built with the Illumina-generated whole genome sequences, and the right counterpart was constructed with the amplicon consensus reads of 51 core genes generated by MinION sequencing. (c) Performance of MinION sequencing with three *L. pneumophila* cells. The blue bar represents the base coverage ratio across 51 core genes, and the orange line represents the Log2 value of SD obtained across 51 core genes. (d) Performance of MinION sequencing with three *L. pneumophila* cells. The y-axis represents the relative base-position (1-based coordinates) where the depth over 20 within each gene (amplicon).

5.3.4 Performance of the 51 cgMLST scheme in determining *L. pneumophila* typing of environmental samples collected from a large hospital

217 environmental samples of six types were initially collected from eight departments of a large hospital located in Shenzhen, China. Among them 134 samples with distinguishable specific amplification band

of 300bp in either pool was subject to subsequent MinION sequencing of the PrimalPCR amplicons (PrimalPCR results are summarized in Table S5.9). All positive controls showed strong specific 300bp bands in both pools and no band appeared in all negative samples, suggesting the absence of contaminating amplicons during PrimalPCR. Nonetheless, I observed a loss of amplification specificity in environmental samples with complexed microbial communities. For example, p-trap and sewage water samples always showed dispersed bands with or without specific 300bp bands. For these samples obtaining dispersed bands with smear 300 bp bands, decreasing the template DNA concentration had no discernible effect in improving amplification specificity. On the other hand, samples with simple uniform community, such as tap water or a surface swab, were unlikely to contain non-specific bands. Nevertheless, the successful amplification of these samples was constrained by the sparse presence of *L. pneumophila* cells in these clean environments. Following the PrimalPCR, DNA of around 20 samples were barcoded and pooled to generate sequencing library for one MinION run. Seven MinION runs were performed to sequence the 145 amplicon samples (Table S5.10). Despite the normalisation during the library preparation, sharp variation in data size (Figure S5.2) and base coverage (Figure S5.3) were observed among samples, and only 64 samples had sufficient pseudo-genome coverage (> 20% of core gene length) for subsequent *L. pneumophila* typing.

MinION sequencing of the PrimalPCR amplicons from 64 environmental samples reveals that an average of 54.4% of sequenced reads could be aligned to the references of 51 *L. pneumophila* core genes (Figure S5.5), indicating non-specific amplification during PrimalPCR of the scheme. Such mapping ratio varied significantly from 100% to as low as 2.84% across different sample types, and this variation was not correlated with the number of reads sequenced. To further explore the factors that caused the alignment fluctuation, the mapping ratio of different sample categories is summarized in Figure 5.4a. The positive *L. pneumophila* strains had the best alignment, with an average mapping ratio of 98%, whereas tap water, surface swab, air dust, sewage inlet, p-trap water, and sewage outlet showed descending mapping ratio of 73.2%, 62.3%, 61.7%, 37.2%, 24.1% and 4.23% respectively. The loss of alignment ratio was consistent with the level of non-specific amplification observed in electrophoresis analysis. This suggests that samples with a relatively complex community composition were more likely to generate a higher proportion of off-target reads. In contrast, samples with lower biomass or sparser populations exhibited

a more favorable mapping ratio of nanopore reads. This demonstrates the priority for using the PrimalSeq-based 51cgMLST scheme in tracking *L. pneumophila* dissemination in clean environments.

Among the 64 sequenced environment samples, the average SD of 51 core genes was 2927× with the average base coverage of 70%. Despite the fact that all samples had acquired sufficient SD, not all 51 core genes in the scheme were able to be detected in an environment sample (Figure 5.4b). In fact, full coverage of the 51 genes was only observed in 24 samples including 7 positive controls, 7 p-trap water samples, 4 tap water samples, 3 sewage inlet samples, 2 air dust samples and 1 surface swab sample. The gene number detected in the rest of 40 samples fluctuated from 50 to 31, which was associated with the level of non-specificity of the PrimalPCR. Investigating genes with consistently low coverage (below 84%) in these 64 amplicon samples as highlighted in Figure S5.3, *lpg2331* with base coverage ranging from 22.2% to 65.3% and *SBT_lpg0627* with base coverage varying from 42% to 65.1% were identified. Furthermore, delving deeper into the relative base positions where the depth exceeded 20× within these two genes (Figure S5.4), it was observed that for *lpg2331*, regions spanning approximately the first 17% and the middle 50% to 60% within the genome were consistently missing across all amplicons. Concerning *SBT_lpg0627*, the initial 30% region within the genome was consistently absent across all amplicons. As stated above, although these regions could have been amplified during the primer design phase, their absence might be attributed to competition during multiplex-PCR or potentially to unexpected mutations occurring within these regions.

As for the further analysis of sequencing efficiency, pseudo-genome was constructed by concatenating consensus reads of the 51 core genes. In order to ensure reliable *L. pneumophila* typing, bases with lower than 20× coverage was masked with “N” before getting the consensus reads. For the environment samples, the pseudo-genome covered 51% of the 51 core genes length on average, resulting in an effective average alignment length of 20,967 bp for subsequent *L. pneumophila* typing. The top performer was the tap water (tap 4 in Figure S5.3), boasting a mapping ratio of 96.0% and pseudo-genome coverage of 74.0%, while the poorest performer was the air-dust sample (air11.2 in Figure S5.3), exhibiting a mapping ratio of 44.8% and pseudo-genome coverage of 20.0%. On average, tap water, air-dust, and surface samples exhibited relatively higher pseudo-genome coverage levels of 46.9%, 43.3%, and 42.3%, respectively. In contrast, the pseudo-genome coverage was lower for p-trap and wastewater

outlet samples, at 41.1% and 26.3%, respectively. Evidently, the distribution of pseudo-genome coverage did not align with that of the mapping ratio (Figure 5.4a). Indeed, it is conspicuous that the pseudo-genome generated from the inlet sample exhibited a coverage of 60% of the core genes. This coverage was higher compared to the clean environments of tap water, surface swab, and air dust, despite these environments showing a higher reads-to-cg mapping ratio. The higher pseudo-genome coverage of inlet samples at comparably lower SD and PrimalPCR specificity indicates a more unified primer efficiency. This could be attributed to the relatively more intact DNA extracted from the inlet water, which contained a reasonably high biomass content (Figure S5.7). Interestingly, similar trend was also observed in the PrimalSeq scheme of SARS-CoV-2 of different environmental mediums. Despite of the evident high proportion of off-target reads (reads mapping ratio was as low as 11%) caused by non-specific amplification, the inlet of a municipal wastewater treatment plant also showed significantly higher breadth of coverage across all primer schemes (Lin et al. 2021).

To explore the reasons behind the low pseudo-genome coverage, two scenarios can be identified. The first scenario was the low alignment percentage due to the low *L. pneumophila* biomass (e.g. sur14 in Figure S5.3, with alignment percentage of 5.69%, and pseudo-genome coverage of 20.0%). The second scenario was the high alignment percentage with uneven amplification (e.g. sur12.2 in Figure S5.3, with alignment percentage of 70.4%, but pseudo-genome coverage of 21.0%). This uneven amplification could manifest as a higher number of alignments for certain targets compared to others, a phenomenon often linked to over-amplification during multiplex PCR. This skewed amplification was able to result in a disproportionate representation of certain genomic regions in the sequencing data. Moreover, in cases where regions exhibited a low number of alignments, this could be attributed to the mutations that had occurred in the *L. pneumophila* genome. These mutations might prevent the amplification of specific regions using the current primer scheme, leading to reduced alignment and coverage in those areas.

Additionally, dissimilar to the scheme used to target virus in the complex environments, the application of PrimalSeq of viruses was typically able to cover nearly the entire sequence of the targeted coding regions from clinical samples (e.g. serum and urine) with high-specific amplification (Quick et al. 2017; Hill et al. 2019; Tyson et al. 2020; Stubbs et al. 2020; Park et al. 2021). However, given the paucity of reports on application of PrimalSeq-based schemes involving diverse environmental samples, a more

systematic comparison is required to dissect and elucidate the molecular machinery underpinning the high core-gene coverage of a specific sample that conversely showed evident level of non-specific implication.

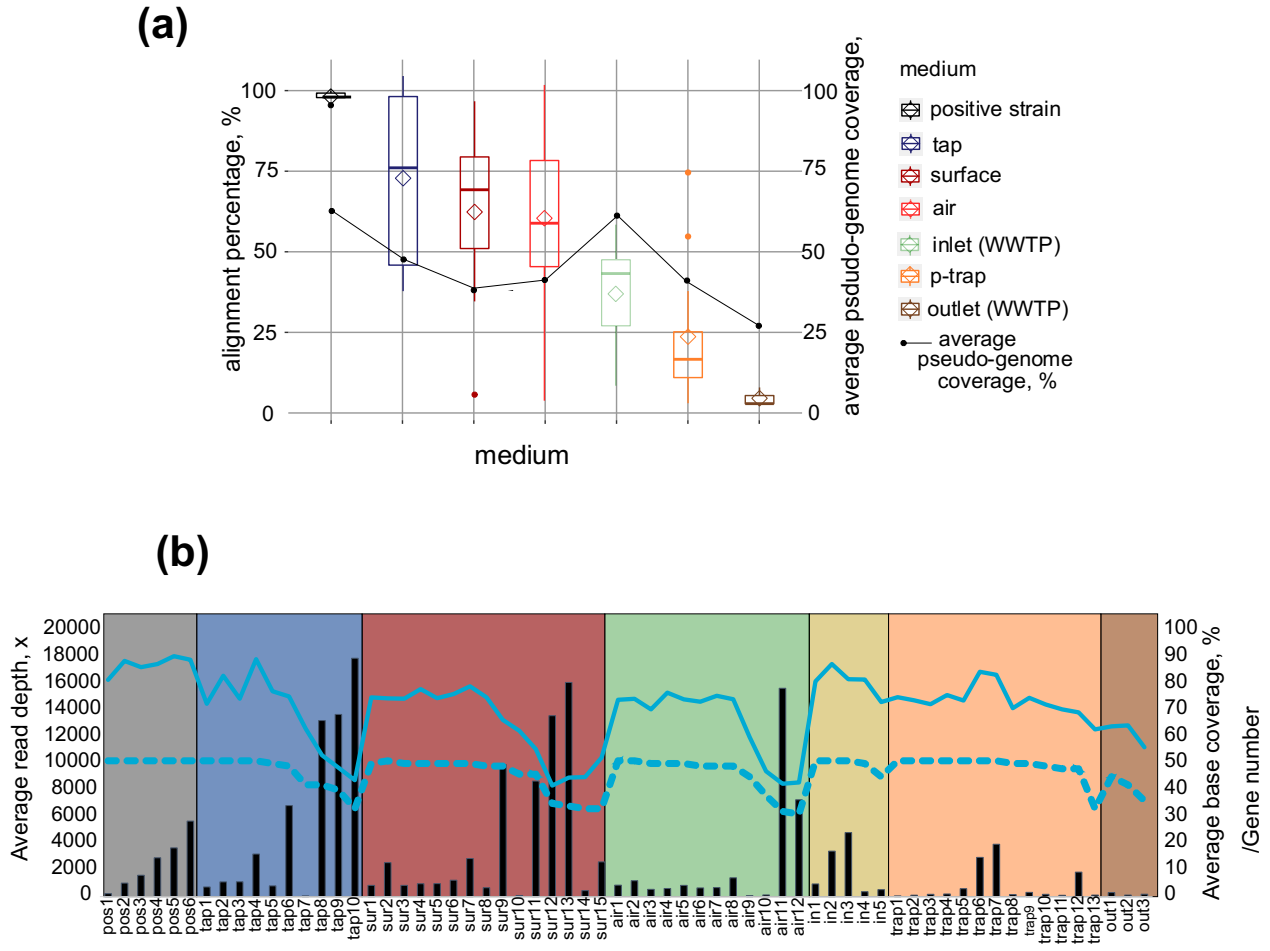


Figure 5.4 (a) Distribution of mapping ratio and average pseudo-genome coverage across six different medium types in the hospital. The medium types are distinguished by different colors. The box plot displays the dataset of mapping ratio, and the line shows the average pseudo-genome coverage. (b) Distribution of average read depth, average base coverage, and number of core genes recovered by amplicon MinION sequencing across 64 hospital environmental samples. The black bars explain the average read depth, the solid blue line represents the average base coverage, and the dash blue line stands

for the number of core genes recovered by amplicon MinION sequencing. Different medium types of 64 samples are filled with different background colors.

5.3.5 51-cgMLST scheme permits the tracing and inferring of potential STs of *L. pneumophila* within the hospital setting

Phylogeny inferred based on the 51-cgMLST scheme reveals the presence of six major *L. pneumophila* clades within the hospital environmental samples collected from six medium types across nine departments (Figure 5.5). Evident variation on the average reads-to-cg mapping ratio and pseudo-genome coverage was observed among these clades (Figure S5.8). Clade B, C, E and F, showed relatively higher level of mapping ratio of over 50%, whereas less than 40% nanopore Fastq reads of clade A and D could be aligned to *L. pneumophila* reference (Figure S5.8). In addition, as unveiled by the violin plots in Figure 5.5, there was a close genetic relatedness among 8 samples within Clade D and 24 samples within Clade F. In Clade D, clusters D1 and D2 demonstrated average pairwise SNP distances of 0.166 and 0.181, respectively, while in Clade F, the average pairwise SNP distance was 0.194. These minimal distances signify that the *L. pneumophila* STs within these specific clades were closely related genetically.

As displayed in Figure 5.5, samples of the water medium were included in every *L. pneumophila* clade detected, hinting the common source of the *L. pneumophila* contamination was the water. The confident clusters (bootstrap value > 0.9) consisting of water, air and surface samples were observed in Clade B, Clade D and C providing evidence for the possible *L. pneumophila* transmission route via water-air-surface path in the hospital environment. Many studies have reported plausible *L. pneumophila* transmission from water to air by aerosol-generating systems such as cooling towers, evaporative condensers, plumbing equipment humidifiers, respiratory therapy, and whirlpool baths (Legionella: Drinking Water Health Advisory 2001; Prussin, Schwake, and Marr 2017; Talapko et al. 2022). However, hitherto, the transmission study involving the route via water-air-surface has not been revealed. In this circumstance, this study has provided the rare-filed evidence for *L. pneumophila* transmission across medium boundary. Apparently, a discernible seasonal preference for *L. pneumophila* STs was also observed. In detail, STs in Cluster 1, Clade B, and Clade C were predominantly identified (88.9%) during the summer months, whereas STs in Clade D, Clade E, and Clade F showed a higher prevalence (85.7%)

during winter. The inclination for specific STs to be present in particular temperature ranges has been previously demonstrated as well (van der Kooij et al. 2016).

Appreciably, regarding the Cluster 1 (11 samples), *L. pneumophila* diversity was only observed in the water medium among sewage plant, blood-test room, gynecology and emergency departments, demonstrating a water-borne transmission of *L. pneumophila* within the hospital. Noteworthy, STs of cluster 1 have not been resolved by the ESGLI scheme, whereas these unknown STs are of particular concern because they exhibited cautious infectious characteristics such as the water-medium environment as the common source and being particularly suited to the summer season when *L. pneumophila* outbreaks were most likely to occur. The discovery of this frequently detected unresolved *L. pneumophila* lineage in the hospital setting also demonstrates the value of sophisticated cgMLST protocol in epidemiology source tracking, and meanwhile the unknown STs indicates the further effort required of ST identification for the outbreak assessment in the future.

With respect to the overall perspective, the scattered ST distribution in the tree reflects the *L. pneumophila* variation in the hospital. Samples of clade D, clade E, clade F and two sporadic cases (highlighted in the grey box in Figure 5.5) show confident clustering with strains of ST36. Given the high detection frequency of these clades (total 37 strains), ST36 was inferred as the predominant *L. pneumophila* type within the hospital, especially among wintertime. Previous studies have also reported *L. pneumophila* outbreak due to ST36. Mercante et al. (2016) found ST36 as the prevalent type in 1976 Legionnaires' Disease Outbreak in PhiladeLphia, USA, by WGS technique. Additionally, a study on typing of clinical and environmental isolates of *L. pneumophila* serotype in the USA from 1982 to 2012, also showed ST36 was one of the dominant STs responsible for both outbreak-associated and sporadic cases (Kozak-Muiznieks et al. 2014). Intriguingly, ST36 are less prevalent in *L. pneumophila* studies conducted in China. Most studies (Guo et al. 2014; Qin et al. 2014; Guo et al. 2015; Teng et al. 2017; Jiang et al. 2021) claimed that ST1 existed in all cities and was the most predominant ST in China. For example, Qin et al. (2014) conducted SBT on 164 *L. pneumophila* serogroup 1 isolates from cooling towers, hot springs, and drinking water in the six cities of Shanghai, Beijing, Shenzhen, Wuxi, Jinan, and Shijiazhuang over a 7-year period. Most of the isolates belonged to ST1, accounting for 49.4% of the 81 isolates. Nevertheless, *L. pneumophila* ST1 did not exist in the hospital of this study. The speculation for

the disparity of ST36 strains was that the ST determination in these prior studies were dependent on the culture-based method, and the culturing difficulty may contribute to an overlook of ST36.

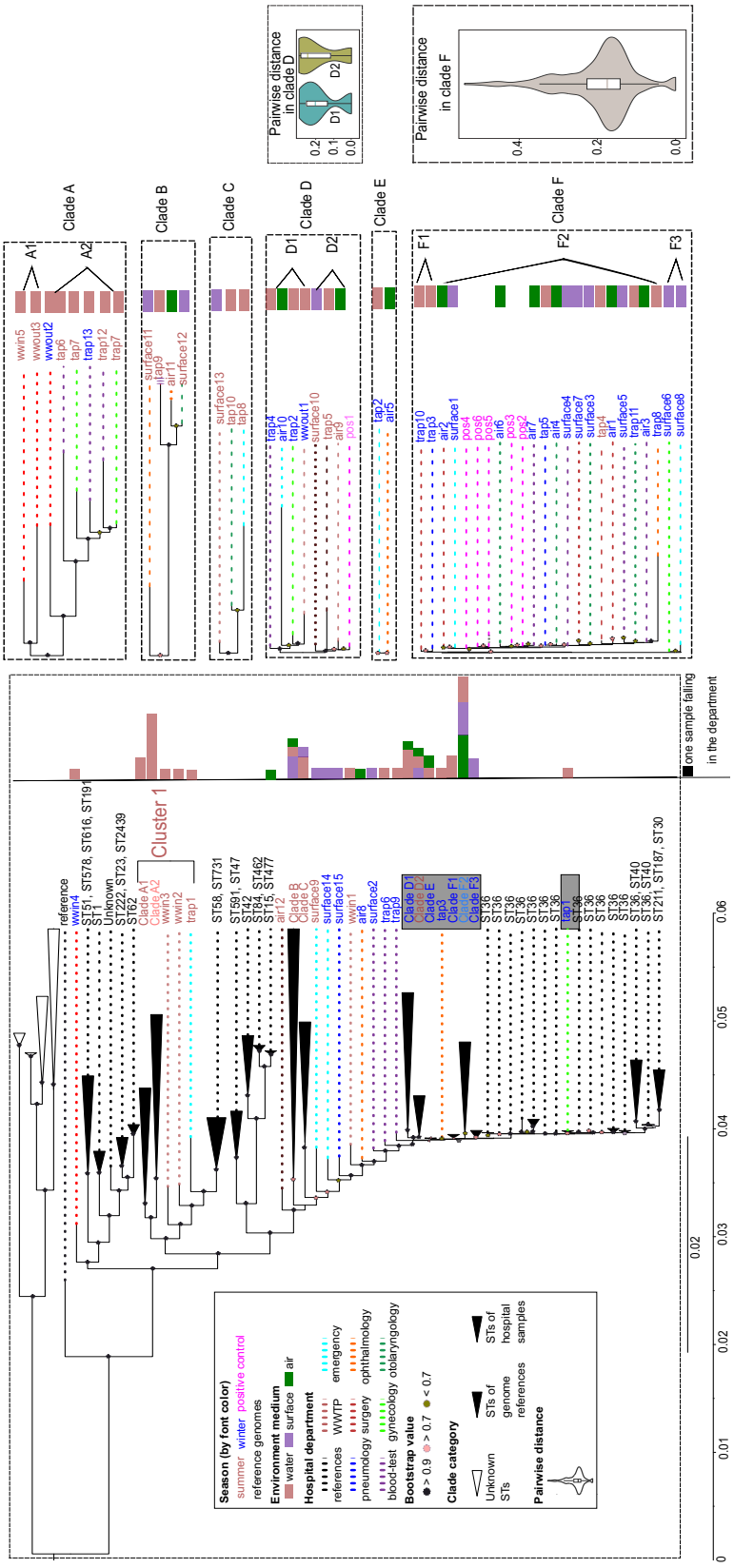


Figure 5.5 Phylogenetic tree constructed based on pseudo-genome generated by MinION sequencing of the 51cgMLST scheme across 64 hospital environmental samples. The left part reflects the whole-tree structure, and the right part is the expanded tree structure of each clade. The different font colors represent different seasons. Different dash line colors represent different hospital departments. Bar chart next to the corresponding environment sample shows composition of medium type within the clade. The different rectangle colors represent different medium types. Solid-circle on the branches represent different bootstrap levels. STs clusters with unknown STs, *L. pneumophila* references and real hospital samples are collapsed into blank, gray and black triangles, respectively. The pairwise SNP distances within the clade D and clade F were shown in the violin plots, where the inside box represents the median and quartile values of the pairwise SNP distance.

5.4 Conclusions

L. pneumophila can exist for a long period in low-temperature environments, especially in the aqueous environment. Therefore, long-term monitoring and genetic analyses should be performed on *L. pneumophila* from the environments, particularly from water supply systems. Overall, the PrimalPCR-based 51-cgMLST scheme coupled with the fast MinION sequencing approach described here represents a viable culture-free solution for assisting epidemiological surveillance of *L. pneumophila* in low resource settings across different environments in the hospital. Despite the challenges posed by the high diversity level of environmental samples, the PrimalPCR method still demonstrated its capability in terms of rapid, portable, and relatively affordable sequencing of *L. pneumophila*. On the one hand, it was able to recover 51 core genes from clean environmental samples, such as tap water, even when they contained as few as three *L. pneumophila* cells. On the other hand, the obtained consensus sequences from these samples were proved to be adequate for precise ST determination, showcasing the method's potential in epidemiological studies. This successful deployment in the hospital setting discovered the predominant ST36 type occurring frequently in wintertime, and meanwhile reveals the clear evidence for contamination of *L. pneumophila* with the water as the potential source. Therefore, the protocol and analytical procedure introduced here will facilitate a better tracing of transmission source and STs in *L. pneumophila* outbreak incidences.

Chapter Six

Discussion

The objectives set in *Chapter 1* have been accomplished by successfully characterising microbes and antibiotic resistance markers directly from low-biomass (or low-target gene) environmental samples in hospital settings, without the need for culturing. These samples included tap water, p-trap water, wastewater inlet, wastewater outlet, air dust, and surfaces.

Chapter 3 integrated V4-V5 *16S rRNA* Illumina amplicon sequencing and HT-qPCR to achieve taxonomic resolution at the genus level, with a LOD of 14.1 copies of the *16S rRNA* gene and 2.6e-07 copies of the ARG per *16S rRNA* copy. This approach facilitated the examination of bacterial community dynamics and co-occurrence of ARGs/MGEs, revealing the seasonal variations as a key driver shaping the beta diversity, highlighting the mobility of ARGs and underscoring the importance of monitoring antibiotic resistance, especially concerning vancomycin and beta-lactam resistance in *Acinetobacter*, *Pseudomonas* spp. and *Legionella* spp. in man-made water systems. In *Chapters 4* and *Chapter 5*, tiling multiplex PCR and MinION sequencing were accommodated, achieving a LOD of 3 copies per reaction. In *Chapter 4,* this not only identified the ST of *P. aeruginosa* through a seven-gene-based MLST but also predicted antibiotic resistance involving 93 markers, particularly due to genetic mutations. In *Chapter 5*, the pathogenicity, contamination pathways, and probable sources of *L. pneumophila* strains were investigated using the 51-gene-based cgMLST scheme, offering greater discriminatory power than MLST and a more straightforward method than whole-genome analysis.

In summary, this thesis provides the valuable insights into studying microbes that were challenging or impossible to culture in vitro, and offers compelling evidence endorsing the use of culture-free targeted sequencing technologies for more efficient, sensitive, and rapid environmental epidemiological studies from extremely low-biomass environmental samples.

6.1 Risks and challenge in the hospital environment

The healthcare environment, particularly the water and air-conditioner systems, can act as a reservoir for various nosocomial pathogens, such as *K. pneumoniae*, *Legionella* spp., *Pseudomonas* spp., and *S. maltophilia*. These pathogens have a higher prevalence in hospitals due to their ability to colonise

complex water systems and their potential transmission through aerosol routes within the healthcare facility. The presence of immunocompromised patients and invasive medical devices further increases the risk of infection with these pathogenic organisms.

In addition to being reservoirs for pathogens, healthcare-associated environments can also harbor ARGs. For example, ARGs like *vanB*, *bla-OXA$_{10}$* as well as *mcr-1*, conferring resistance to vancomycin, beta-lactams and peptide were found to be potentially carried by *Acinetobacter*, *Legionella* spp, and *Pseudomonas* spp, respectively. The presence of ARGs associated with these pathogens adds to the challenges of effective treatment, and the widespread mobility of ARGs through HGT pathways facilitated by MGEs (e.g. *tnpA 201* and *Tp614*) has prompted bacteria to acquire more resistance genes from other bacterial species, leading to the rapid spread of antibiotic resistance. The hospital environment is particularly conducive to HGT due to the high prevalence of antibiotic usage, dense microbial populations, and frequent contact between patients and healthcare workers.

However, the retrieval of genetic information from trace amounts of target DNA extracted from hospital environmental samples, such as portable water, surface, and air samples, poses a significant obstacle. These samples often contain low-input or degraded DNA, making it challenging to obtain adequate and high-quality genetic material for analysis. Overcoming this limitation is crucial for a comprehensive understanding of the genetic profiles of pathogens and the associated antibiotic resistance in healthcare environments.

6.2 Solutions for the challenge of sequencing low-input environmental samples

Here I investigated environmental samples, such as hospital tap-water samples and surface samples all fell under the category of low-biomass samples. Working with low-input or degraded DNA samples requires specialised approaches and techniques to overcome the obstacles associated with limited DNA materials. Thereby, I employed a comprehensive approach combining multiple sequencing technologies and methodologies to detect and analyse trace amounts of pathogens and ARGs in different environmental samples sensitively and accurately.

Firstly, I integrated *16S rRNA* amplicon sequencing and HT-qPCR technologies to evaluate the presence of pathogens and ARGs in water and air dust samples. This approach enabled me to successfully capture the overall bacterial diversity, even when the bacterial biomass was as low as 14.1 copies of the *16S rRNA* gene. Additionally, I was able to detect ARG copies at an even lower level, as low as 2.6e-07 copies per *16S rRNA* copy. Moreover, I developed PrimalSeq panels specifically designed for the fast detection of *L. pneumophila* and *P. aeruginosa*. For the 51-cgMLST *L. pneumophila* panel, the LOD for retrieving the STs of *L. pneumophila* was three copies per reaction. This panel enabled me to accurately identify and characterise the genetic diversity of *L. pneumophila* strains present in the samples with at least 20K MinION reads. Further, I employed a 103-gene *P. aeruginosa* panel, which successfully detected and characterised low-abundance antibiotic resistance even when delivered by as few as three cells of *P. aeruginosa*. This high sensitivity in detecting ARGs provides valuable information about the presence and potential spread of antibiotic resistance in *P. aeruginosa* populations with at least 100K MinION reads. By employing these comprehensive sequencing and detection methodologies, I was able to gain valuable insights into the pathogenic bacteria and resistome of variable samples, especially of low-biomass samples, uncovering the genetic basis of bacterial epidemiology and antibiotic resistance at various levels of abundance.

6.3 Sensitive targeted enrichment by PrimalPCR-based MinION sequencing

The application of multiplex PCR-based sequencing has significantly advanced rapid targeted sequencing approaches by allowing the simultaneous amplification and sequencing of multiple target regions within a single reaction. This technology has proven valuable in various research fields as described in the above chapters. Based on the PrimalScheme approach, I customised this to work with *L. pneumophila* and *P. aeruginosa*. To my knowledge, this was the first application of PrimalScheme for simultaneous bacterial gene identification and typing. An important aspect to achieve the high sensitivity and specificity of these schemes was the multiple primers designed for short amplicons. Another equally critical respect for the high sensitivity relies on the improved touchdown PCR program, making it more robust suited to low quality and load samples.

With the PrimalSeq workflow, the 51 conservative core genes that belong to 16 essential COG types of *L. pneumophila* panel were involved. The pure strain of *L. pneumophila* has validated the sensitivity and reliability of the panel to recover the STs accurately from as few as three cell numbers per reaction. With the real environmental samples in the hospital (LOD: below 37.583 Ct), the 51-cgMLST panel has achieved the STs phylogeny inference and *L. pneumophila* tracing. Nevertheless, one of the limitations was that as the number of primers in a single pool increased, the PCR efficiency was affected, leading to potential dropouts or uneven amplification of target sequences. When applying the *L. pneumophila* panel to complex environmental samples with high microbial biodiversity in my study, off-target amplification occurred. This off-target amplification, which amplified non-target sequences, were able to contribute to decreased multiplex-PCR efficiency and sequencing performance. This consequently demonstrated the priority for using the PrimalSeq-based 51cgMLST scheme in tracking *L. pneumophila* dissemination in clean environments. However, there is an assumption that increasing tiling pools to assign a significant number of primers in the future research might be beneficial to alleviate such off-target issues.

In a novel application of the *P. aeruginosa* panel, I expanded the panel to include 103 genes, which encompassed 10 housekeeping genes and 93 ARGs. Due to the high GC distribution of the primers and target sequences, I incorporated a GC enhancer into the touch-down PCR solution mix. The LOD of three cells of *P. aeruginosa* allowed us to determine the ST of the *P. aeruginosa* strain using seven MLST genes and simultaneously identify antibiotic resistance conferred by *P. aeruginosa*. Beyond profiling the ARGs, this approach provided higher sequence resolution compared to traditional methods, as it enabled us to generate a comprehensive catalogue of genetic polymorphisms, particularly SNPs, which could be used to predict antibiotic resistance caused by mechanisms such as the over-expression of efflux pumps, impermeability, target modification, and enzyme-mediated antimicrobial inactivation.

In summary, the PrimalPCR-based MinION sequencing panels offer several advantages over traditional PCR sequencing methods that require separate reactions for each target. Firstly, the panels allow for the detection and analysis of targets with as few as three bacterial cells, making them highly suitable for samples with limited DNA availability. This is particularly valuable when working with rare or low-

abundance targets. Next, by enabling the amplification and sequencing of multiple targets in a single reaction, the panels save time and resources. The use of portable MinION devices further enhances efficiency by providing rapid on-site sequencing capabilities. This streamlined workflow enables researchers to obtain comprehensive results within no more than three days for 24 samples, all manageable by a single person. This efficiency makes it ideal for field research and time-sensitive experiments. Addtionally, the panels enable the detection of multiple mutations simultaneously. This capability is highly advantageous in various fields, including genomics, microbiology, and clinical research. In addition to the superior advantages over traditional PCR, PrimalSeq panels also shine in circumstances where conducting direct WGS from environmental samples poses challenges, such as difficulties in bacterial culturing or unsuccessful WGA processes.

Overall, by utilising the Primalseq panels, I can gain a precise understanding of pathogen populations, including their spread and evolution, as well as the acquisition of antibiotic resistance. The ability to detect and analyse multiple targets in a single experiment with extremely low initial biomass enhances the efficiency and depth of genetic analysis, opening up more possibilities for research and clinical applications (e.g. with the sputum samples).

6.4 STs allocation for epidemic tracing

In the context of disease outbreak surveillance and management, the ability to quickly type and track infectious diseases is crucial. Determining STs allows for comparisons and characterisation of strains and isolates, providing a high level of discrimination without the need for full-genome sequencing. This level of resolution is often sufficient for pathogen characterisation and epidemiological investigations. High-risk STs have played a significant role in the spread of the current pandemic and other infectious diseases. Identifying and tracking these can provide valuable insights into the transmission dynamics and epidemiology of the disease.

In molecular epidemiology, MLST-based approaches have proven superior when the culture-dependent methods face challenges in discovering novel STs due to the limitations of conventional culturing

techniques. MLST can overcome these limitations by directly analysing the genetic material of the target pathogen, providing a comprehensive and accurate characterisation of strains. Furthermore, there is a trade-off between discriminatory power and epidemiological concordance when using MLST. In cases where the conventional seven-gene MLST scheme cannot differentiate strains within an ST, the use of cgMLST schemes, such as the one implemented with studied 51 genes, can provide better discrimination among epidemiologically unrelated isolates. This approach offers improved resolution while maintaining cost-effectiveness and requiring smaller scale bioinformatic analysis.

In the case of *L. pneumophila*, I found that water systems including tap water, p-trap water, and sewage in the hospital were highly contaminated with ST36 of *L. pneumophila*, particularly during the winter season. This suggests that there might be specific factors or conditions during the winter months that promoted the proliferation and survival of ST36 in the hospital water systems, and then spread farther via the potential water-air-surface pathway. Furthermore, I discovered a cluster of novel STs which have not been resolved by the ESGLI scheme. These STs were prevalent in the water-medium environment, particularly during the summer season. This finding indicates that different STs of *L. pneumophila* may have distinct preferences or adaptations to specific environmental conditions. In this case, the characteristics of the summer season, such as higher temperatures and potentially different pH levels, may favor the growth and persistence of these novel STs in the water systems (Zayed et al. 2020). Moreover, in the study of *P. aeruginosa*, I only discovered the ST549 prevalent in two water samples. While ST549 may not have been previously recognised as a high-risk type, the presence of ARGs in the strains associated with this ST suggests a potential concern regarding antibiotic resistance.

6.5 Acquisition of antibiotic resistance

The acquisition of AR can be promoted by HGT pathways, and genetic variations such as SNPs within specific genes. To be explicit, on the one hand, the HGT pathways are mediated by MGEs like plasmids and transposons, which carry genes encoding altered target binding sites or enzymes capable of inactivating antibiotic. One the other hand, SNPs have the potential to alter antibiotic target binding sites or the expression of porins, leading to antibiotic drug resistance. The emergence of bacteria with these

resistance mechanisms is an intractable issue, particularly in the hospital environment and clinical application. In the hospital environment, besides the HGT, the transmission and accumulation of AR can be facilitated by other factors including bacterial carriage and dissemination, close contact between people, patient movement, adequate environmental reservoirs (e.g. p-trap, contaminated surface, or medical device) or selective pressures (e.g. antibiotic usage). To effectively monitor and predict the potential risks of antibiotic resistance, it is crucial to advance profiling methods. However, it is important to note that the resistome, which refers to the collection of ARGs in a given environment, represents only a small fraction of the overall metagenomic content. This poses challenges in investigating low-abundance ARGs in various environmental settings.

Considering the lower cost and computation burden, I designed targeted technologies to facilitate more specific and efficient detection of ARGs. One such approach was HT-qPCR, which used 109 target primer sets. With HT-qPCR, I was able to detect 107 ARGs and MGEs with the lowest abundance being 2.6e-07 copies per *16S rRNA* copy. The various statistical analyses conducted using the HT-qPCR results offer compelling evidence of ARGs variations and mobility within the hospital environment. Notably, the robust correlation observed between the abundance of ARGs and MGEs suggests the possible HGT of ARGs facilitated by MGEs. Furthermore, the identified co-occurrence relationships among diverse genera (such as *Bacillus*, *Escherichia-Shigella*, *Klebsiella*, *Neisseria*, *Staphylococcus*, *Stenotrophomonas*, and *Streptococcus*) and *sat4*/*Tp614* strengthen the argument that *Tp614* may harbor *sat4* and conveyed it to these pathogen-related genera. Additionally, the noteworthy co-occurrence between bacterial communities and ARGs/MGEs provides further evidence of ARGs transmission facilitated by bacterial carriage. However, despite HT-qPCR provided an abundance information of ARGs, it lacked the capability to recover the nucleotide sequences of the target genes, limiting further exploration based on the genetic mechanisms.

To address this constraint, I designed targeted multiplex-PCR sequencing as a complementary technology. Targeted sequencing allowed it to restore the nucleotide sequences of the specific ARGs of interest, enabling more comprehensive genetic analysis. In total, the PrimalSeq panel of *P. aeruginosa* discovered ten ARGs that conferred resistance to six different antibiotic classes. The presence of mobile ARGs

including *APH(3'')-Ib* and *APH(6)-Id* whose sequences could be fully mapped to the genomes of multiple species, indicates the likelihood of these genes being acquired through HGT pathways. Consequently, despite the sensitive detection confirmed their presence in the particular sample, the precise bacterial host of these ARGs could not be conclusively identified. Importantly, this panel also demonstrates the novelty by enabling deep scanning of SNPs using only three *P. aeruginosa* cells, thereby providing valuable insights into the mechanisms underlying antibiotic resistance, including enzyme inactivation, impermeability, and overexpression of efflux pumps. Specifically, I identified nonsynonymous SNPs in the *gyrA* gene within the QRDR. Examples of these SNPs include ACC to ATC in the 248-nt motif and GAC to GGC in the 260-nt motif. These SNPs have the potential to cause amino acid changes, resulting in threonine being replaced by isoleucine and aspartic acid being replaced by glycine, respectively. Such alterations in the *gyrA* gene may contribute to fluoroquinolone resistance in *P. aeruginosa* strains. Additionally, I observed other SNPs in the *mexR* and *nfxB* genes that may also be associated with fluoroquinolone resistance. The SNP in the *mexR* gene led to an amino acid alteration from leucine to proline, while the SNP in the *nfxB* gene resulted in an amino acid change from asparagine to aspartic acid. These variations in MexR and NfxB proteins could potentially play a role in the development of fluoroquinolone resistance. Furthermore involving the carbapenem resistance, I observed a mutation in the *oprD* gene. Loss of OprD from the outer membrane is a significant mechanism of carbapenem resistance in *P. aeruginosa*. I identified three common mutations that potentially contributed to carbapenem resistance in the tested *P. aeruginosa* strains. These mutations could result in changes in pore specificity and conformation, as they involved the replacement of a positively charged amino acid with a neutral-polarity amino acid. In addition to alterations in OprD, the overexpression of the MexEF-OprN efflux pump, regulated by the *mexT* or *mexS* gene, can also contribute to carbapenem resistance. We detected an SNP in the 514-nt motif of the *mexT* gene, which aligned with the *mexT* SNP observed in other carbapenem-resistant isolates.

Based on the findings provided, this thesis integrating several advanced technologies can effectively expand the resistome by detecting hundreds of low-abundance ARGs. Furthermore, this thesis highlights the importance of SNP-related analysis, which relies on targeted enrichment techniques, to determine antibiotic resistance based on nucleotide mutations. By focusing on specific ARGs and their genetic

variations, this approach allows for a more detailed assessment of antibiotic resistance and provides a cost- and time-efficient tool for studying the variation and evolution of specific ARGs in both clinical and natural environments.

6.6 Transmission and source tracking of bacterial communities and antibiotic resistance

In the thesis, the co-occurrence analyses between ARGs, MGEs and bacterial communities offer a statistical framework for predicting HGT events and bacterial carriage of ARGs. In detail, a correlation coefficient of 1 unequivocally supports the mobility of ARGs in the data. Otherwise, further validation involving comparisons of genetic relatedness or experimental assays (as detailed in *section 6.7*) is necessary to confirm the prediction. With regard to the source inference, utilising a Bayesian approach for tracking the source provides a straightforward means of assessing the presence of both microbes and ARGs in a given sink, originating from potential source environments. Despite the utilisation of Gibbs sampling to consider all conceivable assignments of test samples to various source environments, the accuracy of the source-tracking estimates is heavily dependent on the comprehensiveness of the source environments used during training. This method hinges on evaluating the relative abundance data of bacterial communities or specific genes obtained through OTUs classification or qPCR results, respectively. While it is possible to determine the proportions of the hypothesized sink and sources, unraveling the mechanisms through which bacterial communities or ARGs traverse to the sink, whether via HGT or human carriers, presents a persistent challenge.

Delving into the genetic relatedness or similarity, such as utilising MLST or SNP distance, among targeted genes like core genes or ARGs within the studied organisms, offers a practical approach to not only identifying common sources but also elucidating potential transmission pathways of bacteria or AR contamination. To elaborate, MLST and SNP analysis entail the identification of multiple housekeeping genes and single nucleotide changes within the genetic sequence, respectively. By comparing these STs or SNPs in the genomes of bacteria across different samples, the genetic relatedness can be determined, and a common source can be traced back with the minimal SNPs. Furthermore, constructing phylogenetic trees based on the genetic sequences of interest, not only aids in identifying common sources by

observing the clustering of sequences from diverse samples but also hints at the horizontal transfer of ARGs. If gene sequences from distinct species cluster together on the phylogenetic tree instead of aligning with the species' evolutionary history, it implies potential horizontal transfer of ARGs. Specifically, on the one hand, by analysing the shared genes of MGEs like plasmid or transposon, the movement of them among organisms can be understood. On the other hand, by examining gene cassettes like ARGs within the integron among organisms, their distribution and association with other genetic elements can be elucidated. These in-depth genetic analyses provide valuable insights into the origins and dissemination patterns of bacterial communities and antibiotic resistance, serving as a significant augmentation to bolster the findings derived from statistical analyses, including co-occurrence analysis and Bayesian-based source inference.

6.7 Limitations and future work

The technologies and panels devised in the thesis for achieving culture-free detection of bacteria and antibiotic resistance from low-biomass environmental samples are tailored for situations where traditional culture-based methods are impractical. These tools are not only suitable for clinical applications like rapid diagnostic testing in hospitals but also for outbreak investigations where the swift exploration of the pathogen, antibiotic resistance, virulence factors, and evolutionary relationships is of paramount importance within a short timeframe. While the unique advantages of these tools facilitate their widespread adoption across various fields, they do come with limitations. In *16S rRNA* amplicon sequencing, the taxonomy (OTU) resolution is constrained by focusing on only a few hypervariable regions rather than the entire *16S rRNA* gene. While these regions simplify and ensure confidence in assessing bacterial diversity, achieving species-level and strain-level resolution is also vital. Furthermore, the reliance of PrimalSeq panels (or amplicon-based panels) on the reference alignment and variant calling to generate consensus reads makes it challenging to determine multiple lineages within a sample. Additionally, although phylogenetic trees are preferred for illustrating evolutionary relationships concerning genetic mutation events, they may not effectively capture recombination events. Another limitation is that while the genotype of antibiotic resistance acts as an indicator for the phenotype of

antibiotic resistance, it may not always guarantee accurate predictions of drug sensitivity in real-world scenarios.

In future research, I anticipate that targeted technologies continue playing a significant role in recovering valuable genetic information from extremely low biomass samples, even in situations where culturing might present challenges. Particularly, with the continuous advancements in MinION sequencing technology, such as the introduction of new reagent kits, flow cells, and bioinformatic pipelines, it is indeed reasonable to expect that targeted MinION sequencing can provide competitive performance compared to targeted Illumina sequencing. Additionally, the rapid and portable traits of MinION sequencing further enhance its appeal. Nevertheless, while targeted technologies have their strengths in focusing on specific areas of interest sensitively, it is important to note that they are limited to retrieving information within the boundaries of known databases or pre-defined interests. In fact, there is a trade-off between the ability to discover novel genetic information and the requirement for higher biomass for sequencing. Complementary approaches such as WGS and metagenomic sequencing that can uncover new genomic elements typically demand a higher quantity of genetic material to generate sufficient sequencing coverage and depth, especially for the complex environmental samples during the MinION sequencing. Additionally, *de novo*-assembly-based technologies have the potential to outperform amplicon-based sequencing in identifying multiple strains within a sample. As a result, the battle between low biomass samples and the exploration of novel genetic insights or the identification of multiple strains highlights the need for optimisation and development of low-input WGS techniques that can bridge this gap, such as adding the background or carrier lambda DNA during the WGS nanopore library preparation or improving the specification of selective amplification of specific whole genome during the WGA process.

Building upon the panels with *L. pneumophila* and *P. aeruginosa*, expanding the development of PrimalSeq panels on more pathogenic bacteria and antibiotic resistance holds great promise. (cg)MLST studies of pathogenic bacteria are one of the important applications, which will remain a vital approach for conducting epidemiological investigations during disease outbreaks, particularly in the culture-free application for the low-biomass samples. However, continuous efforts to expand the STs databases for

various pathogenic bacteria are necessary. By expanding these databases through ongoing MLST research, I can improve strain typing accuracy, and advance the understanding of pathogenic bacteria, ultimately supporting effective outbreak control and prevention measures.

The critical issue for the extensive development of PrimalSeq panels is addressing amplification competition among hundreds of target genes, which is crucial in tiling multiplex-PCR for environmental samples with high biodiversity. Increasing the number of primer pools is a potential strategy to alleviate amplification bias and improve coverage of diverse targeted genes. By dividing the primers into three or four pools, amplification bias can be reduced, allowing for a more even and comprehensive amplification of DNA from various strains present in the environmental sample. This approach helps to mitigate the overrepresentation of certain target that may occur when using a smaller number of primer pools. In addition to increasing the number of primer pools, optimising primer design is essential to minimise amplification bias. Careful selection of primers that target conserved regions across different genes ensures that a wide range of bacterial strains can be captured. Additionally, incorporating degenerate primers that target non-conserved regions can enhance the coverage and representation of diverse genetic variants within the targeted genes.

Another limitation of the PrimalSeq panel is its weakness in providing direct insights into the phenotypic expression of resistance. Certainly, the presence of a resistance gene does not guarantee its functional activity, and phenotypic testing, such as antimicrobial susceptibility testing, is still necessary to confirm resistance patterns. In addition, to dig into the which HGT contributed to the transfer of ARGs, conducting experiments with various assays like conjugation assays that include donor, recipient, and selection strains to track genetic material transfer, or transposition assays to evaluate the transposition frequency of transposons, can be highly advantageous.

Considering these factors, traditional culturing methods remain essential in complementing targeted molecular approaches in the future study. When dealing with complex hospital environmental samples that may contain various resistance mechanisms, combining phenotypic testing and SNP analysis using PrimalSeq panels can provide a comprehensive understanding of resistance mechanisms and their

acquisition as well as dissemination patterns in *P. aeruginosa*. Additionally, expanding the inclusion of more environmental samples from hospital settings can provide insights into the prevalence and persistence of antibiotic resistance *P. aeruginosa* strains in healthcare environments.

In summary, optimising the amplification efficiency in high-biodiversity environmental samples, exploring the low-input WGS technologies and supplementing the phenotypic test with molecular approaches are crucial strategies for advancing future targeted research. With such improvements, the discovery of unknown STs and novel mechanisms of antibiotic resistance can be significantly enhanced. This knowledge is particularly important for outbreak assessment in hospitals, as it enables a better understanding of the microbial population dynamics, transmission patterns, and helps in implementing effective measures to control and manage the outbreak.

References

Abbara, Salam, Aurélia Pitsch, Sébastien Jochmans, Kyann Hodjat, Pascale Cherrier, Mehran Monchi, Christophe Vinsonneau, and Sylvain Diamantis. 2019. “Impact of a Multimodal Strategy Combining a New Standard of Care and Restriction of Carbapenems, Fluoroquinolones and Cephalosporins on Antibiotic Consumption and Resistance of Pseudomonas Aeruginosa in a French Intensive Care Unit.” *International Journal of Antimicrobial Agents* 53 (4): 416–22.

Adrian, Jonas, Patrizia Bonsignore, Sebastian Hammer, Tancred Frickey, and Christof R. Hauck. 2019. “Adaptation to Host-Specific Bacterial Pathogens Drives Rapid Evolution of a Human Innate Immune Receptor.” *Current Biology: CB* 29 (4): 616–30.e5.

Ahmed, Omar B. 2022. “Detection of Antibiotic Resistance Genes in Pseudomonas Aeruginosa by Whole Genome Sequencing.” *Infection and Drug Resistance* 15 (November):6703–9.

Ahsanuddin, Sofia, Ebrahim Afshinnekoo, Jorge Gandara, Mustafa Hakyemezoğlu, Daniela Bezdan, Samuel Minot, Nick Greenfield, and Christopher E. Mason. 2017. “Assessment of REPLI-G Multiple Displacement Whole Genome Amplification (WGA) Techniques for Metagenomic Applications.” *Journal of Biomolecular Techniques: JBT* 28 (1): 46–55.

Aires, J. R., T. Köhler, H. Nikaido, and P. Plésiat. 1999. “Involvement of an Active Efflux System in the Natural Resistance of Pseudomonas Aeruginosa to Aminoglycosides.” *Antimicrobial Agents and Chemotherapy* 43 (11): 2624–28.

Alberts, Bruce. 2017. *Molecular Biology of the Cell*. Garland Science.

Al-Trad, Esra’a I., Ainal Mardziah Che Hamzah, Suat Moi Puah, Kek Heng Chua, Muhamad Zarul Hanifah, Qasim Ayub, Prasit Palittapongarnpim, Stephen M. Kwong, Ching Hoong Chew, and Chew Chieng Yeo. 2023. “Complete Genome Sequence and Analysis of a ST573 Multidrug-Resistant

Methicillin-Resistant Staphylococcus Aureus SauR3 Clinical Isolate from Terengganu, Malaysia." *Pathogens* 12 (3). https://doi.org/10.3390/pathogens12030502.

Alcock, Brian P., Amogelang R. Raphenya, Tammy T. Y. Lau, Kara K. Tsang, Mégane Bouchard, Arman Edalatmand, William Huynh, et al. 2020. "CARD 2020: Antibiotic Resistome Surveillance with the Comprehensive Antibiotic Resistance Database." *Nucleic Acids Research* 48 (D1): D517–25.

Alhayek, A., and A. K. H. Hirsch. 2023. "Bacterial Metalloproteases as Promising Drug Targets for Antivirulence Agents."

Alikhan, Nabil-Fareed, Nicola K. Petty, Nouri L. Ben Zakour, and Scott A. Beatson. 2011. "BLAST Ring Image Generator (BRIG): Simple Prokaryote Genome Comparisons." *BMC Genomics* 12 (August): 402.

Alio, Ifey, Raphael Moll, Tim Hoffmann, Uwe Mamat, Ulrich E. Schaible, Kai Pappenfort, Malik Alawi, et al. 2023. "Stenotrophomonas Maltophilia Affects the Gene Expression Profiles of the Major Pathogens Pseudomonas Aeruginosa and Staphylococcus Aureus in an in Vitro Multispecies Biofilm Model." *Microbiology Spectrum* 11 (6): e0085923.

Allen, Heather K., Justin Donato, Helena Huimi Wang, Karen A. Cloud-Hansen, Julian Davies, and Jo Handelsman. 2010. "Call of the Wild: Antibiotic Resistance Genes in Natural Environments." *Nature Reviews. Microbiology* 8 (4): 251–59.

Alloui, Tarek, Imane Boussebough, Allaoua Chaoui, Ahmed Zakaria Nouar, and Mohamed Chaouki Chettah. 2015. "Usearch: A Meta Search Engine Based on a New Result Merging Strategy." In *2015 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)*, 01:531–36.

Almahmoud, Iyad, Elisabeth Kay, Dominique Schneider, and Max Maurin. 2009. “Mutational Paths towards Increased Fluoroquinolone Resistance in Legionella Pneumophila.” *The Journal of Antimicrobial Chemotherapy* 64 (2): 284–93.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. “Basic Local Alignment Search Tool.” *Journal of Molecular Biology* 215 (3): 403–10.

Amábile-Cuevas, Carlos F. 1995. *Origin, Evolution and Spread of Antibiotic Resistance Genes*. R.G. Landes Company.

Andersson, Dan I., and Diarmaid Hughes. 2012. “Evolution of Antibiotic Resistance at Non-Lethal Drug Concentrations.” *Drug Resistance Updates: Reviews and Commentaries in Antimicrobial and Anticancer Chemotherapy* 15 (3): 162–72.

Andrews, Kimberly R., Jeffrey M. Good, Michael R. Miller, Gordon Luikart, and Paul A. Hohenlohe. 2016. “Harnessing the Power of RADseq for Ecological and Evolutionary Genomics.” *Nature Reviews. Genetics* 17 (2): 81–92.

Anscombe, Catherine, Raju V. Misra, and Saheer Gharbia. 2018. “Whole Genome Amplification and Sequencing of Low Cell Numbers Directly from a Bacteria Spiked Blood Model.” *bioRxiv.* https://doi.org/10.1101/153965

Antimicrobial Resistance Collaborators. 2022. “Global Burden of Bacterial Antimicrobial Resistance in 2019: A Systematic Analysis.” *The Lancet* 399 (10325): 629–55.

Arber, Werner. 2014. “Horizontal Gene Transfer among Bacteria and Its Role in Biological Evolution.” *Life* 4 (2): 217–24.

Aronesty, E. 2013. “TOBioJ: Comparison of Sequencing Utility Programs.”

Ashton, Philip M., Satheesh Nair, Tim Dallman, Salvatore Rubino, Wolfgang Rabsch, Solomon Mwaigwisya, John Wain, and Justin O'Grady. 2015. "MinION Nanopore Sequencing Identifies the Position and Structure of a Bacterial Antibiotic Resistance Island." *Nature Biotechnology* 33 (3): 296–300.

Atassi, Giancarlo, Rachel Medernach, Marc Scheetz, Sophia Nozick, Nathaniel J. Rhodes, Megan Murphy-Belcaster, Katherine R. Murphy, Arghavan Alisoltani, Egon A. Ozer, and Alan R. Hauser. 2023. "Genomics of Aminoglycoside Resistance in Pseudomonas Aeruginosa Bloodstream Infections at a United States Academic Hospital." *Microbiology Spectrum* 11 (3): e0508722.

Avershina, Ekaterina, Stephan A. Frye, Jawad Ali, Arne M. Taxt, and Rafi Ahmad. 2022. "Ultrafast and Cost-Effective Pathogen Identification and Resistance Gene Detection in a Clinical Setting Using Nanopore Flongle Sequencing." *Frontiers in Microbiology* 13 (March): 822402.

Ayyildiz, Dilara, and Silvano Piazza. 2019. "Introduction to Bioinformatics." *Methods in Molecular Biology* 1986: 1–15.

Baggesen, D. L., G. Sørensen, E. M. Nielsen, and H. C. Wegener. 2010. "Phage Typing of Salmonella Typhimurium – Is It Still a Useful Tool for Surveillance and Outbreak Investigation?" *Eurosurveillance* 15 (4): 19471.

Baghal Asghari, Farzaneh, Mahnaz Nikaeen, and Hossein Mirhendi. 2013. "Rapid Monitoring of Pseudomonas Aeruginosa in Hospital Water Systems: A Key Priority in Prevention of Nosocomial Infection." *FEMS Microbiology Letters* 343 (1): 77–81.

Baker, Jonathon L. 2022. "Using Nanopore Sequencing to Obtain Complete Bacterial Genomes from Saliva Samples." *mSystems* 7 (5): e0049122.

Balloux, Francois, and Lucy van Dorp. 2017. “Q&A: What Are Pathogens, and What Have They Done to and for Us?” *BMC Biology* 15 (1): 91.

Baquero, Fernando, José-Luis Martínez, and Rafael Cantón. 2008. “Antibiotics and Antibiotic Resistance in Water Environments.” *Current Opinion in Biotechnology* 19 (3): 260–65.

Barbau-Piednoir, Elodie, Sophie Bertrand, Jacques Mahillon, Nancy H. Roosens, and Nadine Botteldoorn. 2013. “SYBR®Green qPCR Salmonella Detection System Allowing Discrimination at the Genus, Species and Subspecies Levels.” *Applied Microbiology and Biotechnology* 97 (22): 9811–24.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. “Gephi: An Open Source Software for Exploring and Manipulating Networks.” *Proceedings of the International AAAI Conference on Web and Social Media* 3 (1): 361–62.

Bäumler, Andreas, and Ferric C. Fang. 2013. “Host Specificity of Bacterial Pathogens.” *Cold Spring Harbor Perspectives in Medicine* 3 (12): a010041.

Baxevanis, Andreas D., Gary D. Bader, and David S. Wishart. 2020. *Bioinformatics*. John Wiley & Sons.
Beceiro, Alejandro, María Tomás, and Germán Bou. 2013. “Antimicrobial Resistance and Virulence: A Successful or Deleterious Association in the Bacterial World?” *Clinical Microbiology Reviews* 26 (2): 185–230.

Ben Maamar, Sarah, Jinglin Hu, and Erica M. Hartmann. 2020. “Implications of Indoor Microbial Ecology and Evolution on Antibiotic Resistance.” *Journal of Exposure Science & Environmental Epidemiology* 30 (1): 1–15.

Bentivoglio, M., and P. Pacini. 1995. “Filippo Pacini: A Determined Observer.” *Brain Research Bulletin* 38 (2): 161–65.

Benton, D. 1996. “Bioinformatics--Principles and Potential of a New Multidisciplinary Tool.” *Trends in Biotechnology* 14 (8): 261–72.

Berglund, Björn. 2015. “Environmental Dissemination of Antibiotic Resistance Genes and Correlation to Anthropogenic Contamination with Antibiotics.” *Infection Ecology & Epidemiology* 5 (September): 28564.

Bernardet, J-F, P. Segers, M. Vancanneyt, F. Berthe, K. Kersters, and P. Vandamme. 1996. “Cutting a Gordian Knot: Emended Classification and Description of the Genus Flavobacterium, Emended Description of the Family Flavobacteriaceae, and Proposal of Flavobacterium Hydatis Nom. Nov. (Basonym, Cytophaga Aquatilis Strohl and Tait 1978).” *International Journal of Systematic Bacteriology* 46 (1): 128–48.

Bhullar, Kirandeep, Nicholas Waglechner, Andrew Pawlowski, Kalinka Koteva, Eric D. Banks, Michael D. Johnston, Hazel A. Barton, and Gerard D. Wright. 2012. “Antibiotic Resistance Is Prevalent in an Isolated Cave Microbiome.” *PloS One* 7 (4): e34953.

Bigot, Yves. 2012. Mobile Genetic Elements: Protocols and Genomic Applications. *Humana Press*.

Blagodatskikh, Konstantin A., Vladimir M. Kramarov, Ekaterina V. Barsova, Alexey V. Garkovenko, Dmitriy S. Shcherbo, Andrew A. Shelenkov, Vera V. Ustinova, et al. 2017. “Improved DOP-PCR (iDOP-PCR): A Robust and Simple WGA Method for Efficient Amplification of Low Copy Number Genomic DNA.” *PloS One* 12 (9): e0184507.

Blair, Jessica M. A., Grace E. Richmond, and Laura J. V. Piddock. 2014. “Multidrug Efflux Pumps in Gram-Negative Bacteria and Their Role in Antibiotic Resistance.” *Future Microbiology* 9 (10): 1165–77.

Blair, Jessica M. A., Mark A. Webber, Alison J. Baylay, David O. Ogbolu, and Laura J. V. Piddock. 2015. “Molecular Mechanisms of Antibiotic Resistance.” *Nature Reviews. Microbiology* 13 (1): 42–51.

Bochner, Barry R. 2009. “Global Phenotypic Characterization of Bacteria.” *FEMS Microbiology Reviews* 33 (1): 191–205.

Bokulich, Nicholas A., David A. Mills, and Mark A. Underwood. 2013. “Surface Microbes in the Neonatal Intensive Care Unit: Changes with Routine Cleaning and over Time.” Journal of Clinical *Microbiology* 51 (8): 2617–24.

Bonadonna, Lucia, Rossella Briancesco, and Anna Maria Coccia. 2017. “Analysis of Microorganisms in Hospital Environments and Potential Risks.” In *Indoor Air Quality in Healthcare Facilities*, edited by Stefano Capolongo, Gaetano Settimo, and Marco Gola, 53–62. Cham: Springer International Publishing.

Borchardt, J., J. H. Helbig, and P. C. Lück. 2008. “Occurrence and Distribution of Sequence Types among Legionella Pneumophila Strains Isolated from Patients in Germany: Common Features and Differences to Other Regions of the World.” *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 27 (1): 29–36.

Bosch, Thijs, Sjoerd M. Euser, Fabian Landman, Jacob P. Bruin, Ed P. IJzerman, Jeroen W. den Boer, and Leo M. Schouls. 2015. “Whole-Genome Mapping as a Novel High-Resolution Typing Tool for Legionella Pneumophila.” *Journal of Clinical Microbiology* 53 (10): 3234–38.

Brachi, Benjamin, Geoffrey P. Morris, and Justin O. Borevitz. 2011. “Genome-Wide Association Studies in Plants: The Missing Heritability Is in the Field.” *Genome Biology* 12 (10): 232.

Boyce, John M. 2007. “Environmental Contamination Makes an Important Contribution to Hospital Infection.” *The Journal of Hospital Infection* 65 Suppl 2 (June): 50–54.

Briese, Thomas, Janusz T. Paweska, Laura K. McMullan, Stephen K. Hutchison, Craig Street, Gustavo Palacios, Marina L. Khristova, et al. 2009. “Genetic Detection and Characterization of Lujo Virus, a New Hemorrhagic Fever-Associated Arenavirus from Southern Africa.” *PLoS Pathogens* 5 (5): e1000455.

Brinkwirth, Simon, Olaniyi Ayobami, Tim Eckmanns, and Robby Markwart. 2021. “Hospital-Acquired Infections Caused by Enterococci: A Systematic Review and Meta-Analysis, WHO European Region, 1 January 2010 to 4 February 2020.” *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 26 (45).

Brown, James H., James F. Gillooly, Andrew P. Allen, Van M. Savage, and Geoffrey B. West. 2004. “Toward a Metabolic Theory of Ecology.” *Ecology* 85 (7): 1771–89.

Brown, Joshua Randal. 2010. *The Design and Synthesis of Novel Antimicrobial Agents for Use in the Battle Against Bacterial Resistance*. University of Tennessee Health Science Center.

Brown, Sam P., Daniel M. Cornforth, and Nicole Mideo. 2012. “Evolution of Virulence in Opportunistic Pathogens: Generalism, Plasticity, and Control.” *Trends in Microbiology* 20 (7): 336–42.

Bruchmann, Sebastian, Andreas Dötsch, Bianka Nouri, Iris F. Chaberny, and Susanne Häussler. 2013. “Quantitative Contributions of Target Alteration and Decreased Drug Accumulation to Pseudomonas Aeruginosa Fluoroquinolone Resistance.” *Antimicrobial Agents and Chemotherapy* 57 (3): 1361–68.

Brüssow, Harald, and Roger W. Hendrix. 2002. “Phage Genomics: Small Is Beautiful.” *Cell* 108 (1): 13–16.

Bryson, John R., Lauren Andres, Aksel Ersoy, and Louise Reardon. 2021. Living with Pandemics: Places, People and Policy. *Edward Elgar Publishing*.

Bulloch, W. 1960. “The History of Bacteriology.” *The History of Bacteriology.* https://www.cabdirect.org/cabdirect/abstract/19611100022.

Cabot, Gabriel, Sebastian Bruchmann, Xavier Mulet, Laura Zamorano, Bartolomé Moyà, Carlos Juan, Susanne Haussler, and Antonio Oliver. 2014. “Pseudomonas Aeruginosa Ceftolozane-Tazobactam Resistance Development Requires Multiple Mutations Leading to Overexpression and Structural Modification of AmpC.” *Antimicrobial Agents and Chemotherapy* 58 (6): 3091–99.

Cabot, Gabriel, Carla López-Causapé, Alain A. Ocampo-Sosa, Lea M. Sommer, María Ángeles Domínguez, Laura Zamorano, Carlos Juan, et al. 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): 7415–23.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. “BLAST+: Architecture and Applications.” *BMC Bioinformatics* 10 (December): 421.

Cangelosi, Gerard A., Nancy E. Freitag, Merry Shannon Riley-Buckley, American Society for Microbiology, and American Academy of Microbiology. 2004. *From Outside to Inside: Environmental Microorganisms as Human Pathogens*.

Cao, Boyang, Rongrong Li, Songjin Xiong, Fangfang Yao, Xiangqian Liu, Min Wang, Lu Feng, and Lei Wang. 2011. “Use of a DNA Microarray for Detection and Identification of Bacterial Pathogens Associated with Fishery Products.” *Applied and Environmental Microbiology* 77 (23): 8219–25.

Cao, Guoqing, Shaotao Tang, Dehua Yang, Wenjia Shi, Xiaorong Wang, Hua Wang, Chen Li, Jia Wei, and Ling Ma. 2020. “The Potential Transmission of SARS-CoV-2 from Patients with Negative RT-PCR Swab Tests to Others: Two Related Clusters of COVID-19 Outbreak.” *Japanese Journal of Infectious Diseases* 73 (6): 399–403.

Caporaso, J. Gregory, Justin Kuczynski, Jesse Stombaugh, Kyle Bittinger, Frederic D. Bushman, Elizabeth K. Costello, Noah Fierer, et al. 2010. “QIIME Allows Analysis of High-Throughput Community Sequencing Data.” *Nature Methods* 7 (5): 335–36.

Carter, Jean-Michel, and Shobbir Hussain. 2017. “Robust Long-Read Native DNA Sequencing Using the ONT CsgG Nanopore System.” *Wellcome Open Research* 2 (April): 23.

Casewell, Mark, Christian Friis, Enric Marco, Paul McMullin, and Ian Phillips. 2003. “The European Ban on Growth-Promoting Antibiotics and Emerging Consequences for Human and Animal Health.” *The Journal of Antimicrobial Chemotherapy* 52 (2): 159–61.

Cassone, Marco, Julia Mantey, Kyle J. Gontjes, Bonnie J. Lansing, Kristen E. Gibson, Joyce Wang, and Lona Mody. 2021. “Seasonal Patterns in Incidence and Antimicrobial Resistance of Common Bacterial Pathogens in Nursing Home Patients and Their Rooms.” *Frontiers in Public Health* 9 (July): 671428.

Castresana, J. 2000. “Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis.” *Molecular Biology and Evolution* 17 (4): 540–52.

Castro-Escarpulli, Graciela, Maribel Nayelli, Gildardo Rivera Sánchez, Virgilio Bocanegra-Garcia, Xianwu Guo, Sara R. Juárez-Enríquez, Julieta Luna-Herrera, et al. n.d. “Identification and Typing Methods for the Study of Bacterial Infections: A Brief Review and Mycobacterial as Case of Study.” Accessed June 13, 2023. http://www.rdcb.cbg.ipn.mx/bitstream/20.500.12273/674/1/358.pdf.

Chadwick, Derek J., and Jamie A. Goode. 2008. *Antibiotic Resistance: Origins, Evolution, Selection and Spread*. John Wiley & Sons.

Chamberlain, J. S., R. A. Gibbs, J. E. Ranier, P. N. Nguyen, and C. T. Caskey. 1988. “Deletion Screening of the Duchenne Muscular Dystrophy Locus via Multiplex DNA Amplification.” *Nucleic Acids Research* 16 (23): 11141–56.

Chancey, Scott T., Dorothea Zähner, and David S. Stephens. 2012. “Acquired Inducible Antimicrobial Resistance in Gram-Positive Bacteria.” *Future Microbiology* 7 (8): 959–78.

Chang, Soju, Dawn M. Sievert, Jeffrey C. Hageman, Matthew L. Boulton, Fred C. Tenover, Frances Pouch Downes, Sandip Shah, et al. 2003. “Infection with Vancomycin-Resistant Staphylococcus Aureus Containing the vanA Resistance Gene.” *The New England Journal of Medicine* 348 (14): 1342–47.

Chawla, Hitikk, Purnima Anand, Kritika Garg, Neeru Bhagat, Shivani G. Varmani, Tanu Bansal, Andrew J. McBain, and Ruchi Gulati Marwah. 2023. “A Comprehensive Review of Microbial Contamination in the Indoor Environment: Sources, Sampling, Health Risks, and Mitigation Strategies.” *Frontiers in Public Health* 11 (November):1285393.

Chen, Hanbo, and Paul C. Boutros. 2011. “VennDiagram: A Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R.” *BMC Bioinformatics* 12 (January): 35.

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor.” *Bioinformatics* 34 (17): i884–90.

Chen, Yihan, Jian-Qiang Su, Junya Zhang, Peng Li, Hongjie Chen, Bo Zhang, Karina Yew-Hoong Gin, and Yiliang He. 2019. “High-Throughput Profiling of Antibiotic Resistance Gene Dynamic in a Drinking Water River-Reservoir System.” *Water Research* 149 (February): 179–89.

Cheng, Hang, Yuhong Sun, Qing Yang, Minggui Deng, Zhijian Yu, Gang Zhu, Jiuxin Qu, Lei Liu, Liang Yang, and Yu Xia. 2022. “A Rapid Bacterial Pathogen and Antimicrobial Resistance Diagnosis Workflow

Using Oxford Nanopore Adaptive Sequencing Method." *Briefings in Bioinformatics* 23 (6). https://doi.org/10.1093/bib/bbac453.

China Meteorological Administration. 2024.

Choe, Young June, Michael A. Smit, and Leonard A. Mermel. 2019. "Seasonality of Respiratory Viruses and Bacterial Pathogens." *Antimicrobial Resistance and Infection Control* 8 (July): 125.

Choi, Chris H., Valeriy Poroyko, So Watanabe, Duo Jiang, James Lane, Marcella deTineo, Fuad M. Baroody, Robert M. Naclerio, and Jayant M. Pinto. 2014. "Seasonal Allergic Rhinitis Affects Sinonasal Microbiota." *American Journal of Rhinology & Allergy* 28 (4): 281–86.

Chow, J. W., and D. M. Shlaes. 1991. "Imipenem Resistance Associated with the Loss of a 40 kDa Outer Membrane Protein in Enterobacter Aerogenes." *The Journal of Antimicrobial Chemotherapy* 28 (4): 499–504.

Clarridge, Jill E., 3rd. 2004. "Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases." *Clinical Microbiology Reviews* 17 (4): 840–62, table of contents.

Coleman, William B., and Gregory J. Tsongalis. 2016. *Diagnostic Molecular Pathology: A Guide to Applied Molecular Testing*. Academic Press.

Collen, Alanna. 2015. 10% Human: How Your Body's Microbes Hold the Key to Health and Happiness. *HarperCollins*.

Collins, A. C., M. Ashenafi, A. A. Saunders, and W. M. Byrnes. 2007. “Cloning and Expression of Streptomycin Inactivating Enzymes APH(6)-Ia and APH(6)-Id.” *Cellular and Molecular Biology* 53 (3): 74–79.

Colomer-Lluch, Marta, Juan Jofre, and Maite Muniesa. 2011. “Antibiotic Resistance Genes in the Bacteriophage DNA Fraction of Environmental Samples.” *PloS One* 6 (3): e17549.

Colque, Claudia A., Andrea G. Albarracín Orio, Pablo E. Tomatis, Gina Dotta, Diego M. Moreno, Laura G. Hedemann, Rachel A. Hickman, et al. 2022. “Longitudinal Evolution of the Pseudomonas-Derived Cephalosporinase (PDC) Structure and Activity in a Cystic Fibrosis Patient Treated with β-Lactams.” *mBio* 13 (5): e0166322.

Connell, Sean R., Dobryan M. Tracz, Knud H. Nierhaus, and Diane E. Taylor. 2003. “Ribosomal Protection Proteins and Their Mechanism of Tetracycline Resistance.” *Antimicrobial Agents and Chemotherapy* 47 (12): 3675–81.

Cornaglia, G., A. Mazzariol, R. Fontana, and G. Satta. 1996. “Diffusion of Carbapenems through the Outer Membrane of Enterobacteriaceae and Correlation of Their Activities with Their Periplasmic Concentrations.” *Microbial Drug Resistance* 2 (2): 273–76.

Crane, Sally L., Josie van Dorst, Grant C. Hose, Catherine K. King, and Belinda C. Ferrari. 2018. “Microfluidic qPCR Enables High Throughput Quantification of Microbial Functional Genes but Requires Strict Curation of Primers.” *Frontiers of Environmental Science & Engineering in China* 6. https://doi.org/10.3389/fenvs.2018.00145.

Cui, Yujun, Chang Yu, Yanfeng Yan, Dongfang Li, Yanjun Li, Thibaut Jombart, Lucy A. Weinert, et al. 2013. “Historical Variations in Mutation Rate in an Epidemic Pathogen, Yersinia Pestis.” *Proceedings of the National Academy of Sciences of the United States of America* 110 (2): 577–82.

Cunha, Burke A., Almudena Burillo, and Emilio Bouza. 2016. “Legionnaires’ Disease.” *The Lancet* 387 (10016): 376–85.

Cunha, Cheston B., and Burke A. Cunha. 2017. *Legionnaire’s Disease, An Issue of Infectious Disease Clinics of North America*. Elsevier Health Sciences.

Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and BCFtools.” *GigaScience* 10 (2). https://doi.org/10.1093/gigascience/giab008.

D’Alessandro, Daniela, and Gaetano Maria Fara. 2017. “Hospital Environments and Epidemiology of Healthcare-Associated Infections.” In Indoor Air Quality in Healthcare Facilities, edited by Stefano Capolongo, Gaetano Settimo, and Marco Gola, 41–52. Cham: *Springer International Publishing*.

Dai, Dongjuan, Connor Brown, Helmut Bürgmann, D. G. Joakim Larsson, Indumathi Nambi, Tong Zhang, Carl-Fredrik Flach, Amy Pruden, and Peter J. Vikesland. 2022. “Long-Read Metagenomic Sequencing Reveals Shifts in Associations of Antibiotic Resistance Genes with Mobile Genetic Elements from Sewage to Activated Sludge.” *Microbiome* 10 (1): 20.

Damas, Marcelo Silva Folhas, Roumayne Lopes Ferreira, Emeline Boni Campanini, Gabriela Guerrera Soares, Leslie Camelo Campos, Pedro Mendes Laprega, Andrea Soares da Costa, et al. 2022. “Whole Genome Sequencing of the Multidrug-Resistant Chryseobacterium Indologenes Isolated from a Patient in Brazil.” *Frontiers of Medicine* 9 (July): 931379.

Darby, Elizabeth M., Eleftheria Trampari, Pauline Siasat, Maria Solsona Gaya, Ilyas Alav, Mark A. Webber, and Jessica M. A. Blair. 2023. “Molecular Mechanisms of Antibiotic Resistance Revisited.” Nature Reviews. *Microbiology* 21 (5): 280–95.

David, Sophia, Massimo Mentasti, Rediat Tewolde, Martin Aslett, Simon R. Harris, Baharak Afshar, Anthony Underwood, Norman K. Fry, Julian Parkhill, and Timothy G. Harrison. 2016. “Evaluation of an Optimal Epidemiological Typing Scheme for Legionella Pneumophila with Whole-Genome Sequence Data Using Validation Guidelines.” *Journal of Clinical Microbiology* 54 (8): 2135–48.

Davies, D. G., M. R. Parsek, J. P. Pearson, B. H. Iglewski, J. W. Costerton, and E. P. Greenberg. 1998. “The Involvement of Cell-to-Cell Signals in the Development of a Bacterial Biofilm.” *Science* 280 (5361): 295–98.

Davison, J. 1999. “Genetic Exchange between Bacteria in the Environment.” *Plasmid* 42 (2): 73–91.

D’Costa, Vanessa M., Christine E. King, Lindsay Kalan, Mariya Morar, Wilson W. L. Sung, Carsten Schwarz, Duane Froese, et al. 2011. “Antibiotic Resistance Is Ancient.” *Nature* 477 (7365): 457–61.

D’Costa, Vanessa M., Katherine M. McGrann, Donald W. Hughes, and Gerard D. Wright. 2006. “Sampling the Antibiotic Resistome.” *Science* 311 (5759): 374–77.

Dé, Indra, Kenneth V. I. Rolston, and Xiang Y. Han. 2004. “Clinical Significance of Roseomonas Species Isolated from Catheter and Blood Samples: Analysis of 36 Cases in Patients with Cancer.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 38 (11): 1579–84.

Deamer, David, Mark Akeson, and Daniel Branton. 2016. “Three Decades of Nanopore Sequencing.” *Nature Biotechnology* 34 (5): 518–24.

Dean, Frank B., John R. Nelson, Theresa L. Giesler, and Roger S. Lasken. 2001. “Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification.” *Genome Research* 11 (6): 1095–99.

Del Barrio-Tofiño, Ester, Carla López-Causapé, Gabriel Cabot, Alba Rivera, Natividad Benito, Concepción Segura, María Milagro Montero, et al. 2017. “Genomics and Susceptibility Profiles of Extensively Drug-Resistant Pseudomonas Aeruginosa Isolates from Spain.” *Antimicrobial Agents and Chemotherapy* 61 (11). https://doi.org/10.1128/AAC.01589-17.

Del Barrio-Tofiño, Ester, Carla López-Causapé, and Antonio Oliver. 2020. “Pseudomonas Aeruginosa Epidemic High-Risk Clones and Their Association with Horizontally-Acquired β-Lactamases: 2020 Update.” *International Journal of Antimicrobial Agents* 56 (6): 106196.

Đermić, Damir, Sven Ljubić, Maja Matulić, Alfredo Procino, Maria Chiara Feliciello, Đurđica Ugarković, and Isidoro Feliciello. 2023. “Reverse Transcription-Quantitative PCR (RT-qPCR) without the Need for Prior Removal of DNA.” *Scientific Reports* 13 (1): 11470.

Diard, Médéric, and Wolf-Dietrich Hardt. 2017. “Evolution of Bacterial Virulence.” *FEMS Microbiology Reviews* 41 (5): 679–97.

diCenzo, George C., and Turlough M. Finan. 2017. “The Divided Bacterial Genome: Structure, Function, and Evolution.” *Microbiology and Molecular Biology Reviews: MMBR* 81 (3). https://doi.org/10.1128/MMBR.00019-17.

Didelot, Xavier, Rory Bowden, Daniel J. Wilson, Tim E. A. Peto, and Derrick W. Crook. 2012. “Transforming Clinical Microbiology with Bacterial Genome Sequencing.” *Nature Reviews. Genetics* 13 (9): 601–12.

Dineen, S. M., R. Aranda 4th, D. L. Anders, and J. M. Robertson. 2010. “An Evaluation of Commercial DNA Extraction Kits for the Isolation of Bacterial Spore DNA from Soil.” *Journal of Applied Microbiology* 109 (6): 1886–96.

Ding, Huijun, Min Qiao, Jiayou Zhong, Yongguan Zhu, Chunjing Guo, Qianqian Zhang, Ping Yang, et al. 2020. “Characterization of Antibiotic Resistance Genes and Bacterial Community in Selected Municipal and Industrial Sewage Treatment Plants beside Poyang Lake.” *Water Research* 174 (May): 115603.

Dixon, Philip. 2003. “VEGAN, a Package of R Functions for Community Ecology.” *Journal of Vegetation Science: Official Organ of the International Association for Vegetation Science* 14 (6): 927–30.

Drmanac, R., I. Labat, I. Brukner, and R. Crkvenjakov. 1989. “Sequencing of Megabase plus DNA by Hybridization: Theory of the Method.” *Genomics* 4 (2): 114–28.

Drmanac, Radoje, Snezana Drmanac, Gloria Chui, Robert Diaz, Aaron Hou, Hui Jin, Paul Jin, et al. 2002. “Sequencing by Hybridization (SBH): Advantages, Achievements, and Opportunities.” *Advances in Biochemical Engineering/biotechnology* 77: 75–101.

Dworkin. 2010. Outbreak Investigations Around the World: Case Studies in Infectious Disease Field Epidemiology. *Jones & Bartlett Publishers*.

Edgar, Robert C. 2010. “Search and Clustering Orders of Magnitude Faster than BLAST.” *Bioinformatics* 26 (19): 2460–61.

Ehrlich, Paul, and Sahachiro Hata. 1911. The Experimental Chemotherapy of Spirilloses: (syphilis, Relapsing Fever, Spirillosis of Fowls, Framboesia). *Rebman*.

El Amin, Nagwa, Christian G. Giske, Shah Jalal, Berit Keijser, Göran Kronvall, and Bengt Wretlind. 2005. “Carbapenem Resistance Mechanisms in Pseudomonas Aeruginosa: Alterations of Porin OprD and Efflux Proteins Do Not Fully Explain Resistance Patterns Observed in Clinical Isolates.” *APMIS: Acta Pathologica, Microbiologica, et Immunologica Scandinavica* 113 (3): 187–96.

Eliopoulos, George M., and Marilyn C. Roberts. 2003. “Tetracycline Therapy: Update.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 36 (4): 462–67.

ElRakaiby, Marwa T., Soha Gamal-Eldin, Magdy A. Amin, and Ramy K. Aziz. 2019. “Hospital Microbiome Variations As Analyzed by High-Throughput Sequencing.” *Omics: A Journal of Integrative Biology* 23 (9): 426–38.

Enany, Shymaa, and Laura Crotty Alexander. 2017. *The Rise of Virulence and Antibiotic Resistance in Staphylococcus Aureus*. BoD – Books on Demand.

Escudero, José Antonio, Céline Loot, Aleksandra Nivina, and Didier Mazel. 2015. “The Integron: Adaptation On Demand.” *Microbiology Spectrum* 3 (2): MDNA3–0019 – 2014.

European Centre for Disease Prevention and Control. Legionnaires’ disease. In: ECDC. Annual Epidemiological Report for 2021. Stockholm: ECDC; 2023.

Evans, Daniel R., Marissa P. Griffith, Alexander J. Sundermann, Kathleen A. Shutt, Melissa I. Saul, Mustapha M. Mustapha, Jane W. Marsh, Vaughn S. Cooper, Lee H. Harrison, and Daria Van Tyne. 2020. “Systematic Detection of Horizontal Gene Transfer across Genera among Multidrug-Resistant Bacteria in a Single Hospital.” *eLife* 9 (April). https://doi.org/10.7554/eLife.53886.

Everett, M. J., Y. F. Jin, V. Ricci, and L. J. Piddock. 1996. “Contributions of Individual Mechanisms to Fluoroquinolone Resistance in 36 Escherichia Coli Strains Isolated from Humans and Animals.” *Antimicrobial Agents and Chemotherapy* 40 (10): 2380–86.

Faino, Luigi, Valeria Scala, Alessio Albanese, Vanessa Modesti, Alessandro Grottoli, Nicoletta Pucci, Andrea Doddi, et al. 2021. “Nanopore Sequencing for the Detection and Identification of Xylella

Fastidiosa Subspecies and Sequence Types from Naturally Infected Plant Material." *Plant Pathology* 70 (8): 1860–70

Fang, Yicheng, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang, and Wenbin Ji. 2020. "Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR." *Radiology* 296 (2): E115–17.

Fan, Yong, and Oluf Pedersen. 2021. "Gut Microbiota in Human Metabolic Health and Disease." *Nature Reviews. Microbiology* 19 (1): 55–71.

Farahi, Rosetta Moshirian, Ahya Abdi Ali, and Sara Gharavi. 2018. "Characterization of gyrA and parC Mutations in Ciprofloxacin-Resistant Pseudomonas Aeruginosa Isolates from Tehran Hospitals in Iran." *Iranian Journal of Microbiology* 10 (4): 242–49.

Farghaly, Mai, Michael F. Hynes, Mohammad Nazari, Sylvia Checkley, and Karen Liljebjelke. 2023. "Examination of the Horizontal Gene Transfer Dynamics of an Integrative and Conjugative Element Encoding Multidrug Resistance in Histophilus Somni." *Canadian Journal of Microbiology* 69 (3): 123–35.

Faria, N. R., M. U. G. Kraemer, S. C. Hill, J. Goes de Jesus, R. S. Aguiar, F. C. M. Iani, J. Xavier, et al. 2018. "Genomic and Epidemiological Monitoring of Yellow Fever Virus Transmission Potential." *Science* 361 (6405): 894–99.

Faria, N. R., J. Quick, I. M. Claro, J. Thézé, J. G. de Jesus, M. Giovanetti, M. U. G. Kraemer, et al. 2017. "Establishment and Cryptic Transmission of Zika Virus in Brazil and the Americas." *Nature* 546 (7658): 406–10.

Ferretti, Joseph, and Werner Köhler. 2016. "History of Streptococcal Research." In *Streptococcus Pyogenes: Basic Biology to Clinical Manifestations*, edited by Joseph J. Ferretti, Dennis L. Stevens, and Vincent A. Fischetti. Oklahoma City (OK): University of Oklahoma Health Sciences Center.

Fitch, W. M. 1966. “An Improved Method of Testing for Evolutionary Homology.” *Journal of Molecular Biology* 16 (1): 9–16.

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, and J. M. Merrick. 1995. “Whole-Genome Random Sequencing and Assembly of Haemophilus Influenzae Rd.” *Science* 269 (5223): 496–512.

Floridia-Yapur, Noelia, Fanny Rusman, Patricio Diosque, and Nicolás Tomasini. 2021. “Genome Data vs MLST for Exploring Intraspecific Evolutionary History in Bacteria: Much Is Not Always Better.” *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 93 (September): 104990.

Food and Agriculture Organization of the United Nations. 2013. *State of Food and Agriculture (SOFA) 2012: Investing in Agriculture for a Better Future (Chinese Edition)*. FAO.

Food and Agriculture Organization of the United Nations, World Health Organization, World Organisation for Animal Health, and United Nations Environment Programme. 2023. *A One Health Priority Research Agenda for Antimicrobial Resistance*. Food & Agriculture Org.

Fouad, Ashraf F., Jody Barry, Melissa Caimano, Michael Clawson, Qiang Zhu, Rachaele Carver, Karsten Hazlett, and Justin D. Radolf. 2002. “PCR-Based Identification of Bacteria Associated with Endodontic Infections.” *Journal of Clinical Microbiology* 40 (9): 3223–31.

Francisco, Alexandre P., Miguel Bugalho, Mário Ramirez, and João A. Carriço. 2009. “Global Optimal eBURST Analysis of Multilocus Typing Data Using a Graphic Matroid Approach.” *BMC Bioinformatics* 10 (May): 152.

Franco, Lauren C., Windy Tanner, Christine Ganim, Terri Davy, Jonathan Edwards, and Rodney Donlan. 2020. “A Microbiological Survey of Handwashing Sinks in the Hospital Built Environment Reveals Differences in Patient Room and Healthcare Personnel Sinks.” *Scientific Reports* 10 (1): 8234.

Franco-Duarte, Ricardo, Lucia Černáková, Snehal Kadam, Karishma S. Kaushik, Bahare Salehi, Antonio Bevilacqua, Maria Rosaria Corbo, et al. 2019. “Advances in Chemical and Biological Methods to Identify Microorganisms—From Past to Present.” *Microorganisms*. https://doi.org/10.3390/microorganisms7050130.

Fredslund, Jakob, and Mette Lange. 2007. “Primique: Automatic Design of Specific PCR Primers for Each Sequence in a Family.” *BMC Bioinformatics* 8 (October): 369.

Frost, Laura S., Raphael Leplae, Anne O. Summers, and Ariane Toussaint. 2005. “Mobile Genetic Elements: The Agents of Open Source Evolution.” *Nature Reviews. Microbiology* 3 (9): 722–32.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012. “CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data.” *Bioinformatics* 28 (23): 3150–52.

Fuks, Garold, Michael Elgart, Amnon Amir, Amit Zeisel, Peter J. Turnbaugh, Yoav Soen, and Noam Shental. 2018. “Combining 16S rRNA Gene Variable Regions Enables High-Resolution Microbial Community Profiling.” *Microbiome* 6 (1): 17.

Gadberry, Michael D., Simon T. Malcomber, Andrew N. Doust, and Elizabeth A. Kellogg. 2005. “Primaclade--a Flexible Tool to Find Conserved PCR Primers across Multiple Species.” *Bioinformatics* 21 (7): 1263–64.

Gaia, Valeria, Norman K. Fry, Baharak Afshar, P. Christian Lück, Hélène Meugnier, Jerome Etienne, Raffaele Peduzzi, and Timothy G. Harrison. 2005. “Consensus Sequence-Based Scheme for

Epidemiological Typing of Clinical and Environmental Isolates of Legionella Pneumophila." *Journal of Clinical Microbiology* 43 (5): 2047–52.

Gaia, Valeria, Norman K. Fry, Timothy G. Harrison, and Raffaele Peduzzi. 2003. "Sequence-Based Typing of Legionella Pneumophila Serogroup 1 Offers the Potential for True Portability in Legionellosis Outbreak Investigation." *Journal of Clinical Microbiology* 41 (7): 2932–39.

Gatzeva-Topalova, Petia Z., Andrew P. May, and Marcelo C. Sousa. 2005. "Structure and Mechanism of ArnA: Conformational Change Implies Ordered Dehydrogenase Mechanism in Key Enzyme for Polymyxin Resistance." *Structure* 13 (6): 929–42.

Gea-Izquierdo, Enrique, Ruth Gil-Prieto, Valentín Hernández-Barrera, Gil Rodríguez-Caravaca, and Ángel Gil-de-Miguel. 2023. "Legionellosis-Associated Hospitalization in Spain from 2002 to 2021." *Microorganisms* 11 (7). https://doi.org/10.3390/microorganisms11071693.

Genthe, Nicholas A., James B. Thoden, and Hazel M. Holden. 2016. "Structure of the Escherichia Coli ArnA N-Formyltransferase Domain in Complex with N(5) -Formyltetrahydrofolate and UDP-Ara4N." Protein Science: *A Publication of the Protein Society* 25 (8): 1555–62.

Genuer, Robin, and Jean-Michel Poggi. 2020. Random Forests with R. Springer Nature.

Gerba, Charles P. 2009. "Environmentally Transmitted Pathogens." *Environmental Microbiology*. https://doi.org/10.1016/b978-0-12-370519-8.00022-5.

Gest, Howard. 2004. "The Discovery of Microorganisms by Robert Hooke and Antoni Van Leeuwenhoek, Fellows of the Royal Society." *Notes and Records of the Royal Society of London* 58 (2): 187–201.

Gibson, Claire, Susanne A. Kraemer, Natalia Klimova, Bing Guo, and Dominic Frigon. 2023. “Antibiotic Resistance Gene Sequencing Is Necessary to Reveal the Complex Dynamics of Immigration from Sewers to Activated Sludge.” *Frontiers in Microbiology* 14 (April): 1155956.

Gill M. J., Simjee S., Al-Hattawi K., Robertson B. D., Easmon C. S. F., and Ison C. A. 1998. “Gonococcal Resistance to β-Lactams and Tetracycline Involves Mutation in Loop 3 of the Porin Encoded at thepenB Locus.” *Antimicrobial Agents and Chemotherapy* 42 (11): 2799–2803.

Golshani, Z., and A. Sharifzadeh. 2013. “Prevalence of blaOxa10 Type Beta-Lactamase Gene in Carbapenemase Producing Pseudomonas Aeruginosa Strains Isolated from Patients in Isfahan.” *Jundishapur Journal of Microbiology*. https://brieflands.com/articles/jjm-56346.html.

Gomez-Valero, Laura, and Carmen Buchrieser. 2013. “Genome Dynamics in Legionella: The Basis of Versatility and Adaptation to Intracellular Replication.” *Cold Spring Harbor Perspectives in Medicine* 3 (6). https://doi.org/10.1101/cshperspect.a009993.

Gong, Liang, Chee-Hong Wong, Jennifer Idol, Chew Yee Ngan, and Chia-Lin Wei. 2019. “Ultra-Long Read Sequencing for Whole Genomic DNA Analysis.” *Journal of Visualised Experiments*: JoVE, no. 145 (March). https://doi.org/10.3791/58954.

Gong, Liang, Chee-Hong Wong, Wei-Chung Cheng, Harianto Tjong, Francesca Menghi, Chew Yee Ngan, Edison T. Liu, and Chia-Lin Wei. 2018. “Picky Comprehensively Detects High-Resolution Structural Variants in Nanopore Long Reads.” *Nature Methods* 15 (6): 455–60.

González-Pastor, José E., Errett C. Hobbs, and Richard Losick. 2003. “Cannibalism by Sporulating Bacteria.” *Science* 301 (5632): 510–13.

Gopinath, Subash C. B., Thean-Hock Tang, Yeng Chen, Marimuthu Citartan, and Thangavel Lakshmipriya. 2014. “Bacterial Detection: From Microscope to Smartphone.” *Biosensors & Bioelectronics* 60 (October): 332–42.

Grote, Alexandra, and Ashlee M. Earl. 2022. “Within-Host Evolution of Bacterial Pathogens during Persistent Infection of Humans.” *Current Opinion in Microbiology* 70 (December): 102197.

Grubaugh, Nathan D., Jason T. Ladner, Moritz U. G. Kraemer, Gytis Dudas, Amanda L. Tan, Karthik Gangavarapu, Michael R. Wiley, et al. 2017. “Genomic Epidemiology Reveals Multiple Introductions of Zika Virus into the United States.” *Nature* 546 (7658): 401–5.

Grubaugh, Nathan D., Jason T. Ladner, Philippe Lemey, Oliver G. Pybus, Andrew Rambaut, Edward C. Holmes, and Kristian G. Andersen. 2019. “Tracking Virus Outbreaks in the Twenty-First Century.” *Nature Microbiology* 4 (1): 10–19.

Guo, Feng, and Tong Zhang. 2013. “Biases during DNA Extraction of Activated Sludge Samples Revealed by High Throughput Sequencing.” *Applied Microbiology and Biotechnology* 97 (10): 4607–16.

Guo, Jingyu, Ting Liang, Chaohui Hu, Ruichen Lv, Xianwei Yang, Yujun Cui, Youtao Song, Ruifu Yang, Qingyi Zhu, and Yajun Song. 2015. “Sequence Types Diversity of Legionella Pneumophila Isolates from Environmental Water Sources in Guangzhou and Jiangmen, China.” *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 29 (January): 35–41.

Guo, Y. M., J. K. Zhou, H. X. Zhang, L. Y. Qin, D. Zhao, J. U. Hui-Dong, and Others. 2014. “Comparison among SBT, PFGE, AFLP Molecular Typing Methods of Legionella Pneumophila.” *Journal of Microbiology* 34 (1): 72–77.

Hammerum, Anette M., Kasper Thystrup Karstensen, Louise Roer, Hülya Kaya, Mikkel Lindegaard, Lone Jannok Porsbo, Anne Kjerulf, et al. 2024. “Surveillance of Vancomycin-Resistant Enterococci

Reveals Shift in Dominating Clusters from vanA to vanB Enterococcus Faecium Clusters, Denmark, 2015 to 2022." *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 29 (23). https://doi.org/10.2807/1560-7917.ES.2024.29.23.2300633.

Hammond, Maria, Felix Homa, Helene Andersson-Svahn, Thijs J. G. Ettema, and Haakan N. Joensson. 2016. "Picodroplet Partitioned Whole Genome Amplification of Low Biomass Samples Preserves Genomic Diversity for Metagenomic Analysis." *Microbiome* 4 (1): 52.

Hao, Qinjian, Hongmei Wu, and Qiang Wang. 2020. "Difficulties in False Negative Diagnosis of Coronavirus Disease 2019: A Case Report." https://www.researchsquare.com/article/rs-17319/latest.pdf.

Halstead, F. D., J. Quick, M. Niebel, M. Garvey, N. Cumley, R. Smith, T. Neal, et al. 2021. "Pseudomonas Aeruginosa Infection in Augmented Care: The Molecular Ecology and Transmission Dynamics in Four Large UK Hospitals." *The Journal of Hospital Infection* 111 (May): 162–68.

Harrison, T. G., B. Afshar, N. Doshi, N. K. Fry, and J. V. Lee. 2009. "Distribution of Legionella Pneumophila Serogroups, Monoclonal Antibody Subgroups and DNA Sequence Types in Recent Clinical and Environmental Isolates from England and Wales (2000-2008)." *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 28 (7): 781–91.

Hartmann, Erica M., Roxana Hickey, Tiffany Hsu, Clarisse M. Betancourt Román, Jing Chen, Randall Schwager, Jeff Kline, et al. 2016. "Antimicrobial Chemicals Are Associated with Elevated Antibiotic Resistance Genes in the Indoor Dust Microbiome." *Environmental Science & Technology* 50 (18): 9807–15.

Hasan, Mohammad Rubayet, Faheem Mirza, Hamad Al-Hail, Sathyavathi Sundararaju, Thabisile Xaba, Muhammad Iqbal, Hashim Alhussain, Hadi Mohamad Yassine, Andres Perez-Lopez, and Patrick Tang.

2020. “Correction: Detection of SARS-CoV-2 RNA by Direct RT-qPCR on Nasopharyngeal Specimens without Extraction of Viral RNA.” *PloS One* 15 (10): e0240343.

Harding, Christian M., Seth W. Hennon, and Mario F. Feldman. 2018. “Uncovering the Mechanisms of Acinetobacter Baumannii Virulence.” *Nature Reviews. Microbiology* 16 (2): 91–102.

Hashimoto, Mayuko, Haruka Hasegawa, and Sumio Maeda. 2019. “High Temperatures Promote Cell-to-Cell Plasmid Transformation in Escherichia Coli.” *Biochemical and Biophysical Research Communications* 515 (1): 196–200.

Hawkey, Peter M. 2003. “Mechanisms of Quinolone Action and Microbial Response.” *The Journal of Antimicrobial Chemotherapy* 51 Suppl 1 (May): 29–35.

Hawkey, P. M. 1998. “The Origins and Molecular Basis of Antibiotic Resistance.” *BMJ* 317 (7159): 657–60.

He, Gui-Xin, Teruo Kuroda, Takehiko Mima, Yuji Morita, Tohru Mizushima, and Tomofusa Tsuchiya. 2004. “An H(+)-Coupled Multidrug Efflux Pump, PmpM, a Member of the MATE Family of Transporters, from Pseudomonas Aeruginosa.” *Journal of Bacteriology* 186 (1): 262–65.

He, Fei, Wanjun Zhou, Ren Cai, Tizhen Yan, and Xiangmin Xu. 2018. “Systematic Assessment of the Performance of Whole-Genome Amplification for SNP/CNV Detection and β-Thalassemia Genotyping.” *Journal of Human Genetics* 63 (4): 407–16.

Health Organization, World. n.d. “Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. Geneva: WHO; 2017.” *Google Scholar There Is No Corresponding Record for*.

Helekal, David, Matt Keeling, Yonatan H. Grad, and Xavier Didelot. 2023. “Estimating the Fitness Cost and Benefit of Antimicrobial Resistance from Pathogen Genomic Data.” *Journal of the Royal Society, Interface / the Royal Society* 20 (203): 20230074.

Hembach, Norman, Gabriele Bierbaum, Christiane Schreiber, and Thomas Schwartz. 2022. “Facultative Pathogenic Bacteria and Antibiotic Resistance Genes in Swine Livestock Manure and Clinical Wastewater: A Molecular Biology Comparison.” *Environmental Pollution* 313 (November): 120128.

Hepp, Crystal M., Jill Hager Cocking, Michael Valentine, Steven J. Young, Dan Damian, Kimberly E. Samuels-Crow, Krystal Sheridan, et al. 2018. “Phylogenetic Analysis of West Nile Virus in Maricopa County, Arizona: Evidence for Dynamic Behavior of Strains in Two Major Lineages in the American Southwest.” *PloS One* 13 (11): e0205801.

Higgins, P. G., A. C. Fluit, D. Milatovic, J. Verhoef, and F-J Schmitz. 2003. “Mutations in GyrA, ParC, MexR and NfxB in Clinical Isolates of Pseudomonas Aeruginosa.” *International Journal of Antimicrobial Agents* 21 (5): 409–13.

Hill, Sarah C., Jocelyne Neto de Vasconcelos, Bernardo Gutierrez Granja, Julien Thézé, Domingos Jandondo, Zoraima Neto, Marinela Mirandela, et al. 2019. “Early Genomic Detection of Cosmopolitan Genotype of Dengue Virus Serotype 2, Angola, 2018.” *Emerging Infectious Diseases* 25 (4): 784–87.

Holden, Matthew T. G., Edward J. Feil, Jodi A. Lindsay, Sharon J. Peacock, Nicholas P. J. Day, Mark C. Enright, Tim J. Foster, et al. 2004. “Complete Genomes of Two Clinical Staphylococcus Aureus Strains: Evidence for the Rapid Evolution of Virulence and Drug Resistance.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (26): 9786–91.

Holland, P. M., R. D. Abramson, R. Watson, and D. H. Gelfand. 1991. “Detection of Specific Polymerase Chain Reaction Product by Utilising the 5’----3' Exonuclease Activity of Thermus Aquaticus DNA

Polymerase." *Proceedings of the National Academy of Sciences of the United States of America* 88 (16): 7276–80.

Holmes, Edward C., Gytis Dudas, Andrew Rambaut, and Kristian G. Andersen. 2016. "The Evolution of Ebola Virus: Insights from the 2013-2016 Epidemic." *Nature* 538 (7624): 193–200.

Hong, Yaming, Ruoyun Ji, Zhenyu Wang, Jiaojie Gu, Xinan Jiao, and Qiuchun Li. 2023. "Development and Application of a Multiplex PCR Method to Differentiate Salmonella Enterica Serovar Typhimurium from Its Monophasic Variants in Pig Farms." *Food Microbiology* 109 (February): 104135.

Hooke, R. 1845. "Milestones." *Phil. Trans. R. Soc. Lond*. http://light.ece.illinois.edu/ECE564/_OK_Lectures/00_Timeline_milestones.pdf.

Horcajada JP, Montero M, Oliver A, Sorlí L, Luque S, Gómez-Zorrilla S, Benito N, Grau S. 2019. Epidemiology and treatment of multidrug-resistant and extensively drug-resistant Pseudomonas aeruginosa infections. *Clin Microbiol Rev* 32:e00031-19. https://doi.org/10.1128/CMR.00031-19.

Hospodsky, Denina, Jing Qian, William W. Nazaroff, Naomichi Yamamoto, Kyle Bibby, Hamid Rismani-Yazdi, and Jordan Peccia. 2012. "Human Occupancy as a Source of Indoor Airborne Bacteria." *PloS One* 7 (4): e34867.

Hota, Bala. 2004. "Contamination, Disinfection, and Cross-Colonization: Are Hospital Surfaces Reservoirs for Nosocomial Infection?" *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 39 (8): 1182–89.

Houldcroft, Charlotte J., Mathew A. Beale, and Judith Breuer. 2017. "Clinical and Biological Insights from Viral Genome Sequencing." *Nature Reviews. Microbiology* 15 (3): 183–92.

Huang, Yao-Ting, Po-Yu Liu, and Pei-Wen Shih. 2021. “Homopolish: A Method for the Removal of Systematic Errors in Nanopore Sequencing by Homologous Polishing.” *Genome Biology* 22 (1): 95.

Huber, François, Hans Peter Lang, Daniela Lang, Daniel Wüthrich, Vladimira Hinić, Christoph Gerber, Adrian Egli, and Ernst Meyer. 2021. “Rapid and Ultrasensitive Detection of Mutations and Genes Relevant to Antimicrobial Resistance in Bacteria.” *Global Challenges (Hoboken, NJ)* 5 (2): 2000066.

Huggins, Lucas G., Vito Colella, Ushani Atapattu, Anson V. Koehler, and Rebecca J. Traub. 2022. “Nanopore Sequencing Using the Full-Length 16S rRNA Gene for Detection of Blood-Borne Bacteria in Dogs Reveals a Novel Species of Hemotropic Mycoplasma.” *Microbiology Spectrum* 10 (6): e0308822.

Humphries, Romney M. 2020. “Update on Susceptibility Testing: Genotypic and Phenotypic Methods.” *Clinics in Laboratory Medicine* 40 (4): 433–46.

Huovinen, P., L. Sundström, G. Swedberg, and O. Sköld. 1995. “Trimethoprim and Sulfonamide Resistance.” *Antimicrobial Agents and Chemotherapy* 39 (2): 279–89.

Huson, Daniel H., Alexander F. Auch, Ji Qi, and Stephan C. Schuster. 2007. “MEGAN Analysis of Metagenomic Data.” *Genome Research* 17 (3): 377–86.

Hyatt, Doug, Gwo-Liang Chen, Philip F. Locascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. 2010. “Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification.” *BMC Bioinformatics* 11 (March): 119.

Ince, Dilek, Xiamei Zhang, L. Christine Silver, and David C. Hooper. 2002. “Dual Targeting of DNA Gyrase and Topoisomerase IV: Target Interactions of Garenoxacin (BMS-284756, T-3811ME), a New Desfluoroquinolone.” *Antimicrobial Agents and Chemotherapy* 46 (11): 3370–80.

Ishii, Satoshi, Takahiro Segawa, and Satoshi Okabe. 2013. “Simultaneous Quantification of Multiple Food- and Waterborne Pathogens by Use of Microfluidic Quantitative PCR.” *Applied and Environmental Microbiology* 79 (9): 2891–98.

Isidro, Joana, Vítor Borges, Miguel Pinto, Daniel Sobral, João Dourado Santos, Alexandra Nunes, Verónica Mixão, et al. 2022. “Phylogenomic Characterization and Signs of Microevolution in the 2022 Multi-Country Outbreak of Monkeypox Virus.” *Nature Medicine* 28 (8): 1569–72.

Jaillard, Magali, Alex van Belkum, Kyle C. Cady, David Creely, Dee Shortridge, Bernadette Blanc, E. Magda Barbu, et al. 2017. “Correlation between Phenotypic Antibiotic Susceptibility and the Resistome in Pseudomonas Aeruginosa.” *International Journal of Antimicrobial Agents* 50 (2): 210–18.

Jain, Miten, Sergey Koren, Karen H. Miga, Josh Quick, Arthur C. Rand, Thomas A. Sasani, John R. Tyson, et al. 2018. “Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads.” *Nature Biotechnology* 36 (4): 338–45.

James, Grace Lymie, Mohd Talib Latif, Mohd Noor Mat Isa, Mohd Faizal Abu Bakar, Nurul Yuziana Mohd Yusuf, William Broughton, Abdul Munir Murad, and Farah Diba Abu Bakar. 2021. “Metagenomic Datasets of Air Samples Collected during Episodes of Severe Smoke-Haze in Malaysia.” *Data in Brief* 36 (June): 107124.

Jeanty, Céline, Dan Longrois, Paul-Michel Mertes, Daniel R. Wagner, and Yvan Devaux. 2010. “An Optimised Protocol for Microarray Validation by Quantitative PCR Using Amplified Amino Allyl Labeled RNA.” *BMC Genomics* 11 (October): 542.

Jensen, Carsten Strøby, Søren Beck Nielsen, and Lars Fynbo. 2018. Risking Antimicrobial Resistance: A Collection of One-Health Studies of Antibiotics and Its Social and Health Consequences. *Springer*.

Jensen, E. C., H. S. Schrader, B. Rieland, T. L. Thompson, K. W. Lee, K. W. Nickerson, and T. A. Kokjohn. 1998. “Prevalence of Broad-Host-Range Lytic Bacteriophages of Sphaerotilus Natans, Escherichia Coli, and Pseudomonas Aeruginosa.” *Applied and Environmental Microbiology* 64 (2): 575–80.

Jia, Hong-Ling, Peng Li, Hong-Jie Liu, Jia-Yong Zhong, Peng-Zhe Qin, Wen-Zhe Su, Ying-Feng Zheng, et al. 2021. “Genomic Elucidation of a COVID-19 Resurgence and Local Transmission of SARS-CoV-2 in Guangzhou, China.” *Journal of Clinical Microbiology* 59 (8): e0007921.

Jiang, Luxi, Sihong Zhao, Xu Cai, Deguang Mu, Xianghua Zhang, Jian Kang, Li Zhao, and Yu Chen. 2021. “Sequence-Based Typing of Clinical and Environmental Legionella Pneumophila Isolates in Shenyang, China.” *Enfermedades Infecciosas Y Microbiologia Clinica* 39 (8): 383–89.

Jiang, S. C., and J. H. Paul. 1998. “Gene Transfer by Transduction in the Marine Environment.” *Applied and Environmental Microbiology* 64 (8): 2780–87.

Jiao, Xiaoyang, Wenyan Guo, Xin Li, Fen Yao, Mi Zeng, Yumeng Yuan, Xiaoling Guo, et al. 2023. “New Insight into the Microbiome, Resistome, and Mobilome on the Dental Waste Water in the Context of Heavy Metal Environment.” *Frontiers in Microbiology* 14 (April): 1106157.

Johnsborg, Ola, and Leiv Sigve Håvarstein. 2009. “Regulation of Natural Genetic Transformation and Acquisition of Transforming DNA in Streptococcus Pneumoniae.” *FEMS Microbiology Reviews* 33 (3): 627–42.

Johnson, Jethro S., Daniel J. Spakowicz, Bo-Young Hong, Lauren M. Petersen, Patrick Demkowicz, Lei Chen, Shana R. Leopold, et al. 2019. “Evaluation of 16S rRNA Gene Sequencing for Species and Strain-Level Microbiome Analysis.” *Nature Communications* 10 (1): 5029.

Johnson, L. Steven, Sean R. Eddy, and Elon Portugaly. 2010. “Hidden Markov Model Speed Heuristic and Iterative HMM Search Procedure.” *BMC Bioinformatics* 11 (August): 431.

Johnson, Mark, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, and Thomas L. Madden. 2008. “NCBI BLAST: A Better Web Interface.” *Nucleic Acids Research* 36 (Web Server issue): W5–9.

Jolley, Keith A., and Martin C. J. Maiden. 2010. “BIGSdb: Scalable Analysis of Bacterial Genome Variation at the Population Level.” *BMC Bioinformatics* 11 (December): 595.

Jolley, Keith A., James E. Bray, and Martin C. J. Maiden. 2018. “Open-Access Bacterial Population Genomics: BIGSdb Software, the PubMLST.org Website and Their Applications.” *Wellcome Open Research 3* (September): 124.

Jong, Mui-Choo, Jian-Qiang Su, Joshua T. Bunce, Colin R. Harwood, Jason R. Snape, Yong-Guan Zhu, and David W. Graham. 2018. “Co-Optimisation of Sponge-Core Bioreactors for Removing Total Nitrogen and Antibiotic Resistance Genes from Domestic Wastewater.” *The Science of the Total Environment* 634 (September): 1417–23.

Juhas, Mario. 2015. “Horizontal Gene Transfer in Human Pathogens.” *Critical Reviews in Microbiology* 41 (1): 101–8.

Kafetzopoulou, L. E., S. T. Pullan, P. Lemey, M. A. Suchard, D. U. Ehichioya, M. Pahlmann, A. Thielebein, et al. 2019. “Metagenomic Sequencing at the Epicenter of the Nigeria 2018 Lassa Fever Outbreak.” *Science* 363 (6422): 74–77.

Kameoka, Shoichiro, Daisuke Motooka, Satoshi Watanabe, Ryuichi Kubo, Nicolas Jung, Yuki Midorikawa, Natsuko O. Shinozaki, Yu Sawai, Aya K. Takeda, and Shota Nakamura. 2021. “Benchmark of 16S rRNA Gene Amplicon Sequencing Using Japanese Gut Microbiome Data from the V1–V2 and V3–V4 Primer Sets.” *BMC Genomics* 22 (1): 1–10.

Kao, Cheng-Yen, Shu-Sheng Chen, Kuei-Hsiang Hung, Hsiu-Mei Wu, Po-Ren Hsueh, Jing-Jou Yan, and Jiunn-Jong Wu. 2016. “Overproduction of Active Efflux Pump and Variations of OprD Dominate in Imipenem-Resistant Pseudomonas Aeruginosa Isolated from Patients with Bloodstream Infections in Taiwan.” *BMC Microbiology* 16 (1): 107.

Kao, M. C. 1998. Time to Degree: Rapid Detection of Escherichia Coli Using Nucleic Acid Sequence-Based Amplification and Oligonucleotide Probe for 16S Ribosomal RNA. Cornell University.

Katoh, Kazutaka, Kazuharu Misawa, Kei-Ichi Kuma, and Takashi Miyata. 2002. “MAFFT: A Novel Method for Rapid Multiple Sequence Alignment Based on Fast Fourier Transform.” *Nucleic Acids Research* 30 (14): 3059–66.

Kelsey, Michael. 2013. “Pseudomonas in Augmented Care: Should We Worry?” The Journal of *Antimicrobial Chemotherapy* 68 (12): 2697–2700.

Kentaro, Itokawa, S. Tsuyoshi, H. Masanori, T. Rina, and K. Makoto. 2020. “Aproposal Ofan Alternative Primer fortheARTIC Network’s Multiplex PCRto Improve Coverageof SARS-CoV-2 Genomesequencing.” *bioRxiv*.

Keskin, Emre. 2014. “Detection of Invasive Freshwater Fish Species Using Environmental DNA Survey.” *Biochemical Systematics and Ecology* 56 (October): 68–74.

Kim, Chi Hyun, Hee Young Kang, Bo Ra Kim, Hyejin Jeon, Yoo Chul Lee, Sang Hwa Lee, and Je Chul Lee. 2016. “Mutational Inactivation of OprD in Carbapenem-Resistant Pseudomonas Aeruginosa Isolates from Korean Hospitals.” *Journal of Microbiology* 54 (1): 44–49.

Kim, Sollip, Ki-Won Eom, Chong-Rae Cho, and Tae Hyun Um. 2016. “Comparison of the Ability of Multiplex and Singleplex PCR to Detect Human Respiratory Viruses.” *Laboratory Medicine Online* 6 (4): 240–45.

Kimura, Soichiro, Kazuhiro Tateda, Yoshikazu Ishii, Manabu Horikawa, Shinichi Miyairi, Naomasa Gotoh, Masaji Ishiguro, and Keizo Yamaguchi. 2009. “Pseudomonas Aeruginosa Las Quorum Sensing Autoinducer Suppresses Growth and Biofilm Production in Legionella Species.” *Microbiology* 155 (Pt 6): 1934–39.

Kim, Won-Keun, Jin Sun No, Seung-Ho Lee, Dong Hyun Song, Daesang Lee, Jeong-Ah Kim, Se Hun Gu, et al. 2018. “Multiplex PCR−Based Next-Generation Sequencing and Global Diversity of Seoul Virus in Humans and Rats.” *Emerging Infectious Diseases*. https://doi.org/10.3201/eid2402.171216.

Klassert, Tilman E., Rasmus Leistner, Cristina Zubiria-Barrera, Magdalena Stock, Mercedes López, Robert Neubert, Dominik Driesch, Petra Gastmeier, and Hortense Slevogt. 2021. “Bacterial Colonization Dynamics and Antibiotic Resistance Gene Dissemination in the Hospital Environment after First Patient Occupancy: A Longitudinal Metagenetic Study.” *Microbiome* 9 (1): 169.

Kleanthous, Colin, and Judith P. Armitage. 2015. “The Bacterial Cell Envelope.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 370 (1679). https://doi.org/10.1098/rstb.2015.0019.

Kleyer, Hannah, Robin Tecon, and Dani Or. 2017. “Resolving Species Level Changes in a Representative Soil Bacterial Community Using Microfluidic Quantitative PCR.” *Frontiers in Microbiology* 8 (October): 2017.

Knights, Dan, Justin Kuczynski, Emily S. Charlson, Jesse Zaneveld, Michael C. Mozer, Ronald G. Collman, Frederic D. Bushman, Rob Knight, and Scott T. Kelley. 2011. “Bayesian Community-Wide Culture-Independent Microbial Source Tracking.” *Nature Methods* 8 (9): 761–63.

Koch, Robert. 1910. Die Ätiologie der Milzbrand-Krankheit, begründet auf die Entwicklungsgeschichte des Bacillus Anthracis (1876).

Kohanski, Michael A., Daniel J. Dwyer, and James J. Collins. 2010. “How Antibiotics Kill Bacteria: From Targets to Networks.” *Nature Reviews. Microbiology* 8 (6): 423–35.

Köhler, T., S. F. Epp, L. K. Curty, and J. C. Pechère. 1999. “Characterization of MexT, the Regulator of the MexE-MexF-OprN Multidrug Efflux System of Pseudomonas Aeruginosa.” *Journal of Bacteriology* 181 (20): 6300–6305.

Kolde, R. 2012. “Pheatmap: Pretty Heatmaps.” *R Package Version*.

Koliada, Alexander, Vladyslav Moseiko, Mariana Romanenko, Liubov Piven, Oleh Lushchak, Nadiia Kryzhanovska, Vitaly Guryanov, and Alexander Vaiserman. 2020. “Seasonal Variation in Gut Microbiota Composition: Cross-Sectional Evidence from Ukrainian Population.” *BMC Microbiology* 20 (1): 100.

Kooij, Dick van der, Anke J. Brouwer-Hanzens, Harm R. Veenendaal, and Bart A. Wullings. 2016. “Multiplication of Legionella Pneumophila Sequence Types 1, 47, and 62 in Buffered Yeast Extract Broth and Biofilms Exposed to Flowing Tap Water at Temperatures of 38°C to 42°C.” *Applied and Environmental Microbiology* 82 (22): 6691–6700.

Korbie, Darren J., and John S. Mattick. 2008. “Touchdown PCR for Increased Specificity and Sensitivity in PCR Amplification.” *Nature Protocols* 3 (9): 1452–56.

Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. 2017. “Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-Mer Weighting and Repeat Separation.” *Genome Research* 27 (5): 722–36.

Korotetskiy, Ilya S., Sergey V. Shilov, Tatyana Kuznetsova, Bahkytzhan Kerimzhanova, Nadezhda Korotetskaya, Lyudmila Ivanova, Natalya Zubenko, Raikhan Parenova, and Oleg N. Reva. 2023.

"Analysis of Whole-Genome Sequences of Pathogenic Gram-Positive and Gram-Negative Isolates from the Same Hospital Environment to Investigate Common Evolutionary Trends Associated with Horizontal Gene Exchange, Mutations and DNA Methylation Patterning." *Microorganisms* 11 (2): 323.

Kosmidis, Christos, Bryan D. Schindler, Pauline L. Jacinto, Diixa Patel, Karanpreet Bains, Susan M. Seo, and Glenn W. Kaatz. 2012. "Expression of Multidrug Resistance Efflux Pump Genes in Clinical and Environmental Isolates of Staphylococcus Aureus." *International Journal of Antimicrobial Agents* 40 (3): 204–9.

Kotay, Shireen Meher, Hardik I. Parikh, Katie Barry, Hyun Soon Gweon, William Guilford, Joanne Carroll, and Amy J. Mathers. 2020. "Nutrients Influence the Dynamics of Klebsiella Pneumoniae Carbapenemase Producing Enterobacterales in Transplanted Hospital Sinks." *Water Research* 176 (June): 115707.

Kozak-Muiznieks, Natalia A., Claressa E. Lucas, Ellen Brown, Tracy Pondo, Thomas H. Taylor Jr, Michael Frace, Diane Miskowski, and Jonas M. Winchell. 2014. "Prevalence of Sequence Types among Clinical and Environmental Isolates of Legionella Pneumophila Serogroup 1 in the United States from 1982 to 2012." *Journal of Clinical Microbiology* 52 (1): 201–11.

Kraemer, Moritz U. G., Houriiyah Tegally, David M. Pigott, Abhishek Dasgupta, James Sheldon, Eduan Wilkinson, Marinanicole Schultheiss, et al. 2022. "Tracking the 2022 Monkeypox Outbreak with Epidemiological Data in Real-Time." *The Lancet Infectious Diseases* 22 (7): 941–42.

Kralik, Petr, and Matteo Ricchi. 2017. "A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything." *Frontiers in Microbiology* 8 (February): 108.

Kramer, Axel, Ingeborg Schwebke, and Günter Kampf. 2006. "How Long Do Nosocomial Pathogens Persist on Inanimate Surfaces? A Systematic Review." *BMC Infectious Diseases* 6 (August): 130.

Ksiazek, Thomas G., Dean Erdman, Cynthia S. Goldsmith, Sherif R. Zaki, Teresa Peret, Shannon Emery, Suxiang Tong, et al. 2003. “A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome.” *The New England Journal of Medicine* 348 (20): 1953–66.

Kumar, Ayush, and Herbert P. Schweizer. 2005. “Bacterial Resistance to Antibiotics: Active Efflux and Reduced Uptake.” *Advanced Drug Delivery Reviews* 57 (10): 1486–1513.

Kumar, Ayush, and Herbert P. Schweizer. 2011. “Evidence of MexT-Independent Overexpression of MexEF-OprN Multidrug Efflux Pump of Pseudomonas Aeruginosa in Presence of Metabolic Stress.” *PloS One* 6 (10): e26520.

Kumaraswamy, Monika, Alison Coady, Richard Szubin, Thomas Cs Martin, Bernhard Palsson, Victor Nizet, and Jonathan M. Monk. 2023. “Comprehensive Whole Genome Sequencing with Hybrid Assembly of Multi-Drug Resistant Candida Albicans Isolate Causing Cerebral Abscess.” *Current Research in Microbial Sciences* 4 (January): 100180.

Kyung, Su Min, Jun Ho Lee, Eun-Seo Lee, Cheol-Yong Hwang, and Han Sang Yoo. 2023. “Whole Genome Structure and Resistance Genes in Carbapenemase-Producing Multidrug Resistant ST378 Klebsiella Pneumoniae.” *BMC Microbiology* 23 (1): 323.

Lagier, Jean-Christophe, Perrine Hugon, Saber Khelaifia, Pierre-Edouard Fournier, Bernard La Scola, and Didier Raoult. 2015. “The Rebirth of Culture in Microbiology through the Example of Culturomics to Study Human Gut Microbiota.” *Clinical Microbiology Reviews* 28 (1): 237–64.

Lai, Huei-Yi, and Tim F. Cooper. 2024. “Interaction with a Phage Gene Underlie Costs of a β-Lactamase.” mBio 15 (2): e0277623.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. “Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome.” *Genome Biology* 10 (3): R25.

Lao, Hiu-Yin, Timothy Ting-Leung Ng, Ryan Yik-Lam Wong, Celia Sze-Ting Wong, Lam-Kwong Lee, Denise Sze-Hang Wong, Chloe Toi-Mei Chan, et al. 2022. “The Clinical Utility of Two High-Throughput 16S rRNA Gene Sequencing Workflows for Taxonomic Assignment of Unidentifiable Bacterial Pathogens in Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry.” *Journal of Clinical Microbiology* 60 (1): e0176921.

Lapa, S. A., E. S. Klochikhina, R. A. Miftakhov, A. M. Zolotov, A. S. Zasedatelev, and A. V. Chudinov. 2020. “Multiplex PCR for Identification of Bacterial Pathogens of Infectious Pneumonia.” *Russian Journal of Bioorganic Chemistry* 46 (5): 859–61.

Larsen, Mette V., Salvatore Cosentino, Simon Rasmussen, Carsten Friis, Henrik Hasman, Rasmus Lykke Marvig, Lars Jelsbak, et al. 2012. “Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria.” *Journal of Clinical Microbiology* 50 (4): 1355–61.

Lear, Gavin, Ian Dickie, Jonathan Banks, Stephane Boyer, Hannah L. Buckley, Thomas R. Buckley, Rob Cruickshank, et al. 2018. “Methods for the Extraction, Storage, Amplification and Sequencing of DNA from Environmental Samples.” *New Zealand Journal of Ecology* 42 (1): 10–50A.

Lear, Gavin, Yimin Dong, and Gillian Lewis. 2010. “Comparison of Methods for the Extraction of DNA from Stream Epilithic Biofilms.” *Antonie van Leeuwenhoek* 98 (4): 567–71.

Lee, Jung Hun, Nam-Hoon Kim, Kyung-Min Jang, Hyeonku Jin, Kyoungmin Shin, Byeong Chul Jeong, Dae-Wi Kim, and Sang Hee Lee. 2023. “Prioritization of Critical Factors for Surveillance of the Dissemination of Antibiotic Resistance in Pseudomonas Aeruginosa: A Systematic Review.” *International Journal of Molecular Sciences* 24 (20). https://doi.org/10.3390/ijms242015209.

Lee, M. R., Y. T. Huang, C. H. Liao, T. Y. Chuang, C. K. Lin, S. W. Lee, C. C. Lai, C. J. Yu, and P. R. Hsueh. 2011. “Bacteremia Caused by Brevundimonas Species at a Tertiary Care Hospital in Taiwan,

2000–2010." *European Journal of Clinical Microbiology & Infectious Diseases: Official Publication of the European Society of Clinical Microbiology* 30 (10): 1185–91.

Leekitcharoenphon, Pimlapas, Markus Hans Kristofer Johansson, Patrick Munk, Burkhard Malorny, Magdalena Skarżyńska, Katharina Wadepohl, Gabriel Moyano, et al. 2021. "Genomic Evolution of Antimicrobial Resistance in Escherichia Coli." *Scientific Reports* 11 (1): 15108.

Legionella: Drinking Water Health Advisory. 2001. U.S. Environmental Protection Agency, Office of Science and Technology, Office of Water.

Leigh, Robert J., Chloe McKenna, Robert McWade, Breda Lynch, and Fiona Walsh. 2022. "Comparative Genomics and Pangenomics of Vancomycin-Resistant and Susceptible Enterococcus Faecium from Irish Hospitals." *Journal of Medical Microbiology* 71 (10). https://doi.org/10.1099/jmm.0.001590.

Leung, Marcus H. Y., Xinzhao Tong, and Patrick K. H. Lee. 2019. "Indoor Microbiome and Airborne Pathogens." *Comprehensive Biotechnology*, 96.

Levy, Stuart B., and Bonnie Marshall. 2004. "Antibacterial Resistance Worldwide: Causes, Challenges and Responses." *Nature Medicine* 10 (12 Suppl): S122–29.

Liassine, Nadia, Laetititia Assouvie, Marie-Christine Descombes, Valérie Dénervaud Tendon, Nicolas Kieffer, Laurent Poirel, and Patrice Nordmann. 2016. "Very Low Prevalence of MCR-1/MCR-2 Plasmid-Mediated Colistin Resistance in Urinary Tract Enterobacteriaceae in Switzerland." *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases* 51 (October): 4–5.

Lin, Xuan, Melissa Glier, Kevin Kuchinski, Tenysha Ross-Van Mierlo, David McVea, John R. Tyson, Natalie Prystajecky, and Ryan M. Ziels. 2021. "Assessing Multiplex Tiling PCR Sequencing Approaches for Detecting Genomic Variants of SARS-CoV-2 in Municipal Wastewater." *mSystems* 6 (5): e0106821.

Liu, Hongjie, Jinhui Li, Yanfeng Lin, Xiaochen Bo, Hongbin Song, Kuibiao Li, Peng Li, and Ming Ni. 2022. “Assessment of Two-Pool Multiplex Long-Amplicon Nanopore Sequencing of SARS-CoV-2.” *Journal of Medical Virology* 94 (1): 327–34.

Liu, Na, Lin Liu, and Xinghua Pan. 2014. “Single-Cell Analysis of the Transcriptome and Its Application in the Characterization of Stem Cells and Early Embryos.” *Cellular and Molecular Life Sciences: CMLS* 71 (14): 2707–15.

Liu, Ning, Xing Yan, Meiling Zhang, Lei Xie, Qian Wang, Yongping Huang, Xuguo Zhou, Shengyue Wang, and Zhihua Zhou. 2011. “Microbiome of Fungus-Growing Termites: A New Reservoir for Lignocellulase Genes.” *Applied and Environmental Microbiology* 77 (1): 48–56.

Liu, Ying, Yang Cao, Tao Wang, Qingyang Dong, Junwen Li, and Chao Niu. 2019. “Detection of 12 Common Food-Borne Bacterial Pathogens by TaqMan Real-Time PCR Using a Single Set of Reaction Conditions.” *Frontiers in Microbiology* 10 (February): 222.

Li, An-Dong, Jacob W. Metch, Yulin Wang, Emily Garner, An Ni Zhang, Maria V. Riquelme, Peter J. Vikesland, Amy Pruden, and Tong Zhang. 2018. “Effects of Sample Preservation and DNA Extraction on Enumeration of Antibiotic Resistance Genes in Wastewater.” *FEMS Microbiology Ecology* 94 (2). https://doi.org/10.1093/femsec/fix189.

Li, Bing, Ying Yang, Liping Ma, Feng Ju, Feng Guo, James M. Tiedje, and Tong Zhang. 2015. “Metagenomic and Network Analysis Reveal Wide Distribution and Co-Occurrence of Environmental Antibiotic Resistance Genes.” *The ISME Journal* 9 (11): 2490–2502.

Li, C., D. Debruyne, J. Spencer, V. Kapoor, and L. Y. Liu. 2020. “High Sensitivity Detection of SARS-CoV-2 Using Multiplex PCR and a Multiplex-PCR-Based Metagenomic Method.” *BioRxiv*. https://covid-19.conacyt.mx/jspui/handle/1000/4660.

Li, Heng. 2018. “Minimap2: Pairwise Alignment for Nucleotide Sequences.” *Bioinformatics* 34 (18): 3094–3100.

Li, Xiang, Ziqi Wu, Chenyuan Dang, Miao Zhang, Bixi Zhao, Zhanwen Cheng, Liming Chen, Zhenfeng Zhong, Yuhui Ye, and Yu Xia. 2021. “A Metagenomic-Based Method to Study Hospital Air Dust Resistome.” *Chemical Engineering Journal* 406 (February): 126854.

Li, Yan-Jun, Yun-Fan Yang, Yi-Jing Zhou, Rong-Hua Zhang, Cheng-Wei Liu, Hong Liu, Xiu-Gui Li, Wen Chen, Yan Chen, and Yong-Ning Wu. 2022. “Estimating the Burden of Foodborne Gastroenteritis due to Nontyphoidal Salmonella Enterica, Shigella and Vibrio Parahaemolyticus in China.” *PloS One* 17 (11): e0277203.

Li, Yaying, Hongkai Liao, and Huaiying Yao. 2019. “Prevalence of Antibiotic Resistance Genes in Air-Conditioning Systems in Hospitals, Farms, and Residences.” *International Journal of Environmental Research and Public Health* 16 (5). https://doi.org/10.3390/ijerph16050683.

Li, Yiming, Weiwei Cao, Shuli Liang, Shinji Yamasaki, Xun Chen, Lei Shi, and Lei Ye. 2020. “Metagenomic Characterization of Bacterial Community and Antibiotic Resistance Genes in Representative Ready-to-Eat Food in Southern China.” *Scientific Reports* 10 (1): 15175.

Long, Naiyun, Yi Qiao, Zheyun Xu, Jing Tu, and Zuhong Lu. 2020. “Recent Advances and Application in Whole-Genome Multiple Displacement Amplification.” *Quantitative Biology* 8 (4): 279–94.

López-Causapé, Carla, Rosa Rubio, Gabriel Cabot, and Antonio Oliver. 2018. “Evolution of the Pseudomonas Aeruginosa Aminoglycoside Mutational Resistome In Vitro and in the Cystic Fibrosis Setting.” *Antimicrobial Agents and Chemotherapy* 62 (4). https://doi.org/10.1128/AAC.02583-17.

López, María, Beatriz Rojo-Bezares, Gabriela Chichón, and Yolanda Sáenz. 2022. “Resistance to Fluoroquinolones in Pseudomonas Aeruginosa from Human, Animal, Food and Environmental Origin: The Role of CrpP and Mobilisable ICEs.” *Antibiotics (Basel, Switzerland)* 11 (9). https://doi.org/10.3390/antibiotics11091271.

Luqman, Arif, Jongkon Saising, Yulianto Ade Prasetya, Aparna Viswanathan Ammanath, Andini, Siti Nur Amala, Enny Zulaika, Nengah Dwianita Kuswytasari, Friedrich Goetz, and Anjar Tri Wibowo. 2024. “Detection of Vancomycin Resistant Genes in Intrinsically Antibiotic Resistant Bacteria from the Gut Microbiota of Indonesian Individuals.” *Iranian Journal of Medical Sciences* 49 (5): 302–12.

Ma, Xiuqing, Yanqin Li, Yuan Liang, Yang Liu, Ling Yu, Chunsun Li, Qiqi Liu, and Liangan Chen. 2020. “Development of a DNA Microarray Assay for Rapid Detection of Fifteen Bacterial Pathogens in Pneumonia.” *BMC Microbiology* 20 (1): 177.

Madebo, Chemere, Admasu Haile, Temesgen Eticha, and Fithamlak Solomon. 2022. “Hospital-Based Air-Borne and Surface-Borne Bacterial Pathogens and Their Antimicrobial Profiles in Wolaita Sodo, Southern Ethiopia.” *International Journal of Microbiology* 2022 (October): 5718341.

Mah, Thien-Fah. 2012. “Biofilm-Specific Antibiotic Resistance.” *Future Microbiology* 7 (9): 1061–72.

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, et al. 1998. “Multilocus Sequence Typing: A Portable Approach to the Identification of Clones within Populations of Pathogenic Microorganisms.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (6): 3140–45.

Makise, Masaki, Shinji Mima, Motohiro Koterasawa, Tomofusa Tsuchiya, and Tohru Mizushima. 2002. “Biochemical Analysis of DnaA Protein with Mutations in Both Arg328 and Lys372.” *Biochemical Journal* 362 (Pt 2): 453–58.

Malla, Bikash, Ocean Thakali, Sadhana Shrestha, Takahiro Segawa, Masaaki Kitajima, and Eiji Haramoto. 2022. “Application of a High-Throughput Quantitative PCR System for Simultaneous Monitoring of SARS-CoV-2 Variants and Other Pathogenic Viruses in Wastewater.” *The Science of the Total Environment* 853 (December): 158659.

Manaia, Célia M., Erica Donner, Ivone Vaz-Moreira, and Peiying Hong. 2020. *Antibiotic Resistance in the Environment: A Worldwide Overview*. Springer Nature.

Manuel, Juana G., Hillary B. Heins, Sandra Crocker, Julie A. Neidich, Lisa Sadzewicz, Luke Tallon, and Tychele N. Turner. 2023. “High Coverage Highly Accurate Long-Read Sequencing of a Mouse Neuronal Cell Line Using the PacBio Revio Sequencer.” *bioRxiv*: The Preprint Server for Biology, June. https://doi.org/10.1101/2023.06.06.543940.

Manyi-Loh, Christy, Sampson Mamphweli, Edson Meyer, and Anthony Okoh. 2018. “Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications.” *Molecules* 23 (4). https://doi.org/10.3390/molecules23040795.

Mao, Daqing, Yi Luo, Jacques Mathieu, Qing Wang, Ling Feng, Quanhua Mu, Chunyan Feng, and P. J. J. Alvarez. 2014. “Persistence of Extracellular DNA in River Sediment Facilitates Antibiotic Resistance Gene Propagation.” *Environmental Science & Technology* 48 (1): 71–78.

Markkanen, Melina A., Kaisa Haukka, Katariina M. M. Pärnänen, Victorien Tamegnon Dougnon, Isidore Juste O. Bonkoungou, Zakaria Garba, Halidou Tinto, et al. 2023. “Metagenomic Analysis of the Abundance and Composition of Antibiotic Resistance Genes in Hospital Wastewater in Benin, Burkina Faso, and Finland.” *mSphere* 8 (1): e0053822.

Martin, Andrew P. 2002. “Phylogenetic Approaches for Describing and Comparing the Diversity of Microbial Communities.” *Applied and Environmental Microbiology* 68 (8): 3673–82.

Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” *EMBnet.journal* 17 (1): 10–12.

Martin, Samuel, Darren Heavens, Yuxuan Lan, Samuel Horsfield, Matthew D. Clark, and Richard M. Leggett. 2022. “Nanopore Adaptive Sampling: A Tool for Enrichment of Low Abundance Species in Metagenomic Samples.” *Genome Biology* 23 (1): 11.

Martinez, Jose Luis, Alicia Fajardo, Leonor Garmendia, Alvaro Hernandez, Juan Francisco Linares, Laura Martínez-Solano, and María Blanca Sánchez. 2009. “A Global View of Antibiotic Resistance.” *FEMS Microbiology Reviews* 33 (1): 44–65.

Mason, Christopher E., and Olivier Elemento. 2012. “Faster Sequencers, Larger Datasets, New Challenges.” *Genome Biology* 13 (3): 314.

Matsuo, Yoshiyuki, Shinnosuke Komiya, Yoshiaki Yasumizu, Yuki Yasuoka, Katsura Mizushima, Tomohisa Takagi, Kirill Kryukov, et al. 2021. “Full-Length 16S rRNA Gene Amplicon Analysis of Human Gut Microbiota Using MinION™ Nanopore Sequencing Confers Species-Level Resolution.” *BMC Microbiology* 21 (1): 35.

Maxam, A. M., and W. Gilbert. 1977. “A New Method for Sequencing DNA.” *Proceedings of the National Academy of Sciences of the United States of America* 74 (2): 560–64.

McAdam, Paul R., Charles W. Vander Broek, Diane S. J. Lindsay, Melissa J. Ward, Mary F. Hanson, Michael Gillies, Mick Watson, Joanne M. Stevens, Giles F. Edwards, and J. Ross Fitzgerald. 2014. “Gene Flow in Environmental Legionella Pneumophila Leads to Genetic and Pathogenic Heterogeneity within a Legionnaires’ Disease Outbreak.” *Genome Biology*. https://doi.org/10.1186/s13059-014-0504-1.

McDermott Patrick F., Tyson Gregory H., Kabera Claudine, Chen Yuansha, Li Cong, Folster Jason P., Ayers Sherry L., Lam Claudia, Tate Heather P., and Zhao Shaohua. 2016. “Whole-Genome Sequencing

for Detecting Antimicrobial Resistance in Nontyphoidal Salmonella." *Antimicrobial Agents and Chemotherapy* 60 (9): 5515–20.

McGann, Patrick, Jessica L. Bunin, Erik Snesrud, Seema Singh, Rosslyn Maybank, Ana C. Ong, Yoon I. Kwak, et al. 2016. "Real Time Application of Whole Genome Sequencing for Outbreak Investigation - What Is an Achievable Turnaround Time?" *Diagnostic Microbiology and Infectious Disease* 85 (3): 277–82.

McLoughlin, Kevin S. 2011. "Microarrays for Pathogen Detection and Analysis." *Briefings in Functional Genomics* 10 (6): 342–53.

McOrist, Alexandra L., Michelle Jackson, and Anthony R. Bird. 2002. "A Comparison of Five Methods for Extraction of Bacterial DNA from Human Faecal Samples." *Journal of Microbiological Methods* 50 (2): 131–39.

McPhee, Joseph B., Shawn Lewenza, and Robert E. W. Hancock. 2003. "Cationic Antimicrobial Peptides Activate a Two-Component Regulatory System, PmrA-PmrB, That Regulates Resistance to Polymyxin B and Cationic Antimicrobial Peptides in Pseudomonas Aeruginosa." *Molecular Microbiology* 50 (1): 205–17.

Mercante, Jeffrey W., Shatavia S. Morrison, Heta P. Desai, Brian H. Raphael, and Jonas M. Winchell. 2016. "Genomic Analysis Reveals Novel Diversity among the 1976 Philadelphia Legionnaires' Disease Outbreak Isolates and Additional ST36 Strains." *PloS One* 11 (9): e0164074.

Merwe, R. G. van der, P. D. van Helden, R. M. Warren, S. L. Sampson, and N. C. Gey van Pittius. 2014. "Phage-Based Detection of Bacterial Pathogens." *The Analyst* 139 (11): 2617–26.

Miller, William R., Jose M. Munita, and Cesar A. Arias. 2014. "Mechanisms of Antibiotic Resistance in Enterococci." *Expert Review of Anti-Infective Therapy* 12 (10): 1221–36.

Milne, Iain, Gordon Stephen, Micha Bayer, Peter J. A. Cock, Leighton Pritchard, Linda Cardle, Paul D. Shaw, and David Marshall. 2013. “Using Tablet for Visual Exploration of Second-Generation Sequencing Data.” *Briefings in Bioinformatics* 14 (2): 193–202.

Minei, Ryuhei, Ryo Hoshina, and Atsushi Ogura. 2018. “De Novo Assembly of Middle-Sized Genome Using MinION and Illumina Sequencers.” *BMC Genomics* 19 (1): 700.

Mitchell, Amber Hogan. 2020. Preventing Occupational Exposures to Infectious Disease in Health Care: A Practical Guide. *Springer Nature*.

Miyazaki, Kentaro, Mitsuharu Sato, and Miyuki Tsukuda. 2017. “PCR Primer Design for 16S rRNAs for Experimental Horizontal Gene Transfer Test in Escherichia Coli.” *Frontiers in Bioengineering and Biotechnology* 5 (February): 14.

Moezi, P., M. Kargar, A. Doosti, and M. Khoshneviszadeh. 2019. “Multiplex Touchdown PCR Assay to Enhance Specificity and Sensitivity for Concurrent Detection of Four Foodborne Pathogens in Raw Milk.” *Journal of Applied Microbiology* 127 (1): 262–73.

Moges, Feleke, Mengistu Endris, Yeshambel Belyhun, and Walelegn Worku. 2014. “Isolation and Characterization of Multiple Drug Resistance Bacterial Pathogens from Waste Water in Hospital and Non-Hospital Environments, Northwest Ethiopia.” *BMC Research Notes* 7 (April): 215.

Moran-Gilad, J., K. Prior, E. Yakunin, T. G. Harrison, A. Underwood, T. Lazarovitch, L. Valinsky, et al. 2015. “Design and Application of a Core Genome Multilocus Sequence Typing Scheme for Investigation of Legionnaires’ Disease Incidents.” *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 20 (28). https://doi.org/10.2807/1560-7917.es2015.20.28.21186.

Morero, Natalia R., Mariela R. Monti, and Carlos E. Argaraña. 2011. “Effect of Ciprofloxacin Concentration on the Frequency and Nature of Resistant Mutants Selected from Pseudomonas Aeruginosa mutS and mutT Hypermutators.” *Antimicrobial Agents and Chemotherapy* 55 (8): 3668–76.

Moss, Eli L., Dylan G. Maghini, and Ami S. Bhatt. 2020. “Complete, Closed Bacterial Genomes from Microbiomes Using Nanopore Sequencing.” *Nature Biotechnology* 38 (6): 701–7.

Muhammed, Musemma K., Lukasz Krych, Dennis S. Nielsen, and Finn K. Vogensen. 2017. “A High-Throughput qPCR System for Simultaneous Quantitative Detection of Dairy Lactococcus Lactis and Leuconostoc Bacteriophages.” *PloS One* 12 (3): e0174223.

Muniesa, Maite, Marta Colomer-Lluch, and Juan Jofre. 2013. “Could Bacteriophages Transfer Antibiotic Resistance Genes from Environmental Bacteria to Human-Body Associated Bacterial Populations?” *Mobile Genetic Elements* 3 (4): e25847.

Munita, Jose M., and Cesar A. Arias. 2016. “Mechanisms of Antibiotic Resistance.” *Microbiology Spectrum* 4 (2). https://doi.org/10.1128/microbiolspec.VMBF-0016-2015.

Muziasari, Windi I., Katariina Pärnänen, Timothy A. Johnson, Christina Lyra, Antti Karkman, Robert D. Stedtfeld, Manu Tamminen, James M. Tiedje, and Marko Virta. 2016. “Aquaculture Changes the Profile of Antibiotic Resistance and Mobile Genetic Element Associated Genes in Baltic Sea Sediments.” *FEMS Microbiology Ecology* 92 (4): fiw052.

Nadal-Ribelles, Mariona, Saiful Islam, Wu Wei, Pablo Latorre, Michelle Nguyen, Eulàlia de Nadal, Francesc Posas, and Lars M. Steinmetz. 2019. “Sensitive High-Throughput Single-Cell RNA-Seq Reveals within-Clonal Transcript Correlations in Yeast Populations.” *Nature Microbiology* 4 (4): 683–92.

Nemec, Alexandr, Lucilla Dolzani, Sylvain Brisse, Peterhans van den Broek, and Lenie Dijkshoorn. 2004. "Diversity of Aminoglycoside-Resistance Genes and Their Association with Class 1 Integrons among Strains of Pan-European Acinetobacter Baumannii Clones." *Journal of Medical Microbiology* 53 (Pt 12): 1233–40.

Ng, Pauline C., and Ewen F. Kirkness. 2010. "Whole Genome Sequencing." *Methods in Molecular Biology* 628: 215–26.

Nguyen, Fabian, Agata L. Starosta, Stefan Arenz, Daniel Sohmen, Alexandra Dönhöfer, and Daniel N. Wilson. 2014. "Tetracycline Antibiotics and Resistance Mechanisms." *Biological Chemistry* 395 (5): 559–75.

Nicholls, Samuel M., Joshua C. Quick, Shuiquan Tang, and Nicholas J. Loman. 2019. "Ultra-Deep, Long-Read Nanopore Sequencing of Mock Microbial Community Standards." *GigaScience* 8 (5). https://doi.org/10.1093/gigascience/giz043.

Nimmo, G. R., and J. Z. Bull. 1995. "Comparative Susceptibility of Legionella Pneumophila and Legionella Longbeachae to 12 Antimicrobial Agents." *The Journal of Antimicrobial Chemotherapy* 36 (1): 219–23.

Ni, Ying, Xudong Liu, Zemenu Mengistie Simeneh, Mengsu Yang, and Runsheng Li. 2023. "Benchmarking of Nanopore R10.4 and R9.4.1 Flow Cells in Single-Cell Whole-Genome Amplification and Whole-Genome Shotgun Sequencing." *Computational and Structural Biotechnology Journal* 21 (March): 2352–64.

Nolan, Laura M., Lynne Turnbull, Marilyn Katrib, Sarah R. Osvath, Davide Losa, James J. Lazenby, and Cynthia B. Whitchurch. 2020. "Pseudomonas Aeruginosa Is Capable of Natural Transformation in Biofilms." *Microbiology* 166 (10): 995–1003.

Oberauner, Lisa, Christin Zachow, Stefan Lackner, Christoph Högenauer, Karl-Heinz Smolle, and Gabriele Berg. 2013. “The Ignored Diversity: Complex Bacterial Communities in Intensive Care Units Revealed by 16S Pyrosequencing.” *Scientific Reports* 3: 1413.

Ocampo-Sosa, Alain A., Gabriel Cabot, Cristina Rodríguez, Elena Roman, Fe Tubau, María D. Macia, Bartolomé Moya, et al. 2012. “Alterations of OprD in Carbapenem-Intermediate and -Susceptible Strains of Pseudomonas Aeruginosa Isolated from Patients with Bacteremia in a Spanish Multicenter Study.” *Antimicrobial Agents and Chemotherapy* 56 (4): 1703–13.

Oliver, Antonio, Xavier Mulet, Carla López-Causapé, and Carlos Juan. 2015. “The Increasing Threat of Pseudomonas Aeruginosa High-Risk Clones.” *Drug Resistance Updates: Reviews and Commentaries in Antimicrobial and Anticancer Chemotherapy* 21-22 (August): 41–59.

O’Neill, J., Review on Antimicrobial Resistance, and Wellcome Trust (London, England). 2016. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*. Review on Antimicrobial Resistance.

Opal, Steven M. 2010. “A Brief History of Microbiology and Immunology.” In *Vaccines: A Biography*, edited by Andrew W. Artenstein, 31–56. New York, NY: Springer New York.

Page, Andrew J., Carla A. Cummins, Martin Hunt, Vanessa K. Wong, Sandra Reuter, Matthew T. G. Holden, Maria Fookes, Daniel Falush, Jacqueline A. Keane, and Julian Parkhill. 2015. “Roary: Rapid Large-Scale Prokaryote Pan Genome Analysis.” *Bioinformatics* 31 (22): 3691–93.

Pallerla, Srinivas Reddy, Do Van Dong, Le Thi Kieu Linh, Trinh Van Son, Dao Thanh Quyen, Phan Quoc Hoan, Ngo Tat Trung, et al. 2022. “Diagnosis of Pathogens Causing Bacterial Meningitis Using Nanopore Sequencing in a Resource-Limited Setting.” *Annals of Clinical Microbiology and Antimicrobials* 21 (1): 39.

Pang, Zheng, Renee Raudonis, Bernard R. Glick, Tong-Jun Lin, and Zhenyu Cheng. 2019. “Antibiotic Resistance in Pseudomonas Aeruginosa: Mechanisms and Alternative Therapeutic Strategies.” *Biotechnology Advances* 37 (1): 177–92.

Pappa, Olga, Dimosthenis Chochlakis, Vassilios Sandalakis, Chrysa Dioli, Anna Psaroulaki, and Athena Mavridou. 2020. “Antibiotic Resistance of Legionella Pneumophila in Clinical and Water Isolates—A Systematic Review.” *International Journal of Environmental Research and Public Health* 17 (16): 5809.

Park, Soon Deok, Young Uh, In Ho Jang, Kap Jun Yoon, Hwang Min Kim, and Yon Jung Bae. 2011. “Rhodococcus Erythropolis Septicaemia in a Patient with Acute Lymphocytic Leukaemia.” *Journal of Medical Microbiology* 60 (Pt 2): 252–55.

Parkhill, Julian, and Brendan W. Wren. 2011. “Bacterial Epidemiology and Biology--Lessons from Genome Sequencing.” *Genome Biology* 12 (10): 230.

Park, Kyungmin, Seung-Ho Lee, Jongwoo Kim, Jingyeong Lee, Geum-Young Lee, Seungchan Cho, Seung Ho Lee, et al. 2021. “Multiplex PCR-Based Nanopore Sequencing and Epidemiological Surveillance of Hantaan Orthohantavirus in Apodemus Agrarius, Republic of Korea.” *Viruses* 13 (5). https://doi.org/10.3390/v13050847.

Park, Youjin, Jaeyoung Oh, Sowon Park, Samuth Sum, Wonkeun Song, Jongchan Chae, and Heemyung Park. 2020. “Antimicrobial Resistance and Novel Mutations Detected in the gyrA and parC Genes of Pseudomonas Aeruginosa Strains Isolated from Companion Dogs.” *BMC Veterinary Research* 16 (1): 111.

Parsons, M. B., K. L. F. Cooper, K. A. Kubota, N. Puhr, S. Simington, P. S. Calimlim, D. Schoonmaker-Bopp, et al. 2007. “PulseNet USA Standardized Pulsed-Field Gel Electrophoresis Protocol for Subtyping of Vibrio Parahaemolyticus.” *Foodborne Pathogens and Disease* 4 (3): 285–92.

Partridge, Sally R., Stephen M. Kwong, Neville Firth, and Slade O. Jensen. 2018. “Mobile Genetic Elements Associated with Antimicrobial Resistance.” *Clinical Microbiology Reviews* 31 (4). https://doi.org/10.1128/CMR.00088-17.

Paulus, Gabriela K., Luc M. Hornstra, and Gertjan Medema. 2020. “International Tempo-Spatial Study of Antibiotic Resistance Genes across the Rhine River Using Newly Developed Multiplex qPCR Assays.” *The Science of the Total Environment* 706 (March): 135733.

Pereira da Fonseca, Tairacan Augusto, Rodrigo Pessôa, Alvina Clara Felix, and Sabri Saeed Sanabani. 2016. “Diversity of Bacterial Communities on Four Frequently Used Surfaces in a Large Brazilian Teaching Hospital.” *International Journal of Environmental Research and Public Health* 13 (2): 152.

Perencevich, Eli N., Jessina C. McGregor, Michelle Shardell, Jon P. Furuno, Anthony D. Harris, J. Glenn Morris Jr, David N. Fisman, and Judith A. Johnson. 2008. “Summer Peaks in the Incidences of Gram-Negative Bacterial Infection Among Hospitalized Patients.” *Infection Control and Hospital Epidemiology: The Official Journal of the Society of Hospital Epidemiologists of America* 29 (12): 1124–31.

Perreten, Vincent, Lorianne Vorlet-Fawer, Peter Slickers, Ralf Ehricht, Peter Kuhnert, and Joachim Frey. 2005. “Microarray-Based Detection of 90 Antibiotic Resistance Genes of Gram-Positive Bacteria.” *Journal of Clinical Microbiology* 43 (5): 2291–2302.

Petriglieri, Francesca, Caitlin Singleton, Miriam Peces, Jette F. Petersen, Marta Nierychlo, and Per H. Nielsen. 2021. “‘Candidatus Dechloromonas Phosphoritropha’ and ‘Ca. D. Phosphorivorans’, Novel Polyphosphate Accumulating Organisms Abundant in Wastewater Treatment Systems.” *The ISME Journal* 15 (12): 3605–14.

Phoon, Hannah Y. P., Hazilawati Hussin, Baizurah Mohd Hussain, Shu Yong Lim, Jia Jie Woon, Yi Xian Er, and Kwai Lin Thong. 2018. “Distribution, Genetic Diversity and Antimicrobial Resistance of

Clinically Important Bacteria from the Environment of a Tertiary Hospital in Malaysia." *Journal of Global Antimicrobial Resistance* 14 (September): 132–40.

Pochon, Xavier, Anastasija Zaiko, Grant A. Hopkins, Jonathan C. Banks, and Susanna A. Wood. 2015. "Early Detection of Eukaryotic Communities from Marine Biofilm Using High-Throughput Sequencing: An Assessment of Different Sampling Devices." *Biofouling* 31 (3): 241–51.

Porter, J. R. 1976. "Antony van Leeuwenhoek: Tercentenary of His Discovery of Bacteria." *Bacteriological Reviews* 40 (2): 260–69.

Prakash, O., M. Verma, P. Sharma, M. Kumar, K. Kumari, A. Singh, H. Kumari, et al. 2007. "Polyphasic Approach of Bacterial Classification - An Overview of Recent Advances." *Indian Journal of Microbiology* 47 (2): 98–108.

Prescott, Lansing M., Joanne M. Willey, Linda M. Sherwood, and Christopher J. Woolverton. 2018. *Microbiologie*. De Boeck Superieur.

Price, Morgan N., Paramvir S. Dehal, and Adam P. Arkin. 2010. "FastTree 2--Approximately Maximum-Likelihood Trees for Large Alignments." *PloS One* 5 (3): e9490.

Probst, Alexander J., Thomas Weinmaier, Todd Z. DeSantis, Jorge W. Santo Domingo, and Nicholas Ashbolt. 2015. "New Perspectives on Microbial Community Distortion after Whole-Genome Amplification." *PloS One* 10 (5): e0124158.

Prussin, Aaron J., 2nd, and Linsey C. Marr. 2015. "Sources of Airborne Microorganisms in the Built Environment." *Microbiome* 3 (December): 78.

Prussin, Aaron J., 2nd, David Otto Schwake, and Linsey C. Marr. 2017. “Ten Questions Concerning the Aerosolisation and Transmission of Legionella in the Built Environment.” *Building and Environment* 123 (October): 684–95.

Pühler, Alfred, and Kenneth N. Timmis. 2012. *Advanced Molecular Genetics*. Springer Science & Business Media.

Pumbwe, L., and L. J. Piddock. 2000. “Two Efflux Systems Expressed Simultaneously in Multidrug-Resistant Pseudomonas Aeruginosa.” Antimicrobial Agents and Chemotherapy 44 (10): 2861–64.

Qian, Xun, Santosh Gunturu, Jiarong Guo, Benli Chai, James R. Cole, Jie Gu, and James M. Tiedje. 2021. “Metagenomic Analysis Reveals the Shared and Distinct Features of the Soil Resistome across Tundra, Temperate Prairie, and Tropical Ecosystems.” *Microbiome* 9 (1): 108.

Qin, Tian, Wen Zhang, Wenbin Liu, Haijian Zhou, Hongyu Ren, Zhujun Shao, Ruiting Lan, and Jianguo Xu. 2016. “Population Structure and Minimum Core Genome Typing of Legionella Pneumophila.” *Scientific Reports* 6 (February): 21356.

Qin, Tian, Haijian Zhou, Hongyu Ren, Hong Guan, Machao Li, Bingqing Zhu, and Zhujun Shao. 2014. “Distribution of Sequence-Based Types of Legionella Pneumophila Serogroup 1 Strains Isolated from Cooling Towers, Hot Springs, and Potable Water Systems in China.” *Applied and Environmental Microbiology* 80 (7): 2150–57.

Quainoo, Scott, Jordy P. M. Coolen, Sacha A. F. T. van Hijum, Martijn A. Huynen, Willem J. G. Melchers, Willem van Schaik, and Heiman F. L. Wertheim. 2017. “Whole-Genome Sequencing of Bacterial Pathogens: The Future of Nosocomial Outbreak Analysis.” *Clinical Microbiology Reviews* 30 (4): 1015–63.

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. 2013. “The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools.” *Nucleic Acids Research* 41 (Database issue): D590–96.

Quick, J. 2020. “nCoV-2019 Sequencing Protocol v3 (LoCost). Protocols. Io.”

Quick, Joshua, Nathan D. Grubaugh, Steven T. Pullan, Ingra M. Claro, Andrew D. Smith, Karthik Gangavarapu, Glenn Oliveira, et al. 2017. “Multiplex PCR Method for MinION and Illumina Sequencing of Zika and Other Virus Genomes Directly from Clinical Samples.” *Nature Protocols* 12 (6): 1261–76.

Quick, Joshua, Nicholas J. Loman, Sophie Duraffour, Jared T. Simpson, Ettore Severi, Lauren Cowley, Joseph Akoi Bore, et al. 2016. “Real-Time, Portable Genome Sequencing for Ebola Surveillance.” *Nature* 530 (7589): 228–32.

Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26 (6): 841–42.

Rahman, Md Ramim Tanver, Ismail Fliss, and Eric Biron. 2022. “Insights in the Development and Uses of Alternatives to Antibiotic Growth Promoters in Poultry and Swine Production.” Antibiotics (Basel, Switzerland) 11 (6). https://doi.org/10.3390/antibiotics11060766.

Rambaut, A. (2012). FigTree v 1.3.1. [Online]. Available online at: http://tree.bio.ed.ac.uk/software/figtree/.

Ramirez, Maria S., and Marcelo E. Tolmasky. 2010. “Aminoglycoside Modifying Enzymes.” *Drug Resistance Updates: Reviews and Commentaries in Antimicrobial and Anticancer Chemotherap*y 13 (6): 151–71.

Ramos, Tiffanie, Sandra Dedesko, Jeffrey A. Siegel, Jack A. Gilbert, and Brent Stephens. 2015. “Spatial and Temporal Variations in Indoor Environmental Conditions, Human Occupancy, and Operational Characteristics in a New Hospital Building.” *PloS One* 10 (3): e0118207.

Ranganathan, Sarath C., Billy Skoric, Kay A. Ramsay, Rosemary Carzino, Anne-Marie Gibson, Emily Hart, Jo Harrison, Scott C. Bell, Timothy J. Kidd, and Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST-CF) *. 2013. “Geographical Differences in First Acquisition of Pseudomonas Aeruginosa in Cystic Fibrosis.” *Annals of the American Thoracic Society* 10 (2): 108–14.

Ranjbar, Reza, Payam Behzadi, Ali Najafi, and Raheleh Roudi. 2017. “DNA Microarray for Rapid Detection and Identification of Food and Water Borne Bacteria: From Dry to Wet Lab.” *The Open Microbiology Journal* 11 (November): 330–38.

Ratzow, Sandra, Valeria Gaia, Jürgen Herbert Helbig, Norman K. Fry, and Paul Christian Lück. 2007. “Addition of neuA, the Gene Encoding N-Acylneuraminate Cytidylyl Transferase, Increases the Discriminatory Ability of the Consensus Sequence-Based Scheme for Typing Legionella Pneumophila Serogroup 1 Strains.” *Journal of Clinical Microbiology* 45 (6): 1965–68.

Redgrave, Liam S., Sam B. Sutton, Mark A. Webber, and Laura J. V. Piddock. 2014. “Fluoroquinolone Resistance: Mechanisms, Impact on Bacteria, and Role in Evolutionary Success.” *Trends in Microbiology* 22 (8): 438–45.

Reuter, Sandra, Timothy G. Harrison, Claudio U. Köser, Matthew J. Ellington, Geoffrey P. Smith, Julian Parkhill, Sharon J. Peacock, Stephen D. Bentley, and M. Estée Török. 2013. “A Pilot Study of Rapid Whole-Genome Sequencing for the Investigation of a Legionella Outbreak.” *BMJ Open* 3 (1). https://doi.org/10.1136/bmjopen-2012-002175.

Reygaert, Wanda. 2009. “Methicillin-Resistant Staphylococcus Aureus (MRSA): Molecular Aspects of Antimicrobial Resistance and Virulence.” Clinical Laboratory Science: *Journal of the American Society for Medical Technology* 22 (2): 115–19.

Reygaert, Wanda C. 2018. “An Overview of the Antimicrobial Resistance Mechanisms of Bacteria.” *AIMS Microbiology* 4 (3): 482–501.

Reynolds, Dan, and Marin Kollef. 2021. “The Epidemiology and Pathogenesis and Treatment of Pseudomonas Aeruginosa Infections: An Update.” *Drugs* 81 (18): 2117–31.

Rhoads, Anthony, and Kin Fai Au. 2015. “PacBio Sequencing and Its Applications.” *Genomics, Proteomics & Bioinformatics* 13 (5): 278–89.

Ribeiro-Gonçalves, Bruno, Alexandre P. Francisco, Cátia Vaz, Mário Ramirez, and João André Carriço. 2016. “PHYLOViZ Online: Web-Based Tool for Visualisation, Phylogenetic Inference, Analysis and Sharing of Minimum Spanning Trees.” *Nucleic Acids Research* 44 (W1): W246–51

Roberts, Marilyn C. 2004. “Resistance to Macrolide, Lincosamide, Streptogramin, Ketolide, and Oxazolidinone Antibiotics.” *Molecular Biotechnology* 28 (1): 47–62.

Robicsek, Ari, Jacob Strahilevitz, George A. Jacoby, Mark Macielag, Darren Abbanat, Chi Hye Park, Karen Bush, and David C. Hooper. 2006. “Fluoroquinolone-Modifying Enzyme: A New Adaptation of a Common Aminoglycoside Acetyltransferase.” *Nature Medicine* 12 (1): 83–88.

Robinson, Esther R., Timothy M. Walker, and Mark J. Pallen. 2013. “Genomics and Outbreak Investigation: From Sequence to Consequence.” *Genome Medicine* 5 (4): 36.

Rocha, Jaqueline, Telma Fernandes, Maria V. Riquelme, Ni Zhu, Amy Pruden, and Célia M. Manaia. 2019. “Comparison of Culture- and Quantitative PCR-Based Indicators of Antibiotic Resistance in

Wastewater, Recycled Water, and Tap Water." *International Journal of Environmental Research and Public Health* 16 (21). https://doi.org/10.3390/ijerph16214217.

Rolain, J. M., R. Canton, and G. Cornaglia. 2012. "Emergence of Antibiotic Resistance: Need for a New Paradigm." *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 18 (7): 615–16.

Rottier, E., and M. Ince. 2003. "Controlling and Preventing Disease: The Role of Water and Environmental Sanitation Interventions." https://books.google.com/books?hl=en&lr=&id=hHCZOxzlCsYC&oi=fnd&pg=PR5&dq=CONTROLLING+AND+PREVENTING+DISEASE&ots=W1pgSNRk0y&sig=c7sfREV6HTd1RdLDpFO9Ul1eQyA.

Ruiz, Joaquim. 2003. "Mechanisms of Resistance to Quinolones: Target Alterations, Decreased Accumulation and DNA Gyrase Protection." *The Journal of Antimicrobial Chemotherapy* 51 (5): 1109–17.

Ryan, K. J., and R. C. G. Sherris. n.d. "Medical Microbiology^." Accessed January 19, 2024. https://booksca.ca/wp-content/uploads/XPreview/Pharmacology/3/sherris-medical-microbiology-7th-edition-by-kenneth-j-ryan.pdf.

Saadatian Farivar, Arezoo, Jamileh Nowroozi, Gita Eslami, and Azar Sabokbar. 2018. "RAPD PCR Profile, Antibiotic Resistance, Prevalence of armA Gene, and Detection of KPC Enzyme in Klebsiella Pneumoniae Isolates." *The Canadian Journal of Infectious Diseases & Medical Microbiology = Journal Canadien Des Maladies Infectieuses et de La Microbiologie Medicale / AMMI Canada 2018* (February): 6183162.

Safronova, V., and I. Tikhonovich. 2012. "Microbes in Applied Research: Current Advances and Challenges."

Saleh, R. M., and A. S. Motib. 2023. “Molecular Detection of OprD and ExoA in Pseudomonas Aeruginosa and Antibiotics Resistance.” AIP Conference Proceedings. https://pubs.aip.org/aip/acp/article-abstract/2475/1/030006/2879546.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. 1985. “Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia.” *Science* 230 (4732): 1350–54.

Sanbongi, Yumiko, Atsuyuki Shimizu, Takahisa Suzuki, Hiroshi Nagaso, Takashi Ida, Kazunori Maebashi, and Naomasa Gotoh. 2009. “Classification of OprD Sequence and Correlation with Antimicrobial Activity of Carbapenem Agents in Pseudomonas Aeruginosa Clinical Isolates Collected in Japan.” *Microbiology and Immunology* 53 (7): 361–67.

Sanger, F., and A. R. Coulson. 1975. “A Rapid Method for Determining Sequences in DNA by Primed Synthesis with DNA Polymerase.” *Journal of Molecular Biology* 94 (3): 441–48.

Sayers, Eric W., Evan E. Bolton, J. Rodney Brister, Kathi Canese, Jessica Chan, Donald C. Comeau, Ryan Connor, et al. 2022. “Database Resources of the National Center for Biotechnology Information.” *Nucleic Acids Research* 50 (D1): D20–26.

Scholz, Miklas, and Byoung-hwa Lee. 2005. “Constructed Wetlands: A Review.” *The International Journal of Environmental Studies* 62 (4): 421–47.

Scholz, P., V. Haring, B. Wittmann-Liebold, K. Ashman, M. Bagdasarian, and E. Scherzinger. 1989. “Complete Nucleotide Sequence and Gene Organization of the Broad-Host-Range Plasmid RSF1010.” *Gene* 75 (2): 271–88.

Schwartz, D. C., and C. R. Cantor. 1984. “Separation of Yeast Chromosome-Sized DNAs by Pulsed Field Gradient Gel Electrophoresis.” *Cell* 37 (1): 67–75.

Schwengers, Oliver, Patrick Barth, Linda Falgenhauer, Torsten Hain, Trinad Chakraborty, and Alexander Goesmann. 2020. “Platon: Identification and Characterization of Bacterial Plasmid Contigs in Short-Read Draft Assemblies Exploiting Protein Sequence-Based Replicon Distribution Scores.” *Microbial Genomics* 6 (10). https://doi.org/10.1099/mgen.0.000398.

Schwarz, Stefan, Corinna Kehrenberg, Benoît Doublet, and Axel Cloeckaert. 2004. “Molecular Basis of Bacterial Resistance to Chloramphenicol and Florfenicol.” *FEMS Microbiology Reviews* 28 (5): 519–42.

Seedy, Fawzy R. El, A. A. Samy, Hala S. H. Salam, Eman A. Khairy, and Aya A. Koraney. 2017. “Polymerase Chain Reaction Detection of Genes Responsible for Multiple Antibiotic Resistance Staphylococcus Aureus Isolated from Food of Animal Origin in Egypt.” *Veterinary World* 10 (10): 1205–11.

Seemann, T. n.d. “Snippy: Rapid Haploid Variant Calling and Core Genome Alignment [Internet]. 2018 [cited 2020 Feb 10].”

Seemann, Torsten. 2014. “Prokka: Rapid Prokaryotic Genome Annotation.” *Bioinformatics* 30 (14): 2068–69.

Seki, Masahide, Eri Katsumata, Ayako Suzuki, Sarun Sereewattanawoot, Yoshitaka Sakamoto, Junko Mizushima-Sugano, Sumio Sugano, et al. 2019. “Evaluation and Application of RNA-Seq by MinION.” *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes* 26 (1): 55–65.

Selway, Caitlin A., Raphael Eisenhofer, and Laura S. Weyrich. 2020. “Microbiome Applications for Pathology: Challenges of Low Microbial Biomass Samples during Diagnostic Testing.” *Hip*

International: The Journal of Clinical and Experimental Research on Hip Pathology and Therapy 6 (2): 97–106.

Sereika, Mantas, Rasmus Hansen Kirkegaard, Søren Michael Karst, Thomas Yssing Michaelsen, Emil Aarre Sørensen, Rasmus Dam Wollenberg, and Mads Albertsen. 2022. "Oxford Nanopore R10.4 Long-Read Sequencing Enables the Generation of near-Finished Bacterial Genomes from Pure Cultures and Metagenomes without Short-Read or Reference Polishing." *Nature Methods* 19 (7): 823–26.

Serpa, Paula Hayakawa, Xianding Deng, Mazin Abdelghany, Emily Crawford, Katherine Malcolm, Saharai Caldera, Monica Fung, et al. 2022. "Metagenomic Prediction of Antimicrobial Resistance in Critically Ill Patients with Lower Respiratory Tract Infections." *Genome Medicine* 14 (1): 74.

Sezmis, Aysha L., Laura C. Woods, Anton Y. Peleg, and Michael J. McDonald. 2023. "Horizontal Gene Transfer, Fitness Costs and Mobility Shape the Spread of Antibiotic Resistance Genes into Experimental Populations of Acinetobacter Baylyi." *Molecular Biology and Evolution* 40 (3). https://doi.org/10.1093/molbev/msad028.

Shaikh, Sibhghatulla, Jamale Fatima, Shazi Shakil, Syed Mohd Danish Rizvi, and Mohammad Amjad Kamal. 2015. "Antibiotic Resistance and Extended Spectrum Beta-Lactamases: Types, Epidemiology and Treatment." *Saudi Journal of Biological Sciences* 22 (1): 90–101.

Sharaby, Yehonatan, Orna Nitzan, Ingrid Brettar, Manfred G. Höfle, Avi Peretz, and Malka Halpern. 2019. "Antimicrobial Agent Susceptibilities of Legionella Pneumophila MLVA-8 Genotypes." *Scientific Reports* 9 (1): 6138.

Shen, Wei, Shuai Le, Yan Li, and Fuquan Hu. 2016. "SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation." *PloS One* 11 (10): e0163962.

Shi, Jinhong, Yan Yan, Matthew G. Links, Longhai Li, Jo-Anne R. Dillon, Michael Horsch, and Anthony Kusalik. 2019. “Antimicrobial Resistance Genetic Factor Identification from Whole-Genome Sequence Data Using Deep Feature Selection.” *BMC Bioinformatics* 20 (Suppl 15): 535.

Silva, Kalinga Pavan T., Ganesh Sundar, and Anupama Khare. 2023. “Efflux Pump Gene Amplifications Bypass Necessity of Multiple Target Mutations for Resistance against Dual-Targeting Antibiotic.” *Nature Communications* 14 (1): 3402.

Simar, Shelby R., Blake M. Hanson, and Cesar A. Arias. 2021. “Techniques in Bacterial Strain Typing: Past, Present, and Future.” *Current Opinion in Infectious Diseases* 34 (4): 339–45.

Singleton, Paul, and Diana Sainsbury. 1981. *Introduction to Bacteria: For Students in the Biological Sciences*. Wiley.

Sint, Daniela, Lorna Raso, and Michael Traugott. 2012. “Advances in Multiplex PCR: Balancing Primer Efficiencies and Improving Detection Success.” *Methods in Ecology and Evolution / British Ecological Society* 3 (5): 898–905.

Sipos, Rita, Anna J. Székely, Márton Palatinszky, Sára Révész, Károly Márialigeti, and Marcell Nikolausz. 2007. “Effect of Primer Mismatch, Annealing Temperature and PCR Cycle Number on 16S rRNA Gene-Targetting Bacterial Community Analysis.” *FEMS Microbiology Ecology* 60 (2): 341–50.

Slatko, Barton E., Andrew F. Gardner, and Frederick M. Ausubel. 2018. “Overview of Next-Generation Sequencing Technologies.” *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* 122 (1): e59.

Smillie, Chris, M. Pilar Garcillán-Barcia, M. Victoria Francia, Eduardo P. C. Rocha, and Fernando de la Cruz. 2010. “Mobility of Plasmids.” *Microbiology and Molecular Biology Reviews: MMBR* 74 (3): 434–52.

Smith, Martin R. 2020. “Information Theoretic Generalized Robinson-Foulds Metrics for Comparing Phylogenetic Trees.” *Bioinformatics* 36 (20): 5007–13.

Sobel, Mara L., Didier Hocquet, Lily Cao, Patrick Plesiat, and Keith Poole. 2005. “Mutations in PA3574 (nalD) Lead to Increased MexAB-OprM Expression and Multidrug Resistance in Laboratory and Clinical Isolates of Pseudomonas Aeruginosa.” *Antimicrobial Agents and Chemotherapy* 49 (5): 1782–86.

Sobel, Mara L., Shadi Neshat, and Keith Poole. 2005. “Mutations in PA2491 (mexS) Promote MexT-Dependent mexEF-oprN Expression and Multidrug Resistance in a Clinical Strain of Pseudomonas Aeruginosa.” *Journal of Bacteriology* 187 (4): 1246–53.

Sondi, Ivan, and Branka Salopek-Sondi. 2004. “Silver Nanoparticles as Antimicrobial Agent: A Case Study on E. Coli as a Model for Gram-Negative Bacteria.” *Journal of Colloid and Interface Science* 275 (1): 177–82.

Song, Zewei, Dan Schlatter, Peter Kennedy, Linda L. Kinkel, H. Corby Kistler, Nhu Nguyen, and Scott T. Bates. 2015. “Effort versus Reward: Preparing Samples for Fungal Community Characterization in High-Throughput Sequencing Surveys of Soils.” *PloS One* 10 (5): e0127234.

Spellberg, Brad, and David N. Gilbert. 2014. “The Future of Antibiotics and Resistance: A Tribute to a Career of Leadership by John Bartlett.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 59 Suppl 2 (Suppl 2): S71–75.

Spits, Claudia, Cédric Le Caignec, Martine De Rycke, Lindsey Van Haute, André Van Steirteghem, Inge Liebaers, and Karen Sermon. 2006. “Whole-Genome Multiple Displacement Amplification from Single Cells.” *Nature Protocols* 1 (4): 1965–70.

Stapleton, Paul D., and Peter W. Taylor. 2002. “Methicillin Resistance in Staphylococcus Aureus: Mechanisms and Modulation.” *Science Progress* 85 (Pt 1): 57–72.

Stedtfeld, Robert D., Xueping Guo, Tiffany M. Stedtfeld, Hongjie Sheng, Maggie R. Williams, Kristin Hauschild, Santosh Gunturu, et al. 2018. “Primer Set 2.0 for Highly Parallel qPCR Array Targeting Antibiotic Resistance Genes and Mobile Genetic Elements.” *FEMS Microbiology Ecology* 94 (9). https://doi.org/10.1093/femsec/fiy130.

Stocker, Roman. 2012. “Marine Microbes See a Sea of Gradients.” *Science* 338 (6107): 628–33.

Stoeck, Thorsten, David Bass, Markus Nebel, Richard Christen, Meredith D. M. Jones, Hans-Werner Breiner, and Thomas A. Richards. 2010. “Multiple Marker Parallel Tag Environmental DNA Sequencing Reveals a Highly Complex Eukaryotic Community in Marine Anoxic Water.” *Molecular Ecology* 19 Suppl 1 (March): 21–31.

Strommenger, Birgit, Christiane Schmidt, Guido Werner, Beate Roessle-Lorch, Till T. Bachmann, and Wolfgang Witte. 2007. “DNA Microarray for the Detection of Therapeutically Relevant Antibiotic Resistance Determinants in Clinical Isolates of Staphylococcus Aureus.” *Molecular and Cellular Probes* 21 (3): 161–70.

Stone, Edel, Katrina Campbell, Irene Grant, and Olivia McAuliffe. 2019. “Understanding and Exploiting Phage-Host Interactions.” *Viruses* 11 (6). https://doi.org/10.3390/v11060567.

Stubbs, Samuel C. B., Barbara A. Blacklaws, Benediktus Yohan, Frilasita A. Yudhaputri, Rahma F. Hayati, Brian Schwem, Edsel M. Salvaña, et al. 2020. “Assessment of a Multiplex PCR and Nanopore-Based Method for Dengue Virus Sequencing in Indonesia.” *Virology Journal* 17 (1): 24.

Su, Haochang, Xiaojuan Hu, Wujie Xu, Yu Xu, Guoliang Wen, and Yucheng Cao. 2022. “Diversity, Abundances and Distribution of Antibiotic Resistance Genes and Virulence Factors in the South China Sea Revealed by Metagenomic Sequencing.” *The Science of the Total Environment* 814 (March): 152803.

Sukhum, Kimberley V., Erin P. Newcomer, Candice Cass, Meghan A. Wallace, Caitlin Johnson, Jeremy Fine, Steven Sax, et al. 2022. “Antibiotic-Resistant Organisms Establish Reservoirs in New Hospital Built Environments and Are Related to Patient Blood Infection Isolates.” *Communication & Medicine* 2 (June): 62.

Sun, Yuhong, Zhanwen Cheng, Xiang Li, Qing Yang, Bixi Zhao, Ziqi Wu, and Yu Xia. 2023. “Genome Enrichment of Rare and Unknown Species from Complicated Microbiomes by Nanopore Selective Sequencing.” *Genome Research* 33 (4): 612–21.

Tahmasebi, Hamed, Sanaz Dehbashi, and Mohammad Reza Arabestani. 2020. “Co-Harboring of Mcr-1 and β-Lactamase Genes in Pseudomonas Aeruginosa by High-Resolution Melting Curve Analysis (HRMA): Molecular Typing of Superbug Strains in Bloodstream Infections (BSI).” *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases* 85 (November):104518.

Takenouchi, T., E. Sakagawa, and M. Sugawara. 1999. “Detection of gyrA Mutations among 335 Pseudomonas Aeruginosa Strains Isolated in Japan and Their Susceptibilities to Fluoroquinolones.” *Antimicrobial Agents and Chemotherapy* 43 (2): 406–9.

Tokuda, Maho, and Masaki Shintani. 2024. “Microbial Evolution through Horizontal Gene Transfer by Mobile Genetic Elements.” *Microbial Biotechnology* 17 (1): e14408.

Torrens, Gabriel, Sara Belén Hernández, Juan Alfonso Ayala, Bartolome Moya, Carlos Juan, Felipe Cava, and Antonio Oliver. 2019. “Regulation of AmpC-Driven β-Lactam Resistance in Pseudomonas

Aeruginosa: Different Pathways, Different Signaling." *mSystems* 4 (6). https://doi.org/10.1128/mSystems.00524-19.

Talapko, Jasminka, Erwin Frauenheim, Martina Juzbašić, Matej Tomas, Suzana Matić, Melita Jukić, Marija Samardžić, and Ivana Škrlec. 2022. "Legionella pneumophila—Virulence Factors and the Possibility of Infection in Dental Practice." *Microorganisms* 10 (2): 255.

Taneja, Neelam, Garima Sangar, Goutam Chowdhury, T. Ramamurthy, Arti Mishra, Meenakshi Singh, and Meera Sharma. 2012. "Molecular Epidemiology of Vibrio Cholerae Causing Outbreaks & Sporadic Cholera in Northern India." *The Indian Journal of Medical Research* 136 (4): 656–63.

Tang, Patrick H. P., Andrew Worster, Jocelyn A. Srigley, and Cheryl L. Main. 2011. "Examination of Staphylococcal Stethoscope Contamination in the Emergency Department (pilot) Study (EXSSCITED Pilot Study)." *CJEM* 13 (4): 239–44.

Tankovic, J., B. Perichon, J. Duval, and P. Courvalin. 1996. "Contribution of Mutations in gyrA and parC Genes to Fluoroquinolone Resistance of Mutants of Streptococcus Pneumoniae Obtained in Vivo and in Vitro." *Antimicrobial Agents and Chemotherapy* 40 (11): 2505–10.

Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin. 2000. "The COG Database: A Tool for Genome-Scale Analysis of Protein Functions and Evolution." *Nucleic Acids Research* 28 (1): 33–36.

Teng, Q. I., Yi Liu, and Others. 2017. "Gene Typing of Legionella Pneumophila Serogroup 1 in Sichuan Province, China." *Zhongguo Ren Shou Gong Huan Bing Za Zhi = Chinese Journal of Zoonoses* 33 (9): 784–88.

Terranova, W., M. L. Cohen, and D. W. Fraser. 1978. "1974 Outbreak of Legionnaires' Disease Diagnosed in 1977. Clinical and Epidemiological Features." *The Lancet* 2 (8081): 122–24.

Terrazos Miani, Miguel A., Loïc Borcard, Sonja Gempeler, Christian Baumann, Pascal Bittel, Stephen L. Leib, Stefan Neuenschwander, and Alban Ramette. 2023. “NASCarD (Nanopore Adaptive Sampling with Carrier DNA): A Rapid, PCR-Free Method for Whole Genome Sequencing of Pathogens in Clinical Samples.” *bioRxiv*. https://doi.org/10.1101/2023.03.10.23287094.

Thiolas, Aurélie, Charléric Bornet, Anne Davin-Régli, Jean-Marie Pagès, and Claude Bollet. 2004. “Resistance to Imipenem, Cefepime, and Cefpirome Associated with Mutation in Omp36 Osmoporin of Enterobacter Aerogenes.” *Biochemical and Biophysical Research Communications* 317 (3): 851–56.

Thisted Lambertz, S., and M-L Danielsson-Tham. 2005. “Identification and Characterization of Pathogenic Yersinia Enterocolitica Isolates by PCR and Pulsed-Field Gel Electrophoresis.” *Applied and Environmental Microbiology* 71 (7): 3674–81.

Thomsen, Philip Francis, Jos Kielgast, Lars Lønsmann Iversen, Peter Rask Møller, Morten Rasmussen, and Eske Willerslev. 2012. “Detection of a Diverse Marine Fish Fauna Using Environmental DNA from Seawater Samples.” *PloS One* 7 (8): e41732.

Thurlow, Charles M., Sandeep J. Joseph, Lilia Ganova-Raeva, Samantha S. Katz, Lara Pereira, Cheng Chen, Alyssa Debra, et al. 2022. “Selective Whole-Genome Amplification as a Tool to Enrich Specimens with Low Treponema Pallidum Genomic DNA Copies for Whole-Genome Sequencing.” *mSphere* 7 (3): e0000922.

Tijet, Nathalie, Patrick Tang, Mya Romilowych, Carla Duncan, Victoria Ng, David N. Fisman, Frances Jamieson, Donald E. Low, and Cyril Guyard. 2010. “New Endemic Legionella Pneumophila Serogroup I Clones, Ontario, Canada.” *Emerging Infectious Diseases* 16 (3): 447–54.

Tolosi, Roberta, Lisa Carraro, Andrea Laconi, and Alessandra Piccirillo. 2021. “Optimisation of Five qPCR Protocols toward the Detection and the Quantification of Antimicrobial Resistance Genes in Environmental Samples.” *MethodsX* 8 (August): 101488

Travers, Kevin J., Chen-Shan Chin, David R. Rank, John S. Eid, and Stephen W. Turner. 2010. “A Flexible and Efficient Template Format for Circular Consensus Sequencing and SNP Detection.” *Nucleic Acids Research* 38 (15): e159.

Treakle, Amy M., Kerri A. Thom, Jon P. Furuno, Sandra M. Strauss, Anthony D. Harris, and Eli N. Perencevich. 2009. “Bacterial Contamination of Health Care Workers’ White Coats.” *American Journal of Infection Control* 37 (2): 101–5.

Treepong, P., V. N. Kos, C. Guyeux, D. S. Blanc, X. Bertrand, B. Valot, and D. Hocquet. 2018. “Global Emergence of the Widespread Pseudomonas Aeruginosa ST235 Clone.” *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 24 (3): 258–66.

Turner, S., K. M. Pryer, V. P. Miao, and J. D. Palmer. 1999. “Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis.” *The Journal of Eukaryotic Microbiology* 46 (4): 327–38.

Tyson, Gregory H., Patrick F. McDermott, Cong Li, Yuansha Chen, Daniel A. Tadesse, Sampa Mukherjee, Sonya Bodeis-Jones, et al. 2015. “WGS Accurately Predicts Antimicrobial Resistance in Escherichia Coli.” *The Journal of Antimicrobial Chemotherapy* 70 (10): 2763–69.

Tyson, John R., Phillip James, David Stoddart, Natalie Sparks, Arthur Wickenhagen, Grant Hall, Ji Hyun Choi, et al. 2020. “Improvements to the ARTIC Multiplex PCR Method for SARS-CoV-2 Genome Sequencing Using Nanopore.” *bioRxiv : The Preprint Server for Biology*, September. https://doi.org/10.1101/2020.09.04.283077.

Unemo, M., R. Ballard, C. Ison, D. Lewis, and F. Ndowa. 2013. “Laboratory Diagnosis of Sexually Transmitted Infections, Including Human Immunodeficiency Virus.” https://apps.who.int/iris/bitstream/handle/10665/85343/?sequence=1.

Urakawa, Hidetoshi, Willm Martens-Habbena, and David A. Stahl. 2010. “High Abundance of Ammonia-Oxidizing Archaea in Coastal Waters, Determined Using a Modified DNA Extraction Method.” *Applied and Environmental Microbiology* 76 (7): 2129–35.

Van Acker, Heleen, Patrick Van Dijck, and Tom Coenye. 2014. “Molecular Mechanisms of Antimicrobial Tolerance and Resistance in Bacterial and Fungal Biofilms.” *Trends in Microbiology* 22 (6): 326–33.

Váradi, Linda, Jia Lin Luo, David E. Hibbs, John D. Perry, Rosaleen J. Anderson, Sylvain Orenga, and Paul W. Groundwater. 2017. “Methods for the Detection and Identification of Pathogenic Bacteria: Past, Present, and Future.” *Chemical Society Reviews* 46 (16): 4818–32.

Vedantam, G., G. G. Guay, N. E. Austria, S. Z. Doktor, and B. P. Nichols. 1998. “Characterization of Mutations Contributing to Sulfathiazole Resistance in Escherichia Coli.” *Antimicrobial Agents and Chemotherapy* 42 (1): 88–93.

Venbrux, Marc, Sam Crauwels, and Hans Rediers. 2023. “Current and Emerging Trends in Techniques for Plant Pathogen Detection.” *Frontiers in Plant Science* 14 (May): 1120968.

Vereecke, Nick, Nadine Botteldoorn, Charlotte Brossé, Caroline Bonckaert, Hans Nauwynck, Freddy Haesebrouck, Filip Boyen, Dominiek Maes, and Sebastiaan Theuns. 2023. “Predictive Power of Long-Read Whole-Genome Sequencing for Rapid Diagnostics of Multidrug-Resistant Brachyspira Hyodysenteriae Strains.” *Microbiology Spectrum* 11 (1): e0412322.

Viehweger, Adrian, Mike Marquet, Martin Hölzer, Nadine Dietze, Mathias W. Pletz, and Christian Brandt. 2023. “Nanopore-Based Enrichment of Antimicrobial Resistance Genes - a Case-Based Study.” *GigaByte (Hong Kong, China)* 2023 (January): gigabyte75.

Vrancianu, Corneliu Ovidiu, Laura Ioana Popa, Coralia Bleotu, and Mariana Carmen Chifiriuc. 2020. “Targeting Plasmids to Limit Acquisition and Transmission of Antimicrobial Resistance.” *Frontiers in Microbiology* 11 (May):761.

Wang, Tiantian, Peilong Li, Qiuchen Qi, Shujun Zhang, Yan Xie, Jing Wang, Shibiao Liu, et al. 2023. “A Multiplex Blood-Based Assay Targeting DNA Methylation in PBMCs Enables Early Detection of Breast Cancer.” *Nature Communications* 14 (1): 4724

Wang, Xiaoming, Yi Gao, Xiao Liu, Naiyan Sun, Jinhu Huang, and Liping Wang. 2021. “First Report of the Plasmid-Mediated fosB Gene in Enterococcus Faecalis from Pigs.” *Genes* 12 (11). https://doi.org/10.3390/genes12111684.

Wang, Xiaoyu, Yapeng Liu, Hongna Liu, Wenjing Pan, Jie Ren, Xiangming Zheng, Yimin Tan, et al. 2022. “Recent Advances and Application of Whole Genome Amplification in Molecular Diagnosis and Medicine.” *MedComm* 3 (1): e116.

Wang, Yiqing, and Tal Dagan. 2024. “The Evolution of Antibiotic Resistance Islands Occurs within the Framework of Plasmid Lineages.” *Nature Communications* 15 (1): 4555.

Wang, Yong, Xiushuo Zhang, Jing Xu, Xiangyu Sun, Xiaolong Zhao, Hongsheng Li, Yanping Liu, et al. 2022. “The Development of Microscopic Imaging Technology and Its Application in Micro- and Nanotechnology.” *Frontiers in Chemistry* 10 (July): 931169.

Wang, Yunhao, Yue Zhao, Audrey Bollas, Yuru Wang, and Kin Fai Au. 2021. “Nanopore Sequencing Technology, Bioinformatics and Applications.” *Nature Biotechnology* 39 (11): 1348–65.

Waseem, Hassan, Sana Jameel, Jafar Ali, Hamza Saleem Ur Rehman, Isfahan Tauseef, Uzma Farooq, Asif Jamal, and Muhammad Ishtiaq Ali. 2019. “Contributions and Challenges of High Throughput qPCR for Determining Antimicrobial Resistance in the Environment: A Critical Review.” *Molecules* 24 (1). https://doi.org/10.3390/molecules24010163.

Waterhouse, Andrew M., James B. Procter, David M. A. Martin, Michèle Clamp, and Geoffrey J. Barton. 2009. “Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench.” *Bioinformatics* 25 (9): 1189–91.

Webber, M. A., and L. J. V. Piddock. 2003. “The Importance of Efflux Pumps in Bacterial Antibiotic Resistance.” *The Journal of Antimicrobial Chemotherapy* 51 (1): 9–11.

Wei, Hua, and Leiv Sigve Håvarstein. 2012. “Fratricide Is Essential for Efficient Gene Transfer between Pneumococci in Biofilms.” *Applied and Environmental Microbiology* 78 (16): 5897–5905.

Wellington, Elizabeth M. H., Alistair B. Boxall, Paul Cross, Edward J. Feil, William H. Gaze, Peter M. Hawkey, Ashley S. Johnson-Rollings, et al. 2013. “The Role of the Natural Environment in the Emergence of Antibiotic Resistance in Gram-Negative Bacteria.” *The Lancet Infectious Diseases* 13 (2): 155–65.

Whittle, Gabrielle, Terence R. Whitehead, Nathan Hamburger, Nadja B. Shoemaker, Michael A. Cotta, and Abigail A. Salyers. 2003. “Identification of a New Ribosomal Protection Type of Tetracycline Resistance Gene, tet(36), from Swine Manure Pits.” *Applied and Environmental Microbiology* 69 (7): 4151–58.

Wick, Ryan R., Louise M. Judd, Claire L. Gorrie, and Kathryn E. Holt. 2017. “Unicycler: Resolving Bacterial Genome Assemblies from Short and Long Sequencing Reads.” *PLoS Computational Biology* 13 (6): e1005595.

Wick, Ryan R., Louise M. Judd, and Kathryn E. Holt. 2019. “Performance of Neural Network Basecalling Tools for Oxford Nanopore Sequencing.” *Genome Biology* 20 (1): 129.

Wick, Ryan R., Mark B. Schultz, Justin Zobel, and Kathryn E. Holt. 2015. “Bandage: Interactive Visualisation of de Novo Genome Assemblies.” *Bioinformatics* 31 (20): 3350–52.

Wickham, Hadley. 2016. ggplot2: Elegant Graphics for Data Analysis. *Springer*.

Wikramaratna, Paul S., Robert S. Paton, Mahan Ghafari, and José Lourenço. 2020. “Estimating the False-Negative Test Probability of SARS-CoV-2 by RT-PCR.” *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 25 (50). https://doi.org/10.2807/1560-7917.ES.2020.25.50.2000568.

Wilder-Smith, Annelies, Calvin J. Chiew, and Vernon J. Lee. 2020. “Can We Contain the COVID-19 Outbreak with the Same Measures as for SARS?” *The Lancet Infectious Diseases* 20 (5): e102–7.

Wilson, Daniel N., Vasili Hauryliuk, Gemma C. Atkinson, and Alex J. O’Neill. 2020. “Target Protection as a Key Antibiotic Resistance Mechanism.” *Nature Reviews. Microbiology* 18 (11): 637–48.

Winn, W. C., Jr. 1988. “Legionnaires Disease: Historical Perspective.” *Clinical Microbiology Reviews* 1 (1): 60–81.

Wölfel, Roman, Victor M. Corman, Wolfgang Guggemos, Michael Seilmaier, Sabine Zange, Marcel A. Müller, Daniela Niemeyer, et al. 2020. “Virological Assessment of Hospitalized Patients with COVID-2019.” *Nature* 581 (7809): 465–69.

Wu, Dong, Ling Jin, Jiawen Xie, Hang Liu, Jue Zhao, Dan Ye, and Xiang-Dong Li. 2022. “Inhalable Antibiotic Resistomes Emitted from Hospitals: Metagenomic Insights into Bacterial Hosts, Clinical Relevance, and Environmental Risks.” *Microbiome* 10 (1): 19.

Wu, Minghong, Jiajia Xiang, Chenjing Que, Fenfen Chen, and Gang Xu. 2015. “Occurrence and Fate of Psychiatric Pharmaceuticals in the Urban Water System of Shanghai, China.” *Chemosphere* 138 (November): 486–93.

Wu, Xiaofang, Yongying Zhu, Wei Yan, Peng Zhang, and Liping Chen. 2024. “Pathogenic Characteristics of the Vibrio Parahaemolyticus Which Caused a Gastroenteritis Outbreak Event in Huzhou.” *FEMS Microbiology Letters* 371 (January). https://doi.org/10.1093/femsle/fnad130.

Wu, Ziqi, You Che, Chenyuan Dang, Miao Zhang, Xuyang Zhang, Yuhong Sun, Xiang Li, Tong Zhang, and Yu Xia. 2022. “Nanopore-Based Long-Read Metagenomics Uncover the Resistome Intrusion by Antibiotic Resistant Bacteria from Treated Wastewater in Receiving Water Body.” *Water Research* 226 (November):119282.

Xia, Yu, An-Dong Li, Yu Deng, Xiao-Tao Jiang, Li-Guan Li, and Tong Zhang. 2017. “MinION Nanopore Sequencing Enables Correlation between Resistome Phenotype and Genotype of Coliform Bacteria in Municipal Sewage.” *Frontiers in Microbiology* 8 (October): 2105.

Xie, Xingzhi, Zheng Zhong, Wei Zhao, Chao Zheng, Fei Wang, and Jun Liu. 2020. “Chest CT for Typical Coronavirus Disease 2019 (COVID-19) Pneumonia: Relationship to Negative RT-PCR Testing.” *Radiology* 296 (2): E41–45.

Xu, Jingwei, and Ye Xiang. 2017. “Membrane Penetration by Bacterial Viruses.” *Journal of Virology* 91 (13). https://doi.org/10.1128/JVI.00162-17.

Xu, Ying, Hongbing Luo, Zhili Wang, Hon-Ming Lam, and Cheng Huang. 2022. “Oxford Nanopore Technology: Revolutionizing Genomics Research in Plants.” *Trends in Plant Science* 27 (5): 510–11.

Yamada, Tomoyuki, Haruhiko Soma, and Shinichi Morishita. 2006. “PrimerStation: A Highly Specific Multiplex Genomic PCR Primer Design Server for the Human Genome.” *Nucleic Acids Research* 34 (Web Server issue): W665–69.

Yamin, Dina, Vuk Uskoković, Abubakar Muhammad Wakil, Mohammed Dauda Goni, Shazana Hilda Shamsuddin, Fatin Hamimi Mustafa, Wadha A. Alfouzan, et al. 2023. “Current and Future Technologies for the Detection of Antibiotic-Resistant Bacteria.” *Diagnostics (Basel, Switzerland)* 13 (20). https://doi.org/10.3390/diagnostics13203246.

Yan, Jinyuan, Henri Estanbouli, Chen Liao, Wook Kim, Jonathan M. Monk, Rayees Rahman, Mini Kamboj, Bernhard O. Palsson, Weigang Qiu, and Joao B. Xavier. 2019. “Systems-Level Analysis of NalD Mutation, a Recurrent Driver of Rapid Drug Resistance in Acute Pseudomonas Aeruginosa Infection.” *PLoS Computational Biology* 15 (12): e1007562.

Yan, Shigan, Wencheng Zhang, Chengyu Li, Xu Liu, Liping Zhu, Leilei Chen, and Baowei Yang. 2021. “Serotyping, MLST, and Core Genome MLST Analysis of Salmonella Enterica From Different Sources in China During 2004–2019.” *Frontiers in Microbiology* 12. https://doi.org/10.3389/fmicb.2021.688614.

Yang, Dong, Wenxin Wang, Liang Zhao, Lei Rao, and Xiaojun Liao. 2023. “Resuscitation of Viable but Nonculturable Bacteria Promoted by ATP-Mediated NAD+ Synthesis.” *Journal of Advertising Research*, August. https://doi.org/10.1016/j.jare.2023.08.002.

Yang, Jun, Jinyi Xiang, Yu Xie, Kaifeng Yu, Junnan Li, Haoyan Wang, Peng Li, Karina Yew-Hoong Gin, and Yiliang He. 2022. “Removal Behavior and Key Drivers of Antibiotic Resistance Genes in Two Full-Scale Leachate Treatment Plants.” *Water Research* 226 (November): 119239.

Yang, Shengzhi, Xin Gao, Jianghong Meng, Anyun Zhang, Yingmin Zhou, Mei Long, Bei Li, et al. 2018. "Metagenomic Analysis of Bacteria, Fungi, Bacteriophages, and Helminths in the Gut of Giant Pandas." *Frontiers in Microbiology* 9 (July): 1717.

Yang, Wan, Subbiah Elankumaran, and Linsey C. Marr. 2012. "Relationship between Humidity and Influenza A Viability in Droplets and Implications for Influenza's Seasonality." *PloS One* 7 (10): e46789.

Yang, Ying, Xiaotao Jiang, Benli Chai, Liping Ma, Bing Li, Anni Zhang, James R. Cole, James M. Tiedje, and Tong Zhang. 2016. "ARGs-OAP: Online Analysis Pipeline for Antibiotic Resistance Genes Detection from Metagenomic Data Using an Integrated Structured ARG-Database." *Bioinformatics* 32 (15): 2346–51.

Yang, Yu, You Che, Lei Liu, Chunxiao Wang, Xiaole Yin, Yu Deng, Chao Yang, and Tong Zhang. 2022. "Rapid Absolute Quantification of Pathogens and ARGs by Nanopore Sequencing." *The Science of the Total Environment* 809 (February): 152190.

Yang, Ziheng, and Bruce Rannala. 2012. "Molecular Phylogenetics: Principles and Practice." *Nature Reviews. Genetics* 13 (5): 303–14.

Yao, Kuan, Narjol González-Escalona, and Maria Hoffmann. 2021. "Multiple Displacement Amplification as a Solution for Low Copy Number Plasmid Sequencing." *Frontiers in Microbiology* 12 (February): 617487.

Ye, Fenxia, Ge Peng, and Ying Li. 2011. "Influences of Influent Carbon Source on Extracellular Polymeric Substances (EPS) and Physicochemical Properties of Activated Sludge." *Chemosphere* 84 (9): 1250–55.

Yoon, Eun-Jeong, Catherine Grillot-Courvalin, and Patrice Courvalin. 2017. “New Aminoglycoside-Modifying Enzymes APH(3’)-VIII and APH(3')-IX in Acinetobacter Rudis and Acinetobacter Gerneri.” *The Journal of Antibiotics* 70 (4): 400–403.

Yu, Qian, Yuanyuan Wang, Zhiqin Gao, Hong Yang, Siyu Liu, Jingwen Tan, and Lianjuan Yang. 2023. “DNA Microarray Chip Assay in New Use: Early Diagnostic Value in Cutaneous Mycobacterial Infection.” *Frontiers in Cellular and Infection Microbiology* 13 (July): 1183078.

Zaheer, Rahat, Shaun R. Cook, Ruth Barbieri, Noriko Goji, Andrew Cameron, Aaron Petkau, Rodrigo Ortega Polo, et al. 2020. “Surveillance of Enterococcus Spp. Reveals Distinct Species and Antimicrobial Resistance Diversity across a One-Health Continuum.” *Scientific Reports* 10 (1): 3937.

Zaki, Ali M., Sander van Boheemen, Theo M. Bestebroer, Albert D. M. E. Osterhaus, and Ron A. M. Fouchier. 2012. “Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia.” *The New England Journal of Medicine* 367 (19): 1814–20.

Zayed, Ashraf R., Marina Pecellin, Alaa Salah, Hanna Alalam, Suha Butmeh, Michael Steinert, Rene Lesnik, Ingrid Brettar, Manfred G. Höfle, and Dina M. Bitar. 2020. “Characterization of Legionella Pneumophila Populations by Multilocus Variable Number of Tandem Repeats (MLVA) Genotyping from Drinking Water and Biofilm in Hospitals from Different Regions of the West Bank.” *Pathogens* 9 (11). https://doi.org/10.3390/pathogens9110862.

Zeng, Lin, and Shouguang Jin. 2003. “aph(3’)-IIb, a Gene Encoding an Aminoglycoside-Modifying Enzyme, Is under the Positive Control of Surrogate Regulator HpaA.” *Antimicrobial Agents and Chemotherapy* 47 (12): 3867–76.

Zhang, Chi, Leshan Xiu, Yamei Li, Liying Sun, Yizhun Li, Yaling Zeng, Feng Wang, and Junping Peng. 2021. “Multiplex PCR and Nanopore Sequencing of Genes Associated with Antimicrobial Resistance in Neisseria Gonorrhoeae Directly from Clinical Samples.” *Clinical Chemistry* 67 (4): 610–20.

Zhang, Fan, Shouwang Hu, Jian Huang, Hui Wang, Zheng Wen, Geng Yongyao, and Shengqi Wang. 2006. “Development and Clinical Evaluation of Oligonucleotide Microarray for HLA-AB Genotyping.” *Pharmacogenomics* 7 (7): 973–85.

Zhao, Wenxuan, Wei Zeng, Bo Pang, Ming Luo, Yao Peng, Jialiang Xu, Biao Kan, Zhenpeng Li, and Xin Lu. 2023. “Oxford Nanopore Long-Read Sequencing Enables the Generation of Complete Bacterial and Plasmid Genomes without Short-Read Sequencing.” *Frontiers in Microbiology* 14 (May): 1179966.

Zheng, Jinshui, Stijn Wittouck, Elisa Salvetti, Charles M. A. P. Franz, Hugh M. B. Harris, Paola Mattarelli, Paul W. O’Toole, et al. 2020. “A Taxonomic Note on the Genus Lactobacillus: Description of 23 Novel Genera, Emended Description of the Genus Lactobacillus Beijerinck 1901, and Union of Lactobacillaceae and Leuconostocaceae.” *International Journal of Systematic and Evolutionary Microbiology* 70 (4): 2782–2858.

Zhou, Jizhong, Ye Deng, Lina Shen, Chongqing Wen, Qingyun Yan, Daliang Ning, Yujia Qin, et al. 2016. “Temperature Mediates Continental-Scale Diversity of Microbes in Forest Soils.” *Nature Communications* 7 (July): 12083.

Zhou, Xue, Min Qiao, Jian-Qiang Su, and Yong-Guan Zhu. 2019. “High-Throughput Characterization of Antibiotic Resistome in Soil Amended with Commercial Organic Fertilizers.” *Journal of Soils and Sediments*. https://doi.org/10.1007/s11368-018-2064-6.

Zhou, Zhen-Chao, Yang Liu, Ze-Jun Lin, Xin-Yi Shuai, Lin Zhu, Lan Xu, Ling-Xuan Meng, Yu-Jie Sun, and Hong Chen. 2021. “Spread of Antibiotic Resistance Genes and Microbiota in Airborne Particulate Matter, Dust, and Human Airways in the Urban Hospital.” *Environment International* 153 (August): 106501.

Zhu, Dong, Jing Ding, Yue Yin, Xin Ke, Patrick O'Connor, and Yong-Guan Zhu. 2020. "Effects of Earthworms on the Microbiomes and Antibiotic Resistomes of Detritus Fauna and Phyllospheres." *Environmental Science & Technology* 54 (10): 6000–6008.

Zong, Chenghang, Sijia Lu, Alec R. Chapman, and X. Sunney Xie. 2012. "Genome-Wide Detection of Single-Nucleotide and Copy-Number Variations of a Single Human Cell." *Science* 338 (6114): 1622–26.

Zou, Wen, Hailin Tang, Weizhong Zhao, Joe Meehan, Steven L. Foley, Wei-Jiun Lin, Hung-Chia Chen, Hong Fang, Rajesh Nayak, and James J. Chen. 2013. "Data Mining Tools for Salmonella Characterization: Application to Gel-Based Fingerprinting Analysis." *BMC Bioinformatics* 14 Suppl 14 (October): S15.

Appendix 1 Supplementary tables

Table S3.1 Samples and barcoded primers information for *16S rRNA* amplicon sequencing

sample number	barcode+primer, forward (5'-3')	barcode+primer, reverse (5'-3')	department	medium	season2
K04	GGCTACGTGCCAGCMGCCGCGGTAA	CACCGGCCGTCAATTCMTTTRAGTTT	Blood	water	early-summer
L04	GGCTACGTGCCAGCMGCCGCGGTAA	CACGATCCGTCAATTCMTTTRAGTTT	Blood	water	early-summer
J04	GGCTACGTGCCAGCMGCCGCGGTAA	CACTCACCGTCAATTCMTTTRAGTTT	Emergency	water	early-summer
Q04	GGCTACGTGCCAGCMGCCGCGGTAA	CAGGCGCCGTCAATTCMTTTRAGTTT	Emergency	air	early-summer
M04	GGCTACGTGCCAGCMGCCGCGGTAA	CATGGCCCGTCAATTCMTTTRAGTTT	gynecology	water	early-summer
H04	GGCTACGTGCCAGCMGCCGCGGTAA	CATTTTCCGTCAATTCMTTTRAGTTT	ophthalmology	water	early-summer
N04	GGCTACGTGCCAGCMGCCGCGGTAA	CCAACACCGTCAATTCMTTTRAGTTT	ophthalmology	air	early-summer
I04	GGCTACGTGCCAGCMGCCGCGGTAA	CGGAATCCGTCAATTCMTTTRAGTTT	otolaryngology	water	early-summer
R04	GGCTACGTGCCAGCMGCCGCGGTAA	CTAGCTCCGTCAATTCMTTTRAGTTT	otolaryngology	air	early-summer
O04	GGCTACGTGCCAGCMGCCGCGGTAA	CTATACCCGTCAATTCMTTTRAGTTT	pneumology	air	early-summer
N26	CTTGTAGTGCCAGCMGCCGCGGTAA	CACCGGCCGTCAATTCMTTTRAGTTT	Blood	water	early-summer
O26	CTTGTAGTGCCAGCMGCCGCGGTAA	CACGATCCGTCAATTCMTTTRAGTTT	Blood	water	early-summer
L26	CTTGTAGTGCCAGCMGCCGCGGTAA	CACTCACCGTCAATTCMTTTRAGTTT	emergency	water	early-summer
P26	CTTGTAGTGCCAGCMGCCGCGGTAA	CAGGCGCCGTCAATTCMTTTRAGTTT	gynecology	water	early-summer
A26	CTTGTAGTGCCAGCMGCCGCGGTAA	CATGGCCCGTCAATTCMTTTRAGTTT	sewage	water	early-summer

J26	CTTGTAGTGCCAGCMGCCGCGGTAA	CATTTTCCGTCAATTCMTTTRAGTTT	ophthalmology	water	early-summer
K26	CTTGTAGTGCCAGCMGCCGCGGTAA	CCAACACCGTCAATTCMTTTRAGTTT	otolaryngology	water	early-summer
B26	CTTGTAGTGCCAGCMGCCGCGGTAA	CGGAATCCGTCAATTCMTTTRAGTTT	sewage	water	early-summer
R26	CTTGTAGTGCCAGCMGCCGCGGTAA	CTAGCTCCGTCAATTCMTTTRAGTTT	pneumology	air	early-summer
ptrap	CTTGTAGTGCCAGCMGCCGCGGTAA	CTATACCCGTCAATTCMTTTRAGTTT	surgery	water	early-summer
M26	GGCTACGTGCCAGCMGCCGCGGTAA	CACCGGCCGTCAATTCMTTTRAGTTT	surgery	water	early-summer
N09	GGCTACGTGCCAGCMGCCGCGGTAA	CACGATCCGTCAATTCMTTTRAGTTT	Blood	water	late-summer
O09	GGCTACGTGCCAGCMGCCGCGGTAA	CACTCACCGTCAATTCMTTTRAGTTT	Blood	water	late-summer
L09	GGCTACGTGCCAGCMGCCGCGGTAA	CAGGCGCCGTCAATTCMTTTRAGTTT	emergency	water	late-summer
S09	GGCTACGTGCCAGCMGCCGCGGTAA	CATGGCCCGTCAATTCMTTTRAGTTT	emergency	air	late-summer
P09	GGCTACGTGCCAGCMGCCGCGGTAA	CATTTTCCGTCAATTCMTTTRAGTTT	gynecology	water	late-summer
J09	GGCTACGTGCCAGCMGCCGCGGTAA	CCAACACCGTCAATTCMTTTRAGTTT	ophthalmology	water	late-summer
Q09	GGCTACGTGCCAGCMGCCGCGGTAA	CGGAATCCGTCAATTCMTTTRAGTTT	ophthalmology	air	late-summer
K09	GGCTACGTGCCAGCMGCCGCGGTAA	CTAGCTCCGTCAATTCMTTTRAGTTT	otolaryngology	water	late-summer
U09	GGCTACGTGCCAGCMGCCGCGGTAA	CTATACCCGTCAATTCMTTTRAGTTT	otolaryngology	air	late-summer
M09	CTTGTAGTGCCAGCMGCCGCGGTAA	CACCGGCCGTCAATTCMTTTRAGTTT	pneumology	water	late-summer

R09	CTTGTAGTGCCAGCMGCCGCGGTAA	CACGATCCGTCAATTCMTTTRAGTTT	pneumology	air	late-summer
T09	CTTGTAGTGCCAGCMGCCGCGGTAA	CACTCACCGTCAATTCMTTTRAGTTT	surgery	air	late-summer
N28	CTTGTAGTGCCAGCMGCCGCGGTAA	CAGGCGCCGTCAATTCMTTTRAGTTT	Blood	water	late-summer
O28	CTTGTAGTGCCAGCMGCCGCGGTAA	CATGGCCCGTCAATTCMTTTRAGTTT	Blood	water	late-summer
L28	CTTGTAGTGCCAGCMGCCGCGGTAA	CATTTTCCGTCAATTCMTTTRAGTTT	emergency	water	late-summer
P28	CTTGTAGTGCCAGCMGCCGCGGTAA	CCAACACCGTCAATTCMTTTRAGTTT	gynecology	water	late-summer
A28	CTTGTAGTGCCAGCMGCCGCGGTAA	CGGAATCCGTCAATTCMTTTRAGTTT	sewage	water	late-summer
J28	CTTGTAGTGCCAGCMGCCGCGGTAA	CTAGCTCCGTCAATTCMTTTRAGTTT	ophthalmology	water	late-summer
K28	CTTGTAGTGCCAGCMGCCGCGGTAA	CTATACCCGTCAATTCMTTTRAGTTT	otolaryngology	water	late-summer
B28	AGTCAAGTGCCAGCMGCCGCGGTAA	CACCGGCCGTCAATTCMTTTRAGTTT	sewage	water	late-summer
M28	AGTCAAGTGCCAGCMGCCGCGGTAA	CACGATCCGTCAATTCMTTTRAGTTT	surgery	water	late-summer
O30	GGCTACGTGCCAGCMGCCGCGGTAA	CACCGGCCGTCAATTCMTTTRAGTTT	Blood	water	winter
P30	GGCTACGTGCCAGCMGCCGCGGTAA	CACGATCCGTCAATTCMTTTRAGTTT	Blood	water	winter
M30	GGCTACGTGCCAGCMGCCGCGGTAA	CACTCACCGTCAATTCMTTTRAGTTT	emergency	water	winter
Q30	GGCTACGTGCCAGCMGCCGCGGTAA	CAGGCGCCGTCAATTCMTTTRAGTTT	gynecology	water	winter
K30	GGCTACGTGCCAGCMGCCGCGGTAA	CATGGCCCGTCAATTCMTTTRAGTTT	ophthalmology	water	winter

L30	GGCTACGTGCCAGCMGCCGCGGTAA	CATTTTCCGTCAATTCMTTTRAGTTT	otolaryngology	water	winter
R30	GGCTACGTGCCAGCMGCCGCGGTAA	CCAACACCGTCAATTCMTTTRAGTTT	pneumology	water	winter
O12	GGCTACGTGCCAGCMGCCGCGGTAA	CGGAATCCGTCAATTCMTTTRAGTTT	Blood	water	winter
P12	GGCTACGTGCCAGCMGCCGCGGTAA	CTAGCTCCGTCAATTCMTTTRAGTTT	Blood	water	winter
M12	GGCTACGTGCCAGCMGCCGCGGTAA	CTATACCCGTCAATTCMTTTRAGTTT	emergency	water	winter
Q12	CTTGTAGTGCCAGCMGCCGCGGTAA	CACCGGCCGTCAATTCMTTTRAGTTT	gynecology	water	winter
K12	CTTGTAGTGCCAGCMGCCGCGGTAA	CACGATCCGTCAATTCMTTTRAGTTT	ophthalmology	water	winter
L12	CTTGTAGTGCCAGCMGCCGCGGTAA	CACTCACCGTCAATTCMTTTRAGTTT	otolaryngology	water	winter
R12	CTTGTAGTGCCAGCMGCCGCGGTAA	CAGGCGCCGTCAATTCMTTTRAGTTT	pneumology	water	winter
N12	CTTGTAGTGCCAGCMGCCGCGGTAA	CATGGCCCGTCAATTCMTTTRAGTTT	surgery	water	winter
O01	CTTGTAGTGCCAGCMGCCGCGGTAA	CATTTTCCGTCAATTCMTTTRAGTTT	Blood	water	winter
P01	CTTGTAGTGCCAGCMGCCGCGGTAA	CCAACACCGTCAATTCMTTTRAGTTT	Blood	water	winter
Q01	CTTGTAGTGCCAGCMGCCGCGGTAA	CGGAATCCGTCAATTCMTTTRAGTTT	gynecology	water	winter
K01	CTTGTAGTGCCAGCMGCCGCGGTAA	CTAGCTCCGTCAATTCMTTTRAGTTT	ophthalmology	water	winter
L01	CTTGTAGTGCCAGCMGCCGCGGTAA	CTATACCCGTCAATTCMTTTRAGTTT	otolaryngology	water	winter
R01	AGTCAAGTGCCAGCMGCCGCGGTAA	CACCGGCCGTCAATTCMTTTRAGTTT	pneumology	water	winter

N01	AGTCAAGTGCCAGCMGCCGCGGTAA	CACGATCCGTCAATTCMTTTRAGTTT	surgery	water	winter

Table S3.2 Primer sets utilised in HT-qPCR

Gene	Forward	Reverse	Drug	Assay
16s rRNA	GGCTACGTGCCAGC	CACCGGCCGTCAATT	16S	AY1
ant6-ib	AGAACATCCGACAGCACGTTC	CCAACCTTCCATGAAATCATTCGC	Aminoglycoside	AY101
aac(6)-iic	CAGTCTTTGGCTAATCCATCACAG	AACGAACCCGGCCTTCTC	Aminoglycoside	AY104
aac3-Via	GTGTCCGTCGCCAAGGA	GGTGACGGCCTTGTCGA	Aminoglycoside	AY3
armA	TCTTCGACGAATGAAAGAGTCG	GCTAATGGATTGAAGCCACAACC	Aminoglycoside	AY4
aac(3)-iid_iii_iif_iia_iie	CGATGGTCGCGGTTGGTC	TCGGCGTAGTGCAATGCG	Aminoglycoside	AY5
sat4	GAATGGGCAAAGCATAAAAACTTG	CCGATTTTGAAACCACAATTATGATA	Aminoglycoside	AY50
aph6ia	CGCTGGGAGCTGAAGAGG	AGCATCGTGCTGCTCTCC	Aminoglycoside	AY8
spec_aph	GGTGCTGATATGAATGCCTTTGG	CATTGGGCGCATCAATAAATGG	Aminoglycoside	AY82
aadA7	CACTCCGCGCCTTGGA	TGTGGCGGGCTCGAAG	Aminoglycoside	AY83
aph3-III	CAGAAGGCAATGTCATACCACTTG	GACAGCCGCTTAGCCGAA	Aminoglycoside	AY92
floR	AACCCGCCCTCTGGATCA	GCCGTCGAGAAGAAGACGAA	Amphenicol	AY30
catA1	GGGTGAGTTTCACCAGTTTTGATT	CACCTTGTCGCCTTGCGTATA	Amphenicol	AY62
catB3	GCACTCGATGCCTTCCAAAA	AGAGCCGATCCAAACGTCAT	Amphenicol	AY67
catB8	CACTCGACGCCTTCCAAAG	CCGAGCCTATCCAGACATCATT	Amphenicol	AY69
ceoA	ATCAACACGGACCAGGACAAG	GGAAAGTCCGCTCACGATGA	Amphenicol	AY72
cmlA1	TAGGAAGCATCGGAACGTTGAT	CAGACCGAGCACGACTGTTG	Amphenicol	AY75
cmlA5	GCGCTCTTCGAGGATTCG	CCGCCCAAGCAGAAGTAGAC	Amphenicol	AY77
blaPSE	TTGTGACCTATTCCCCTGTAATAGAA	TGCGAAGCACGCATCATC	Beta Lactam	AY14
KPC	GCCGCCAATTTGTTGCTGAA	GCCGGTCGTGTTTCCCTTT	Beta Lactam	AY34
mecA	GGTTACGGACAAGGTGAAATACTGAT	TGTCTTTTAATAAGTGAGGTGCGTTAATA	Beta Lactam	AY38

nonmobile_blaADC	GGTATGGCTGTGGGTGTTATTCA	AGGCAAGGTTACCACTTGTATACG	Beta Lactam	AY45
bla-ACT	AAGCCGCTCAAGCTGGA	GCCATATCCTGCACGTTGG	Beta Lactam (ESBLs)	AY100
blaOXA10	CGACCGAGTATGTACCTGCTTC	TCAAGTCCAATACGACGAGCTA	Beta Lactam (ESBLs)	AY12
blaOXY-1	AAAGGTGACCGCATTCGC	CCAGCGTCAGCTTGCG	Beta Lactam (ESBLs)	AY13
blaVEB	CCCGATGCAAAGCGTTATG	GAAAGATTCCCTTTATCTATCTCAGACAA	Beta Lactam (ESBLs)	AY19
nonmobile_blaBEL	ATGTCCATGGCACAGACTGTG	CCTGTCTTGTCACCCGTTACC	Beta Lactam (ESBLs)	AY44
ampC	CTGGCGCATACCTGGATTAC	GCCAGTTCAGCATCTCCCA	Beta Lactam (ESBLs)	AY7
qepA_1_2	GGGCATCGCGCTGTTC	GCGCATCGGTGAAGCC	Fluoroquinolone	AY17
qnrA	AGGATTTCTCACGCCAGGATT	CCGCTTTCAATGAAACTGCAA	Fluoroquinolone	AY18
QnrB4	TCACCACCCGCACCTG	GGATATCTAAATCGCCCAGTTCC	Fluoroquinolone	AY46
QnrVC1_VC3_VC6	CTCACATCAGGACTTGCAAGAA	ATGAAGCATCTCGAAGATCAGC	Fluoroquinolone	AY47
pmrA	TTTGCAGGTTTTGTTCCTAATGC	GCAGAGCCTGATTTCTCCTTTG	Fluoroquinolone	AY49
qnrS2	TCCCGAGCAAACTTTGCCAA	GGTGAGTCCCTATCCAGCGA	Fluoroquinolone	AY90
QnrS1_S3_S5	CCACTTTGATGTCGCAGATCTTC	CCCTCTCCATATTGGCATAGGAAA	Fluoroquinolone	AY99
int1-a-marko	CGAAGTCGAGGCATTTCTGTC	GCCTTCCAGAAAACCGAGGA	Integrase	AY111
intl2	TGCTTTTCCCACCCTTACC	GACGGCTACCCTCTGTTATCTC	Integrase	AY113
intl3	CAGGTGCTGGGCATGGA	CCTGGGCAGCATCACCA	Integrase	AY16
intl3_339	GCCACCACTTGTTTGAGGA	GGATGTCTGTGCCTGCTTG	Integrase	AY26
intI1F165_clinical	CGAACGAGTGGCGGAGGGTG	TACCCGAGAGCTTGGCACCCA	integrase	AY98
nimE	TGCGCCAAGATAGGGCATA	GTCGTGAATTCGGCAGGTTTA	MDR	AY36
marR	GCTGTTGATGACATTGCTCACA	CGGCGTACTGGTGAAGCTAAC	MDR	AY37
merA-marko	GTGCCGTCCAAGATCATG	GGTGGAAGTCCAGTAGGGTGA	MDR	AY39
mdtH	ATGCTGGCTGTACAAGTGATG	CACTCCAGCGGGCGATA	MDR-chromo	AY91
cefa_qacelta	TAGTTGGCGAAGTAATCGCAAC	TGCGATGCCATAACCGATTATG	MDR-mobile	AY20

qacF/H	TCGCAACATCCG CATTAAAA	ATGGATTTCAGAACC AGAGAAAGAAA	MDR-mobile	AY84
czcA	GCCTTGTTCATCG GCGAAC	GGCAATGTCGCCTTC GTTC	MDR-mobile	AY85
cadc	CGCTCTGTGTCAG GATGAAGAG	CTTTCTTATGTGCTA GGGCGATCA	MDR-mobile	AY87
arsA	CAGGTCAGCCGC ATCAACC	GCCTGAAACACGGC AATTTCTTC	MDR-mobile	AY88
ere(B)	TCGTATATGGCG GGCGTAGTA	GGTCCAAGATGGGT GAATGCA	MLSB	AY24
erm(35)	CCTTCAGTCAGA ACCGGCAA	GCTGATTTGACAGTT GGTGGTG	MLSB	AY25
ermT	GTTCACTAGCACT ATTTTTAATGACA GAAGT	GAAGGGTGTCTTTTT AATACAATTAACGA	MLSB	AY27
erm(Y)	TTGTCTTTGAAAG TGAAGCAACAGT	TAACGCTAGAGAAC GATTTGTATTGAG	MLSB	AY28
lnuA	TGACGCTCAACA CACTCAAAAA	TTCATGCTTAAGTTC CATACGTGAA	MLSB	AY35
mphA	TCAGCGGGATGA TCGACTG	GAGGGCGTAGAGGG CGTA	MLSB	AY40
mphB	CGCAGCGCTTGA TCTTGTAG	TTACTGCATCCATAC GCTGCTT	MLSB	AY41
msr(E)	CGGCAGATGGTC TGAGCTTAAA	CGCACTCTTCCTGCA TAAAGGA	MLSB	AY42
vga(A)LC	GTGAAGATGTCT CGGGTACAATTG	GAAATACCAGGATT CCCATGCAC	MLSB	AY79
fomB	CACATCGACGGC GTATTCTG	CAACGAGCATACCG ACATCG	other	AY10
mcr-1	CACATCGACGGC GTATTCTG	CAACGAGCATACCG ACATCG	other	AY102
fabK	CAGGAGCAGGAA ATCCAAGC	CCAGCTTCCATTCCT TCTGC	other	AY29
fosB	CTTGCAGGCCTAT GGATTGC	TCTGTTCTCAAGTGT GCCAGTA	other	AY33
nisB	GGGAGAGTTGCC GATGTTGTA	AGCCACTCGTTAAAG GGCAAT	other	AY43
mcr-2	CGGCGTACTTTA AGCGTTATGATG	GCATTTGGCATACCA TGCAGATAG	other	AY86
bacA	ATCCGCGGCACC CTGA	CCTGCTTGATGGACT TGATGAAGA	other	AY9
ARR-3	GATCGTCTTCGA ACGGTCCTG	TTTGGCGATTGGTGA CTTGCT	other	AY95
pAKD1-IncP-1β	GGTAAGATTACC GATAAACT	GTTCGTGAAGAAGA TGTA	plasmid	AY103
tra-A	AAGTGTTCAGGG TGCTTCTGCGC	GTCATGTACATGATG ACCAAAA	plasmid	AY114
pBS228-IncP-1α	CAATCCATCGAC AATCAC	GACAATCAGCTACTT CAC	plasmid	AY15

trb-C	CGGYATWCCGSCSACRCTGCG	GCCACCTGYSBGCAGTCMCC	plasmid	AY21
traN	GCTTGGCGGTCAGCAATT	TTAGGAATAACAATCGCTACACCTTTA	plasmid	AY32
sulIII-marko	TCCGTTCAGCGAATTGGTGCAG	TTCGTTCACGCCTTACACCAGC	Sulfonamide	AY2
folA	CGAGCAGTTCCTGCCAAAG	CCCAGTCATCCGGTTCATAATC	Sulfonamide	AY31
sul1	GCCGATGAGATCAGACGTATTG	CGCATAGCGCTGGGTTTC	Sulfonamide	AY51
strB	GCTCGGTCGTGAGAACAATCT	CAATTTCGGTCGCCTGGTAGT	Sulfonamide	AY52
sul2	TCATCTGCCAAACTCGTCGTTA	GTCAAAGAACGCCGCAATGT	Sulfonamide	AY53
sulA/folP	CAGGCTCGTAAATTGATAGCAGAAG	CTTTCCTTGCGAATCGCTTT	Sulfonamide	AY54
sulIII	CGCGCTCAAGGCAGATG	GGGAATGCCATCTGCCTTG	Sulfonamide	AY80
tetT	CCATATAGAGGTTCCACCAAATCC	TGACCCTATTGGTAGTGGTTCTATTG	Tetracycline	AY11
tet(32)	CCATTACTTCGGACAACGGTAGA	CAATCTCTGTGAGGGCATTTAACA	Tetracycline	AY55
tet(36)	AGAATACTCAGCAGAGGTCAGTTCCT	TGGTAGGTCGATAACCCGAAAAT	Tetracycline	AY56
tetM	GGAGCGATTACAGAATTAGGAAGC	TCCATATGTCCTGGCGTGTC	Tetracycline	AY57
tetO	CAACATTAACGGAAAGTTTATTGTATACCA	TTGACGCTCCAAATTCATTGTATC	Tetracycline	AY58
tetQ	CGCCTCAGAAGTAAGTTCATACACTAAG	TCGTTCATGCGGATATTATCAGAAT	Tetracycline	AY59
tetU	GTGGCAAAGCAACGGATTG	TGCGGGCTTGCAAACTATC	Tetracycline	AY60
tetX	AAATTTGTTACCGACACGGAAGTT	CATAGCTGAAAAAATCCAGGACAGTT	Tetracycline	AY61
tetW	ATGAACATTCCCACCGTTATCTTT	ATATCGGCGGAGAGCTTATCC	Tetracycline	AY89
IS613	AGGTTCGGACTCAATGCAACA	TTCAGCACATACCGCCTTGAT	Transposase	AY105
tnpA_201	GCCGCACTGTCGATTTTTATC	GCGGGATCTGCCACTTCTT	Transposase	AY106
tnpA_203	GGGCGGGTCGATTGAAA	GTGGGCGGGATCTGCTT	Transposase	AY107
tnpA_205	GAAACCGATGCTACAATATCCAATTT	CAGCACCGTTTGCAGTGTAAG	Transposase	AY108

tnpA_207	AATTGATGCGGACGGCTTAA	CACCGGCCGTCAATT	Transposase	AY109
Tp614	GGAAATCAACGGCATCCAGTT	CATCCATGCGCTTTTGTCTCT	Transposase	AY110
dfrA1	GGAATGGCCCTGATATTCCA	AGTCTTGCGTCCAACCAACAG	trimethoprim	AY22
dfrA12	CCTCTACCGAACCGTCACACA	GCGACAGCGTTGAAACAACTAC	trimethoprim	AY23
dfra5	CCATGGAGTGCCAAAGGTG	CACCTTTGGCACTCCATGG	trimethoprim	AY78
dfra14	CGGATCATGTCATTGTTTCAGG	ATGTTAGAGGCGAAGTCTTGG	trimethoprim	AY81
dfra7	GTAATCGGTAGTGGTCCTGA	ATCAGGACCACTACCGATTAC	trimethoprim	AY93
dfrG	TCAATCGGAAGAGCCTTACCTGA	TGGGCAAATACCTCATTCCATTCC	trimethoprim	AY94
dfrA27	GCCGCTCAGGATCGGTA	GTCGAGATATGTAGCGTGTCG	trimethoprim	AY97
vanB	TTGTCGGCGAAGTGGATCA	AGCCTTTTTCCGGCTCGTT	Vancomycin	AY63
vanC	CCTGCCACAATCGATCGTT	CGGCTTCATTCGGCTTGATA	Vancomycin	AY65
vanC2/vanC3	TGACTGTCGGTGCTTGTGA	GATAGAGCAGCTGAGCTTGTTC	Vancomycin	AY66
vanHB	GAGGTTTCCGAGGCGACAA	CTCTCGGCGGCAGTCGTAT	Vancomycin	AY68
vanSA	CGCGTCATGCTTTCAAAATTC	TCCGCAGAAAGCTCAATTTGTT	Vancomycin	AY70
vanSB	GAAGATAAAGAGGGAAGCGTACTC	CCGAATTGTCAGCCCTTGATAA	Vancomycin	AY71
vanTC	ACAGTTGCCGCTGGTGAAG	CGTGGCTGGTCGATCAAAA	Vancomycin	AY73
vanWB	CGGACAAAGATACCCCCTATAAAG	AAATAGTAAATTGCTCATCTGGCACAT	Vancomycin	AY74
vanYD	AAGGCGATACCCTGACTGTCA	ATTGCCGGACGGAAGCA	Vancomycin	AY76

Table S3.3 HT-qPCR results of 217 samples, in which the gene name of each assay number is shown in Table S3.2
(the left table: 64 samples with Ct below or equal to 31;
the right table: 153 samples with Ct over 31)

Samples	Average Ct	Assay
A26	13.368	AY1
A28	13.835	AY1
B26	16.040	AY1
B28	13.004	AY1
H04	26.521	AY1
I04	15.546	AY1
J04	15.046	AY1
J09	20.217	AY1
J26	14.259	AY1
J28	13.922	AY1
K01	14.084	AY1
K04	17.054	AY1
K09	15.626	AY1
K12	14.899	AY1
K26	15.326	AY1
K28	13.939	AY1
K30	15.458	AY1
L01	19.168	AY1
L04	15.478	AY1
L09	14.479	AY1
L12	14.980	AY1
L26	15.712	AY1
L28	16.325	AY1
L30	13.965	AY1
M04	25.888	AY1
M09	13.607	AY1
M12	15.312	AY1
M26	19.777	AY1
M28	17.331	AY1
M30	14.492	AY1
N01	20.022	AY1
N04	27.167	AY1
N09	13.916	AY1
N12	14.044	AY1
N26	14.401	AY1
N28	16.131	AY1
neg	34.454	AY1
O01	17.428	AY1
O04	25.897	AY1
O09	14.111	AY1
O12	14.625	AY1
O26	15.201	AY1
O28	18.619	AY1
O30	14.712	AY1
P01	19.193	AY1
P09	12.691	AY1
P12	14.462	AY1

Samples	Assay	Average Ct
s65	AY1	33.858
s66	AY1	35.933
s67	AY1	34.065
s68	AY1	34.692
s69	AY1	35.198
s70	AY1	35.804
s71	AY1	32.339
s72	AY1	35.798
s73	AY1	34.280
s74	AY1	35.512
s75	AY1	34.318
s76	AY1	33.563
s77	AY1	35.854
s78	AY1	33.206
s79	AY1	33.006
s80	AY1	34.731
s81	AY1	32.879
s82	AY1	34.179
s83	AY1	34.909
s84	AY1	33.511
s85	AY1	35.246
s86	AY1	32.742
s87	AY1	33.506
s88	AY1	35.555
s89	AY1	32.990
s90	AY1	35.936
s91	AY1	32.273
s92	AY1	35.760
s93	AY1	35.098
s94	AY1	32.257
s95	AY1	32.081
s96	AY1	34.553
s97	AY1	32.135
s98	AY1	32.932
s99	AY1	33.688
s100	AY1	32.097
s101	AY1	33.985
s102	AY1	34.005
s103	AY1	33.180
s104	AY1	34.534
s105	AY1	32.912
s106	AY1	32.464
s107	AY1	33.314
s108	AY1	33.846
s109	AY1	35.676
s110	AY1	34.107
s111	AY1	35.158

P26	14.222	AY1
P28	17.089	AY1
P30	15.738	AY1
pos1	11.272	AY1
Q01	16.922	AY1
Q04	24.931	AY1
Q09	21.022	AY1
Q12	19.071	AY1
Q30	19.880	AY1
R01	17.918	AY1
R04	23.791	AY1
R09	19.841	AY1
R12	13.496	AY1
R26	15.584	AY1
R30	15.421	AY1
S09	22.435	AY1
T09	19.895	AY1
U09	22.275	AY1
ptrap	13.180	AY1
A28	22.229	AY10
B28	20.293	AY10
H04	31.491	AY10
I04	29.646	AY10
J09	30.269	AY10
J28	31.790	AY10
K01	31.730	AY10
K09	27.951	AY10
K12	30.933	AY10
K28	30.160	AY10
K30	32.388	AY10
L01	33.567	AY10
L09	29.296	AY10
L12	34.726	AY10
L28	27.781	AY10
L30	32.251	AY10
M09	26.495	AY10
M12	29.583	AY10
M28	28.641	AY10
M30	26.811	AY10
N09	32.547	AY10
O09	32.608	AY10
Q01	33.619	AY10
Q04	31.299	AY10
R04	29.213	AY10
R12	29.371	AY10
ptrap	26.824	AY10
A28	29.177	AY100
B26	30.949	AY100

s112	AY1	32.554
s113	AY1	32.405
s114	AY1	35.731
s115	AY1	34.463
s116	AY1	35.591
s117	AY1	34.745
s118	AY1	33.229
s119	AY1	34.067
s120	AY1	35.803
s121	AY1	33.280
s122	AY1	32.465
s123	AY1	35.984
s124	AY1	34.690
s125	AY1	32.584
s126	AY1	33.890
s127	AY1	34.778
s128	AY1	34.324
s129	AY1	32.630
s130	AY1	33.414
s131	AY1	35.436
s132	AY1	34.772
s133	AY1	33.761
s134	AY1	35.386
s135	AY1	35.604
s136	AY1	35.572
s137	AY1	35.778
s138	AY1	35.560
s139	AY1	33.248
s140	AY1	34.857
s141	AY1	35.201
s142	AY1	34.587
s143	AY1	33.366
s144	AY1	35.192
s145	AY1	32.255
s146	AY1	33.373
s147	AY1	34.979
s148	AY1	35.414
s149	AY1	33.440
s150	AY1	32.081
s151	AY1	33.530
s152	AY1	35.098
s153	AY1	33.410
s154	AY1	35.726
s155	AY1	32.980
s156	AY1	34.150
s157	AY1	32.389
s158	AY1	34.945
s159	AY1	34.095

B28	27.550	AY100
I04	28.239	AY100
J04	31.326	AY100
J09	28.944	AY100
J26	30.891	AY100
J28	28.532	AY100
K01	31.073	AY100
K09	30.362	AY100
K12	31.592	AY100
K26	31.089	AY100
K28	28.930	AY100
K30	31.411	AY100
L01	32.446	AY100
L04	28.972	AY100
L09	32.013	AY100
L12	28.190	AY100
L26	33.698	AY100
L28	34.313	AY100
L30	29.966	AY100
M09	31.652	AY100
M12	32.643	AY100
M26	33.886	AY100
M28	33.143	AY100
N09	32.290	AY100
N12	31.601	AY100
N28	33.097	AY100
neg	35.703	AY100
O01	33.216	AY100
O09	27.397	AY100
O26	28.673	AY100
O28	29.223	AY100
P01	31.818	AY100
P09	33.509	AY100
P12	30.999	AY100
P26	32.330	AY100
P28	34.021	AY100
P30	30.735	AY100
Q01	32.205	AY100
Q12	32.835	AY100
R01	33.286	AY100
R09	33.790	AY100
R12	31.130	AY100
R30	30.623	AY100
ptrap	31.889	AY100
A26	29.624	AY101
H04	33.293	AY101
L26	32.016	AY101
R26	30.499	AY101

s160	AY1	34.927
s161	AY1	33.776
s162	AY1	32.497
s163	AY1	35.937
s164	AY1	32.168
s165	AY1	33.792
s166	AY1	33.873
s167	AY1	33.745
s168	AY1	32.824
s169	AY1	35.064
s170	AY1	32.609
s171	AY1	33.863
s172	AY1	33.316
s173	AY1	32.418
s174	AY1	33.909
s175	AY1	34.134
s176	AY1	34.833
s177	AY1	32.921
s178	AY1	34.078
s179	AY1	34.028
s180	AY1	32.315
s181	AY1	33.995
s182	AY1	33.045
s183	AY1	33.296
s184	AY1	33.148
s185	AY1	32.767
s186	AY1	34.974
s187	AY1	33.906
s188	AY1	35.441
s189	AY1	34.218
s190	AY1	35.157
s191	AY1	35.596
s192	AY1	35.177
s193	AY1	32.983
s194	AY1	34.859
s195	AY1	32.377
s196	AY1	34.506
s197	AY1	34.027
s198	AY1	32.563
s199	AY1	35.171
s200	AY1	32.599
s201	AY1	34.652
s202	AY1	34.832
s203	AY1	34.149
s204	AY1	35.311
s205	AY1	35.067
s206	AY1	34.409
s207	AY1	35.507

A26	28.363	AY102
A28	31.830	AY102
B26	34.073	AY102
B28	33.861	AY102
H04	31.894	AY102
I04	31.202	AY102
J26	32.797	AY102
K01	28.672	AY102
K04	31.251	AY102
K09	31.257	AY102
K12	31.875	AY102
K30	27.391	AY102
L01	32.460	AY102
L09	22.308	AY102
M09	30.676	AY102
M12	31.145	AY102
N09	31.027	AY102
N26	30.841	AY102
neg	32.386	AY102
O09	31.954	AY102
O12	31.889	AY102
O26	31.804	AY102
O28	31.855	AY102
P01	31.396	AY102
P12	33.025	AY102
P26	30.867	AY102
P30	32.091	AY102
pos1	32.804	AY102
Q30	31.905	AY102
R01	32.369	AY102
R09	31.835	AY102
S09	31.850	AY102
T09	30.703	AY102
ptrap	28.413	AY102
A26	30.610	AY103
A28	28.451	AY103
B26	29.282	AY103
B28	27.031	AY103
I04	33.123	AY103
J26	32.857	AY103
J28	33.265	AY103
K04	31.105	AY103
K09	31.242	AY103
K12	33.752	AY103
K26	34.209	AY103
K30	33.554	AY103
L01	29.842	AY103
L04	32.273	AY103

s208	AY1	32.193
s209	AY1	32.783
s210	AY1	32.535
s211	AY1	35.604
s212	AY1	33.299
s213	AY1	33.230
s214	AY1	35.726
s215	AY1	35.310
s216	AY1	34.535
s217	AY1	33.709

L12	26.770	AY103
L28	30.108	AY103
L30	24.284	AY103
M09	32.918	AY103
M12	32.830	AY103
M30	33.200	AY103
N09	32.768	AY103
N12	34.590	AY103
O01	28.946	AY103
O09	30.176	AY103
O12	33.281	AY103
O26	30.816	AY103
O30	33.482	AY103
P01	26.884	AY103
P09	32.157	AY103
P26	34.326	AY103
P28	33.733	AY103
P30	28.174	AY103
Q01	31.368	AY103
R01	32.891	AY103
R12	34.987	AY103
R26	34.986	AY103
R30	31.991	AY103
ptrap	32.893	AY103
L04	31.718	AY104
M26	30.047	AY104
N26	29.728	AY104
neg	33.602	AY104
O12	30.652	AY104
A26	31.218	AY105
A28	27.592	AY105
B26	30.340	AY105
B28	27.821	AY105
H04	31.829	AY105
neg	34.991	AY105
A26	34.787	AY106
A28	32.154	AY106
B26	34.721	AY106
J26	30.281	AY106
K01	30.535	AY106
K12	30.236	AY106
K30	27.863	AY106
L09	22.765	AY106
M09	34.993	AY106
N12	34.116	AY106

O01	30.446	AY106
O12	33.854	AY106
O28	33.703	AY106

P01	33.284	AY106
R09	33.186	AY106
ptrap	30.641	AY106
A26	31.121	AY107
A28	29.379	AY107
B28	28.041	AY107
H04	29.946	AY107
A26	30.608	AY108
A28	32.462	AY108
B26	33.516	AY108
B28	30.962	AY108
H04	33.169	AY108
I04	31.838	AY108
J09	32.523	AY108
J26	33.147	AY108
J28	31.156	AY108
K26	32.131	AY108
K30	33.291	AY108
L26	32.908	AY108
M09	33.518	AY108
M30	33.450	AY108
N09	31.293	AY108
neg	35.770	AY108
O09	30.134	AY108
O26	32.837	AY108
O30	31.667	AY108
P12	32.198	AY108
P30	33.184	AY108
Q09	32.918	AY108
R12	34.185	AY108
R30	32.390	AY108
ptrap	33.898	AY108
A28	31.742	AY109
K28	32.945	AY109
K30	34.200	AY109
L09	31.363	AY109
L28	34.119	AY109
M09	31.419	AY109
M30	32.409	AY109
N04	31.453	AY109
N28	29.932	AY109

neg	34.780	AY109
O28	29.799	AY109
P12	33.448	AY109
Q01	29.815	AY109
Q04	27.377	AY109
Q09	28.529	AY109
R04	30.859	AY109

R09	30.773	AY109
T09	25.308	AY109
U09	30.879	AY109
H04	30.524	AY11
I04	25.504	AY11
J04	21.506	AY11
J09	28.279	AY11
J26	27.452	AY11
J28	28.700	AY11
K01	28.834	AY11
K04	30.181	AY11
K09	28.987	AY11
K12	31.424	AY11
K26	25.845	AY11
K30	27.314	AY11
L01	28.653	AY11
L04	30.702	AY11
L09	26.361	AY11
L12	30.561	AY11
L28	29.616	AY11
M04	30.548	AY11
M09	21.154	AY11
M12	32.646	AY11
M26	26.550	AY11
M28	23.433	AY11
N01	29.154	AY11
N04	29.322	AY11
N09	28.997	AY11
N12	23.325	AY11
N28	32.041	AY11
neg	35.211	AY11
O01	28.444	AY11
O04	29.654	AY11
O09	28.481	AY11
O12	26.453	AY11
O30	27.163	AY11
P01	33.215	AY11

P09	27.787	AY11
P12	32.221	AY11
P26	30.237	AY11
P30	30.074	AY11
pos1	29.320	AY11
Q01	26.124	AY11
Q04	30.708	AY11
Q09	30.682	AY11
Q12	32.383	AY11
R01	32.328	AY11
R04	29.514	AY11

R26	29.596	AY11
R30	32.882	AY11
S09	29.479	AY11
U09	29.603	AY11
ptrap	20.234	AY11
U09	28.91635	AY110
T09	29.69579	AY110
R09	30.73498	AY110
Q09	30.280055	AY110
Q04	28.74749	AY110
P09	34.60537	AY110
O09	30.99207	AY110
N26	31.51768	AY110
N09	29.85241	AY110
N01	31.68983	AY110
M26	31.25518	AY110
M09	31.54157	AY110
M04	30.44983	AY110
L09	32.16042	AY110
J28	34.69422	AY110
J09	31.1949021	AY110
I04	30.33784	AY110
H04	29.91233	AY110
B28	23.74044	AY110
B26	24.23469	AY110
A28	23.28924	AY110
A26	20.028305	AY110
ptrap	30.0278	AY110
T09	24.208225	AY111

S09	26.71543	AY111
R30	18.940465	AY111
R26	20.43536	AY111
R12	22.17793	AY111
R09	25.428205	AY111
R01	18.07561	AY111
Q30	20.69637	AY111
Q12	20.599165	AY111
Q09	25.476185	AY111
Q04	26.483145	AY111
Q01	17.680825	AY111
P30	19.333875	AY111
P28	18.637015	AY111
P26	17.32768	AY111

P12	17.256905	AY111
P09	16.35049	AY111
P01	22.944315	AY111
O30	19.693965	AY111
O28	18.7417	AY111
O26	19.43254	AY111
O12	18.75121	AY111
O09	18.50597	AY111
O04	25.208955	AY111
O01	17.9525	AY111
N28	17.889315	AY111
N26	18.416785	AY111
N12	18.20041	AY111
N09	18.266225	AY111
N01	21.897915	AY111
M30	22.127995	AY111
M28	18.44016	AY111
M26	22.144315	AY111
M12	21.403615	AY111
M09	16.432615	AY111
L30	20.561465	AY111
L28	19.880285	AY111
L26	22.519155	AY111
L12	22.04959	AY111
L09	19.90012	AY111

L04	19.691835	AY111
L01	20.97023	AY111
K30	19.755535	AY111
K28	19.467215	AY111
K26	20.56254	AY111
K12	20.53534	AY111
K09	22.973845	AY111
K04	21.702315	AY111
K01	21.194615	AY111
J28	19.73706	AY111
J26	20.315625	AY111
J04	16.814755	AY111
I04	20.41312	AY111
I04	19.950705	AY111
H04	28.2548	AY111
B28	16.26624	AY111
B26	18.203925	AY111
A28	17.40035	AY111

A26	17.25844	AY111
ptrap	15.392515	AY111
P09	29.98924	AY113
K30	31.38675	AY113
K01	28.36639	AY113
J04	30.15779	AY113
B28	29.46295	AY113
A28	28.133845	AY113
A26	27.648015	AY113
P26	32.50604	AY114
P12	34.65764	AY114
L30	34.45335	AY114
L26	31.393	AY114
K30	32.45169	AY114
K12	31.36443	AY114
K01	32.521075	AY114
J28	32.75215	AY114
I04	34.47803	AY114
B28	33.06394	AY114
A26	31.707	AY12
A28	29.706	AY12

B26	34.599	AY12
B28	29.916	AY12
I04	33.142	AY12
J09	34.029	AY12
J28	34.257	AY12
K01	29.109	AY12
K09	34.557	AY12
K12	29.168	AY12
K28	32.184	AY12
K30	26.564	AY12
L01	34.863	AY12
L04	30.895	AY12
L09	30.879	AY12
L12	32.355	AY12
L26	32.671	AY12
L30	29.286	AY12
M12	33.810	AY12
N09	31.977	AY12
N26	28.933	AY12
neg	34.265	AY12
O04	34.513	AY12
O12	33.231	AY12
O26	29.306	AY12
O30	32.877	AY12
P09	30.331	AY12
P12	34.300	AY12
P26	29.943	AY12
P28	30.220	AY12
P30	34.120	AY12
Q01	27.275	AY12
Q12	31.732	AY12
Q30	29.911	AY12
R26	23.864	AY12
R30	31.918	AY12
S09	29.446	AY12
T09	31.275	AY12
ptrap	33.856	AY12
A26	27.884	AY13
B26	29.954	AY13
B28	29.127	AY13
H04	32.195	AY13
I04	28.570	AY13
J04	30.069	AY13
J09	29.890	AY13

J26	29.922	AY13
J28	28.396	AY13
K01	26.397	AY13
K04	32.766	AY13
K12	27.327	AY13
K26	31.544	AY13
K28	23.020	AY13
K30	25.839	AY13
L01	27.911	AY13
L04	29.987	AY13
L09	27.558	AY13
L12	24.933	AY13
L28	27.517	AY13
L30	22.260	AY13
M12	29.272	AY13
M26	34.661	AY13
M30	21.578	AY13
N01	29.625	AY13
N04	31.057	AY13
N09	29.052	AY13
N12	27.325	AY13
N26	29.755	AY13
N28	29.485	AY13
neg	33.705	AY13
O01	27.645	AY13
O04	30.788	AY13
O09	29.712	AY13
O12	28.667	AY13
O26	31.396	AY13

O28	28.776	AY13
O30	27.580	AY13
P01	27.096	AY13
P09	28.783	AY13
P12	28.262	AY13
P26	30.800	AY13
P30	29.797	AY13
Q01	29.036	AY13
Q04	30.731	AY13
Q09	31.699	AY13
Q12	30.196	AY13
Q30	30.901	AY13
R04	28.172	AY13
R09	27.857	AY13
R26	30.321	AY13

S09	34.177	AY13
T09	31.560	AY13
U09	29.023	AY13
ptrap	28.375	AY13
A26	24.667	AY14
A28	29.082	AY14
B28	28.445	AY14
I04	21.795	AY14
J04	28.734	AY14
K28	29.526	AY14
L04	27.092	AY14
L09	28.377	AY14
L30	28.091	AY14
M30	28.369	AY14
N04	30.558	AY14
P09	25.923	AY14
P26	26.188	AY14
P28	30.519	AY14
Q04	29.630	AY14
Q09	28.144	AY14
R04	30.554	AY14
R09	29.479	AY14
R12	29.792	AY14
T09	28.772	AY14
ptrap	29.585	AY14
B28	24.293	AY15
H04	25.537	AY15
I04	25.659	AY15
J04	26.020	AY15
J09	24.252	AY15
K01	25.703	AY15
K04	25.901	AY15
K12	25.627	AY15

K30	25.316	AY15
L01	25.325	AY15
L04	27.072	AY15
L09	26.106	AY15
L26	26.933	AY15
L30	24.861	AY15
M04	25.279	AY15
M09	25.488	AY15
M26	26.690	AY15
M28	24.648	AY15
M30	22.476	AY15

N01	25.813	AY15
N04	29.224	AY15
N09	26.059	AY15
N26	25.974	AY15
N28	25.270	AY15
neg	34.042	AY15
O01	26.672	AY15
O04	29.347	AY15
O12	25.327	AY15
O28	25.669	AY15
O30	24.778	AY15
P01	26.538	AY15
P09	25.823	AY15
P12	26.327	AY15
P28	26.421	AY15
pos1	26.721	AY15
Q01	23.488	AY15
Q04	28.480	AY15
Q09	26.580	AY15
Q12	26.451	AY15
Q30	27.173	AY15
R01	25.292	AY15
R09	26.932	AY15
R12	25.865	AY15
S09	27.321	AY15
T09	28.324	AY15
U09	28.073	AY15
A26	24.995	AY16
A28	24.686	AY16
B26	26.272	AY16
B28	24.363	AY16
I04	30.552	AY16
J09	31.358	AY16
J28	30.303	AY16
K26	30.586	AY16
K30	31.587	AY16
L01	29.908	AY16

L09	28.689	AY16
L12	30.353	AY16
L28	30.933	AY16
M09	22.689	AY16
M12	18.838	AY16
M26	22.207	AY16
M28	22.624	AY16

N01	31.082	AY16
N09	31.703	AY16
N12	24.865	AY16
N26	30.859	AY16
neg	33.694	AY16
O01	30.429	AY16
P09	26.371	AY16
P12	28.977	AY16
P26	27.386	AY16
P28	24.575	AY16
Q01	25.075	AY16
Q04	30.391	AY16
R04	29.734	AY16
R12	31.769	AY16
R30	28.687	AY16
T09	30.565	AY16
ptrap	18.666	AY16
A26	17.835	AY17
A28	19.664	AY17
B26	20.770	AY17
B28	19.863	AY17
I04	31.472	AY17
J04	23.847	AY17
J09	30.107	AY17
J26	27.597	AY17
J28	29.396	AY17
K01	27.884	AY17
K04	29.303	AY17
K09	28.774	AY17
K12	28.080	AY17
K26	27.747	AY17
K28	23.501	AY17
K30	25.593	AY17
L01	31.294	AY17
L04	20.121	AY17
L09	22.901	AY17
L12	29.841	AY17
L26	25.479	AY17
L28	28.739	AY17
L30	25.110	AY17
M09	27.569	AY17

M26	26.028	AY17
M28	25.416	AY17
M30	30.337	AY17

N04	30.145	AY17
N12	26.812	AY17
N26	29.289	AY17
N28	28.808	AY17
O01	26.072	AY17
O09	18.362	AY17
O12	22.709	AY17
O26	18.801	AY17
O28	19.285	AY17
O30	22.400	AY17
P01	23.411	AY17
P09	30.321	AY17
P12	21.049	AY17
P26	29.046	AY17
P30	19.668	AY17
Q01	27.798	AY17
Q09	30.299	AY17
R26	31.050	AY17
R30	27.299	AY17
T09	29.260	AY17
ptrap	21.808	AY17
A28	33.107	AY18
B28	32.257	AY18
J28	33.893	AY18
K09	34.589	AY18
K30	32.832	AY18
L09	31.636	AY18
L30	33.045	AY18
M04	33.707	AY18
M09	34.841	AY18
M30	34.365	AY18
N09	30.302	AY18
neg	33.523	AY18
O12	32.963	AY18
O28	34.900	AY18
P01	33.247	AY18
Q04	30.292	AY18
Q09	30.512	AY18
R04	30.653	AY18
R30	33.649	AY18
T09	26.731	AY18
U09	30.285	AY18
A28	31.856	AY19
B28	34.068	AY19
K09	32.982	AY19

L01	29.923	AY19
L09	31.050	AY19
M09	30.924	AY19
N01	29.732	AY19
N09	33.188	AY19
O09	33.668	AY19
O30	32.337	AY19
P09	34.018	AY19
R30	34.624	AY19
ptrap	30.965	AY19
A26	19.638	AY2
A28	21.414	AY2
B26	21.241	AY2
B28	19.055	AY2
I04	24.045	AY2
J04	23.426	AY2
J09	27.590	AY2
J26	24.796	AY2
J28	26.106	AY2
K01	23.582	AY2
K04	31.668	AY2
K09	26.599	AY2
K12	21.429	AY2
K26	26.846	AY2
K28	21.471	AY2
K30	21.782	AY2
L01	23.089	AY2
L04	24.806	AY2
L09	23.201	AY2
L26	23.335	AY2
L28	23.645	AY2
L30	21.997	AY2
M04	30.775	AY2
M09	25.538	AY2
M12	23.225	AY2
M26	31.021	AY2
M28	24.982	AY2
M30	23.497	AY2
N01	23.343	AY2
N09	26.267	AY2
N28	26.741	AY2
neg	35.094	AY2
O01	25.592	AY2
O12	26.015	AY2
O26	26.325	AY2

O28	24.957	AY2
O30	25.587	AY2
P28	27.812	AY2

Q01	26.208	AY2
Q04	29.529	AY2
Q09	32.276	AY2
Q12	28.682	AY2
Q30	28.739	AY2
R04	29.046	AY2
R26	29.092	AY2
R30	23.079	AY2
S09	28.945	AY2
T09	28.612	AY2
ptrap	28.119	AY2
A26	33.620	AY20
B28	33.002	AY20
J04	33.718	AY20
J26	33.836	AY20
K28	33.374	AY20
L09	34.541	AY20
M09	34.274	AY20
N09	32.089	AY20
N26	32.314	AY20
O09	31.870	AY20
O26	33.962	AY20
P26	31.092	AY20
R26	34.369	AY20
A26	32.070	AY21
A26	19.730	AY22
A28	20.635	AY22
B26	20.891	AY22
B28	18.603	AY22
I04	23.091	AY22
J04	19.857	AY22
J09	25.655	AY22
J26	23.041	AY22
J28	22.399	AY22
K01	25.254	AY22
K04	25.477	AY22
K09	27.634	AY22
K12	23.587	AY22
K26	22.523	AY22
K28	23.000	AY22
K30	23.124	AY22

L01	23.721	AY22
L04	24.889	AY22
L09	22.618	AY22
L26	25.766	AY22
L28	24.472	AY22
L30	26.453	AY22
M09	18.276	AY22

M12	20.489	AY22
M26	22.323	AY22
M28	20.336	AY22
M30	27.860	AY22
N01	24.807	AY22
N09	25.708	AY22
N12	20.337	AY22
N26	28.325	AY22
N28	22.653	AY22
neg	32.080	AY22
O01	21.438	AY22
O09	27.851	AY22
O12	25.137	AY22
O26	25.632	AY22
O28	24.875	AY22
O30	26.960	AY22
P01	27.172	AY22
P09	19.448	AY22
P12	24.724	AY22
P26	20.452	AY22
P28	22.440	AY22
P30	24.821	AY22
Q01	26.242	AY22
Q04	29.844	AY22
Q12	26.471	AY22
R01	20.655	AY22
R12	24.319	AY22
R26	22.296	AY22
R30	22.670	AY22
S09	27.651	AY22
T09	26.605	AY22
ptrap	16.367	AY22
A26	20.082	AY23
A28	24.992	AY23
B26	25.754	AY23
B28	24.436	AY23
I04	30.484	AY23

J04	30.415	AY23
J28	30.974	AY23
K04	31.092	AY23
K09	28.876	AY23
K28	31.560	AY23
L04	32.814	AY23
M26	29.186	AY23
N26	30.502	AY23
neg	33.724	AY23
pos1	31.020	AY23
Q09	31.045	AY23

R09	31.927	AY23
T09	31.288	AY23
U09	30.072	AY23
ptrap	27.288	AY23
L12	33.747	AY24
L30	33.661	AY24
A26	23.254	AY25
A28	23.397	AY25
B26	27.041	AY25
B28	24.985	AY25
I04	34.552	AY25
J04	31.614	AY25
J09	30.407	AY25
J28	34.786	AY25
K12	33.327	AY25
K28	33.239	AY25
K30	31.522	AY25
L09	29.398	AY25
L28	32.840	AY25
L30	32.953	AY25
M30	32.514	AY25
N12	33.118	AY25
P09	32.763	AY25
Q09	31.045	AY25
T09	31.316	AY25
A26	19.337	AY26
B26	19.318	AY26
J04	18.747	AY26
J26	19.342	AY26
K01	32.628	AY26
K04	18.944	AY26
K12	33.112	AY26
K26	19.477	AY26

L04	19.348	AY26
L26	19.265	AY26
M26	19.399	AY26
M30	33.182	AY26
N12	33.512	AY26
N26	19.497	AY26
neg	35.583	AY26
O01	33.354	AY26
O12	32.934	AY26
O26	19.330	AY26
P26	19.633	AY26
pos1	18.672	AY26
R01	34.145	AY26
R12	34.314	AY26
R26	19.203	AY26

R30	33.661	AY26
ptrap	33.488	AY26
A26	22.779	AY27
A28	22.485	AY27
B26	25.120	AY27
B28	22.483	AY27
K01	28.440	AY27
K12	22.213	AY27
K28	25.785	AY27
L09	25.876	AY27
L28	26.647	AY27
L30	27.114	AY27
M12	24.660	AY27
M28	30.963	AY27
M30	27.373	AY27
P01	32.953	AY27
Q04	28.755	AY27
R04	30.810	AY27
R26	28.438	AY27
T09	28.798	AY27
U09	29.819	AY27
A26	32.840	AY28
A28	32.079	AY28
B26	32.668	AY28
B28	35.716	AY28
H04	32.580	AY28
I04	33.702	AY28
J04	32.844	AY28
J09	32.966	AY28

J26	32.574	AY28
J28	35.016	AY28
K01	34.569	AY28
K04	33.720	AY28
K09	34.921	AY28
K12	33.311	AY28
K26	34.507	AY28
K28	32.094	AY28
K30	33.538	AY28
L01	32.296	AY28
L04	35.327	AY28
L09	35.827	AY28
L12	33.638	AY28
L26	32.693	AY28
L28	35.213	AY28
L30	35.429	AY28
M04	34.585	AY28
M09	35.228	AY28
M12	32.218	AY28

M26	32.644	AY28
M28	32.339	AY28
M30	34.263	AY28
N01	35.523	AY28
N04	35.237	AY28
N09	33.402	AY28
N12	33.717	AY28
N26	33.353	AY28
N28	34.113	AY28
neg	33.319	AY28
O01	35.771	AY28
O04	34.628	AY28
O09	33.173	AY28
O12	33.105	AY28
O26	34.757	AY28
O28	33.536	AY28
O30	33.112	AY28
P01	34.413	AY28
P09	33.245	AY28
P12	35.223	AY28
P26	33.132	AY28
P28	34.856	AY28
P30	35.479	AY28
pos1	35.947	AY28
Q01	33.681	AY28

Q04	35.481	AY28
Q09	33.090	AY28
Q12	35.818	AY28
Q30	35.611	AY28
R01	35.497	AY28
R04	32.906	AY28
R09	33.018	AY28
R12	33.066	AY28
R26	34.793	AY28
R30	34.156	AY28
S09	33.489	AY28
T09	34.990	AY28
U09	32.969	AY28
ptrap	32.535	AY28
A26	23.905	AY29
A28	24.930	AY29
B26	28.417	AY29
B28	25.839	AY29
K09	32.635	AY29
K28	29.537	AY29
L12	29.044	AY29
L28	32.175	AY29
L30	30.415	AY29

N26	29.977	AY29
P09	31.180	AY29
Q04	32.174	AY29
R04	30.385	AY29
T09	30.050	AY29
ptrap	31.154	AY29
A26	25.490	AY3
A28	25.898	AY3
B26	27.147	AY3
B28	24.674	AY3
I04	32.647	AY3
J04	34.634	AY3
J09	31.287	AY3
J26	32.453	AY3
J28	30.965	AY3
K01	31.942	AY3
K09	33.576	AY3
K12	30.247	AY3
K28	32.826	AY3
K30	28.502	AY3
L01	24.897	AY3

L12	30.611	AY3
L28	28.187	AY3
L30	30.741	AY3
M09	25.923	AY3
M28	23.197	AY3
N01	26.176	AY3
N04	31.763	AY3
N09	32.636	AY3
N12	23.183	AY3
N26	30.847	AY3
N28	31.793	AY3
neg	33.543	AY3
O12	29.247	AY3
O30	28.403	AY3
P01	32.229	AY3
P09	27.607	AY3
P26	34.903	AY3
P28	34.131	AY3
P30	34.919	AY3
Q01	27.663	AY3
Q12	33.001	AY3
R01	34.739	AY3
R12	32.936	AY3
R30	32.470	AY3
U09	32.856	AY3
ptrap	31.552	AY3
A26	22.470	AY30

A28	22.563	AY30
B26	24.712	AY30
B28	22.483	AY30
J04	30.432	AY30
J28	30.480	AY30
K01	29.473	AY30
K09	28.421	AY30
K12	22.002	AY30
K28	25.787	AY30
L09	26.188	AY30
L28	25.897	AY30
L30	26.603	AY30
M12	24.630	AY30
M28	30.125	AY30
M30	29.026	AY30
N04	29.684	AY30
N12	29.077	AY30

R26	28.937	AY30
R30	29.247	AY30
S09	29.781	AY30
T09	30.399	AY30
A26	24.058	AY31
A28	23.867	AY31
B26	27.827	AY31
B28	26.361	AY31
H04	32.077	AY31
I04	33.474	AY31
J26	34.445	AY31
K01	33.865	AY31
K04	33.460	AY31
K28	32.795	AY31
L04	34.183	AY31
L09	28.175	AY31
L12	34.192	AY31
L28	30.788	AY31
M30	28.986	AY31
N12	30.815	AY31
N26	32.883	AY31
O01	33.925	AY31
O12	32.195	AY31
O28	33.922	AY31
R12	33.755	AY31
R26	33.312	AY31
R30	33.112	AY31
ptrap	29.840	AY31
A26	29.910	AY32
A28	30.844	AY32
B28	31.686	AY32

J09	31.020	AY32
L01	34.651	AY32
L12	33.890	AY32
N01	34.499	AY32
R12	34.011	AY32
A26	28.344	AY33
A28	30.326	AY33
B28	30.447	AY33
J28	34.294	AY33
K01	34.121	AY33
K04	34.097	AY33
K12	32.333	AY33
N12	34.129	AY33

O12	34.817	AY33
P01	34.869	AY33
T09	31.461	AY33
A26	27.841	AY34
A28	29.884	AY34
B26	29.563	AY34
B28	29.750	AY34
I04	31.222	AY34
J04	31.573	AY34
J09	31.029	AY34
J26	30.339	AY34
J28	31.494	AY34
K01	30.306	AY34
K09	32.272	AY34
K28	28.776	AY34
K30	29.276	AY34
L09	23.321	AY34
L26	31.598	AY34
M30	32.109	AY34
N01	33.014	AY34
neg	35.598	AY34
O04	32.719	AY34
P26	30.785	AY34
Q09	29.664	AY34
R26	31.803	AY34
S09	30.579	AY34
T09	29.753	AY34
A26	17.838	AY35
A28	19.794	AY35
B26	21.207	AY35
B28	19.422	AY35
I04	29.111	AY35
J04	23.446	AY35
J09	28.589	AY35
J26	29.027	AY35

K09	28.821	AY35
K26	29.140	AY35
K28	24.668	AY35
L01	29.234	AY35
L09	26.860	AY35
L12	26.964	AY35
L26	30.120	AY35
L28	27.170	AY35
L30	27.590	AY35

M09	20.042	AY35
M12	29.566	AY35
M26	21.278	AY35
M28	19.690	AY35
M30	29.754	AY35
N01	29.007	AY35
N04	30.698	AY35
N12	24.459	AY35
N26	27.494	AY35
O04	29.612	AY35
O09	33.648	AY35
P01	33.227	AY35
P12	29.727	AY35
P26	29.590	AY35
Q01	22.765	AY35
Q30	29.519	AY35
R01	30.491	AY35
R12	28.888	AY35
R26	29.872	AY35
R30	27.850	AY35
T09	30.641	AY35
ptrap	15.783	AY35
A26	27.221	AY36
A28	28.695	AY36
B26	27.595	AY36
B28	27.797	AY36
H04	30.642	AY36
I04	30.781	AY36
J04	30.499	AY36
J26	31.592	AY36
K04	34.221	AY36
K09	30.310	AY36
K12	32.750	AY36
K26	31.296	AY36
K30	34.430	AY36
L01	30.157	AY36
L04	30.484	AY36
L09	30.360	AY36
L26	29.746	AY36

L28	32.751	AY36
L30	34.859	AY36
M04	29.586	AY36
M09	28.611	AY36
M12	31.778	AY36

M28	31.442	AY36
M30	31.720	AY36
N12	33.074	AY36
N26	30.303	AY36
N28	30.468	AY36
neg	35.705	AY36
O09	30.233	AY36
O26	31.289	AY36
O28	34.293	AY36
P09	31.154	AY36
P26	31.483	AY36
pos1	30.062	AY36
Q09	31.322	AY36
R01	33.512	AY36
R09	32.059	AY36
R12	31.933	AY36
R26	31.106	AY36
R30	26.797	AY36
S09	30.471	AY36
T09	33.622	AY36
ptrap	32.689	AY36
A26	31.458	AY37
A28	31.291	AY37
B26	32.946	AY37
B28	31.937	AY37
H04	29.856	AY37
I04	29.152	AY37
J09	33.378	AY37
J26	30.745	AY37
J28	32.965	AY37
K01	31.401	AY37
K04	33.920	AY37
K09	28.175	AY37
K12	32.745	AY37
K26	31.035	AY37
K28	32.751	AY37
L01	32.348	AY37
L04	34.797	AY37
L09	34.072	AY37
L12	29.324	AY37
L30	29.354	AY37
M09	30.600	AY37
M28	34.904	AY37

M30	30.721	AY37

N01	31.945	AY37
N04	29.353	AY37
N09	31.894	AY37
N26	31.343	AY37
N28	29.142	AY37
O01	29.853	AY37
O04	31.284	AY37
O09	32.383	AY37
O12	25.668	AY37
O28	33.617	AY37
O30	27.709	AY37
P09	31.376	AY37
P26	30.801	AY37
P28	31.581	AY37
P30	30.314	AY37
Q01	32.183	AY37
Q04	25.807	AY37
Q09	29.115	AY37
Q12	31.040	AY37
R04	28.267	AY37
R09	32.423	AY37
R12	29.615	AY37
R26	29.765	AY37
R30	30.093	AY37
S09	30.502	AY37
T09	25.172	AY37
U09	27.321	AY37
ptrap	31.045	AY37
A26	24.961	AY38
A28	25.056	AY38
B26	26.488	AY38
B28	24.562	AY38
H04	31.780	AY38
I04	31.340	AY38
J04	31.809	AY38
K09	30.856	AY38
K26	30.553	AY38
K28	28.906	AY38
L12	28.423	AY38
L30	28.106	AY38
N26	33.640	AY38
O26	31.971	AY38
P09	34.642	AY38
Q04	29.085	AY38
T09	29.267	AY38

U09	31.345	AY38
A26	29.072	AY39

A28	29.023	AY39
B26	29.474	AY39
B28	29.406	AY39
H04	29.975	AY39
I04	29.818	AY39
J04	28.782	AY39
J28	29.394	AY39
K01	30.871	AY39
K04	29.294	AY39
K09	28.621	AY39
K12	28.919	AY39
K26	29.493	AY39
K28	28.517	AY39
L01	30.609	AY39
L04	30.142	AY39
L09	30.167	AY39
L12	29.636	AY39
L26	30.692	AY39
L28	30.504	AY39
L30	30.131	AY39
M04	29.926	AY39
M09	31.656	AY39
M12	30.579	AY39
M26	28.474	AY39
M28	30.188	AY39
M30	29.510	AY39
N04	29.317	AY39
N12	30.080	AY39
N26	29.070	AY39
N28	31.014	AY39
neg	33.892	AY39
O04	33.374	AY39
O09	28.477	AY39
O12	31.412	AY39
O26	29.928	AY39
O28	30.518	AY39
O30	32.373	AY39
P01	31.013	AY39
P09	29.278	AY39
P12	30.169	AY39
P26	29.936	AY39
P30	30.616	AY39

pos1	29.866	AY39
Q01	30.717	AY39
Q04	28.387	AY39
Q09	31.499	AY39
Q12	30.659	AY39
Q30	28.869	AY39

R01	30.332	AY39
R04	29.921	AY39
R09	29.679	AY39
R12	29.411	AY39
R26	29.296	AY39
R30	30.570	AY39
S09	29.921	AY39
T09	28.036	AY39
U09	28.109	AY39
ptrap	30.511	AY39
A26	21.881	AY4
A28	22.137	AY4
B26	22.894	AY4
B28	21.197	AY4
H04	32.431	AY4
I04	21.617	AY4
J04	21.629	AY4
J09	21.270	AY4
J26	20.687	AY4
J28	19.099	AY4
K01	20.875	AY4
K04	24.273	AY4
K09	21.529	AY4
K12	20.672	AY4
K26	21.973	AY4
K30	19.924	AY4
L01	20.798	AY4
L04	21.426	AY4
L09	20.308	AY4
L12	21.162	AY4
L28	24.888	AY4
L30	20.417	AY4
M04	34.464	AY4
M09	20.649	AY4
M12	24.411	AY4
M26	24.908	AY4
M28	20.657	AY4
M30	26.944	AY4

N01	20.810	AY4
N04	31.087	AY4
N09	20.625	AY4
N12	20.276	AY4
N26	22.257	AY4
neg	33.944	AY4
O01	21.950	AY4
O04	30.313	AY4
O09	21.585	AY4
O12	21.563	AY4

O26	22.681	AY4
O28	21.744	AY4
O30	21.101	AY4
P01	21.984	AY4
P09	18.258	AY4
P12	22.701	AY4
P26	21.260	AY4
P30	22.917	AY4
pos1	31.216	AY4
Q01	21.912	AY4
Q04	26.456	AY4
Q09	27.116	AY4
Q12	24.658	AY4
Q30	26.281	AY4
R01	19.819	AY4
R04	27.992	AY4
R12	20.712	AY4
R26	21.735	AY4
R30	20.651	AY4
S09	26.177	AY4
U09	28.383	AY4
ptrap	19.242	AY4
A26	22.646	AY40
A28	22.625	AY40
B26	22.294	AY40
B28	20.170	AY40
I04	24.541	AY40
J04	25.682	AY40
J09	24.772	AY40
J26	22.677	AY40
J28	23.470	AY40
K01	24.055	AY40
K04	25.673	AY40
K09	26.867	AY40

K12	24.669	AY40
K26	27.802	AY40
K28	22.164	AY40
K30	23.295	AY40
L01	26.980	AY40
L04	30.502	AY40
L09	23.209	AY40
L12	27.793	AY40
L26	23.807	AY40
L28	26.747	AY40
L30	22.591	AY40
M04	30.632	AY40
M09	30.139	AY40
M12	27.391	AY40

M26	31.130	AY40
M28	29.317	AY40
M30	23.544	AY40
N01	26.883	AY40
N09	29.901	AY40
N12	24.426	AY40
N28	29.315	AY40
O01	26.298	AY40
O09	31.388	AY40
O12	23.798	AY40
O26	30.800	AY40
O30	25.426	AY40
P09	28.779	AY40
P12	29.093	AY40
P26	34.196	AY40
P28	29.747	AY40
Q04	27.622	AY40
Q09	29.550	AY40
R01	27.427	AY40
R04	28.971	AY40
R09	29.117	AY40
R12	29.988	AY40
R26	18.273	AY40
R30	25.309	AY40
S09	22.877	AY40
T09	22.521	AY40
U09	29.433	AY40
ptrap	29.932	AY40
A26	29.823	AY41
A28	31.933	AY41

B28	32.974	AY41
J28	34.709	AY41
K12	34.322	AY41
K28	32.509	AY41
K30	33.351	AY41
L12	34.100	AY41
L30	32.516	AY41
N26	34.265	AY41
N28	34.830	AY41
O12	30.734	AY41
O30	33.039	AY41
P09	32.948	AY41
pos1	29.593	AY41
R30	34.494	AY41
ptrap	33.615	AY41
I04	33.870	AY42
K01	31.677	AY42
K12	33.103	AY42

K30	32.146	AY42
L12	32.666	AY42
M30	31.691	AY42
N12	32.189	AY42
neg	34.295	AY42
O30	31.646	AY42
P12	32.262	AY42
Q12	32.517	AY42
R12	31.193	AY42
R30	32.317	AY42
S09	30.001	AY42
T09	31.062	AY42
U09	29.758	AY42
ptrap	32.694	AY42
A26	17.258	AY43
A28	17.400	AY43
B26	18.204	AY43
B28	16.266	AY43
H04	28.255	AY43
I04	20.182	AY43
J04	16.815	AY43
J09	27.077	AY43
J26	20.316	AY43
J28	19.737	AY43
K01	21.195	AY43
K04	21.702	AY43

K09	22.974	AY43
K12	20.535	AY43
K26	20.563	AY43
K28	19.467	AY43
K30	19.756	AY43
L01	20.970	AY43
L04	19.692	AY43
L09	19.900	AY43
L12	22.050	AY43
L26	22.519	AY43
L28	19.880	AY43
L30	20.561	AY43
M04	26.963	AY43
M09	16.433	AY43
M12	21.404	AY43
M26	22.144	AY43
M28	18.440	AY43
M30	22.128	AY43
N01	21.898	AY43
N04	29.280	AY43
N09	18.266	AY43
N12	18.200	AY43

N26	18.417	AY43
N28	17.889	AY43
neg	35.608	AY43
O01	17.953	AY43
O04	25.209	AY43
O09	18.506	AY43
O12	18.751	AY43
O26	19.433	AY43
O28	18.742	AY43
O30	19.694	AY43
P01	22.944	AY43
P09	16.350	AY43
P12	17.257	AY43
P26	17.328	AY43
P28	18.637	AY43
P30	19.334	AY43
pos1	27.721	AY43
Q01	17.681	AY43
Q04	26.483	AY43
Q09	25.476	AY43
Q12	20.599	AY43
Q30	20.696	AY43

R01	18.076	AY43
R04	26.938	AY43
R09	25.428	AY43
R12	22.178	AY43
R26	20.435	AY43
R30	18.940	AY43
S09	26.715	AY43
T09	24.208	AY43
U09	26.998	AY43
ptrap	15.282	AY43
A26	24.730	AY44
A28	28.793	AY44
B28	27.704	AY44
I04	26.694	AY44
J04	33.197	AY44
J28	31.604	AY44
K01	32.736	AY44
K09	31.263	AY44
K12	34.120	AY44
K28	27.195	AY44
L09	30.161	AY44
L12	29.753	AY44
L28	29.685	AY44
L30	28.907	AY44
M12	33.042	AY44
M28	32.722	AY44

M30	33.253	AY44
neg	32.394	AY44
O01	29.712	AY44
O04	31.803	AY44
O09	31.916	AY44
P09	31.874	AY44
Q12	33.364	AY44
Q30	33.275	AY44
R12	32.070	AY44
T09	32.397	AY44
ptrap	33.055	AY44
A26	27.648	AY45
A28	28.134	AY45
B28	29.463	AY45
J04	30.158	AY45
K01	28.366	AY45
K30	31.387	AY45
P09	29.989	AY45

A26	24.480	AY46
B26	24.189	AY46
B28	22.121	AY46
I04	24.308	AY46
J09	23.292	AY46
J26	23.398	AY46
J28	22.312	AY46
K01	22.418	AY46
K04	25.676	AY46
K09	23.237	AY46
K28	22.838	AY46
K30	20.783	AY46
L01	21.781	AY46
L04	22.540	AY46
L12	23.693	AY46
L26	25.312	AY46
L28	24.678	AY46
L30	22.502	AY46
M09	22.522	AY46
M12	22.523	AY46
M28	24.891	AY46
M30	22.542	AY46
N01	23.844	AY46
N04	31.271	AY46
neg	33.737	AY46
O01	24.845	AY46
O04	32.270	AY46
O09	21.030	AY46
O26	22.229	AY46
O28	23.953	AY46

P09	21.743	AY46
P12	22.873	AY46
P30	23.527	AY46
Q09	29.933	AY46
Q12	26.767	AY46
R01	24.834	AY46
R04	31.291	AY46
R30	24.197	AY46
S09	28.229	AY46
ptrap	23.099	AY46
A26	24.454	AY47
A28	28.851	AY47
B26	27.414	AY47
B28	27.577	AY47

J04	31.971	AY47
J09	31.153	AY47
J26	29.908	AY47
K26	29.164	AY47
K28	31.338	AY47
L04	31.163	AY47
L26	33.943	AY47
L30	30.469	AY47
O26	32.524	AY47
pos1	30.404	AY47
S09	32.757	AY47
A26	18.704	AY49
A28	21.716	AY49
B26	24.262	AY49
B28	22.227	AY49
H04	30.269	AY49
K26	26.872	AY49
K30	30.753	AY49
L09	30.002	AY49
L28	28.245	AY49
L30	30.385	AY49
M04	30.942	AY49
M30	29.687	AY49
neg	35.254	AY49
O26	30.690	AY49
P09	30.193	AY49
P12	29.381	AY49
pos1	32.491	AY49
Q04	30.339	AY49
S09	33.139	AY49
U09	30.180	AY49
ptrap	28.748	AY49
A26	24.210	AY5
A28	22.348	AY5

B26	24.145	AY5
I04	21.643	AY5
J04	24.576	AY5
J26	21.283	AY5
K04	24.510	AY5
K26	23.265	AY5
K28	20.293	AY5
L12	23.720	AY5
L30	21.298	AY5
M12	23.513	AY5

M26	26.335	AY5
N04	30.622	AY5
N09	19.776	AY5
N12	19.016	AY5
neg	35.880	AY5
O01	23.704	AY5
O28	23.400	AY5
P12	23.320	AY5
P26	22.451	AY5
P30	24.069	AY5
Q12	26.400	AY5
Q30	28.685	AY5
R01	24.650	AY5
R04	32.015	AY5
R26	26.181	AY5
R30	24.764	AY5
S09	28.661	AY5
T09	26.487	AY5
A26	30.887	AY50
I04	31.595	AY50
J04	31.893	AY50
K12	31.217	AY50
K26	34.018	AY50
K28	29.994	AY50
L04	30.632	AY50
L09	30.862	AY50
L30	30.348	AY50
M09	30.568	AY50
N01	31.305	AY50
N04	27.743	AY50
N09	28.449	AY50
O04	27.954	AY50
O09	31.470	AY50
O30	30.741	AY50
P09	28.913	AY50
P12	32.057	AY50
P26	29.806	AY50
P28	31.476	AY50

Q04	24.212	AY50
Q09	26.824	AY50
Q12	30.201	AY50
R04	28.015	AY50
R09	28.886	AY50
R12	31.034	AY50

S09	28.470	AY50
T09	22.087	AY50
U09	25.601	AY50
ptrap	32.721	AY50
A26	32.780	AY51
A28	32.168	AY51
B28	31.598	AY51
K09	32.103	AY51
K28	24.966	AY51
L01	34.049	AY51
L09	30.983	AY51
L26	33.997	AY51
L28	32.598	AY51
L30	25.627	AY51
N04	32.331	AY51
O09	34.388	AY51
O28	32.680	AY51
P30	34.670	AY51
R01	31.982	AY51
R04	33.019	AY51
R12	30.104	AY51
R30	30.362	AY51
T09	32.358	AY51
U09	33.059	AY51
A26	28.345	AY52
A28	30.564	AY52
I04	30.070	AY52
J09	31.570	AY52
J26	29.697	AY52
J28	32.883	AY52
K01	30.727	AY52
K09	34.752	AY52
K30	31.711	AY52
L01	32.128	AY52
L04	33.011	AY52
L09	34.315	AY52
L26	34.899	AY52
L30	34.068	AY52
M09	30.316	AY52
M12	34.405	AY52
M28	32.748	AY52
M30	33.947	AY52

N01	31.474	AY52
N04	30.639	AY52

N12	30.028	AY52
N26	34.568	AY52
neg	35.053	AY52
O01	29.072	AY52
O04	31.334	AY52
O26	30.973	AY52
O28	31.549	AY52
P12	31.210	AY52
P28	31.296	AY52
P30	32.959	AY52
pos1	30.184	AY52
Q04	33.422	AY52
Q12	31.938	AY52
R01	30.972	AY52
R09	30.476	AY52
R12	31.382	AY52
R26	33.485	AY52
T09	30.689	AY52
U09	29.216	AY52
ptrap	32.873	AY52
M09	31.952	AY53
A26	22.710	AY54
A28	22.873	AY54
B26	26.850	AY54
B28	24.730	AY54
I04	32.553	AY54
J04	32.086	AY54
K04	33.919	AY54
K09	32.646	AY54
K30	34.534	AY54
L04	33.780	AY54
L09	32.345	AY54
L12	34.040	AY54
L30	32.395	AY54
M09	33.074	AY54
N26	30.581	AY54
O09	30.557	AY54
O26	34.293	AY54
P09	30.768	AY54
P12	29.562	AY54
P26	34.448	AY54
pos1	24.435	AY54
Q01	34.475	AY54
T09	30.589	AY54
U09	30.529	AY54

I04	29.517	AY55

K09	32.764	AY55
K28	31.236	AY55
L01	32.242	AY55
N04	30.243	AY55
N09	30.857	AY55
O04	31.932	AY55
O28	32.145	AY55
P12	30.288	AY55
Q04	27.131	AY55
Q09	30.241	AY55
R04	31.193	AY55
R09	31.314	AY55
S09	30.633	AY55
T09	25.355	AY55
U09	28.570	AY55
A26	19.197	AY56
A28	18.729	AY56
B26	20.858	AY56
B28	19.258	AY56
H04	32.479	AY56
I04	18.328	AY56
J04	17.590	AY56
J09	18.239	AY56
J26	16.966	AY56
J28	17.401	AY56
K01	15.551	AY56
K04	20.648	AY56
K09	19.218	AY56
K12	16.893	AY56
K26	18.901	AY56
K28	17.220	AY56
K30	16.414	AY56
L01	20.208	AY56
L04	19.057	AY56
L09	18.732	AY56
L12	17.209	AY56
L26	19.466	AY56
L28	19.975	AY56
L30	16.020	AY56
M09	16.649	AY56
M12	16.122	AY56
M26	23.958	AY56
M28	17.289	AY56

M30	21.501	AY56
N01	21.416	AY56
N04	28.122	AY56
N09	20.961	AY56
N12	16.578	AY56

N26	19.934	AY56
N28	21.837	AY56
neg	32.587	AY56
O01	17.193	AY56
O04	26.400	AY56
O09	17.751	AY56
O12	17.424	AY56
O26	18.727	AY56
O28	18.085	AY56
O30	19.942	AY56
P01	18.453	AY56
P09	16.156	AY56
P12	17.518	AY56
P26	17.224	AY56
P28	18.419	AY56
P30	17.850	AY56
pos1	29.685	AY56
Q01	18.409	AY56
Q04	28.232	AY56
Q09	25.423	AY56
Q12	21.730	AY56
Q30	24.440	AY56
R01	16.828	AY56
R04	26.990	AY56
R09	24.613	AY56
R12	14.823	AY56
R26	18.083	AY56
R30	17.664	AY56
S09	25.536	AY56
T09	24.625	AY56
U09	29.652	AY56
ptrap	16.125	AY56
A26	22.320	AY57
A28	22.193	AY57
B26	23.913	AY57
B28	22.779	AY57
J04	28.493	AY57
J09	25.593	AY57
J26	24.448	AY57

J28	22.629	AY57
K01	25.843	AY57
K04	30.605	AY57
K12	23.226	AY57
K26	29.043	AY57
K28	26.065	AY57
K30	24.704	AY57
L04	24.943	AY57
L12	25.696	AY57

L26	28.663	AY57
L28	23.598	AY57
L30	23.071	AY57
M04	32.568	AY57
M12	24.868	AY57
M26	32.116	AY57
M30	20.877	AY57
N01	30.351	AY57
N09	25.396	AY57
N26	27.885	AY57
N28	27.370	AY57
neg	35.712	AY57
O01	26.062	AY57
O04	31.760	AY57
O12	24.637	AY57
O26	26.614	AY57
O28	27.431	AY57
O30	26.896	AY57
P01	23.664	AY57
P09	23.785	AY57
P12	24.911	AY57
P26	27.190	AY57
P28	27.751	AY57
P30	24.486	AY57
Q01	28.129	AY57
Q04	33.160	AY57
Q09	31.040	AY57
Q12	30.689	AY57
R01	25.720	AY57
R09	34.376	AY57
R12	27.436	AY57
R26	29.115	AY57
R30	24.999	AY57
S09	34.601	AY57
T09	30.714	AY57

U09	33.844	AY57
A26	28.442	AY58
A28	30.854	AY58
B28	34.025	AY58
H04	29.390	AY58
I04	31.421	AY58
J26	30.909	AY58
J28	32.393	AY58
K04	30.602	AY58
K09	31.125	AY58
L09	34.827	AY58
L30	32.089	AY58
M09	33.420	AY58

neg	35.110	AY58
O09	29.501	AY58
Q09	29.729	AY58
A26	17.709	AY59
A28	18.563	AY59
B26	18.412	AY59
B28	14.992	AY59
I04	24.125	AY59
J04	22.142	AY59
J26	26.399	AY59
J28	28.932	AY59
K01	26.460	AY59
K04	29.891	AY59
K09	26.619	AY59
K12	26.774	AY59
K26	25.942	AY59
K28	18.701	AY59
K30	23.247	AY59
L01	28.821	AY59
L04	25.594	AY59
L09	23.420	AY59
L12	20.317	AY59
L26	23.372	AY59
L28	22.469	AY59
L30	21.424	AY59
M09	27.813	AY59
M12	21.126	AY59
M28	27.906	AY59
M30	23.019	AY59
N01	29.145	AY59
N04	29.297	AY59

N09	30.388	AY59
N12	24.254	AY59
O01	30.305	AY59
O04	30.446	AY59
O09	25.513	AY59
O12	29.805	AY59
O26	25.743	AY59
O28	27.884	AY59
P01	29.265	AY59
P09	23.178	AY59
P12	27.853	AY59
P26	25.997	AY59
P28	27.666	AY59
P30	23.818	AY59
Q01	30.861	AY59
Q04	28.938	AY59
Q09	30.104	AY59

Q12	26.602	AY59
Q30	26.656	AY59
R01	26.188	AY59
R04	28.907	AY59
R09	30.727	AY59
R12	20.724	AY59
R26	27.658	AY59
R30	21.015	AY59
S09	29.216	AY59
T09	26.776	AY59
U09	28.640	AY59
ptrap	27.295	AY59
A26	31.041	AY60
K28	30.509	AY60
L04	30.358	AY60
L12	33.507	AY60
L26	30.604	AY60
L30	31.607	AY60
Q09	34.041	AY60
R09	30.415	AY60
S09	34.445	AY60
T09	31.230	AY60
U09	34.485	AY60
H04	34.736	AY61
J26	32.998	AY61
J28	34.499	AY61
L01	33.016	AY61

L04	32.039	AY61
L09	31.054	AY61
L26	31.324	AY61
M04	31.570	AY61
M26	34.157	AY61
M30	33.136	AY61
N09	34.355	AY61
O30	33.699	AY61
P30	30.895	AY61
Q04	32.095	AY61
ptrap	30.785	AY61
A26	25.151	AY62
A28	29.564	AY62
B26	29.150	AY62
B28	29.960	AY62
H04	27.379	AY62
I04	22.702	AY62
J04	29.654	AY62
K28	30.930	AY62
L04	27.126	AY62
L09	29.622	AY62

L26	30.120	AY62
L28	33.262	AY62
L30	28.864	AY62
M30	32.193	AY62
P09	26.333	AY62
P26	28.821	AY62
P28	31.274	AY62
pos1	31.662	AY62
Q04	33.297	AY62
Q09	29.227	AY62
Q30	30.136	AY62
R04	32.658	AY62
R09	30.373	AY62
T09	28.219	AY62
ptrap	30.654	AY62
M09	31.775	AY63
A28	25.040	AY65
B28	23.061	AY65
J09	29.795	AY65
L09	27.239	AY65
L28	26.799	AY65
M12	25.814	AY65
M30	28.687	AY65

Q04	30.101	AY65
S09	30.149	AY65
T09	30.094	AY65
U09	30.250	AY65
A26	29.372	AY66
A28	29.323	AY66
B26	31.534	AY66
B28	29.340	AY66
I04	30.834	AY66
J04	29.794	AY66
J09	30.527	AY66
J26	30.677	AY66
J28	28.828	AY66
K04	30.441	AY66
K09	30.644	AY66
K26	30.155	AY66
K28	30.806	AY66
L04	31.116	AY66
L09	30.414	AY66
L26	29.528	AY66
M09	30.666	AY66
M26	30.732	AY66
N09	29.738	AY66
N26	29.221	AY66
neg	33.654	AY66

O09	31.958	AY66
P26	30.369	AY66
pos1	29.819	AY66
Q09	28.581	AY66
R09	30.839	AY66
R26	29.109	AY66
S09	29.012	AY66
U09	29.597	AY66
A26	33.420	AY67
I04	30.169	AY67
K28	30.096	AY67
L12	29.899	AY67
L30	30.895	AY67
M12	31.637	AY67
M26	30.377	AY67
N26	31.693	AY67
neg	33.175	AY67
O09	34.304	AY67
P01	31.604	AY67

Q04	31.786	AY67
R09	30.229	AY67
R12	32.082	AY67
R30	31.921	AY67
S09	29.588	AY67
T09	30.058	AY67
U09	28.037	AY67
A26	21.758	AY68
A28	21.348	AY68
B26	22.741	AY68
B28	21.243	AY68
I04	19.018	AY68
J04	22.383	AY68
J09	21.414	AY68
J26	20.984	AY68
J28	21.549	AY68
K01	21.612	AY68
K04	29.592	AY68
K09	20.158	AY68
K12	21.571	AY68
K26	18.401	AY68
K28	18.655	AY68
K30	22.403	AY68
L01	22.790	AY68
L04	20.292	AY68
L09	22.533	AY68
L12	20.705	AY68
L26	20.087	AY68
L28	25.480	AY68

L30	17.945	AY68
M09	18.590	AY68
M12	25.292	AY68
M26	24.868	AY68
M28	16.864	AY68
M30	27.268	AY68
N01	24.703	AY68
N04	29.822	AY68
N12	17.372	AY68
N26	29.236	AY68
N28	29.491	AY68
neg	33.287	AY68
O01	32.304	AY68
O09	17.486	AY68
O12	34.150	AY68

O26	17.894	AY68
O28	19.644	AY68
O30	30.863	AY68
P01	22.041	AY68
P09	21.347	AY68
P12	23.107	AY68
P26	22.509	AY68
P28	25.302	AY68
P30	18.589	AY68
Q01	23.104	AY68
Q04	29.543	AY68
Q12	25.249	AY68
Q30	29.103	AY68
R01	22.888	AY68
R04	30.114	AY68
R12	22.814	AY68
R26	27.307	AY68
R30	25.114	AY68
T09	28.521	AY68
U09	29.234	AY68
ptrap	20.561	AY68
A26	31.833	AY69
A28	30.369	AY69
B26	32.760	AY69
B28	28.578	AY69
I04	29.950	AY69
J04	30.522	AY69
K04	32.230	AY69
K12	31.676	AY69
K30	29.894	AY69
L01	30.187	AY69
L04	31.926	AY69
L09	27.045	AY69

L12	27.847	AY69
L26	34.336	AY69
M09	28.966	AY69
M12	31.533	AY69
M28	30.648	AY69
M30	27.313	AY69
N01	30.382	AY69
N09	30.773	AY69
N12	30.080	AY69
N26	32.333	AY69
N28	33.570	AY69

neg	35.403	AY69
O01	34.542	AY69
O12	32.867	AY69
O26	30.333	AY69
O28	32.686	AY69
O30	32.658	AY69
P12	30.997	AY69
P26	33.106	AY69
P28	33.634	AY69
P30	30.841	AY69
Q09	34.092	AY69
Q12	34.232	AY69
Q30	34.834	AY69
R01	30.859	AY69
R09	34.825	AY69
R26	33.274	AY69
R30	30.363	AY69
S09	32.267	AY69
ptrap	31.754	AY69
A26	23.362	AY7
A28	24.593	AY7
B26	25.629	AY7
B28	25.624	AY7
J09	30.757	AY7
K28	33.648	AY7
N09	31.789	AY7
O30	34.597	AY7
Q04	30.235	AY7
R09	32.830	AY7
A26	33.088	AY70
A28	26.190	AY70
B26	33.394	AY70
B28	31.441	AY70
I04	33.317	AY70
J04	32.259	AY70
J09	33.249	AY70
K09	32.383	AY70

K12	33.890	AY70
K30	31.636	AY70
L04	33.000	AY70
L12	33.140	AY70
M09	30.937	AY70
M12	30.810	AY70
M28	33.511	AY70

M30	32.259	AY70
N09	32.794	AY70
O09	31.152	AY70
O12	33.300	AY70
O26	31.964	AY70
O30	31.401	AY70
P09	31.072	AY70
P12	32.822	AY70
P26	32.811	AY70
P28	33.057	AY70
P30	33.832	AY70
R01	34.222	AY70
U09	33.872	AY70
B26	31.694	AY71
L09	33.886	AY71
neg	35.106	AY71
U09	31.348	AY71
A26	22.201	AY72
A28	22.605	AY72
B26	25.000	AY72
B28	24.238	AY72
L09	28.656	AY72
L26	30.907	AY72
L28	30.708	AY72
M09	32.607	AY72
M12	28.529	AY72
M28	31.778	AY72
M30	28.803	AY72
P09	30.502	AY72
R04	29.788	AY72
R26	30.529	AY72
S09	33.481	AY72
T09	29.730	AY72
ptrap	30.814	AY72
A26	20.383	AY73
A28	23.056	AY73
B26	25.544	AY73
B28	23.274	AY73
J04	29.495	AY73
K28	29.111	AY73
L09	28.286	AY73

L26	29.487	AY73
L28	26.572	AY73
M12	27.994	AY73

M30	26.001	AY73
Q04	30.376	AY73
T09	31.403	AY73
A26	29.048	AY74
A28	31.579	AY74
N09	30.826	AY74
neg	32.778	AY74
R09	32.435	AY74
A26	20.392	AY75
A28	24.782	AY75
B26	26.289	AY75
B28	24.595	AY75
H04	31.803	AY75
I04	29.356	AY75
J04	31.537	AY75
J09	32.140	AY75
J26	32.345	AY75
J28	30.380	AY75
K04	31.758	AY75
K09	30.142	AY75
K26	30.129	AY75
K28	30.114	AY75
K30	34.921	AY75
L04	30.823	AY75
L09	30.218	AY75
L26	31.458	AY75
M04	30.001	AY75
M09	32.547	AY75
M26	30.172	AY75
N04	30.840	AY75
N09	31.193	AY75
N26	30.304	AY75
neg	32.502	AY75
O09	31.013	AY75
O26	30.968	AY75
P09	29.977	AY75
P12	32.792	AY75
P26	30.244	AY75
P30	33.762	AY75
pos1	30.415	AY75
Q04	29.042	AY75
Q09	31.481	AY75
R04	30.162	AY75
R09	32.119	AY75
R26	32.106	AY75

S09	29.904	AY75
T09	27.899	AY75
U09	29.790	AY75
ptrap	30.902	AY75
B28	27.304	AY76
I04	28.396	AY76
J04	30.348	AY76
J09	27.477	AY76
J26	27.383	AY76
J28	27.116	AY76
K01	26.946	AY76
K04	29.369	AY76
K12	26.610	AY76
K28	27.297	AY76
K30	26.813	AY76
L01	27.204	AY76
L04	28.149	AY76
L12	27.960	AY76
L30	29.199	AY76
M04	30.693	AY76
M09	27.561	AY76
M12	27.319	AY76
M26	31.727	AY76
M30	29.265	AY76
N04	34.846	AY76
N28	33.522	AY76
neg	34.893	AY76
O28	30.461	AY76
O30	29.822	AY76
P01	28.472	AY76
P12	28.937	AY76
P28	32.022	AY76
P30	29.842	AY76
Q04	33.134	AY76
Q09	32.599	AY76
R01	30.420	AY76
R09	32.976	AY76
R12	26.740	AY76
R26	28.233	AY76
S09	33.499	AY76
J09	31.263	AY77
L09	30.952	AY77
M04	29.433	AY77
pos1	31.089	AY77
ptrap	31.457	AY77

A26	19.186	AY78
A28	18.685	AY78
B26	20.953	AY78

B28	18.146	AY78
H04	30.324	AY78
I04	21.013	AY78
J04	20.970	AY78
J09	24.547	AY78
J26	23.318	AY78
J28	23.605	AY78
K01	24.207	AY78
K04	22.735	AY78
K09	24.012	AY78
K12	21.711	AY78
K26	21.652	AY78
K28	19.485	AY78
K30	21.619	AY78
L01	22.408	AY78
L04	20.257	AY78
L09	20.495	AY78
L12	21.708	AY78
L26	22.481	AY78
L28	23.019	AY78
L30	19.472	AY78
M04	30.752	AY78
M09	16.955	AY78
M12	23.590	AY78
M26	21.741	AY78
M28	19.160	AY78
M30	24.500	AY78
N01	21.748	AY78
N04	26.895	AY78
N09	22.627	AY78
N12	19.115	AY78
N26	24.147	AY78
N28	24.433	AY78
neg	32.115	AY78
O01	22.811	AY78
O04	28.598	AY78
O09	15.919	AY78
O12	22.554	AY78
O26	19.605	AY78
O28	21.529	AY78
O30	21.428	AY78

P01	19.801	AY78
P09	19.608	AY78
P12	21.254	AY78
P26	18.478	AY78
P28	21.649	AY78
P30	18.743	AY78
pos1	31.378	AY78

Q01	16.803	AY78
Q04	24.912	AY78
Q09	25.489	AY78
Q12	22.328	AY78
Q30	23.697	AY78
R01	19.359	AY78
R04	24.761	AY78
R09	27.067	AY78
R12	19.300	AY78
R26	25.902	AY78
R30	23.259	AY78
S09	25.976	AY78
T09	24.037	AY78
U09	26.477	AY78
ptrap	16.278	AY78
A26	30.110	AY79
B26	29.940	AY79
B28	30.600	AY79
J04	30.379	AY79
J28	31.303	AY79
L09	30.575	AY79
neg	35.111	AY79
P09	30.375	AY79
pos1	30.392	AY79
R26	30.176	AY79
A26	19.718	AY8
A28	24.005	AY8
B26	24.966	AY8
B28	23.353	AY8
I04	30.002	AY8
J26	31.079	AY8
K28	30.671	AY8
L09	30.348	AY8
L12	31.314	AY8
L26	30.421	AY8
M04	30.036	AY8
M26	30.196	AY8

N09	29.408	AY8
O04	30.562	AY8
O26	34.263	AY8
pos1	30.440	AY8
Q04	29.738	AY8
R09	30.478	AY8
T09	28.089	AY8
U09	30.137	AY8
ptrap	28.913	AY8
A26	18.942	AY80
A28	19.703	AY80

B26	21.106	AY80
B28	18.120	AY80
I04	24.407	AY80
J04	22.847	AY80
J09	25.167	AY80
J26	24.187	AY80
J28	25.005	AY80
K01	26.150	AY80
K04	26.906	AY80
K09	26.274	AY80
K12	24.057	AY80
K26	25.348	AY80
K28	22.123	AY80
K30	21.402	AY80
L01	21.643	AY80
L04	22.775	AY80
L09	20.706	AY80
L12	26.292	AY80
L26	21.843	AY80
L28	24.859	AY80
L30	22.400	AY80
M04	31.687	AY80
M09	20.516	AY80
M12	25.173	AY80
M26	21.457	AY80
M28	22.132	AY80
M30	23.791	AY80
N01	21.177	AY80
N04	29.019	AY80
N09	25.042	AY80
N12	22.303	AY80
N26	24.847	AY80
N28	25.681	AY80

neg	34.093	AY80
O01	23.262	AY80
O04	28.390	AY80
O09	27.813	AY80
O12	23.816	AY80
O26	29.976	AY80
O28	26.197	AY80
O30	25.499	AY80
P01	25.309	AY80
P09	28.363	AY80
P12	27.596	AY80
P26	26.202	AY80
P28	26.475	AY80
P30	30.076	AY80
pos1	30.955	AY80

Q01	28.025	AY80
Q04	26.300	AY80
Q09	26.034	AY80
Q30	30.396	AY80
R01	27.962	AY80
R04	26.203	AY80
R09	27.926	AY80
R12	26.935	AY80
R26	21.359	AY80
R30	24.086	AY80
S09	25.217	AY80
T09	23.905	AY80
U09	27.600	AY80
ptrap	19.795	AY80
A26	29.857	AY81
A28	30.282	AY81
B26	31.175	AY81
I04	29.442	AY81
J04	29.372	AY81
J09	31.485	AY81
J26	30.360	AY81
K04	31.177	AY81
K12	30.420	AY81
K26	30.287	AY81
K28	27.297	AY81
K30	34.242	AY81
L09	28.990	AY81
L12	30.327	AY81
L30	29.426	AY81

M09	30.226	AY81
M30	30.186	AY81
N04	29.484	AY81
neg	34.738	AY81
O09	30.586	AY81
O12	29.384	AY81
O28	29.547	AY81
P01	31.229	AY81
P09	28.885	AY81
P12	30.326	AY81
P30	31.276	AY81
Q04	29.894	AY81
Q12	31.058	AY81
Q30	30.370	AY81
R09	30.373	AY81
R12	28.711	AY81
S09	30.619	AY81
T09	27.198	AY81
U09	29.466	AY81

ptrap	30.337	AY81
A26	32.851	AY82
A28	28.149	AY82
B26	32.146	AY82
B28	27.605	AY82
J09	30.848	AY82
N01	34.375	AY82
neg	34.819	AY82
pos1	31.115	AY82
Q01	34.266	AY82
ptrap	33.208	AY82
A26	21.709	AY83
A28	24.289	AY83
B26	24.839	AY83
B28	23.584	AY83
I04	30.818	AY83
K01	31.753	AY83
K09	29.644	AY83
K26	30.841	AY83
K28	30.559	AY83
L30	30.983	AY83
M12	30.722	AY83
N12	31.012	AY83
N28	30.830	AY83
O01	30.907	AY83

O09	31.454	AY83
O30	30.682	AY83
P12	29.653	AY83
Q04	31.221	AY83
R01	30.871	AY83
R30	30.804	AY83
T09	28.985	AY83
U09	29.932	AY83
ptrap	30.284	AY83
A26	25.416	AY84
A28	26.598	AY84
B26	30.392	AY84
B28	27.983	AY84
I04	29.153	AY84
J26	27.788	AY84
J28	29.061	AY84
K09	27.747	AY84
K26	26.802	AY84
K28	27.927	AY84
U09	31.752	AY84
A26	20.555	AY85
A28	21.966	AY85
B26	24.390	AY85

B28	22.259	AY85
H04	30.416	AY85
I04	25.509	AY85
J04	27.343	AY85
J09	29.428	AY85
J26	27.474	AY85
J28	28.972	AY85
K04	28.607	AY85
K09	28.308	AY85
K12	29.436	AY85
K26	26.979	AY85
K28	24.611	AY85
L01	29.466	AY85
L04	29.161	AY85
L09	28.183	AY85
L12	27.882	AY85
L26	30.371	AY85
L28	29.038	AY85
L30	27.495	AY85
M04	29.274	AY85
M09	25.499	AY85

M12	27.603	AY85
M26	26.838	AY85
M28	25.464	AY85
M30	27.699	AY85
N01	30.550	AY85
N04	30.182	AY85
N09	29.420	AY85
N12	25.648	AY85
N26	30.632	AY85
neg	32.646	AY85
O04	28.441	AY85
O09	27.834	AY85
O12	29.491	AY85
O26	30.901	AY85
O28	29.494	AY85
O30	29.349	AY85
P09	29.190	AY85
P12	28.575	AY85
P26	28.672	AY85
P30	27.751	AY85
pos1	30.211	AY85
Q01	28.291	AY85
Q04	26.623	AY85
Q09	30.516	AY85
R01	29.391	AY85
R04	33.128	AY85
R09	27.792	AY85

R12	28.929	AY85
R26	30.433	AY85
R30	28.713	AY85
S09	29.550	AY85
T09	25.007	AY85
U09	27.737	AY85
ptrap	21.798	AY85
A26	20.984	AY86
A28	25.265	AY86
B26	25.846	AY86
B28	24.777	AY86
I04	29.363	AY86
K26	30.578	AY86
L09	30.733	AY86
L28	29.815	AY86
M04	30.876	AY86
M09	28.145	AY86

M26	30.071	AY86
M28	27.497	AY86
N09	30.837	AY86
N12	28.133	AY86
neg	32.840	AY86
O09	30.855	AY86
O26	31.021	AY86
P09	30.630	AY86
pos1	30.866	AY86
Q04	28.629	AY86
Q09	29.810	AY86
R30	30.392	AY86
S09	30.480	AY86
T09	27.176	AY86
U09	29.362	AY86
ptrap	25.112	AY86
A26	18.467	AY87
A28	21.983	AY87
B26	25.373	AY87
B28	23.655	AY87
I04	28.000	AY87
J04	29.894	AY87
J09	28.639	AY87
J26	28.790	AY87
J28	28.472	AY87
K01	30.319	AY87
K04	33.868	AY87
K09	30.350	AY87
K12	30.284	AY87
K28	27.633	AY87
K30	29.796	AY87

L12	29.880	AY87
L28	32.997	AY87
L30	28.244	AY87
M04	32.659	AY87
M09	30.696	AY87
M12	29.301	AY87
M30	30.036	AY87
N12	29.652	AY87
O04	30.840	AY87
O09	30.183	AY87
O30	29.485	AY87
P01	34.209	AY87
P12	28.506	AY87

P30	30.279	AY87
Q01	28.735	AY87
Q04	29.108	AY87
Q09	31.003	AY87
R30	30.225	AY87
S09	32.502	AY87
T09	26.560	AY87
U09	30.110	AY87
ptrap	30.996	AY87
A26	29.279	AY88
N01	34.083	AY88
N26	31.977	AY88
O26	32.966	AY88
Q04	30.854	AY88
ptrap	29.593	AY88
A26	27.572	AY89
B28	29.570	AY89
L09	32.840	AY89
M09	31.860	AY89
R09	30.470	AY89
R26	34.482	AY89
A28	27.643	AY9
H04	31.258	AY9
J04	29.744	AY9
J28	27.712	AY9
K09	28.409	AY9
K12	29.185	AY9
K30	28.890	AY9
L12	30.739	AY9
L28	31.239	AY9
M28	30.883	AY9
N01	30.259	AY9
N04	31.377	AY9
neg	34.138	AY9
O04	30.820	AY9

O09	28.467	AY9
O26	29.417	AY9
O30	28.847	AY9
P01	29.301	AY9
P09	28.184	AY9
P28	30.957	AY9
Q04	32.019	AY9
Q09	29.119	AY9
Q12	30.051	AY9

Q30	28.718	AY9
R04	31.339	AY9
R09	28.908	AY9
R30	28.788	AY9
T09	28.676	AY9
U09	29.846	AY9
ptrap	28.548	AY9
A26	27.280	AY90
A28	30.964	AY90
B28	31.342	AY90
J09	31.185	AY90
J26	30.917	AY90
K26	32.640	AY90
K28	22.682	AY90
K30	32.593	AY90
L01	29.283	AY90
L04	30.909	AY90
L09	31.625	AY90
L12	34.927	AY90
L28	28.196	AY90
L30	22.832	AY90
M12	28.822	AY90
M30	31.209	AY90
N09	31.756	AY90
N12	31.132	AY90
O01	29.724	AY90
O12	29.796	AY90
R01	26.097	AY90
R04	31.458	AY90
R12	29.190	AY90
R30	23.976	AY90
T09	32.240	AY90
ptrap	31.829	AY90
A26	18.707	AY91
A28	20.699	AY91
B26	21.330	AY91
B28	20.453	AY91
H04	31.470	AY91
I04	29.315	AY91

J04	25.229	AY91
J09	30.981	AY91
J26	29.070	AY91
J28	23.364	AY91
K04	29.981	AY91

K09	23.591	AY91
K12	24.282	AY91
K26	24.474	AY91
K28	25.662	AY91
K30	29.673	AY91
L01	21.812	AY91
L04	29.509	AY91
L09	28.177	AY91
L12	30.017	AY91
L26	30.616	AY91
L28	27.708	AY91
L30	24.930	AY91
M09	20.234	AY91
M12	30.126	AY91
M26	21.858	AY91
M28	21.565	AY91
M30	28.199	AY91
N01	21.347	AY91
N09	29.967	AY91
N12	20.774	AY91
N26	29.745	AY91
O01	29.693	AY91
O09	28.598	AY91
O26	26.193	AY91
O28	27.807	AY91
O30	30.329	AY91
P01	21.602	AY91
P09	26.988	AY91
P12	26.992	AY91
P26	27.666	AY91
P28	25.192	AY91
Q01	27.914	AY91
Q04	28.856	AY91
Q09	31.030	AY91
Q12	28.812	AY91
Q30	28.718	AY91
R01	27.904	AY91
R04	28.987	AY91
R09	30.140	AY91
R12	29.171	AY91
R26	26.224	AY91
R30	26.258	AY91
S09	27.526	AY91

T09	27.104	AY91

U09	28.853	AY91
ptrap	21.661	AY91
A26	16.928	AY92
A28	16.877	AY92
B26	18.317	AY92
B28	17.101	AY92
H04	26.977	AY92
I04	22.268	AY92
J04	20.144	AY92
J09	28.146	AY92
J26	23.649	AY92
J28	22.906	AY92
K01	20.590	AY92
K04	25.868	AY92
K09	24.551	AY92
K12	22.364	AY92
K26	24.302	AY92
K28	17.848	AY92
K30	21.245	AY92
L01	23.927	AY92
L04	21.565	AY92
L09	20.852	AY92
L12	19.726	AY92
L26	23.571	AY92
L28	20.669	AY92
L30	19.780	AY92
M04	27.582	AY92
M09	22.216	AY92
M12	22.223	AY92
M26	25.951	AY92
M28	21.277	AY92
M30	23.036	AY92
N01	23.681	AY92
N04	28.044	AY92
N09	22.355	AY92
N12	21.268	AY92
N26	20.879	AY92
N28	26.536	AY92
neg	32.439	AY92
O01	22.131	AY92
O04	29.447	AY92
O09	23.739	AY92
O12	20.404	AY92
O26	23.977	AY92
O28	26.314	AY92

O30	21.937	AY92
P01	24.502	AY92

P09	21.339	AY92
P12	22.414	AY92
P26	22.162	AY92
P28	24.758	AY92
P30	21.640	AY92
pos1	27.034	AY92
Q01	22.682	AY92
Q04	24.016	AY92
Q09	26.502	AY92
Q12	24.837	AY92
Q30	25.153	AY92
R01	19.933	AY92
R04	25.230	AY92
R09	26.566	AY92
R12	19.567	AY92
R26	26.569	AY92
R30	18.944	AY92
S09	26.890	AY92
T09	23.280	AY92
U09	25.287	AY92
ptrap	21.865	AY92
A26	20.355	AY93
A28	19.962	AY93
B26	21.948	AY93
B28	20.427	AY93
I04	20.306	AY93
J04	20.868	AY93
J09	27.538	AY93
J26	23.283	AY93
J28	20.728	AY93
K01	24.220	AY93
K04	18.403	AY93
K09	22.715	AY93
K12	23.620	AY93
K26	21.138	AY93
K28	20.053	AY93
K30	22.228	AY93
L01	24.589	AY93
L09	23.038	AY93
L12	21.554	AY93
L26	22.262	AY93
L28	22.467	AY93

L30	19.523	AY93
M04	29.618	AY93
M09	17.466	AY93
M12	22.039	AY93
M26	22.621	AY93
M28	19.839	AY93

M30	25.912	AY93
N01	23.882	AY93
N04	27.523	AY93
N12	19.654	AY93
neg	32.730	AY93
O01	21.093	AY93
O04	27.773	AY93
O12	21.202	AY93
O26	29.734	AY93
O28	27.474	AY93
O30	20.732	AY93
P01	26.170	AY93
P09	19.190	AY93
P12	24.299	AY93
P26	18.349	AY93
P28	21.147	AY93
P30	28.298	AY93
pos1	29.411	AY93
Q01	16.459	AY93
Q04	26.100	AY93
Q09	27.835	AY93
Q12	22.141	AY93
Q30	23.749	AY93
R01	19.147	AY93
R04	25.566	AY93
R09	29.982	AY93
R12	19.130	AY93
R26	25.757	AY93
R30	23.135	AY93
S09	26.019	AY93
T09	24.409	AY93
U09	27.147	AY93
ptrap	16.822	AY93
A26	20.691	AY94
A28	21.233	AY94
B26	23.493	AY94
B28	21.982	AY94
H04	31.043	AY94

I04	27.933	AY94
J04	30.971	AY94
J09	30.541	AY94
J26	28.924	AY94
J28	31.336	AY94
K09	26.641	AY94
K12	34.777	AY94
K26	23.968	AY94
K28	27.451	AY94
K30	34.035	AY94

L01	34.500	AY94
L04	30.812	AY94
L09	27.704	AY94
L12	29.723	AY94
L26	32.452	AY94
L28	29.055	AY94
L30	28.120	AY94
M04	31.173	AY94
M09	26.175	AY94
M26	26.513	AY94
M28	26.780	AY94
N09	30.518	AY94
N12	24.805	AY94
N26	30.027	AY94
neg	35.761	AY94
O01	33.566	AY94
O09	32.166	AY94
O12	30.882	AY94
O26	29.694	AY94
O28	31.553	AY94
P12	29.841	AY94
P26	32.276	AY94
pos1	31.281	AY94
Q01	29.514	AY94
Q04	29.977	AY94
R12	30.482	AY94
R26	29.598	AY94
R30	30.401	AY94
T09	28.875	AY94
U09	30.609	AY94
ptrap	22.412	AY94
A26	22.018	AY95
A28	23.778	AY95
B26	24.892	AY95

B28	22.484	AY95
H04	28.667	AY95
I04	29.864	AY95
J04	28.795	AY95
J09	30.142	AY95
J26	27.991	AY95
J28	28.885	AY95
K01	31.850	AY95
K04	27.690	AY95
K09	28.873	AY95
K12	34.109	AY95
K26	27.575	AY95
K28	28.076	AY95
L04	27.892	AY95

L09	29.068	AY95
L26	28.314	AY95
L30	31.192	AY95
M04	28.352	AY95
M09	28.194	AY95
M12	29.864	AY95
M26	27.633	AY95
N09	27.945	AY95
N26	28.275	AY95
N28	31.186	AY95
neg	33.939	AY95
O01	32.014	AY95
O09	28.767	AY95
O12	28.682	AY95
O26	28.501	AY95
O30	28.856	AY95
P09	29.714	AY95
P26	27.775	AY95
pos1	27.876	AY95
Q04	27.717	AY95
Q09	28.378	AY95
R09	28.208	AY95
R26	28.657	AY95
S09	28.971	AY95
T09	26.450	AY95
U09	28.210	AY95
ptrap	29.500	AY95
B28	33.064	AY97
I04	34.478	AY97
J28	32.752	AY97

K01	32.521	AY97
K12	31.364	AY97
K30	32.452	AY97
L26	31.393	AY97
L30	34.453	AY97
P12	34.658	AY97
P26	32.506	AY97
A26	27.692	AY98
A28	31.794	AY98
B28	30.759	AY98
J04	30.391	AY98
K01	30.624	AY98
K28	31.985	AY98
L09	31.542	AY98
L26	30.482	AY98
M04	32.522	AY98
M09	30.786	AY98
M30	32.091	AY98

N26	31.839	AY98
neg	33.174	AY98
P26	33.549	AY98
pos1	30.667	AY98
Q04	30.103	AY98
Q09	30.567	AY98
R04	28.878	AY98
R09	31.965	AY98
R26	31.753	AY98
T09	28.700	AY98
U09	30.234	AY98
A26	31.234	AY99
B28	34.437	AY99
J28	31.958	AY99
M28	32.883	AY99
T09	31.186	AY99

Table S3.4 *16S rRNA* PCR components

Components	Volume, µL
Taq mix polymerase	25
Forward-primer (10 nM)	1
Reverse-primer (10 nM)	1
DNA Template	1
Nuclease-free water	22

Table S3.5 *16S rRNA* PCR programme

Stage	Temprature,°C	Time	Cycle number
Stage 1	94	5 min	1

Stage 2	94	30s	30
	60	30s	
	72	1min30s	
Stage 3	72	10min	1
Stage 4	4	∞	1

Table S4.1 Testing primers for single-plex high-GC PCR

Primers	Primer length, nt	GC content, %	Tm, °C	Ta, °C	Amplicon length, nt
aroE_1_FWD	17	59	64	65	424
aroE_1_REV	16	63	64		
aroE_2_FWD	19	63	69	70	424
aroE_2_REV	19	63	69		
aroE_3_FWD	20	70	75	75	398

aroE_3_REV	20	70	75		

Table S4.2 PCR components in single-plex high-GC PCR

Component	Volume, µl	Final concentration
Q5 reaction buffer (5x)	5	1x
dNTPs, 10 mM	0.5	200 µM
Q5 DNA polomerase, 0.5 U	0.25	0.02 Unit/µl
Primer (FWD or REV) (10 µM)	1.25	0.025 µM
DNA template	5	1-1000, pg/µl
GC enhancer	5	-
PCR-grade water	6.75	-

Table S4.3 PCR programme of single-plex high-GC PCR

Stage	Cycle number	Denature	Anneal	Extend
1	1	98°C, 30 s	-	-
2	35	98°C, 10 s	65/70/75°C, 30 s	75°C, 12 s

3	1		-	75°C, 2 min

Table S4.4 PCR components in touch-down multiplex-PCR reaction

Q5 reaction buffer (5x)	5	1x
dNTPs, 10 mM	0.5	200 μM
Q5 DNA polomerase, 0.5 U	0.25	0.02 Unit/μl
Primer pool1 or 2 (50 μM)	2.68	0.020 μM per primer
DNA template	5	LOD: 3 copies per reaction
GC enhancer	5	-
PCR-grade water	6.57	-

Table S4.5 PCR programme of touch-down multiplex-PCR reaction

Stage	Cycle number	Denature	Anneal/Extend
1	1	98°C, 30s	-
2	10	98°C, 15s	65°C, 5min
3	25	98°C, 15s	65°C, 5min, delta t = -0.1°C

Table S4.6 103 target genes in the *P. aeruginosa* panel

Target gene	Type	Resistance mechanism
AAC(3)-Ia	aminoglycoside	antibiotic inactivation
AAC(3)-Ib	aminoglycoside	antibiotic inactivation
AAC(3)-Ib/AAC(6')-Ib"	aminoglycoside	antibiotic inactivation
AAC(3)-Ic	aminoglycoside	antibiotic inactivation
AAC(3)-IIIa	aminoglycoside	antibiotic inactivation
AAC(3)-IIIb	aminoglycoside	antibiotic inactivation
AAC(3)-IIIc	aminoglycoside	antibiotic inactivation
AAC(6')-29a	aminoglycoside	antibiotic inactivation
AAC(6')-30/AAC(6')-Ib' fusion protein	aminoglycoside	antibiotic inactivation
AAC(6')-31	aminoglycoside	antibiotic inactivation
AAC(6')-32	aminoglycoside	antibiotic inactivation
AAC(6')-Iai	aminoglycoside	antibiotic inactivation
AAC(6')-Iaj	aminoglycoside	antibiotic inactivation
AAC(6')-Ib'	aminoglycoside	antibiotic inactivation
AAC(6')-Ib10	aminoglycoside	antibiotic inactivation
AAC(6')-Iib	aminoglycoside	antibiotic inactivation
AAC(6')-Iic	aminoglycoside	antibiotic inactivation
AAC(6')-Iid	aminoglycoside	antibiotic inactivation
aadA11	aminoglycoside	antibiotic inactivation
aadA13	aminoglycoside	antibiotic inactivation
aadA15	aminoglycoside	antibiotic inactivation
ANT(2")-Ia	aminoglycoside	antibiotic inactivation
ANT(4')-Iia	aminoglycoside	antibiotic inactivation
ANT(4')-Iib	aminoglycoside	antibiotic inactivation
APH(3")-Ib	aminoglycoside	antibiotic inactivation
APH(3')-Iib	aminoglycoside	antibiotic inactivation
APH(6)-Id	aminoglycoside	antibiotic inactivation
rmtA	aminoglycoside	antibiotic target alteration
rmtC	aminoglycoside	antibiotic target alteration
PmpM	aminoglycoside-fluoroquinolone	efflux overexpression
HMB-1	carbapenem	antibiotic inactivation
SPM-1	carbapenem	antibiotic inactivation

LCR-1	cephalosporin-penam	antibiotic inactivation
OXA-13	cephalosporin-penam	antibiotic inactivation
OXA-18	cephalosporin-penam	antibiotic inactivation
OXA-198	cephalosporin-penam	antibiotic inactivation
OXA-205	cephalosporin-penam	antibiotic inactivation
OXA-3	cephalosporin-penam	antibiotic inactivation
OXA-31	cephalosporin-penam	antibiotic inactivation
OXA-33	cephalosporin-penam	antibiotic inactivation
OXA-36	cephalosporin-penam	antibiotic inactivation
OXA-45	cephalosporin-penam	antibiotic inactivation
OXA-46	cephalosporin-penam	antibiotic inactivation
OXA-5	cephalosporin-penam	antibiotic inactivation
OXA-50	cephalosporin-penam	antibiotic inactivation
dfrA17	diaminopyrimidine	antibiotic target replacement
FosC	fosfomycin	antibiotic inactivation
16srRNA1	housekeeping	housekeeping
16srRNA2	housekeeping	housekeeping
gyrA	housekeeping	housekeeping
mphD	macrolide	antibiotic inactivation
acsA	mlst	mlst
aroE	mlst	mlst
guaA	mlst	mlst
mutL	mlst	mlst
nuoD	mlst	mlst
ppsA	mlst	mlst
trpE	mlst	mlst
BEL-1	monabactam-cephalosporin-penam	antibiotic inactivation
PDC-2	monobactam-carbapenem-cephalosporin	antibiotic inactivation
PDC-73	monobactam-carbapenem-cephalosporin	antibiotic inactivation
AIM-1	multidrug	antibiotic inactivation
GIM-1	multidrug	antibiotic inactivation
IMP-12	multidrug	antibiotic inactivation
IMP-14	multidrug	antibiotic inactivation
IMP-15	multidrug	antibiotic inactivation

IMP-16	multidrug	antibiotic inactivation
IMP-18	multidrug	antibiotic inactivation
IMP-20	multidrug	antibiotic inactivation
IMP-22	multidrug	antibiotic inactivation
IMP-29	multidrug	antibiotic inactivation
IMP-30	multidrug	antibiotic inactivation
IMP-33	multidrug	antibiotic inactivation
IMP-35	multidrug	antibiotic inactivation
IMP-41	multidrug	antibiotic inactivation
IMP-7	multidrug	antibiotic inactivation
IMP-9	multidrug	antibiotic inactivation
KPC-5	multidrug	antibiotic inactivation
oprD	multidrug	permeability
PER-1	multidrug	antibiotic inactivation
VIM-1	multidrug	antibiotic inactivation
VIM-13	multidrug	antibiotic inactivation
VIM-3	multidrug	antibiotic inactivation
VIM-7	multidrug	antibiotic inactivation
CARB-1	penam	antibiotic inactivation
CARB-3	penam	antibiotic inactivation
CARB-4	penam	antibiotic inactivation
arnA	peptide	antibiotic target alteration
basS	peptide	antibiotic target alteration
catB10	phenicol	antibiotic inactivation
catB6	phenicol	antibiotic inactivation
catB7	phenicol	antibiotic inactivation
mexR	regulator	efflux overexpression
MexS	regulator	efflux overexpression
MexT	regulator	efflux overexpression
MexZ	regulator	efflux overexpression
nalC	regulator	efflux overexpression
nalD	regulator	efflux overexpression
nfxB	regulator	efflux overexpression
arr-2	rifamycin	antibiotic inactivation
arr-4	rifamycin	antibiotic inactivation
arr-7	rifamycin	antibiotic inactivation

tet(G)	tetracycline	antibiotic efflux

Table S4.7 Primers of 103 target genes

primer name	seq(5'-3')	pool
16S_rRNA1_1_LEFT	TGAACCTGCTGCTCCTCGACGA	1
16S_rRNA1_1_RIGHT	TCCAGGCCGAGGACCAGGTT	1
16S_rRNA1_2_LEFT	GCGATGGGCGTACCGAAGCT	2
16S_rRNA1_2_RIGHT	CGGTTTCCACGCGCAGGATG	2
16S_rRNA2_1_LEFT	TTCGTCGACCAGCCTCTCTCCC	1
16S_rRNA2_1_RIGHT	AGGCGGATGGATCAGCGGCA	1
16S_rRNA2_2_LEFT	CAGCGAGCGTTGCGAAGTGC	2
16S_rRNA2_2_RIGHT	CGAAGTCGCCCCAGAGTTGCT	2
AAC(3)-Ia _2_LEFT	TCTTTTCGGTCGTGAGTTCGGAGAC	2
AAC(3)-Ia _2_RIGHT	AGGTGGCGGTACTTGGGTCGA	2
AAC(3)-Ia_1_LEFT	TCATTCGCACATGTAGGCTCGGC	1
AAC(3)-Ia_1_RIGHT	GAGCCACTGCGGGATCGTCA	1
AAC(3)-Ib _1_LEFT	TTACGCAGCAGGGCAGTCGC	1
AAC(3)-Ib _1_RIGHT	AGCGACTGCCGGATCGTCAC	1
AAC(3)-Ib _2_LEFT	TCGGCAAAGAGTTTGAGGACATTCCA	2
AAC(3)-Ib _2_RIGHT	TCGAAGTGCATGACGTCTTCCCG	2
AAC(3)-Ib/AAC(6')-Ib" _1_LEFT	CGGCCCTGACGAAATTTCAGCCA	2
AAC(3)-Ib/AAC(6')-Ib" _1_RIGHT	AGAGAGCGACTGCCGGATCGT	2
AAC(3)-Ib/AAC(6')-Ib" _2_LEFT	GTGGCGGTTGAACTTGGCGC	1
AAC(3)-Ib/AAC(6')-Ib" _2_RIGHT	ATCCGTCCCCGCCTCCAAGA	1

AAC(3)-Ib/AAC(6')-Ib" _3_LEFT	CGTTTTAGCGCAAGAGTCCGTCACT	2
AAC(3)-Ib/AAC(6')-Ib" _3_RIGHT	TGTTCGCTCGAATGCCTGGCG	2
AAC(3)-Ic _1_LEFT	AACCAAGATTACCCGCCTCAACTCTC	1
AAC(3)-Ic _1_RIGHT	AGTGCATCACGTCCTCCCGGA	1
AAC(3)-Ic _2_LEFT	GGCGAGGCTTTTGAGGACGCT	2
AAC(3)-Ic _2_RIGHT	AAGTGCATCACGTCCTCCCGG	2
AAC(3)-IIIa _2_LEFT	ATGGCTACGGAGTCGAGTCGCC	2
AAC(3)-IIIa _2_RIGHT	ACGCTGAGTCACCGAACCGTG	2
AAC(3)-IIIa_1_LEFT	CGCATACACACGCGCACCTTG	1
AAC(3)-IIIa_1_RIGHT	AGCGGCGACTCGACTCCGTA	1
AAC(3)-IIIb _2_LEFT	TTCACTGCCGACCACCCGCT	2
AAC(3)-IIIb _2_RIGHT	CGAGGGCGTACCGAACCGTT	2
AAC(3)-IIIb_1_LEFT	ATGGTCCATGCCGCCGTCAG	1
AAC(3)-IIIb_1_RIGHT	AGCAGGGTCAGCGTGTCGAG	1
AAC(3)-IIIc _2_LEFT	ACCGCAGACCATCCCCTCGA	2
AAC(3)-IIIc _2_RIGHT	AGCCGAAGCGCGATTCCAGC	2
AAC(3)-IIIc_1_LEFT	AACCTCTCGTGCTTGCCGCC	1
AAC(3)-IIIc_1_RIGHT	GCCATAGCCGTAGTCGAGGGGA	1
AAC(6')-29a_1_LEFT	TTACGCAGCAGCAGTCGCCC	1
AAC(6')-29a_1_RIGHT	TGGTTCCCAAGCCTTTGCCCA	1
AAC(6')-29a_2_LEFT	CGTTTTAGCGCAAGAGTCCGTCACT	2
AAC(6')-29a_2_RIGHT	TGTTCGCTCGAATGCCTGGCG	2
AAC(6')-30/AAC(6')-Ib' fusion protein _2_LEFT	GCGTTTGCCTCAGACAGGGATGA	2
AAC(6')-30/AAC(6')-Ib' fusion protein _2_RIGHT	TCCCAAGCCTTTGCCCAGTTGT	2

AAC(6')-30/AAC(6')-Ib'_1_LEFT	CCGACCCGTAGAACAAAGTGACGC	1
AAC(6')-30/AAC(6')-Ib'_1_RIGHT	TGTTAGGTGGCGCTCCCGTG	1
AAC(6')-30/AAC(6')-Ib'_3_LEFT	CGTTTTAGCGCAAGAGTCCGTCACT	1
AAC(6')-30/AAC(6')-Ib'_3_RIGHT	TGTTCGCTCGAATGCCTGGCG	1
AAC(6')-31 _2_LEFT	GCCCTGGCGAAAGAGTCCGT	2
AAC(6')-31 _2_RIGHT	TGACGTGTTTGAACCATGTATACGGCC	2
AAC(6')-31_1_LEFT	TGACCGAGCACGACCTTCCGA	1
AAC(6')-31_1_RIGHT	TTTGCGCCACAAAACCCGCC	1
AAC(6')-32 _2_LEFT	GGTCAAACACTACCTGCCGAGGG	2
AAC(6')-32 _2_RIGHT	TGGGCCGTCTGGTGTGGTGA	2
AAC(6')-32_1_LEFT	ACACCCGTTACCTTGCGCCT	1
AAC(6')-32_1_RIGHT	ATGGTTGTTTGGCGCCGGGT	1
AAC(6')-Iai _2_LEFT	GCCCAGAATCATGGCCAACACTCC	2
AAC(6')-Iai _2_RIGHT	TGGCTTATTCTTCCCATTTGCGTTTGG	2
AAC(6')-Iai_1_LEFT	GCCCAGAATCATGGCCAACACTCC	1
AAC(6')-Iai_1_RIGHT	TGGCTTATTCTTCCCATTTGCGTTTGG	1
AAC(6')-Iaj _2_LEFT	TGAAGTCGAAGAGTGTATTGATCTGCCCA	2
AAC(6')-Iaj _2_RIGHT	TCGGGTTTTCTTAGTCCATTTGCATTTGG	2
AAC(6')-Iaj_1_LEFT	AAAACTTGGCCAACTATTCAAAGCGCA	1
AAC(6')-Iaj_1_RIGHT	TCGGGTTTTCTTAGTCCATTTGCATTTGG	1
AAC(6')-Ib' _2_LEFT	CGTTTTAGCGCAAGAGTCCGTCACT	2
AAC(6')-Ib' _2_RIGHT	TGTTCGCTCGAATGCCTGGCG	2
AAC(6')-Ib'_1_LEFT	CGATTCCGTCACACTGCGCCT	1
AAC(6')-Ib'_1_RIGHT	TCGCTCGCAAGTTGCTCGGC	1
AAC(6')-Ib10 _1_LEFT	TGTGAAAGAACAAGACGCTGCCGA	1

AAC(6')-Ib10 _1_RIGHT	GCTGCCTGCGAACTTGCGGT	1
AAC(6')-Ib10 _2_LEFT	ACCTGTGAAAGAACAAGACGCTGCC	2
AAC(6')-Ib10 _2_RIGHT	GCTGCCTGCGAACTTGCGGT	2
AAC(6')-IIb _2_LEFT	GCGCGTTGGCCGAAGAAGGA	2
AAC(6')-IIb _2_RIGHT	TCACCAACGGACGCTCGTGC	2
AAC(6')-IIb_1_LEFT	CCCGGCGTTGTTACTCTGCGT	1
AAC(6')-IIb_1_RIGHT	TCGCGCGCATGTTGTTCGGG	1
AAC(6')-IIc _2_LEFT	GCCTTACATCGCAATGCTAGATGACGAA	2
AAC(6')-IIc _2_RIGHT	GTGCGCAGGCTTTCGAACGC	2
AAC(6')-IIc_1_LEFT	TGTCCGCCAACAATGCCGCA	1
AAC(6')-IIc_1_RIGHT	TTTCGTTACGGCCGGGTCGC	1
AAC(6')-IId _2_LEFT	CGTTTTAGCGCAAGAGTCCGTCACT	2
AAC(6')-IId _2_RIGHT	TGTTCGCTCGAATGCCTGGCG	2
AAC(6')-IId_1_LEFT	ACCTTGCGATGCTCTATGAGTGGC	1
AAC(6')-IId_1_RIGHT	ACCTTGCCTCTCAAACCCCGC	1
aadA11 _1_LEFT	CGCAGTACCCGCCGAGATTTCG	2
aadA11 _1_RIGHT	AACATCGGTTGTGGCGGGCT	2
aadA11 _2_LEFT	TGCAATTCGGGGAGTGGCAGC	1
aadA11 _2_RIGHT	AACGCCGCCAACTGATCCGC	1
aadA11 _3_LEFT	CAGGTTCGGCCGCAGAGGAT	2
aadA11 _3_RIGHT	AGACATCACTGGCATGGCACCAAG	2
aadA13 _1_LEFT	AGGGACTCAGTGACCGCCGA	2
aadA13 _1_RIGHT	CGATGTCGATCGTGGCTGGCT	2
aadA13 _2_LEFT	GCCAAGCGCGAACTGCAATTTGG	1
aadA13 _2_RIGHT	ACTTATCTGCGCGCAAGACCGA	1
aadA13 _3_LEFT	CGAAAGCGAGAGAACATAGCGTGGC	2
aadA13 _3_RIGHT	GCCGAGCACCTTGGTGATCTCG	2
aadA15 _2_LEFT	ACAAAAGCAAGAGAACATAGCGTTGCC	2
aadA15 _2_RIGHT	ACTTGATGATCTCGCCTTTCACAAAGCG	2

aadA15_1_LEFT	AGGTAGTTGGCGTCATCGAGCG	1
aadA15_1_RIGHT	AAGAGTTCCTCCGCCGCTGG	1
acsA_1_LEFT	GCATCCCTGTACCCCGTGCA	1
acsA_1_RIGHT	GCCTCGGGGATCATCGGCAT	1
acsA_2_LEFT	CAAGTTCGCCAACGCCCTGC	2
acsA_2_RIGHT	ATGAACAGCGGGTCCTCGGC	2
acsA_3_LEFT	CCTGATGAAGGTTGCCGGCAGC	1
acsA_3_RIGHT	GAACCGAGCAGACGCAGGCT	1
acsA_4_LEFT	ATCCGCGCGATGATGGCCGA	2
acsA_4_RIGHT	TCGCCGGTGAAGTACATGCCC	2
acsA_5_LEFT	GGACCCTGTTCGGCGACCAT	1
acsA_5_RIGHT	TGCGCCGCATGATCTTGCCC	1
acsA_6_LEFT	AACGTCTCCGGCCACCGCAT	2
acsA_6_RIGHT	GCGATGGGTCTCGATCAGGTGC	2
AIM-1 _1_LEFT	ACGTCGCTTCACCCTGCTGG	2
AIM-1 _1_RIGHT	TCGAATTGCGGGTCGGTGCG	2
AIM-1 _2_LEFT	ACTGCAGAAGGCCACGGGTG	1
AIM-1 _2_RIGHT	ATCCGGTTCCACAGGCCGCT	1
AIM-1 _3_LEFT	GCCGGTCGCCAACATCGTCA	2
AIM-1 _3_RIGHT	TTCCTCGGCCAGGCGCTTCT	2
ANT(2")-Ia _1_LEFT	TGGACACAACGCAGGTCACATTGAT	1
ANT(2")-Ia _1_RIGHT	TCCCAGATGATCGCCTCCCAGC	1
ANT(2")-Ia _2_LEFT	CGCAAGCACGATGATATTGATCTGACGT	2
ANT(2")-Ia _2_RIGHT	TTTCCGCCCCGAGTGAGGTG	2
ANT(4')-IIa _2_LEFT	ACACGGTAGAAGCGCTCGGG	2
ANT(4')-IIa _2_RIGHT	CGCAAGTACCAATCCTGCTTCCAGC	2
ANT(4')-IIa_1_LEFT	ACCTACTGGATCGATCGTCTGCGA	1
ANT(4')-IIa_1_RIGHT	CGCAAGGTCGTCACCCCGAA	1
ANT(4')-IIb _2_LEFT	CGCCGAACCCGAGGTGGAGAAT	2

ANT(4')-IIb _2_RIGHT	TTTCTCGAGCAACCCGGCGT	2
ANT(4')-IIb_1_LEFT	TCGCCCGTTGGGTTGATCGC	1
ANT(4')-IIb_1_RIGHT	CAGGTCATCGCCGCGAGCAA	1
APH(3")-Ib _1_LEFT	TCGCATTCTGACTGGTTGCCTGT	1
APH(3")-Ib _1_RIGHT	TGACGGCATTGCGGGACACC	1
APH(3")-Ib _2_LEFT	GGACGCGCCGTTGATGTGGT	2
APH(3")-Ib _2_RIGHT	AAGGCAAGGCGTTCGCGGTC	2
APH(3')-IIb _2_LEFT	TGGAACGGCGTCTGGACACC	2
APH(3')-IIb _2_RIGHT	AGCCTGAAGTACGCCAGCCG	2
APH(3')-IIb_1_LEFT	ATGCAGCCACCTCCATGCCG	1
APH(3')-IIb_1_RIGHT	TGTCCAGACGCCGTTCCAGG	1
APH(6)-Id _1_LEFT	TGCCGCCTGTTTTTCCTGCTCA	2
APH(6)-Id _1_RIGHT	GCGTTTTGATCATCGCGCGCC	2
APH(6)-Id _2_LEFT	ACCCCTGCCTTCTGCCCTTCT	1
APH(6)-Id _2_RIGHT	AAGCTGCGGAAAGGCACCCA	1
APH(6)-Id _3_LEFT	TGACTACGTCCACGCGGCGA	2
APH(6)-Id _3_RIGHT	TGTCGCACCTGCTTGATCGCG	2
arnA _1_LEFT	AGCCGTCGTCTTCGCCTACCA	1
arnA _1_RIGHT	GGTCACCCCGGTCTGCGTTT	1
arnA _2_LEFT	ACCTGCACGGTTCGCTGCTG	2
arnA _2_RIGHT	GGCGCTCCAGACGATCAGCTT	2
arnA _3_LEFT	TTGTACGACCTGGTGCGCGC	1
arnA _3_RIGHT	CAGCAGGCGTTCGGACAGGT	1
arnA _4_LEFT	ACGCAACTGGCCACCGAGCT	2
arnA _4_RIGHT	GTGGAAGGGAACACCACGCGTT	2
arnA _5_LEFT	CGAGGAAAACCTGCGGATCGTCC	1
arnA _5_RIGHT	GTCGTCGACGTCGGTGAAGCA	1
arnA _6_LEFT	CTGGTCGAAGGCACGCCGAT	2
arnA _6_RIGHT	TCTTGCCGATGGTCTCGCGC	2

aroE_1_LEFT	ATTCGGCAACCCCATCGGCC	1
aroE_1_RIGHT	ATCAGCAACTCCGCCGGGCA	1
aroE_2_LEFT	TGCTGATCGCCAACCGCACG	2
aroE_2_RIGHT	AGCGTCTCCAACACTGGCGC	2
arr-2 _2_LEFT	ACCGTTCTATCATGGAACCAAAGCCA	2
arr-2 _2_RIGHT	TGTAAACCACGGCGCTTTAAGTCCT	2
arr-2_1_LEFT	ACAAGCAGGTGCAAGGACCGT	1
arr-2_1_RIGHT	TGTAAACCACGGCGCTTTAAGTCCT	1
arr-4 _2_LEFT	GCACCAAAGCCAAACTCACGGTT	2
arr-4 _2_RIGHT	CGAGGCCACGACGCTGAAGA	2
arr-4_1_LEFT	TCCCACTTCGCATGACAACTGCT	1
arr-4_1_RIGHT	CGAGGCCACGACGCTGAAGA	1
arr-7 _2_LEFT	ACGGCACCAAAGCAAAACTCGC	2
arr-7 _2_RIGHT	GAGGCCGCGACGCTGAAGAT	2
arr-7_1_LEFT	ACTGGATTCCCACCTCGCACGA	1
arr-7_1_RIGHT	GAGGCCGCGACGCTGAAGAT	1
basS _1_LEFT	GGCTTCGTGCTGTGCTGGCT	1
basS _1_RIGHT	AGGTTCATGGTGTGCTGGCGG	1
basS _2_LEFT	TGCTCCAGCTGAACATCGACGG	2
basS _2_RIGHT	TGGGCCTTGGGGTCTTCCGA	2
basS _3_LEFT	CGCCCATGAAATCCGCACGC	1
basS _3_RIGHT	GTGCTGGGCGCGATTCTCGA	1
basS _4_LEFT	TCGACCTGTGGCTGAAGGCGA	2
basS _4_RIGHT	TATGTGACCGCCCGCTGTCC	2
BEL-1 _1_LEFT	CTGCTGCTCTACCCGTTATTGCTGT	1
BEL-1 _1_RIGHT	AAAGGTGGGCCGCCAACCTC	1
BEL-1 _2_LEFT	ATCCGCCCTCCAACTCAGCG	2
BEL-1 _2_RIGHT	ACGCCACATCGTTACGGCCG	2
BEL-1 _3_LEFT	AGTAAGTCGCCTTGATCGTATTGAACCGA	1

BEL-1 _3_RIGHT	GCAATTAATAACGCCCTTTCCTCGCC	1
CARB-1 _1_LEFT	TGGCATTTTCGCTTTTAATACCATCCGTG	2
CARB-1 _1_RIGHT	CCTTTGGGGCCACCTACAGCA	2
CARB-1 _3_LEFT	CCGTATTGAGCCTGATTTAAATGAAGGTAAGC	2
CARB-1 _3_RIGHT	AGCGCGACTGTGATGTATAAACGTCA	2
CARB-1_2_LEFT	CGTGCTTCGCAACTATGACTACAAGTGA	1
CARB-1_2_RIGHT	ACTCCGAGCACCAAATCCGCC	1
CARB-3 _1_LEFT	TGGCATTTTCGCTTTTAATACCATCCGTG	2
CARB-3 _1_RIGHT	CCTTTGGGGCCACCTACAGCA	2
CARB-3 _3_LEFT	CCGTATTGAGCCTGATTTAAATGAAGGTAAGC	2
CARB-3 _3_RIGHT	AGCGCGACTGTGATGTATAAACGTCA	2
CARB-3_2_LEFT	CGTGCTTCGCAACTATGACTACAAGTGA	1
CARB-3_2_RIGHT	ACTCCGAGCACCAAATCCGCC	1
CARB-4 _2_LEFT	GCAACTATGACGACAAGTGATAATGCAGCA	2
CARB-4 _2_RIGHT	ACTCCGAGCACCAAATCCGCC	2
CARB-4_1_LEFT	TCGCTTTTAATACCGTCTATGGTGTTTGCA	1
CARB-4_1_RIGHT	ATCCGTCACGCTTTCAGGACCTC	1
CARB-4_3_LEFT	CGGAGATAAAGAAACCCGTCTAGACCGT	1
CARB-4_3_RIGHT	TTGCATCATTTCGATCTGCTATTGAAGCC	1
catB10 _2_LEFT	GCACCGTTATGACTGGGCGTCT	2
catB10 _2_RIGHT	AGCCTTGCCAGTATCGATGCAGC	2
catB10_1_LEFT	GGCAAACTGCTGGCCGACCA	1
catB10_1_RIGHT	TTCGTGACCAAAGCGCGGCT	1
catB6 _2_LEFT	CGGCATGACTGGGTAACATCTTTCCC	2
catB6 _2_RIGHT	GCAATCCCACGCCAATGCCTG	2
catB6_1_LEFT	AAAGGGAAACTACTTTCAGAGCAAGTGACT	1
catB6_1_RIGHT	TCGGCTGCCTATTACCGCGC	1
catB7 _2_LEFT	TGGGCGTCGACCTTCCCCTT	2
catB7 _2_RIGHT	GTGGCCTGGCGCTGTTTCCA	2

catB7_1_LEFT	GGGCAACTATTTCGAGAGCCCCTT	1
catB7_1_RIGHT	ACGCCGGGCATGAACATCGC	1
dfrA17 _1_LEFT	TGCAGTGTCAGAAAATGGCGTAATCGG	1
dfrA17 _1_RIGHT	AGGGAATTTGATATCACCTTCGACTTCAACG	1
dfrA17 _2_LEFT	TGGCGTAATCGGTAGTGGTCCTGA	2
dfrA17 _2_RIGHT	CTCAGGCATTATAGGGAATTTGATATCACCTTCG	2
FosC _1_LEFT	TGTTGTCGGACATCACGCGGA	1
FosC _1_RIGHT	CCAACTGGCGCTGCTGCATG	1
FosC _2_LEFT	ACATGCCCTTGCTCACTGGGG	2
FosC _2_RIGHT	TCCAAATTGACGCAACCCGCAC	2
GIM-1 _1_LEFT	GCTTTAGCTCAGGGTCATAAACCGCT	1
GIM-1 _1_RIGHT	AGCCCATTTCCCAGTGTGAATTCGT	1
GIM-1 _2_LEFT	GCCCGTGAAGGAAAGCCGGT	2
GIM-1 _2_RIGHT	GCTTCAGCGGTCGGTTGCATTAATT	2
guaA_1_LEFT	CCCAAGACATTCACGCCCACCG	1
guaA_1_RIGHT	GTGACCTTGTCGCCGTGGCT	1
guaA_2_LEFT	ACGGCATCGAGGACCACGTG	2
guaA_2_RIGHT	ACACGCAGGTCAGTTGGTCGC	2
guaA_3_LEFT	TCCAAGGTCCTGCTCGGCCT	1
guaA_3_RIGHT	TCGAACTGCATGTCCTCCGGC	1
guaA_4_LEFT	GGCAAGGCCCACGTGATCAAGT	2
guaA_4_RIGHT	AGCAGTTCGTAGGGCAGGTGC	2
guaA_5_LEFT	ACATGGTCTACCGCCACCCGT	1
guaA_5_RIGHT	CATTCCCACTCGATAGTCGCCGG	1
gyrA_1_LEFT	ACTGGCCAAAGAAATTCTCCCGGT	1
gyrA_1_RIGHT	CAGCAGTTCGTGGGCCAGCT	1
gyrA_2_LEFT	GGTTCGGTGGACGGCGACAA	2
gyrA_2_RIGHT	GCGCGGGCACGGATGTAGAT	2
gyrA_3_LEFT	CCCACCGCCGGCATCATCAA	1

gyrA_3_RIGHT	TCCTTCAGGTTCAGCGTGCGC	1
gyrA_4_LEFT	GCGTGTTCGGCATCAACGTGG	2
gyrA_4_RIGHT	TTCTCGTGCTCCAGGCCGGT	2
gyrA_5_LEFT	AGTACTACCTGTCGCCGGAGCA	1
gyrA_5_RIGHT	AGCAGGAGGGTCGCATGGCT	1
gyrA_6_LEFT	CCACCGGGATGAAGGACGAGGA	2
gyrA_6_RIGHT	CGGACGGCTGAACTGCACCA	2
gyrA_7_LEFT	CCGACCGGCGCCTACATCTT	1
gyrA_7_RIGHT	GCCGTTGCGCTCGTTGGTGA	1
gyrA_8_LEFT	AGGGGCAGCAGCTGATCTCCA	2
gyrA_8_RIGHT	TCGCCCAGAGATTCGGCAGC	2
HMB-1_1_LEFT	TGCTGCTGACAAACATCGTTCTTGCA	1
HMB-1_1_RIGHT	GCCAGCGCCCGGATAAAAAGC	1
HMB-1_2_LEFT	TCCATTCCGACCTACGCATCAGAACA	2
HMB-1_2_RIGHT	TTCTTCTTTGCAACCGCCTCAAGC	2
IMP-12 _2_LEFT	TCCCACGTATGCATCTGAGTTAACAAATGA	2
IMP-12 _2_RIGHT	AGCCCTTTAACAGCCTGCTCCC	2
IMP-12_1_LEFT	GCCTGATTTGAAAATTGAGAAGCTTGAAGAGG	1
IMP-12_1_RIGHT	TCTTGAGTATGTCCTGGGCCGGG	1
IMP-14 _1_LEFT	CCAGATTTAAAAATTGAGAAGCTTGACGAAGGCG	1
IMP-14 _1_RIGHT	ACCACTACGTTATCTGGAGTGTGCCC	1
IMP-14 _2_LEFT	CGGGCGGAATAGAGTGGCTTAATTCT	2
IMP-14 _2_RIGHT	AGTGATGCATCTCCAACTTCACTGTGAC	2
IMP-15 _2_LEFT	TCTCAATCTATCCCCACGTATGCATCTGA	2
IMP-15 _2_RIGHT	AGCCCTTTAACAGCCTGCTCCC	2
IMP-15_1_LEFT	CGCAGGAGAGTCTTTGCCAGATTTAAAAATTG	1
IMP-15_1_RIGHT	ACGTTATCTGGAGTGTGCCCTGGA	1
IMP-16 _2_LEFT	TCCCACGTATGCATCTGAATTAACAAACGA	2

IMP-16 _2_RIGHT	AGCCCTTTAACAGCCTGCTCCC	2
IMP-16_1_LEFT	TGAGAAGCTTGAAGACGGTGTTTATGTTCA	1
IMP-16_1_RIGHT	TCTTTTCAGGCAACCAAACCACTACGT	1
IMP-18 _2_LEFT	TCTCAATCTATCTCCACGTATGCCTCTGA	2
IMP-18 _2_RIGHT	ACAGCCTGCTCCCATGTGCG	2
IMP-18_1_LEFT	TGCTGCTGCAGATGATTCTTTGCCT	1
IMP-18_1_RIGHT	TCTTGAGTGTGTCCTGGACCTGGA	1
IMP-20 _2_LEFT	AGGGGGAATAGAGTGGCTTAATTCTCAATCT	2
IMP-20 _2_RIGHT	ACGTTTCAAGAGTGATGCGTCCCC	2
IMP-20_1_LEFT	GGCTTTGCCTGATTTAAAAATCGAGAAGCT	1
IMP-20_1_RIGHT	ACTACGTTATCTTGAGTGTGCCCCG	1
IMP-22 _2_LEFT	TCTCAATCAATTCCCACGTATGCATCTGA	2
IMP-22 _2_RIGHT	AGCCCTTTAACAGCCTGCTCCC	2
IMP-22_1_LEFT	CGCAGGAGAGTCTTTGCCCGA	1
IMP-22_1_RIGHT	ACGTTATCTTGAGTGTGCCCGGG	1
IMP-29 _2_LEFT	TCTCAATCTATCCCCACGTATGCATCTGA	2
IMP-29 _2_RIGHT	AGCCCTTTAACAGCCTGCTCCC	2
IMP-29_1_LEFT	GCCGCAGCAGAGTCTTTGCCA	1
IMP-29_1_RIGHT	TCTGGAGTGTGCCCTGGACCA	1
IMP-30 _1_LEFT	CCGCAGCAGAGTCTTTGCCAGAT	1
IMP-30 _1_RIGHT	AGTGTGTCCCGGGCCTGGAT	1
IMP-30 _2_LEFT	CTCGATCTATCCCCACGTATGCATCTGA	2
IMP-30 _2_RIGHT	TCGTTTAACCCTTTAACCGCCTGCT	2
IMP-33 _1_LEFT	GCAGTTGCAGAAGAATCTTTGCCTGATT	1
IMP-33 _1_RIGHT	ACGTTATCTTGAGTGTGCCCTGGG	1
IMP-33 _2_LEFT	CGGCTGGAATAGAGTGGCTTAATTCTCA	2
IMP-33 _2_RIGHT	AGCCCTTTAACAGCCTGCTCCC	2
IMP-35 _1_LEFT	CATCGTTCGAAGAAGTTAACGGTTGGGG	1
IMP-35 _1_RIGHT	GGTTATCTTGAGTGTGACCTGGGCC	1

IMP-35 _2_LEFT	GAGTGGCTTAATTCTCAATCTATTCCCACGT	2
IMP-35 _2_RIGHT	AGCCCTTTAAGAGCTTGTTCCCATGT	2
IMP-41 _1_LEFT	TGCCGCCGGAGAGTCTTTGC	1
IMP-41 _1_RIGHT	TCCTGAGTGTGCCCTGGTCCT	1
IMP-41 _2_LEFT	GCGCAGGTGGAATTGAGTGGCT	2
IMP-41 _2_RIGHT	ACAGCCTGCTCCCAAGTGCG	2
IMP-7 _1_LEFT	CGCAGGAGAGTCTTTGCCAGATTTAAAAATTG	1
IMP-7 _1_RIGHT	ACGTTATCTGGAGCGTGCCCTG	1
IMP-7 _2_LEFT	TCTCAATCTATCCCCACGTATGCATCTGA	2
IMP-7 _2_RIGHT	ACCGTCTGCTCCCATGTAAGCTTC	2
IMP-9 _1_LEFT	GGAGCGTCTTTGCCTGATTTAAAAATTGAGAAG C	1
IMP-9 _1_RIGHT	ACCACTACGTTATCTTGAGTGTGCCC	1
IMP-9 _2_LEFT	GCACGGGTGGAATAGAGTGGCT	2
IMP-9 _2_RIGHT	AGCCCTTTTACCGCCTGTTCCC	2
KPC-5 _1_LEFT	TCGCCGTCTAGTTCTGCTGTCTTGT	2
KPC-5 _1_RIGHT	GCCGCCCAACTCCTTCAGCA	2
KPC-5 _2_LEFT	CGGCCGCCGTGCAATACAGT	1
KPC-5 _2_RIGHT	ATCCTTGTTAGGCGCCCGGG	1
KPC-5 _3_LEFT	AGCTGAACTCCGCCATCCCAG	2
KPC-5 _3_RIGHT	ACTGCCCGTTGACGCCCAAT	2
LCR-1 _1_LEFT	GCACCCTTCTGGCCTTTGGTCT	1
LCR-1 _1_RIGHT	CCCCCACTCGAAGGGCAATCTG	1
LCR-1 _2_LEFT	CCAGGACCAGACTCTAGACAGTGCG	2
LCR-1 _2_RIGHT	CTGCGTTCGTAGGGGCAGGT	2
LCR-1 _3_LEFT	CCCAGCCTATATTCAACAGACAAACTATGGT	1
LCR-1 _3_RIGHT	GCCTTCAGCGCGTCTTTGGC	1
mexR_1_LEFT	AATCCCGACCTGATGCCCGC	1
mexR_1_RIGHT	AGATGCACCAGGGTGGCCTG	1

mexR_2_LEFT	TGGCGGTCTTCCAGCATGTGC	2
mexR_2_RIGHT	AGATGCACCAGGGTGGCCTG	2
mexS_1_LEFT	CGTTTTCATCAGTTTGGCCCGCC	1
mexS_1_RIGHT	CTTCAACCCGGCCAGGTCGA	1
mexS_2_LEFT	CCAGCGTCTACTACACCGGCCT	2
mexS_2_RIGHT	GTGCTTCTTGAACGCGGCGC	2
mexS_3_LEFT	CCGAGGGCAAGGGCGTCAAT	1
mexS_3_RIGHT	GGCCGCGCTTTGGACAGGTT	1
mexT_1_LEFT	TCAGTGATCCTATGCCCCTCCGG	1
mexT_1_RIGHT	GTGCTGGTCGCCGGATCGAA	1
mexT_2_LEFT	GCCGGCGCTGGATTCCATCT	2
mexT_2_RIGHT	CCTTGCGCTTGCGGCCGAAT	2
mexT_3_LEFT	AAGCCGAAGATCCTCCGCGC	1
mexT_3_RIGHT	GCCGATGAACATGCTGATCCGC	1
mexZ_1_LEFT	AGAGGAATCCCAGAAAACCCGCG	1
mexZ_1_RIGHT	AGCGTTGCCCCTGCTTCTCG	1
mexZ_2_LEFT	GTGCCGGCGCTGGATATCCT	2
mexZ_2_RIGHT	GCGTCCGCCAGCAACAGGTA	2
mphD_1_LEFT	GGTCGTTTTTGCTCTTGATACAAAGGGG	1
mphD_1_RIGHT	TCTGTAGCGGGTTTCCAATTGCTCA	1
mphD_2_LEFT	AGTGAGCAATTGGAAACCCGCTACA	2
mphD_2_RIGHT	GCTTTTGCTCCAACGATAAGGCTTTCA	2
mutL_1_LEFT	GCACCGCGTATCCAGCTGCT	1
mutL_1_RIGHT	CTGTACCCGCGGCTGCATGT	1
mutL_2_LEFT	TCGGTAGCGCGCCTGACCAT	2
mutL_2_RIGHT	CTGCGGGAGAAGGTCGGCAA	2
mutL_3_LEFT	ATCTTCGCCCTGCACGAGGC	1
mutL_3_RIGHT	ACGATGCAGGGTGCCATAGAGGA	1
mutL_4_LEFT	GTGCACCCGACCAAGCACGA	2

mutL_4_RIGHT	TCCTGGGCACTCTCGGGCAA	2
mutL_5_LEFT	CCTACAAGGCCTACTTCGCGCC	1
mutL_5_RIGHT	TCCAGCAGGTCGGCGATGAC	1
mutL_6_LEFT	GTTGCTGGTGCCGGAGTCGA	2
mutL_6_RIGHT	TCCAGCTCGTCCAGGCCCAA	2
nalC_1_LEFT	TGCTTCTCCCCGTCTGACCGA	1
nalC_1_RIGHT	ATAGCTCTGCTGCGGCCCCT	1
nalC_2_LEFT	TGAGCGCCACCCTCGAGCAT	2
nalC_2_RIGHT	GGGGCTCTGCGACAGGTGTT	2
nalD_1_LEFT	TGCGACGCACAAAGGAAGATTCTGA	1
nalD_1_RIGHT	TCGCGGGCGAACAACTGCTC	1
nalD_2_LEFT	GCAACTCACCGAGCGCCTGT	2
nalD_2_RIGHT	CGCACCAGGCCACGGAACAT	2
nfxB_1_LEFT	ACCCTGATTTCCCATGACGAGCG	1
nfxB_1_RIGHT	AGTTCGGTGAACACGGCCGC	1
nfxB_2_LEFT	CGTACTGAACCAGATCATCCAGGCC	2
nfxB_2_RIGHT	TGGAGGCGCCATGGAGGAAC	2
nuoD_1_LEFT	TCCGCTCTGTACATCCCGCCT	1
nuoD_1_RIGHT	TGGCCGGTGAAGGTGATGCC	1
nuoD_2_LEFT	GCCCAACGCCAACTGGTACGA	2
nuoD_2_RIGHT	CGATGCGGTCGGTGTAGGGGAT	2
nuoD_3_LEFT	AGATCATCGACTGCGTCCCGGA	1
nuoD_3_RIGHT	AGGAACTCGCGGACCAGCTTG	1
nuoD_4_LEFT	CAGCGCGCCTACAAGGTGGT	2
nuoD_4_RIGHT	TCGCCCATCTTGACCATGCAGC	2
nuoD_5_LEFT	CGCCAGAGCCTGCGGATCAT	1
nuoD_5_RIGHT	CAGCGGTCCACGTCAGCCAT	1
oprD_1_LEFT	TGAAAGTGATGAAGTGGAGCGCCA	1
oprD_1_RIGHT	GTCGGTTGCATCTCGCCCCA	1

oprD_2_LEFT	GGCAAGCCGCGCGATGACTA	2
oprD_2_RIGHT	CGATTGGTCGGATGCCAGTGGG	2
oprD_3_LEFT	CGGCGCCGAACTCGAAGACA	1
oprD_3_RIGHT	TGGTGCCATCGATGTCCTTGCC	1
oprD_4_LEFT	GCAGGTGGCGACTCGATTTTCCT	2
oprD_4_RIGHT	CAGCGGATAGTCGACGATCAGGC	2
OXA-13 _1_LEFT	TCTTGCCACTTTCGCGCATGC	1
OXA-13 _1_RIGHT	ACTCATCACGAACAGCGCCTGC	1
OXA-13 _2_LEFT	ACGCTACTCGCCTGCATCGAC	2
OXA-13 _2_RIGHT	GCGCCGAGCTTTGTCATCACC	2
OXA-13 _3_LEFT	TGGTGATGACAAAGCTCGGCGC	1
OXA-13 _3_RIGHT	ACCAAGCGCTGATGTTCTACCCG	1
OXA-13 _4_LEFT	CGGGTAGAACATCAGCGCTTGGT	2
OXA-13 _4_RIGHT	GCACGATTGCCTCCCTCTTGAAAAG	2
OXA-18 _1_LEFT	GGAGCCTGTCCATGAGCGGAA	2
OXA-18 _1_RIGHT	TGCCGTAATCGAAGCGCTGCA	2
OXA-18 _3_LEFT	CCCGGCAAGCATAACGGCCT	2
OXA-18 _3_RIGHT	AGTTTTCCGACAGGGCCGGC	2
OXA-18_2_LEFT	ATTCGCAGGAGCTGACGCGC	1
OXA-18_2_RIGHT	CGTTAGGCGGGCGAAGACGA	1
OXA-198 _2_LEFT	CAGGTCGGCGCGGAGAAGTAT	2
OXA-198 _2_RIGHT	CGCGTTAGCTTCTGACGTAAGGGC	2
OXA-198_1_LEFT	TGCATAAACACATGAGTAAGCTCTTCATCGC	1
OXA-198_1_RIGHT	TCGATACTTCTCCGCGCCGAC	1
OXA-205 _2_LEFT	CGCAAGACGCTATTTAAAGCAAATTGACTATGGC	2
OXA-205 _2_RIGHT	AGAGCGAAGGATTGCCCGCG	2
OXA-205_1_LEFT	GGCAATCCGATTCCTCACCATACTGC	1
OXA-205_1_RIGHT	GCGTCTTGCGTTGTCCTCTCCG	1

OXA-3 _1_LEFT	GGCAATCCGAATCTTTGCAATACTTTTCTCC	2
OXA-3 _1_RIGHT	GCGTCGAGCCTTGTCTTCACCG	2
OXA-3 _3_LEFT	TGAAGCAAATCGACTATGGCAACGC	2
OXA-3 _3_RIGHT	GCGCTGCGTCCGAGTTGACT	2
OXA-3_2_LEFT	CAGCAATGCGGAATTCTACTGTCTGGA	1
OXA-3_2_RIGHT	AGAGCGAAGGATTGCCCGCA	1
OXA-31 _1_LEFT	ACAGCAGCGCCAGTGCATCA	1
OXA-31 _1_RIGHT	TCGAGCCATGCTTCTGTTAATCCGT	1
OXA-31 _2_LEFT	CGTGGATGCAATTTTCTGTTGTTTGGGT	2
OXA-31 _2_RIGHT	CGACCCCAAGTTTCCTGTAAGTGCG	2
OXA-33 _2_LEFT	CGTGGATGCAATTTTCTGTTGTTTGGGT	2
OXA-33 _2_RIGHT	CGACCCCAAGTTTCCTGTAAGTGCG	2
OXA-33_1_LEFT	ACAGCAGCGCCAGTGCATCA	1
OXA-33_1_RIGHT	TCGAGCCATGCTTCTGTTAATCCGT	1
OXA-36_1_LEFT	ACTGCGTGTCTTTCAAGTACGGCATT	1
OXA-36_1_RIGHT	GCCACCACTGATATTCTGGTTGCCA	1
OXA-36_2_LEFT	CCAGAGAAGTTGGCGAAGTAAGAATGCAG	2
OXA-36_2_RIGHT	AGCCACCAATGATGCCCTCACTT	2
OXA-45 _2_LEFT	AGCCCGGTAAGAGCAACGGC	2
OXA-45 _2_RIGHT	TCGCGAGATCCGGCAATGCC	2
OXA-45_1_LEFT	TGTCATTCTGGGCGCGGCAC	1
OXA-45_1_RIGHT	TGCGTCAGGCCGTTGCTCTT	1
OXA-46 _1_LEFT	TGGCAATCCGATTCTTCACCATACTGC	2
OXA-46 _1_RIGHT	CCCAAACCGTAGAATTTCGCATCGC	2
OXA-46 _2_LEFT	GGCGTTAACCGAAGCTTTGCAGG	1
OXA-46 _2_RIGHT	TCCACCCTACCCACCAGCCA	1
OXA-46 _3_LEFT	GATATCGGAGAGGACAAAGCAAGACGT	2
OXA-46 _3_RIGHT	ACGAAGGATTGCCCGTGCGA	2
OXA-5 _2_LEFT	CCAGAGAAGTTGGCGAAATAAGAATGCAAA	2

OXA-5 _2_RIGHT	AGCCACCAATGATGATGCCTTCACT	2
OXA-5_1_LEFT	GCACCGCGCTCTCAGAGTCT	1
OXA-5_1_RIGHT	AGCCAGAATTTGTCAATGCCTCCCC	1
OXA-50 _2_LEFT	GATGCGCGCCAATGTCTCGC	2
OXA-50 _2_RIGHT	CTTGCCCAGTTCGACGCGCTT	2
OXA-50_1_LEFT	GCCCTCTCCTCTTCAGTGCCCT	1
OXA-50_1_RIGHT	CTGGCCGATTTCCGCGTTGC	1
PDC-2 _1_LEFT	TGAAGGCACTGGTCGACGCC	1
PDC-2 _1_RIGHT	AGTCGCGGATCTGTGCCTGG	1
PDC-2 _2_LEFT	ATCAGCCTGCTCGACCTCGC	2
PDC-2 _2_RIGHT	AGGCGCTCCGGATGCAGGTT	2
PDC-2 _3_LEFT	TGCCAACCTGCATCCGGAGC	1
PDC-2 _3_RIGHT	TGCTCCAGGCCGCTGAGGAT	1
PDC-73_1_LEFT	AGATTCCCCTGCCTGTGCGG	1
PDC-73_1_RIGHT	GCGGTATAGGTCGCGAGGTCGA	1
PDC-73_2_LEFT	AAGATGCGCCTCGACGACCG	2
PDC-73_2_RIGHT	GGTCTTCACCCCGTAGCCTTCG	2
PDC-73_3_LEFT	TATGGCAAGGACGACCGCCC	1
PDC-73_3_RIGHT	TGCGGTTGGCCAGGATCACC	1
PER-1 _1_LEFT	ACTGCCTCGACGCTACTGATGGT	1
PER-1 _1_RIGHT	TCATGCAAAGCAGCTGGTCCACC	1
PER-1 _2_LEFT	GCACAGCGATAACGTGGCCTGT	2
PER-1 _2_RIGHT	CAACCAGCAAGGGCCGTCCA	2
PER-1 _3_LEFT	TGCAGTATCAAAACTGGACCTCGATGAA	1
PER-1 _3_RIGHT	TGGGCTTAGGGCAGAAAGCTTTTTCA	1
PmpM_1_LEFT	CCTTTCCCGTGGCTTGCGCA	1
PmpM_1_RIGHT	TGCCCTTGAGGTACAGCAGGCT	1
PmpM_2_LEFT	GTGGTGGTTGTCGGAGCCGA	2
PmpM_2_RIGHT	CCACTCCCAGCGCGAGAACA	2

PmpM_3_LEFT	CCTCGGCATGCTGTTCTGGGT	1
PmpM_3_RIGHT	ACATCGCGGCGATCTGCTCG	1
PmpM_4_LEFT	GCCTGATGTTGTTGCTGCGCG	2
PmpM_4_RIGHT	CGCTCGTGCTGGCGGATGAA	2
ppsA_1_LEFT	ACGTAGTTTCCCTCGATAAGCTCGGC	1
ppsA_1_RIGHT	TCCACGCCGCGGATATTGAGGA	1
ppsA_2_LEFT	ACCGCCGAAGACCTTCCGGA	2
ppsA_2_RIGHT	TTGCTACCCAGGTTGCGGCG	2
ppsA_3_LEFT	ACGTCCACAAGCCGACCCTG	1
ppsA_3_RIGHT	CGATGGCACGTCCTTCCACCA	1
ppsA_4_LEFT	AAACCGTGAAGAGCCGCGCC	2
ppsA_4_RIGHT	AAGGTCGGGCATGGCGTCGA	2
ppsA_5_LEFT	GCGATACCGGCTTCATCTTCGAAGG	1
ppsA_5_RIGHT	AGAGCTTGCCGCCGATCAGG	1
ppsA_6_LEFT	CATCGTGCGCCTGTCCGACT	2
ppsA_6_RIGHT	TGCCGGAATCGCGATCCAGG	2
ppsA_7_LEFT	CGCGGCGAGAACGGACTGAA	1
ppsA_7_RIGHT	TCGGCGAGGAAGAACCAGGTATCC	1
rmtA _2_LEFT	ACATCTGTATGGGCGTGCGACA	2
rmtA _2_RIGHT	TGTACACAAGCTCTATTCCAATGGTCTTGG	2
rmtA_1_LEFT	TGCCCTAGCGTCCATCCTTTCCT	1
rmtA_1_RIGHT	ACATCGCCCAACCCCTGATGGA	1
rmtC _2_LEFT	TTTACGGAGGCTGGCATGCTTGG	2
rmtC _2_RIGHT	TTGCATGATTCTCAGATCTGACCCAACA	2
rmtC_1_LEFT	AAGTAACAGCCAAAGTACTCACAAGTGGT	1
rmtC_1_RIGHT	TGCCCAAGCATGCCAGCCTC	1
SPM-1 _1_LEFT	CGAGAGCCCTGCTTGGATTCATGG	2
SPM-1 _1_RIGHT	CACGTCTCCGCGCCCATCTT	2
SPM-1 _2_LEFT	ACGCACTTTCATTTGGACGGCAC	1

SPM-1 _2_RIGHT	TGGCATCTCCCAGATAACCAAGTTCCT	1
SPM-1 _3_LEFT	AGCAGCTGAGTTCTATAAAAACGAGGATCTG	2
SPM-1 _3_RIGHT	CGGCCTTTTCCGCGACCTTG	2
tet(G) _1_LEFT	TGCCCTGCTGATCGTCGGTC	2
tet(G) _1_RIGHT	ACCGAGCATGCCACCAAGTGC	2
tet(G) _3_LEFT	GGCCCTCTTTCAAGCCGGCT	2
tet(G) _3_RIGHT	ACAATGAAGTTGCGAAAGGTCTGCG	2
tet(G)_2_LEFT	CGCACGCTGGTTCGGCTACA	1
tet(G)_2_RIGHT	AGCCGGCTTGAAAGAGGGCC	1
trpE_1_LEFT	CGCGAAGAATTCCTGCGGCTG	1
trpE_1_RIGHT	ATATCCGGGTTGCCCAGCGG	1
trpE_2_LEFT	CCTGGTCGGCTACTTCGGTTACG	2
trpE_2_RIGHT	TCGATGGACATGCGCTGCGA	2
trpE_3_LEFT	ACGCGGTAGGAAGGATCAAGGACT	1
trpE_3_RIGHT	GTGACCTTCACCGCGCCGAT	1
trpE_4_LEFT	TGATCGACCTGGGGCGCAAC	2
trpE_4_RIGHT	GCCCGGCGCTTGTTGATGGT	2
trpE_5_LEFT	CCAACGTCATGCACATCGTGTCCA	1
trpE_5_RIGHT	ATTCGACGCTCTGCTCGGCC	1
VIM-1 _1_LEFT	GACCGCGTCTATCATGGCTATTGCG	1
VIM-1 _1_RIGHT	TGCGTACGTTGCCACCCCAG	1
VIM-1 _2_LEFT	CTGGGGTGGCAACGTACGCA	2
VIM-1 _2_RIGHT	GCTTGAGCAAGTCTAGACCGCCC	2
VIM-13 _2_LEFT	TCGACACGCCGTCTAGCCGA	2
VIM-13 _2_RIGHT	ACTCGGCGACTGAGCGATTTGT	2
VIM-13_1_LEFT	ACCGCCTCTCTAATGGCTGTAGCT	1
VIM-13_1_RIGHT	AACCTCGTTCCCCTCTGCCTCG	1
VIM-3_1_LEFT	ACCGCGTCTGTCATGGCTGT	1
VIM-3_1_RIGHT	GCGTGTCGACGGTGATGCGT	1

VIM-3_2_LEFT	ACGCATCACCGTCGACACGC	2
VIM-3_2_RIGHT	ACGTTCGCTGTGTGCTGGAGC	2
VIM-7 _2_LEFT	ACCCTTGACACGCCAGCTGG	2
VIM-7 _2_RIGHT	CACCGGGCGTACTTTGTGCG	2
VIM-7_1_LEFT	GCAGCTTTCTGGTTGGTATCAGTGCA	1
VIM-7_1_RIGHT	CACCTCGTTTCCCGCCGCTT	1

Table S5.1 51 core genes retrieved in the 51-cgMLST scheme

lpg0104	lpg1869
lpg0127	lpg2008
lpg0287	lpg2053
lpg0329	lpg2264
lpg0331	lpg2317
lpg0409	lpg2331
lpg0525	lpg2333
lpg0531	lpg2345
lpg0540	lpg2349
lpg0551	lpg2597
lpg0596	lpg2623
lpg0601	lpg2633
lpg0607	lpg2654
lpg0664	lpg2657
lpg0689	lpg2699
lpg0700	lpg2764
lpg0880	lpg2878
lpg0890	lpg2882
lpg0957	lpg2902
lpg1202	SBT_lpg0467
lpg1302	SBT_lpg0627
lpg1457	SBT_lpg0791
lpg1565	SBT_lpg1340
lpg1576	SBT_lpg2302
lpg1759	SBT_mompS

Table S5.2 Primers of 51 core genes for tiling multiplex PCR

Name	Sequence : (5' to 3')	Pool
SBT_lpg0467_1_LEFT	TGCACCCAAATTATTATTTGTCACCTT	1
SBT_lpg0467_1_RIGHT	CTGTGCAGGATAGCGTAACCAC	1
SBT_lpg0467_2_LEFT	AGTTTGCAGTTTATAAGACAGCATACTG	2
SBT_lpg0467_2_RIGHT	CTTGATCTTCACTAACTTGTTTATTCGCG	2
SBT_lpg0467_3_LEFT	TGCAAGCAGAATTAGGACAGCC	1
SBT_lpg0467_3_RIGHT	ACGATTTCCACCAAATCCCATCC	1
SBT_lpg0467_4_LEFT	AGCGGCCAACAGCAATTATTGA	2
SBT_lpg0467_4_RIGHT	TCGGGAGTCTCTTTACAAGTAAATTGC	2
SBT_lpg0467_5_LEFT	TCAAGTGTTGAAATGTGCTTTATGGAA	1
SBT_lpg0467_5_RIGHT	TGCATTGGCGATCCATCTGATT	1
SBT_lpg0467_6_LEFT	CCGCTTCTCCAACCAATGATGC	2
SBT_lpg0467_6_RIGHT	CATTCATGCCACCAGACTGACC	2
SBT_lpg0467_7_LEFT	TGGTTTCTCTAGGTGTAGGCGG	1
SBT_lpg0467_7_RIGHT	AACCCGCCATAATAATCATCAGCT	1
SBT_lpg0467_8_LEFT	CGGCCCGGAAATAATGAAAGAAGA	2
SBT_lpg0467_8_RIGHT	TTTGGCAGCACTCAACATACCG	2
SBT_lpg0627_1_LEFT	CGCAGCAATGAACATTTTATGAAGAATAAC	1
SBT_lpg0627_1_RIGHT	TACGTTGTGGCTGTCAGGTTTG	1
SBT_lpg0627_2_LEFT	GCCACATTGACGCAAGATCAAAT	2

SBT_lpg0627_2_RIGHT	CAGCATTCTGGTTGCGCATTAG	2
SBT_lpg0791_1_LEFT	TGAAATTGGTGACTGCAGCTGTT	1
SBT_lpg0791_1_RIGHT	GCAGTACGCTTTGCCATCAAATC	1
SBT_lpg0791_2_LEFT	GCTAAAGGCATGCAAGACGCTA	2
SBT_lpg0791_2_RIGHT	CGGTACCATCAATCAGACGACC	2
SBT_lpg0791_3_LEFT	TTGCCAAGTGGTTTGCAATACAAA	1
SBT_lpg0791_3_RIGHT	TAAAGTTTCATTTGGGCCAATAGGTC	1
SBT_lpg1340_1_LEFT	GCTTACAGCCCAACGCAATTTG	1
SBT_lpg1340_1_RIGHT	GAGTTGGCAGCTTGAACGGAAA	1
SBT_lpg1340_2_LEFT	ACGATGGTATATCCCTATCACAGGT	2
SBT_lpg1340_2_RIGHT	TGGCAATACCACCAATAGAAGAAGC	2
SBT_lpg1340_3_LEFT	ATGGTTCTTTCTCTGGCGCAAG	1
SBT_lpg1340_3_RIGHT	CTCCTATCCCGGCATCGTTGAT	1
SBT_lpg1340_4_LEFT	GCAACCAGTATTAACTCTTCTGCCA	2
SBT_lpg1340_4_RIGHT	TGTTACCACCGTTAAGTGACGC	2
SBT_lpg1340_5_LEFT	CTAACCTCGATGTAGCAACGGC	1
SBT_lpg1340_5_RIGHT	GCAACAGTACCACCAATCGTGAT	1
SBT_lpg1340_6_LEFT	CTGCGGGTACTGACGGTTTAAC	2
SBT_lpg1340_6_RIGHT	TGATACGTTTTGCAGGTTGGCT	2
SBT_lpg1340_7_LEFT	AGACACAGTAGGGATAGATTCACTG G	1
SBT_lpg1340_7_RIGHT	ACATCGCTGTACCTGCTTGTTG	1
SBT_lpg2302_1_LEFT	TGTAGCTATCGTAGGTGCCACC	1
SBT_lpg2302_1_RIGHT	TCACTTCAGGGACAACAAGAGGA	1
SBT_lpg2302_2_LEFT	AAGAGTATGCGCCAAAAGCTGT	2
SBT_lpg2302_2_RIGHT	AAAGATCGCCTACCTGAGCAAC	2
SBT_lpg2302_3_LEFT	GCCAATCTATGACGCTGTTGGT	1
SBT_lpg2302_3_RIGHT	AGATGAACCGCTTCAGAATGCC	1
SBT_lpg2302_4_LEFT	CCCGGGAAGAGATGAAGATGGT	2

SBT_lpg2302_4_RIGHT	TGCGCCCTACAAAAACATCATCA	2
SBT_mompS_1_LEFT	CCAGCCGGACTTCTATAAAGTCAA	1
SBT_mompS_1_RIGHT	GCTGCTGCTATTCTGGTTGGTA	1
SBT_mompS_2_LEFT	GCACCAAGCTTAGCTTCTAACTCA	2
SBT_mompS_2_RIGHT	CAGCGCTAACAAGAAAATGCGT	2
SBT_mompS_3_LEFT	AACACCAAAGCCATTGCCAAAT	1
SBT_mompS_3_RIGHT	ATAATGACAGCGATCACTGGGCT	1
lpg0104_1_LEFT	ATGTTGCGTATCCTTTTGTATTTTCTCTT	1
lpg0104_1_RIGHT	TTGCGCATTTTTGACGGTAGGA	1
lpg0104_2_LEFT	CCGAATCGGTAATGGTGGAGTT	2
lpg0104_2_RIGHT	CGGCCACAACGATAGCGATAATAC	2
lpg0127_1_LEFT	CCGTCCATCTAACTCACCCTCA	1
lpg0127_1_RIGHT	CTGGGCAAAAGTCAGGGCTTTA	1
lpg0127_2_LEFT	TGTGAGCGCTAATTGTCTGGAT	2
lpg0127_2_RIGHT	AGGACGAGGCTACTAATCGCTG	2
lpg0127_3_LEFT	ATGATTCCTGAAGCGGCCATTG	1
lpg0127_3_RIGHT	TTTGTTTCAGCTCATGCCACCA	1
lpg0127_4_LEFT	TGAAGCCAGCCGGGATCTAAAT	2
lpg0127_4_RIGHT	CAACCTACATCAGCAGTGCACC	2
lpg0127_5_LEFT	GGTACATACCACAGGCGGCTAT	1
lpg0127_5_RIGHT	GTTTAGCCATTGATCGCCTGCT	1
lpg0127_6_LEFT	GCAAGGAATTGGCGCATCATTG	2
lpg0127_6_RIGHT	GTTTACGAGCAGAGCCTGGTTT	2
lpg0127_7_LEFT	AAAAGTTGGGCAAGGGAAATGC	1
lpg0127_7_RIGHT	CCATCGCCAGTGATGTAATAGCC	1
lpg0127_8_LEFT	TGGCTATCAAATACCCATGGCCT	2
lpg0127_8_RIGHT	CCGGCTCATTTCCCTGTTTCAG	2
lpg0127_9_LEFT	CATCCTAAGGTGGCAGAAGCTG	1

lpg0127_9_RIGHT	CCTGAGGGTTAGCAAGGGTTGT	1
lpg0287_1_LEFT	TCTATAGCACAAATGAATTTAAAAACGGCT	1
lpg0287_1_RIGHT	CCATTGTGCTGCATCGGCTAAT	1
lpg0287_2_LEFT	TGCAGTATCTATATAATGATGGGGAGCA	2
lpg0287_2_RIGHT	CCTTTGCGCGTGTCTACCTTAAT	2
lpg0329_1_LEFT	CGTAAGATCGGTATGACGCGTG	1
lpg0329_1_RIGHT	GCAACAGATATGACTTGACCCGG	1
lpg0329_2_LEFT	GTGGCTGGGCATTTTGCTAAAG	2
lpg0329_2_RIGHT	AAAAACACGACCAGGGGTTTGG	2
lpg0329_3_LEFT	AAGGTAAAGGTTTCGCAGGAACT	1
lpg0329_3_RIGHT	CTCGTGCTTGCTTTTTGACTGC	1
lpg0331_1_LEFT	ATGAACGCTGAGAGATTGATGATGG	2
lpg0331_1_RIGHT	TCTTGGTCTGCATGAAGAGAGACA	2
lpg0409_1_LEFT	ACCTGTTTTAATTTTCGTTTCACCTCA	1
lpg0409_1_RIGHT	GCATAGGAGAAACAACGTCATAACCA	1
lpg0409_2_LEFT	CCAGAGCAATATCAACGAATCAGC	2
lpg0409_2_RIGHT	ACTTGGCCCTAATACCCATTGTTT	2
lpg0409_3_LEFT	TGGAGACATTACTCGACGCACT	1
lpg0409_3_RIGHT	TTTGCGGGGGCATGGATACTAT	1
lpg0525_1_LEFT	GGCAGACGGCGATATCGAAATC	1
lpg0525_1_RIGHT	CCCGCACTAAACATATCCTGGC	1
lpg0525_2_LEFT	TCGAACCCGAACTGATATCTGGT	2
lpg0525_2_RIGHT	TTTCACCCCATATTCCAGGATCAAAA	2
lpg0525_3_LEFT	AAAATGGTTCATATTCTGGTAGATCGCT	1
lpg0525_3_RIGHT	AGTTGCAGAAGATACCGGATGC	1

lpg0531_1_LEFT	TGGCTGATAGTAGAAATGTAATCTTGTCG	1
lpg0531_1_RIGHT	ACGAATTACTGGAAAACCTGGCAA	1
lpg0531_2_LEFT	CTGACGGTATGAATATCAATGGGACT	2
lpg0531_2_RIGHT	GTCCAGCAGGGCCTACAAATTT	2
lpg0531_3_LEFT	GGGCACAATTGGATGGCTTGTA	1
lpg0531_3_RIGHT	CGTCAACATTTGCGTGCGAATT	1
lpg0540_1_LEFT	AATTACCTGGGTTTTGCGTGGTT	1
lpg0540_1_RIGHT	GCAAGGCCAAATACACAGCAAC	1
lpg0540_2_LEFT	TGCAAATTCCTGTAGGATTGACAGT	2
lpg0540_2_RIGHT	CGCCAGCCAACATTACTCACTAA	2
lpg0540_3_LEFT	GTTTCCCCCAACTATGCTTGGAT	1
lpg0540_3_RIGHT	AGCGGTGGGGCCAAATAAAAAT	1
lpg0540_4_LEFT	TCGTTCTGTTAATCGAATCAGTATTTTGGA	2
lpg0540_4_RIGHT	AGATCACACCACATACTGCGGA	2
lpg0540_5_LEFT	ATAGGCTGGGGAATTAGTGGCC	1
lpg0540_5_RIGHT	TGGAATAAGGCCCCGACAAAAA	1
lpg0540_6_LEFT	ACAAACACTGGAGTTGCGATTTC	2
lpg0540_6_RIGHT	TGCCTTATGGGTTTGCAATAAGTCT	2
lpg0551_1_LEFT	ACCCGCCTATTTTATTACACAGGT	1
lpg0551_1_RIGHT	GAGAGTTCTCGTTTGGCAAGCA	1
lpg0551_2_LEFT	GGCAAGGCAACGATCCTTATGT	2
lpg0551_2_RIGHT	TGCCATCTTTTGAGCGAGTTCC	2
lpg0551_3_LEFT	AAACAGATTTCAAGCCATAAAGGATTTGG	1
lpg0551_3_RIGHT	GAGCTGCATCAATAGGCTCACC	1
lpg0596_1_LEFT	GGCTTGCGAGTCCTTAATTCCT	1

lpg0596_1_RIGHT	GCAAGGTACCAAATACAGGCCC	1
lpg0596_2_LEFT	TGCCCCTGTTCTTTTCTTGATCAC	2
lpg0596_2_RIGHT	TGGAAACAATCGCAAGAAAGCCA	2
lpg0596_3_LEFT	CACCTGGTTTACGATTGGTTCTCA	1
lpg0596_3_RIGHT	TCCAGTTAGAAAAATTCCACCATTCCT	1
lpg0601_1_LEFT	GCAGTGAGCAACTAAATTCTCTGC	1
lpg0601_1_RIGHT	CAACCTCATCGAGACTTTTAGGAGC	1
lpg0601_2_LEFT	CTGAATGGTCAAGCGTGCACTA	2
lpg0601_2_RIGHT	AAGGGACAACTGAGCCCAGATA	2
lpg0601_3_LEFT	AGCGAAATTAGCTGAAAAAGGAGTGA	1
lpg0601_3_RIGHT	TTTTCATCTCGCATCGGAGCAG	1
lpg0601_4_LEFT	AGGATCAATGCTGCTTCAACTGG	2
lpg0601_4_RIGHT	AGGTAATGGCTGAACCTGTTTCA	2
lpg0601_5_LEFT	GGTACAAAACTGGTATCCTGGGGA	1
lpg0601_5_RIGHT	CGGCGCTTATCCCTTTCGAAAT	1
lpg0601_6_LEFT	GTTGCCTTAACCAATAATTTTCAACAAGC	2
lpg0601_6_RIGHT	ATACCGCGTTGCTGGCAATAAA	2
lpg0601_7_LEFT	CAAACGCACGTAATTATACTCAGTGC	1
lpg0601_7_RIGHT	AACCTACTGCCCCTTCTAAGCTT	1
lpg0607_1_LEFT	GGTTGCCCTCAGCTTCAATTCA	1
lpg0607_1_RIGHT	AATTCAGGATTATGCCAGCGCC	1
lpg0607_2_LEFT	CCGGAATACCATATGAAACGGCT	2
lpg0607_2_RIGHT	GACACCATTCAAATCATGCTGTGC	2
lpg0607_3_LEFT	ACCTATCAACAGGCTTTTGAAGAGG	1
lpg0607_3_RIGHT	GCCTCGACATCAGTCAACTCATG	1
lpg0607_4_LEFT	TGGCACAAGTAAAAGAGGGTGTG	2

lpg0607_4_RIGHT	TCTATGCTTGGTTGTTTTAAAGCGAG	2
lpg0664_1_LEFT	TACCCTCTTTACTGTCAGCGGA	1
lpg0664_1_RIGHT	CTAAACCCGCTTCGCATCCAAG	1
lpg0664_2_LEFT	GCAGGTGCGAAACGTATCAGTA	2
lpg0664_2_RIGHT	GTGTGACTCCACCATCAACACA	2
lpg0664_3_LEFT	GACACATTAACCTGGTGTGCCC	1
lpg0664_3_RIGHT	AATGCTTGCTAATTGTTCACGCA	1
lpg0689_1_LEFT	AAAAGCTGGAAGTCGCTCTTGC	1
lpg0689_1_RIGHT	TCTTTTGCCAACTCACTGACCA	1
lpg0689_2_LEFT	CTGGCAGAAGCAGTGGATCAAATT	2
lpg0689_2_RIGHT	CCAATGAGCTTTTTCATGAGCCG	2
lpg0700_1_LEFT	AGAGCAAATGAATCACAGTGCAC	1
lpg0700_1_RIGHT	TAAAACCGGTTCCTGTTCCCAC	1
lpg0700_2_LEFT	TGGCATACGGTCAAAGGATGTT	2
lpg0700_2_RIGHT	TATGGCGCACTTTCCAACCATC	2
lpg0700_3_LEFT	TGACTATTATTCTGAATTTACTGCCA ATGC	1
lpg0700_3_RIGHT	CCGTTTCAAAAAGCATGGATTCTGT	1
lpg0880_1_LEFT	GCTGATGGTTACTATCCTGAGCC	1
lpg0880_1_RIGHT	CCAAATAAAGGGTCGCTTTTGGC	1
lpg0880_2_LEFT	GCGCTTGAATATCCCACGAGAAA	2
lpg0880_2_RIGHT	ACACTGCCAGTTTTTGAGCTCA	2
lpg0890_1_LEFT	AGACTCATTTCGATACACGTGCA	1
lpg0890_1_RIGHT	AAAAAGCCGGAAAGTTCCTCCA	1
lpg0890_2_LEFT	TTGAATCAGGCCAAAAAGGATTTGC	2
lpg0890_2_RIGHT	AGCTAGCTTTAGCATGGGATTGG	2
lpg0890_3_LEFT	TCTTTTATTGATATGTCAGTACCCGA GAA	1
lpg0890_3_RIGHT	GATTGTCACCGACCACAACGAC	1

lpg0890_4_LEFT	CCCTCTTGAGCTCGGATTTGAC	2
lpg0890_4_RIGHT	TTTTTCAATCTTGGGGTGGCCG	2
lpg0890_5_LEFT	CAGTTTCCTGGTTCTGAGAAGCT	1
lpg0890_5_RIGHT	ACTGGAATAGAGGCATGGGTCA	1
lpg0890_6_LEFT	CCGCAATATTCTCTTGCCAAGGA	2
lpg0890_6_RIGHT	TAAAGCATGCTCCAGATCGGCA	2
lpg0957_1_LEFT	GGAACTTCTGCGCATTTTATCCAC	1
lpg0957_1_RIGHT	GGGTGTAGACTGTGAGAGAGCT	1
lpg0957_2_LEFT	CCAATATATGTGGCATGAAGACGC	2
lpg0957_2_RIGHT	TAACCAGGCCAGACTCGTTTCA	2
lpg0957_3_LEFT	CTCTCAAACGCAAAATCAGTCAAGT	1
lpg0957_3_RIGHT	GCTGGATGTTGAAAATGGTGAAGC	1
lpg0957_4_LEFT	AACGAAGAACTGCTATTGGGAAGC	2
lpg0957_4_RIGHT	CCAAACCAAAACAAGAATGTCGACT	2
lpg1202_1_LEFT	TTCCAGCTGCTGAAGTTGTTGA	1
lpg1202_1_RIGHT	GGAGCACCAAATACATCTCCAACA	1
lpg1202_2_LEFT	TTTGTTTTGGGAGTTGCAACCG	2
lpg1202_2_RIGHT	GTTGCATCCAACCATTAGCGATT	2
lpg1202_3_LEFT	GGAACAGATTGTCAAAATTGCAGCA	1
lpg1202_3_RIGHT	TCCCCGCAAAAGAAAATAGGCG	1
lpg1202_4_LEFT	TGTCGCTCAGGCTAAGTTTGTG	2
lpg1202_4_RIGHT	CCGAAAATAGTCAAATTGGCGGG	2
lpg1202_5_LEFT	AGCGGTTACGTTGCCAATGAAA	1
lpg1202_5_RIGHT	CAGCCTTGGCTTGTTCGTTCTT	1
lpg1202_6_LEFT	GGGCATTATTGAATTAATTGAGGAA GGA	2
lpg1202_6_RIGHT	AAAAACCGGTTGCGAAAAGAAGT	2
lpg1202_7_LEFT	ATAAATCAAGCAGCGCACGATCT	1
lpg1202_7_RIGHT	GCCCTGATGTTAAAGAGGAGGC	1

lpg1302_1_LEFT	TGCGTATTGCCTTAGTGGTTGAA	1
lpg1302_1_RIGHT	CCCCATTTGACGCAGATGTCTT	1
lpg1302_2_LEFT	CCAATCAGGTGATTCATTTTGATTGTGA	2
lpg1302_2_RIGHT	TTTTCCCCAAGAAGACATTGCCC	2
lpg1302_3_LEFT	TGGAGCTATACGTCCAGGGTTG	1
lpg1302_3_RIGHT	GGATGTTTTCCCGAGCCAACAG	1
lpg1302_4_LEFT	ACCAATGCGTAATATTCATCAATTGCA	2
lpg1302_4_RIGHT	TCTCCCAAAGAAACAAAGGGCC	2
lpg1457_1_LEFT	TGGTAAGAGTAAAAGATACGACTCCG	1
lpg1457_1_RIGHT	CCCAATTGCTCTTCAACATCGTC	1
lpg1457_2_LEFT	GGCTGATTTACTTGCTGACCTGG	2
lpg1457_2_RIGHT	GTTTCAGACAGATGCCCTGCTG	2
lpg1457_3_LEFT	ATGCTCCTTGCGATGGTTGATG	1
lpg1457_3_RIGHT	AGCTGTTCGACAATGCGATTGA	1
lpg1457_4_LEFT	GGCTTTCCGTCATTTGCATCCA	2
lpg1457_4_RIGHT	TCCACAAAGTATGAACCATGCCT	2
lpg1457_5_LEFT	AGCGCAAAAATGTATCATTAGACGAGA	1
lpg1457_5_RIGHT	TATTTCCAATGGGCTGCCACAC	1
lpg1457_6_LEFT	TTCATACCGCAGTAGAGGGTCC	2
lpg1457_6_RIGHT	AATCCAAAGGAGTCACCCCATG	2
lpg1457_7_LEFT	GGTGTCGCAGAAAACGTTACGA	1
lpg1457_7_RIGHT	CCTTGGCTCTTGCGGTCTTTAAA	1
lpg1457_8_LEFT	GGAGACAGGGTTGAAGTACTGAC	2
lpg1457_8_RIGHT	CTCGTCCCAAGCTGGCATAAAG	2
lpg1457_9_LEFT	TGAGAGATGGTCGTGAGTTATTGGA	1

lpg1457_9_RIGHT	ATTTCATCACCGGGAACAGGTTG	1
lpg1457_10_LEFT	AAAACCAGAGGTCACAGGGAGT	2
lpg1457_10_RIGHT	AATAACTCGGAGCGATCAAAGGC	2
lpg1457_11_LEFT	CCAGTGAAAAGCAAAAGCGACG	1
lpg1457_11_RIGHT	TCCAGCACGTTAGGAATTTGCTC	1
lpg1565_1_LEFT	TGTCATCACTAAAATCCAGGGTTACC	1
lpg1565_1_RIGHT	ATAAGTCCGGTAGGGGGCTCAT	1
lpg1565_2_LEFT	GGAACAGTAGATTTCGGTGTTAAAGC	2
lpg1565_2_RIGHT	TGTGACATTCATGCCAATACGGA	2
lpg1565_3_LEFT	ACGAGTGGGTTATATTGGCGAATT	1
lpg1565_3_RIGHT	GCGCTGTGTTGCTCTGAGAAAT	1
lpg1565_4_LEFT	AGCTGGTTTAGGCTGTTGTTGT	2
lpg1565_4_RIGHT	CGTCCAACCTTATCCCAGTCTCT	2
lpg1576_1_LEFT	TCTTATGCTCGAAAGTGATCGCC	1
lpg1576_1_RIGHT	CACTGGCCCAGAAGTTTGTCTG	1
lpg1576_2_LEFT	TGTTCTTATTTTTGGCCCACCTGG	2
lpg1576_2_RIGHT	TTTAATAGATCGCGCTGCAGGC	2
lpg1576_3_LEFT	TGACGAAATTCATAGACTAAGTCCTGT	1
lpg1576_3_RIGHT	TGTAGGGACACCAAGCAAATGC	1
lpg1576_4_LEFT	AAGGGATAGATTTGGTATAGTGCAACG	2
lpg1576_4_RIGHT	ACCACCGTTGAAATGCTCTATCAC	2
lpg1576_5_LEFT	GGCATCATTACTGTTGATATGGCTCA	1
lpg1576_5_RIGHT	CCTGTTCGATCGCACTAAAACCA	1
lpg1759_1_LEFT	ATGGACGGAATAGAGCGAGCAG	1
lpg1759_1_RIGHT	TCCAACCATTTCAGCTTATTCAAACC	1
lpg1759_2_LEFT	ACCGTTCTCATAGAAGCTCTGGG	2

lpg1759_2_RIGHT	AGCTGCTTTTCAATCACATGTCCA	2
lpg1759_3_LEFT	GCGCAGAAAAGAGAGCAGAAGT	1
lpg1759_3_RIGHT	ATGTCACCTCTCGCAATAAAGTTTGT	1
lpg1759_4_LEFT	GGATTGGGATGAGGAGTTAAGCG	2
lpg1759_4_RIGHT	GCTAATCTTGCCATCTTCAGCCA	2
lpg1811_1_LEFT	AGTCATCAAATTTGGCGGCACT	1
lpg1811_1_RIGHT	AAGGAGATCCCGGTTAGCCATT	1
lpg1811_2_LEFT	TGACATTCAGAGCAGTCACCTGA	2
lpg1811_2_RIGHT	CGTATTCACTCTCACAACGGGC	2
lpg1811_3_LEFT	AAGTGAAATGGTATGATGCCAGAGA	1
lpg1811_3_RIGHT	TCCATATTTCACAAGAGGCGGC	1
lpg1811_4_LEFT	GCTGCCAATCCTCATGGTGAAA	2
lpg1811_4_RIGHT	TCCTGAATGTTCTGGCAGGTGA	2
lpg1811_5_LEFT	CTCAATGGGGGCCAAAGTCTTAC	1
lpg1811_5_RIGHT	AACATCCAGGGATAGGGTGACA	1
lpg1811_6_LEFT	TTTCTGGCTGATGTATTCGCCG	2
lpg1811_6_RIGHT	TCATTCGAGGCCAGAGACATCA	2
lpg1811_7_LEFT	TGGTAGGACATCACATCCGAACT	1
lpg1811_7_RIGHT	GTCAGCAAGCGGTCTCTTTCAA	1
lpg1811_8_LEFT	ATCCTCAAATCTTTTATTACTCCAAA AGCT	2
lpg1811_8_RIGHT	ACGCATTCAAAACCTATCCCTTCTTT	2
lpg1811_9_LEFT	TGCAAGACAATTATCTGCTCTGGA	1
lpg1811_9_RIGHT	TGCCAGGATCTATGCGAACGAT	1
lpg1811_10_LEFT	TGCAAACAGGCTGCTATGTGAC	2
lpg1811_10_RIGHT	GAGGCAAGCATCACAGCAGTTT	2
lpg1811_11_LEFT	ACCAAATCCAACCACATTAAGGTCA	1
lpg1811_11_RIGHT	TGACGCCACTTTCAGCTACAAA	1
lpg1811_12_LEFT	CACTTGATGCGCAATTAATGGCA	2

lpg1811_12_RIGHT	GATTCGCAGATTGGGCCAACTA	2
lpg1811_13_LEFT	CCAAAGTAACGCAATGTAAAGAGAAAGG	1
lpg1811_13_RIGHT	CCTGAGCAGGTGGTCTTAAATTGT	1
lpg1869_1_LEFT	TTAGAAAGATTGTGCAGACGTTTAAACT	1
lpg1869_1_RIGHT	TCCTTAACTCACCTTGCCCAAGA	1
lpg1869_2_LEFT	AAGTCGATTGCGTTCTTTTCTTGTT	2
lpg1869_2_RIGHT	TTTTGATGCTTGCAAAAACTCCTGT	2
lpg1869_3_LEFT	GGAGGAATGATTGCTGCAAAACA	1
lpg1869_3_RIGHT	CTGTTCCAACATGGCTTTTGCTG	1
lpg2008_1_LEFT	CCAATCAATACAAAATTGGCGCCA	1
lpg2008_1_RIGHT	ATTGCCGCTCTTGCAGGATAAG	1
lpg2008_2_LEFT	GGCAGATCCCATTAGATCCTGAAAC	2
lpg2008_2_RIGHT	ATCCCATCCATTTCAACCTTGGC	2
lpg2053_1_LEFT	TGGCAAATTCCCGCTAAAACATTT	1
lpg2053_1_RIGHT	TCTTGCAGTGAACAAGTTGCCA	1
lpg2053_2_LEFT	CCTGCTGGATTATGGTGGCAAA	2
lpg2053_2_RIGHT	GGCCAGGAGAACGAATTAACCTG	2
lpg2053_3_LEFT	ATTACCAATCGTCCTGGTCGGG	1
lpg2053_3_RIGHT	GGTGGCTATGAGCTAAATCAAAAGC	1
lpg2053_4_LEFT	TGGCGACACATCATCATCTTCAAG	2
lpg2053_4_RIGHT	CGGTAAGAATTAACAAAGTATCAGCTCCT	2
lpg2264_1_LEFT	AGGCATGCTTATTGAAGAAACCAC	1
lpg2264_1_RIGHT	GCGTAAAATTTCAAGCTCATGATGCA	1
lpg2317_1_LEFT	GAGTATCACCATATTGCATGCAGCT	1
lpg2317_1_RIGHT	ACCTGGCTGTCCCAAGAGTAAA	1
lpg2317_2_LEFT	CTCCGTCCGCAGTCAATGAATC	2

lpg2317_2_RIGHT	CGCTACTTACTTCAACGGCTCC	2
lpg2317_3_LEFT	ACTTCTCCTTTGCATAACAGTTTGGA	1
lpg2317_3_RIGHT	GGAACCATCCCAAAAGAATCAAACAAA	1
lpg2317_4_LEFT	ACTGGGATGGCGGTATATCAAGTA	2
lpg2317_4_RIGHT	AATCGAACCAGCGAAATAGCGG	2
lpg2317_5_LEFT	CACAAGGAAATTCATGATTTAAGAAACACA	1
lpg2317_5_RIGHT	TTGGCAGACATGTCATACAGTGC	1
lpg2331_1_LEFT	ATTGCAAGTTGACTTTTGTGAAGCT	1
lpg2331_1_RIGHT	CCGCACCCTAAATCCAGAATTCG	1
lpg2331_2_LEFT	CCGATGACTATGAGCGTGTTGC	2
lpg2331_2_RIGHT	GCTACTCCAGTGAATGACCTGGT	2
lpg2331_3_LEFT	TTGGCGTCGTAAATGGCCTTTA	1
lpg2331_3_RIGHT	TCCATGTCCATAACCGGATCCA	1
lpg2331_4_LEFT	GCCTGGTCTGCTGCTAATCAAT	2
lpg2331_4_RIGHT	GGCTTGCCCATAAACCACTTCA	2
lpg2333_1_LEFT	GTCCATTATGATTATGGCATTGGTAGC	1
lpg2333_1_RIGHT	GTTGGCAATCGCTCCTGACAAT	1
lpg2333_2_LEFT	TCGCGCATTAGTCCAGTAAAGC	2
lpg2333_2_RIGHT	ATTGCATACCTACCCCTTGCCA	2
lpg2333_3_LEFT	AAAACTGGATAGCCTTGCTGCC	1
lpg2333_3_RIGHT	AGCCAGTAGAGATGGATAGCGC	1
lpg2333_4_LEFT	ACAATTGAGTCTTGATTTAAGTCAGATTGA	2
lpg2333_4_RIGHT	ACCAAAGCTACAGCAAGACTCAA	2
lpg2333_5_LEFT	TGTTTCTCGGGGTTATTCCTTTTTCAA	1

lpg2333_5_RIGHT	AAAACACAAAATAGAATAATCATCAGGCGA	1
lpg2345_1_LEFT	ATTACGCCATCGCCTATCCAGG	1
lpg2345_1_RIGHT	ATACGTCCTGGCGTTCCAACAA	1
lpg2345_2_LEFT	GTGACGATTGCTGTTCTTTGCG	2
lpg2345_2_RIGHT	GACGAATTCGATAGGGCATGGT	2
lpg2345_3_LEFT	TGTTGCGTATGGGTTTCATCGAA	1
lpg2345_3_RIGHT	CCTCAGTACTGCTTTTAGTGCGT	1
lpg2345_4_LEFT	TTGCTTCCGTTCATCAAAAACCG	2
lpg2345_4_RIGHT	TGAGTAACTCGCTCGACATCCA	2
lpg2345_5_LEFT	CGTGAGCGGATTATTGCCCAAT	1
lpg2345_5_RIGHT	TTGTTGACGAGCCATCTGTATCA	1
lpg2345_6_LEFT	AGGAGTCGCGTTTAATTTCCAGC	2
lpg2345_6_RIGHT	TCCTTTTTCTCCTACCCGAGCT	2
lpg2345_7_LEFT	GCTGCTCTTGCTTTATTGCTGC	1
lpg2345_7_RIGHT	GTTGCCAGGTTTGACTCCATGAA	1
lpg2345_8_LEFT	GCGCGAGGAAGCAAGAAAGATT	2
lpg2345_8_RIGHT	CCAAGCCTTGGTTAAGTCCTTCAT	2
lpg2349_1_LEFT	AAAGAAAGTATACCGGAAGTGGCT	1
lpg2349_1_RIGHT	CCTGGAGTCATCATTCCTTGCA	1
lpg2349_2_LEFT	CCTTATGGCGATGACTAATATTTACTATCG	2
lpg2349_2_RIGHT	ATGAACCGCATGAATCACAGCAG	2
lpg2597_1_LEFT	CAACCCCTCCTCGCCTTAAACA	1
lpg2597_1_RIGHT	CGGCTATTTCTTTGAGCAATGCG	1
lpg2597_2_LEFT	CCGGACTTTATAAAAACTGACTTGGACT	2
lpg2597_2_RIGHT	TCCCAAGGCTAAACCACTCACT	2
lpg2597_3_LEFT	TTGGAAGTCGCAATGCTTCTGT	1

lpg2597_3_RIGHT	CCAAAATGCATAATGATAGACCGCT T	1
lpg2597_4_LEFT	CCGGCATTTAAAATTAGCTGAGCA	2
lpg2597_4_RIGHT	TGTTGCAATAAGTAGTGGCAGCC	2
lpg2597_5_LEFT	CGGTTCTTTAATTACAGCAAGAATGG C	1
lpg2597_5_RIGHT	AGTCCACTCGTGACTTGTTCCA	1
lpg2623_1_LEFT	TGCGTACTCTATTGTCATGTTTCCTG	1
lpg2623_1_RIGHT	CTGCGTAAGAAGCTGACGACAA	1
lpg2623_2_LEFT	GGAGGACTATTGGCCAGTCTCTT	2
lpg2623_2_RIGHT	GGTCGTAAGCAGCTTGTAGACG	2
lpg2623_3_LEFT	ACCGAAATAGTGGAGGAGCTCA	1
lpg2623_3_RIGHT	TGTACGCTGGCTATCGTGGATA	1
lpg2633_1_LEFT	TGTTTGTGACTTTTAGTTGTGATGCA T	1
lpg2633_1_RIGHT	GACCACAGCACATCACAATCCT	1
lpg2654_1_LEFT	GTGGAATAGTAGGATTGCCCAATGT	1
lpg2654_1_RIGHT	GAGTTGTCAAAGCAGCGAACGA	1
lpg2654_2_LEFT	GGATTGGTAAAAGGCGCTTCCA	2
lpg2654_2_RIGHT	GGATATCCGGCATCAAGATGTGC	2
lpg2654_3_LEFT	TGGAAAAATCGCTATTAAAAGTAGG GAAA	1
lpg2654_3_RIGHT	CACAAAGGGCTACTATGCTGGC	1
lpg2654_4_LEFT	GCCAATGTCGATGATAACGGCT	2
lpg2654_4_RIGHT	CTTTGCGTATTGTCCAGGCTCTT	2
lpg2654_5_LEFT	CCTGGATTAAATAAAGTGATCCGTGC	1
lpg2654_5_RIGHT	AGCGAAAATGCATGACGTCTCC	1
lpg2657_1_LEFT	CCCAATTGTGGTAAGACAACCCT	1
lpg2657_1_RIGHT	AGTGCTACAACAACGGGTTTACC	1

lpg2657_2_LEFT	TGATTTGGAATACGATTGCATCATTAATGT	2
lpg2657_2_RIGHT	CAGGTCATCCAAAACTTGTTGTGC	2
lpg2657_3_LEFT	TTCCGGCACTTCAGCAATCATT	1
lpg2657_3_RIGHT	TTTTTGATAGCGAGCGTCTGCC	1
lpg2657_4_LEFT	CGAAGGGGATACTTTGATAGGAGAGA	2
lpg2657_4_RIGHT	AAAAATCCTGGAATGCTCCCCC	2
lpg2657_5_LEFT	ACTTGTTTTGCATCGCTTTTTAGCT	1
lpg2657_5_RIGHT	CGCATCGCTTTATCCACGACAA	1
lpg2657_6_LEFT	TATTCCTGTGATTGCCGCGATG	2
lpg2657_6_RIGHT	AAACAATATTATGCCCGCCAGAGG	2
lpg2657_7_LEFT	CGCGATCGATTATTAACTGTGATGATG	1
lpg2657_7_RIGHT	AGCTTACCCGCACGATAAACAA	1
lpg2657_8_LEFT	CCTTTGATTTTGGAGTTACCCGC	2
lpg2657_8_RIGHT	CTCTTTAGCCAGCATGCCAGTT	2
lpg2657_9_LEFT	ATTGTCAATAATAGGTCAGTGGATTACTCC	1
lpg2657_9_RIGHT	GGCAGAAACAGGGTTCCACAAT	1
lpg2657_10_LEFT	AGCTGCTGCCCACTTTGATTTT	2
lpg2657_10_RIGHT	GCCACGACAAATGACCAAACAA	2
lpg2657_11_LEFT	AGCCTACGCTTATTTGCTTTTTGT	1
lpg2657_11_RIGHT	AGCAGCATTTTGCCTCCCCATA	1
lpg2699_1_LEFT	TGGATTTGGTTAGTGAACAAGAGAGC	1
lpg2699_1_RIGHT	TTCCACCATGTTCAGGCCATTC	1
lpg2699_2_LEFT	TCAGCCATCAAGAGTCCAACTTT	2
lpg2699_2_RIGHT	TTTACCCGCGACACTTAATGCA	2

lpg2764_1_LEFT	TGGAAATTCAGAGCGGTCGTGA	1
lpg2764_1_RIGHT	AACATGCCAACTGCTCGACATG	1
lpg2764_2_LEFT	CCGAATACTTTATCAGAAGATGGGGA	2
lpg2764_2_RIGHT	CAGCCCTCTACCTTGACCCATT	2
lpg2878_1_LEFT	CGCATAAAACAAAGGCTTCAGCAA	1
lpg2878_1_RIGHT	CCAGTTTGGGTCTTTGTCTGGT	1
lpg2878_2_LEFT	ATTTCAGGTCTTGCAGACATCGA	2
lpg2878_2_RIGHT	TTCCCTGGTTAACGAGTGCTCT	2
lpg2878_3_LEFT	AGTTCTGTTTAATCATCAAGCAAAGGC	1
lpg2878_3_RIGHT	AAAGGAGGAATGATTTTTCTGAGGGT	1
lpg2878_4_LEFT	CCAACTGCTTGAAACCTTGGAAG	2
lpg2878_4_RIGHT	CTGGCAAACAAGGTCAGAATTTTCAT	2
lpg2878_5_LEFT	ATGATCATTGCATTCGCCTGGT	1
lpg2878_5_RIGHT	AAAAATACAGCATGAGTATTGCCAAAAGTA	1
lpg2882_1_LEFT	ATGTTAGTGACCAGTGCACTGC	1
lpg2882_1_RIGHT	AGTCGCCAATGCCTGATTTTCA	1
lpg2882_2_LEFT	ACCGCCGAAATTAAGTTAAGTCATGA	2
lpg2882_2_RIGHT	AAATCCGTCGGCGAATAAGTGG	2
lpg2882_3_LEFT	TGTTTTTGCCGGACAGGTATGT	1
lpg2882_3_RIGHT	TAAACCCGCCTCAAACCATTCG	1
lpg2882_4_LEFT	GGAGCCAGCCCAGTAGAAAAAG	2
lpg2882_4_RIGHT	TCAAAAGAAACGCCACGCTCAT	2
lpg2882_5_LEFT	CTCGCGATGCTCCCTATTTTGG	1
lpg2882_5_RIGHT	CGCGCGATTTGGACATTTTCTG	1

lpg2882_6_LEFT	TGCTTTGTTCTGGCCAGCTATG	2
lpg2882_6_RIGHT	TTCACTACCTTCCCGACCAGATC	2
lpg2882_7_LEFT	TCGCTGCCAAATTGAATGGTCG	1
lpg2882_7_RIGHT	TGACTTTGTCAGCACAATCCATGA	1
lpg2882_8_LEFT	TATACAACGACTTGCTGAGCGC	2
lpg2882_8_RIGHT	AGAATCCCAGTGTAAAGGCTCAGA	2
lpg2882_9_LEFT	GCCATATGCACAATGGGAATCAAT	1
lpg2882_9_RIGHT	GCAATCCGTAAATCGACTTTGGC	1
lpg2882_10_LEFT	AGGAATCTCTGATGACTACCCCAG	2
lpg2882_10_RIGHT	CGCCAATACCATGCCTTCAGAA	2
lpg2902_1_LEFT	CGGCGATACAGTACAAGAAACTCG	1
lpg2902_1_RIGHT	GTTCAACAGCCATGGGATCCTG	1
lpg2902_2_LEFT	AAATCATTGAACAAATTCCATCTTTTCCTG	2
lpg2902_2_RIGHT	ATTAAATAGTGGTTTTCATCTGGCGG	2

Table S5.3 Fifteen *L. pneumophila* references sequences (pair-end Fastq) from ENA

Sample name	Sample Accession in ENA
Lp-001	SAMEA2743239
Lp-012	SAMEA2743240
Lp-032	SAMEA2743241
Lp-119	SAMEA2781629
Lp-120	SAMEA2743243
Lp-121	SAMEA2743244
Lp-122	SAMEA2743245
Lp-2002694p7	SAMEA2743246
Lp-2002694p8	SAMEA2743247
Lp-282-1	SAMEA2743248
Lp-283	SAMEA2743249
Lp-284	SAMEA2743250
Lp-285	SAMEA2743251
Lp-286-1	SAMEA2743252
Lp-56207	SAMEA2747162

Table S5.4 Summary on the hospital environmental samples collected. The same set of samples were collected from summer (07/2020, 08/2020, 09/2020, 10/2020) and winter (11/2020, 12/2020, 01/2021) seasons of the hospital investigated

	Sewage inlet	Sewage outlet	Tap	P-trap	Surface (door handle)	Air dust
WW Plant	✓	✓				
Ophthalmology			✓	✓	✓	✓
Otolaryngology			✓	✓	✓	✓
Emergency			✓	✓	✓	✓
Blood Test room 1 & room 2			✓	✓	✓	
Surgery			✓	✓	✓	✓
Pneumology			✓	✓	✓	✓
Gynecology			✓	✓	✓	

Table S5.5 PCR touchdown program for PrimalPCR of the 51 cgMLST scheme

Stage	Cycle number	Denature	Anneal/Extend
1	1	98°C, 30s	
2	10	98°C, 15s	65°C, 5min
3	25	98°C, 15s	65°C, 5min, delta t = -0.1°C (Touchdown)

Table S5.6 The components of PrimalPCR used for 51cgMLST scheme

Component	Volume, µl	Final concentration
Q5 reaction buffer (5x)	5	1x
dNTPs, 10 mM	0.5	200 µM
Q5 DNA polomerase, 0.5 U	0.25	0.02 Unit/µl
Primer pool1 or 2 (10 µM)	9.8 or 7.58	0.015 per primer
DNA template	5	Variable
PCR-grade water	4.45 or 6.67	-

Table S5.7 Different PCR programs for the optimisation of multiplex PCR

PCR Program	DNA polymerase	Mode
TouchDown	hot start	A
Non-TouchDown	hot start	B
TouchDown	normal high fidelity	C
Non-TouchDown	normal high fidelity	D

Table S5.8 Different DNA concentrations of *L. pneumophila* strain for the optimisation of multiplex PCR

Accession	1	2	3	4	5	6	7	8	9	10
Genome copy	1400	800	300	140	80	30	14	8	3	0.3

Table S5.9 Multiplex PCR results of 217 hospital environmental samples. Samples are marked with different categrories of: (1) with only specific targeted 300bp band **; (2) specific targeted 300bp band and few non-specific bands *; (3) dispersed bands with obvious specific targeted 300bp band: o ; (4) dispersed bands without target band: w ; (5) no band x

07/2020	Sewage inlet	Sewage outlet	Tap	P-trap	Surface (door handle)	Air dust
WW Plant	Pool1: o	Pooll: x				
	Pool2: **	Pool2:x				
Ophthalmology			Pool1: x	Pool1: x	Pool1: x	Pool1: x
			Pool2: x	Pool2: x	Pool2: **	Pool2: **
Otolaryngology			Pool1: w	Pool1: o	Pool1: **	Pool1: w
			Pool2: **	Pool2: o	Pool2: **	Pool2: **
Emergency			Pool1: x	Pool1: o	Pool1: x	Pool1: **
			Pool2: **	Pool2: w	Pool2: **	Pool2: **
Blood Test room 1			Pool1: x	Pool1: **	Pool1: x	
			Pool2: x	Pool2: o	Pool2: x	
Blood Test room 2			Pool1: x	Pool1: **	-	
			Pool2: x	Pool2: **	-	
Surgery			Pool1: **	Pool1: o	Pool1: x	Pool1: x
			Pool2: **	Pool2: o	Pool2: **	Pool2: **
Pneumology			Pool1: **	Pool1: o	Pool1: x	Pool1: **

			Pool2: **	Pool2: o	Pool2: **	Pool2: **
Gynecology			Pool1: **	Pool1: x	Pool1: x	
			Pool2: x	Pool2: x	Pool2: x	

08/2020	Sewage inlet	Sewage outlet	Tap	P-trap	Surface (door handle)	Air dust
WW Plant	Pool1: o	Pool1: o				
	Pool2: o	Pool2: o				
Ophthalmology			Pool1: **	Pool1: w	Pool1: x	Pool1: *
			Pool2: o	Pool2: w	Pool2:**	Pool2: *
Otolaryngology			Pool1: x	Pool1: w	Pool1: x	Pool1: w
			Pool2: o	Pool2: w	Pool2: **	Pool2: o
Emergency			Pool1: **	Pool1: o	Pool1: **	Pool1: **
			Pool2: *	Pool2: o	Pool2: **	Pool2: **
Blood Test room 1			Pool1: *	Pool1: *	Pool1: x	
			Pool2: o	Pool2: w	Pool2: **	
Blood Test room 2			Pool1: o	Pool1: *	Pool1: x	
			Pool2: *	Pool2: o	Pool2: **	
Surgery			Pool1: o	Pool1: w	-	Pool1: w

			Pool2: w	Pool2: **	-	Pool2: x
Pneumology			Pool1: o	Pool1: w	Pool1: x	Pool1: *
			Pool2: w	Pool2: **	Pool2: **	Pool2: o
Gynecology			Pool1:*	Pool1: *	Pool1: x	
			Pool2: *	Pool2: *	Pool2: **	

09/2020	Sewage inlet	Sewage outlet	Tap	P-trap	Surface (door handle)	Air dust
WW Plant	Pool1:**	Pool1:x				
	Pool2:**	Pool2: **				
Ophthalmology			Pool1: x	Pool1: o	Pool1: x	Pool1: **
			Pool2: **	Pool2: o	Pool2: **	Pool2: **
Otolaryngology			Pool1: x	Pool1: o	-	Pool1: w
			Pool2: **	Pool2: o	-	Pool2: **
Emergency			Pool1: x	Pool1: o	Pool1: x	Pool1: w
			Pool2: **	Pool2: o	Pool2: x	Pool2: **
Blood Test room 1			Pool1: **	Pool1: o	Pool1: x	
			Pool2: **	Pool2: o	Pool2: x	
Blood Test room 2			Pool1: x	Pool1: o	Pool1: x	
			Pool2: **	Pool2: o	Pool2: **	

Surgery			Pool1: **	Pool1: w	Pool1: x	Pool1: w
			Pool2:	Pool2: w	Pool2: **	Pool2: **
Pneumology			Pool1:**	Pool1: w	Pool1: x	Pool1: o
			Pool2: w	Pool2: w	Pool2: x	Pool2: **
Gynecology			Pool1: x	Pool1: o	Pool1: x	
			Pool2: w	Pool2: o	Pool2: x	

10/2020	Sewage inlet	Sewage outlet	Tap	P-trap	Surface (door handle)	Air dust
WW Plant	Pool1: *	Pool1: o				
	Pool2: *	Pool2: o				
Ophthalmology			Pool1: **	Pool1: w	Pool1: x	Pool1: w
			Pool2: x	Pool2: w	Pool2: x	Pool2: w
Otolaryngology			Pool1: **	Pool1: o	Pool1: x	Pool1: o
			Pool2: x	Pool2: w	Pool2: x	Pool2: o
Emergency			Pool1: x	Pool1: o	Pool1: *	Pool1: x
			Pool2: x	Pool2: w	Pool2: *	Pool2: w
Blood Test room 1			Pool1: x	Pool1: w	Pool1: x	
			Pool2: x	Pool2: w	Pool2: x	
			Pool1: x	Pool1: w	Pool1: x	

Blood Test room 2			Pool2: x	Pool2: w	Pool2: **	
Surgery			Pool1: x	Pool1: w	Pool1: x	Pool1: w
			Pool2: x	Pool2: o	Pool2: x	Pool2: w
Pneumology			Pool1	-	Pool1: x	Pool1: w
			Pool2	-	Pool2: x	Pool2: w
Gynecology			Pool1: x	Pool1: x	Pool1: x	
			Pool2: **	Pool2: w	Pool2: x	

11/2020	Sewage inlet	Sewage outlet	Tap	P-trap	Surface (door handle)	Air dust
WW Plant	Pool1: **	Pool1: **				
	Pool2: w	Pool2: w				
Ophthalmology			Pool1: x	Pool1: *	Pool1: x	Pool1: o
			Pool2: **	Pool2: o	Pool2: x	Pool2: **
Otolaryngology			Pool1: x	Pool1: *	Pool1: x	Pool1: x
			Pool2: x	Pool2: w	Pool2: x	Pool2: x
Emergency			Pool1: x	Pool1: o	Pool1: o	Pool1: x
			Pool2: x	Pool2: w	Pool2:	Pool2: x
			Pool1: x	Pool1: w	Pool1: x	

Blood Test room 1			Pool2: x	Pool2: o	Pool2: w	
Blood Test room 2			Pool1: x	Pool1: w	Pool1: x	
			Pool2: x	Pool2: w	Pool2: **	
Surgery			Pool1: x	Pool1: w	Pool1: **	Pool1: **
			Pool2: x	Pool2: o	Pool2: **	Pool2: x
Pneumology			Pool1: x	Pool1: w	Pool1: o	Pool1: x
			Pool2: x	Pool2: w	Pool2: w	Pool2: x
Gynecology			Pool1: x	Pool1: w	Pool1: w	
			Pool2: x	Pool2: w	pool2: x	

12/2020	Sewage inlet	Sewage outlet	Tap	P-trap	Surface (door handle)	Air dust
WW Plant	Pool1: x	Pool1: x				
	Pool2: x	Pool2: w				
Ophthalmology			Pool1: x	Pool1: w	Pool1: x	Pool1: x
			Pool2: x	Pool2: w	Pool2: x	Pool2: x
Otolaryngology			Pool1: x	Pool1: **	Pool1: w	Pool1: x
			Pool2: x	Pool2: **	Pool2: **	Pool2: x
Emergency			Pool1: x	Pool1: w	Pool1: o	Pool1: **

			Pool2: x	Pool2: o	Pool2: o	Pool2: **
Blood Test room 1			Pool1: x	Pool1: w	Pool1: x	
			Pool2: x	Pool2: w	Pool2: **	
Blood Test room 2			Pool1: x	Pool1: **	Pool1: w	
			Pool2: x	Pool2: **	Pool2: **	
Surgery			Pool1: x	Pool1: o	Pool1: x	Pool1: **
			Pool2: x	Pool2: w	Pool2: x	Pool2: **
Pneumology			Pool1: x	Pool1: o	Pool1: x	Pool1: x
			Pool2: x	Pool2: o	Pool2: x	Pool2: x
Gynecology			Pool1: x	Pool1: w	Pool1: x	
			Pool2: x	Pool2: w	Pool2: x	

01/2021:	Sewage inlet	Sewage outlet	Tap	P-trap	Surface (door handle)	Air dust
WW Plant	Pool1: w	Pool1: w				
	Pool2: x	Pool2: o				
Ophthalmology			Pool1: x	Pool1: w	Pool1: w	Pool1: x
			Pool2: **	Pool2: **	Pool2: **	Pool2: x
Otolaryngology			Pool1: x	Pool1:	Pool1: x	Pool1: **
			Pool2: x	Pool2: **	Pool2: **	Pool2: **

Emergency			Pool1: **	Pool1: x	Pool1: x	Pool1: x
			Pool2: **	Pool2: x	Pool2: **	Pool2; x
Blood Test room 1			Pool1: x	Pool1: o	Pool1: x	
			Pool2: x	Pool2: w	Pool2: **	
Blood Test room 2			Pool1: x	Pool1: o	Pool1: **	
			Pool2: x	Pool2: w	Pool2: **	
Surgery			Pool1: x	Pool1: o	Pool1: x	Pool1: **
			Pool2: x	Pool2: o	Pool2: **	Pool2: **
Pneumology			Pool1: **	Pool1: o	Pool1: x	Pool1: x
			Pool2: **	Pool2: w	Pool2: x	Pool2: x
Gynecology			Pool1: x	Pool1: o	Pool1: o	
			Pool2: **	Pool2: w	Pool2: **	

Table S5.10 Seven MinION sequencing run of 134 amplicon derived from 51 cgMLST scheme

MinION sequencing run	Barcode number	Avaliable pores	Input amplicon library, ng	Duration time, h	Yielded number and percent-age of pass Fastq	Reads yielded, million
1st	5	754	15	24	5.7G	2.99
2nd	62	1036	15 (1st); two reloading of 15 for each	40	987Mbp, 92.5%	3.91
3rd	13	953	20	11	6.4 Gbp, 82.1%	6.22
4th	12	512	20	47	5.5 Gbp, 78.6%	5.44
5th	16	1072	20	14	7.1Gbp, 93.3%	6.76
6th	26	675	20	42	6.3 Gbp, 80.8%	6.62

Resequencing	19	1227	20	47.7	16 Gbp, 93.6%	14.88

Table S5.11 qPCR results of 25 hospital samples

Medium	Date	Departement	Ct value
Tap water	08/2020	Surgery	-
Tap water	08/2020	Emergency	-
Tap water	11/2020	Pneumology	-
Tap water	09/2020	Blood-test	31.531
Tap water	09/2020	Surgery	31.369
Tap water	01/2021	Gynecology	-
Tap water	09/2020	Ophthalmology	30.497
Surface	07/2020	Blood-test	-
Surface	09/2020	Surgery	33.521
Surface	08/2020	Ophthalmology	31.238
Surface	08/2020	Otolaryngology	-
Surface	11/2020	Emergency	-
Surface	12/2020	Otolaryngology	-
P-trap water	07/2020	Emergency	37.583
P-trap water	09/2020	Gynecology	-
P-trap water	11/2020	Surgery	-
P-trap water	01/2021	Otolaryngology	-
P-trap water	11/2020	Blood-test	-
Outlet	01/2021	Sewage	-

Inlet	08/2020	Sewage	-
Inlet	09/2020	Sewage	-
Inlet	12/2021	Sewage	-
Air	10/2020	Otolaryngology	-
Air	08/2020	Surgery	-
Air	12/2020	Emergency	-

Appendix 2 Supplementary figures

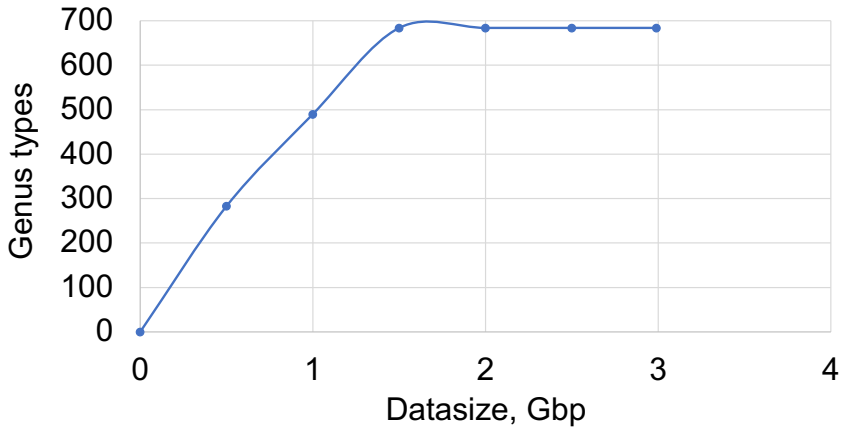


Figure S3.1 Rarefaction analysis of genus types based on increasing datasize.

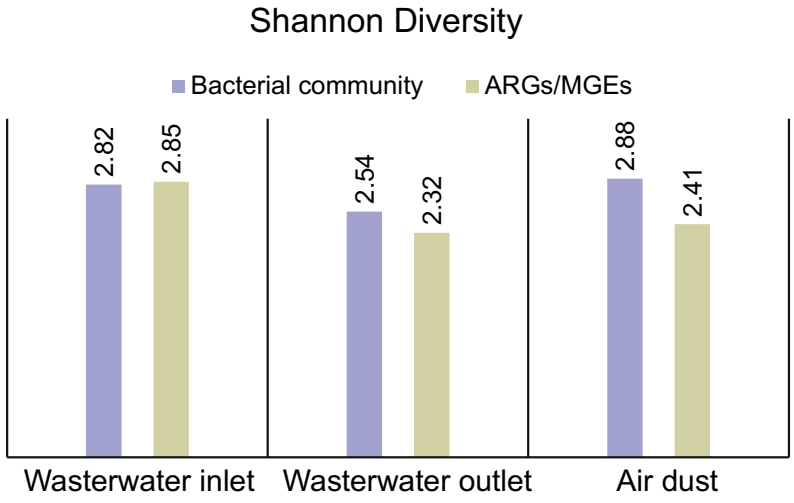


Figure S3.2 Average Shannon diverity of wastewater inlet, wastewater outlet and air-dust samples.

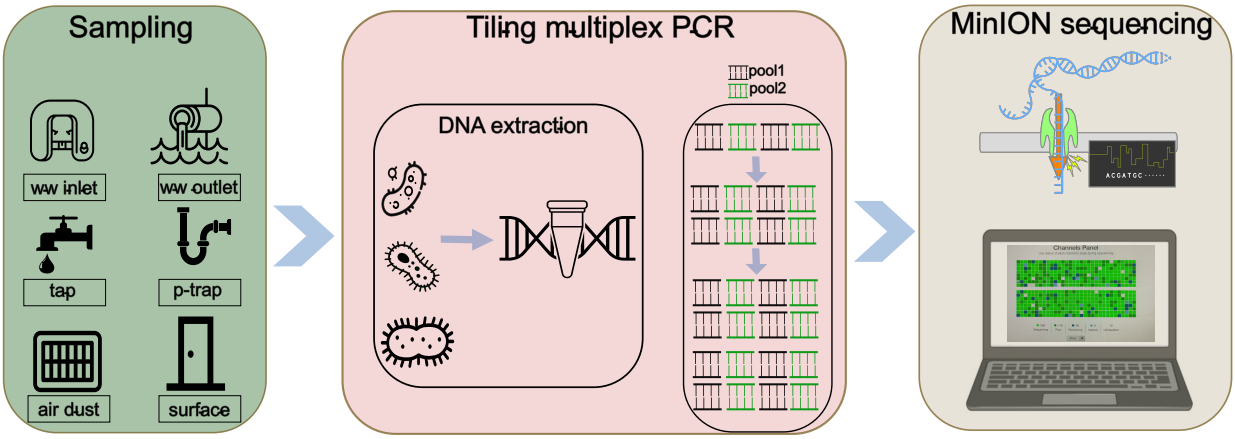


Figure S5.1 The lab workflow of the 51-cgMLST integrating with multiplex-PCR based MinION sequencing.

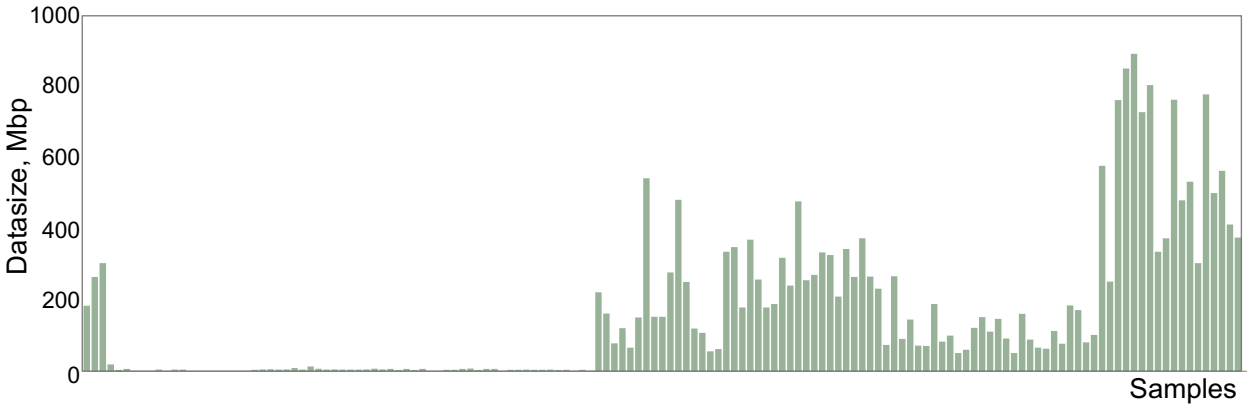
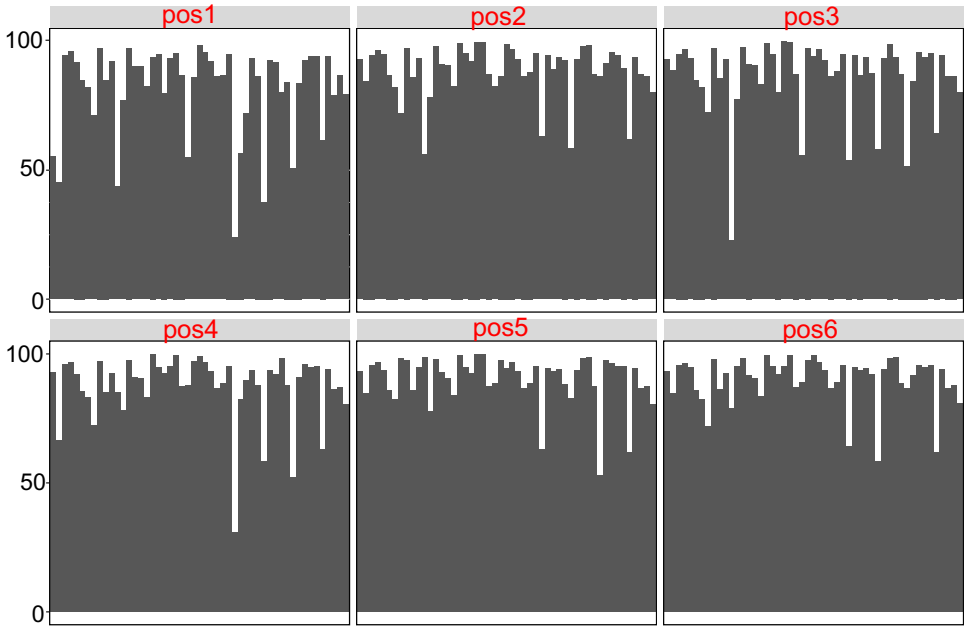
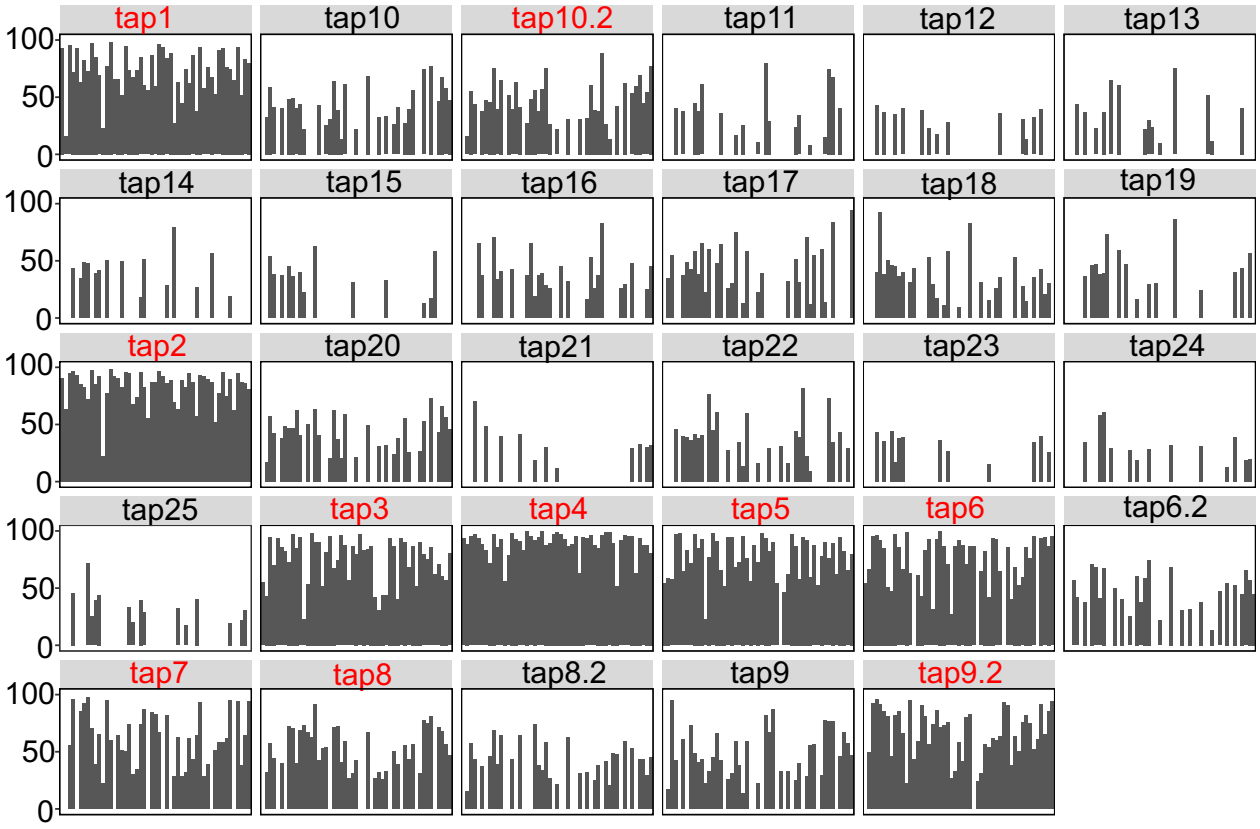
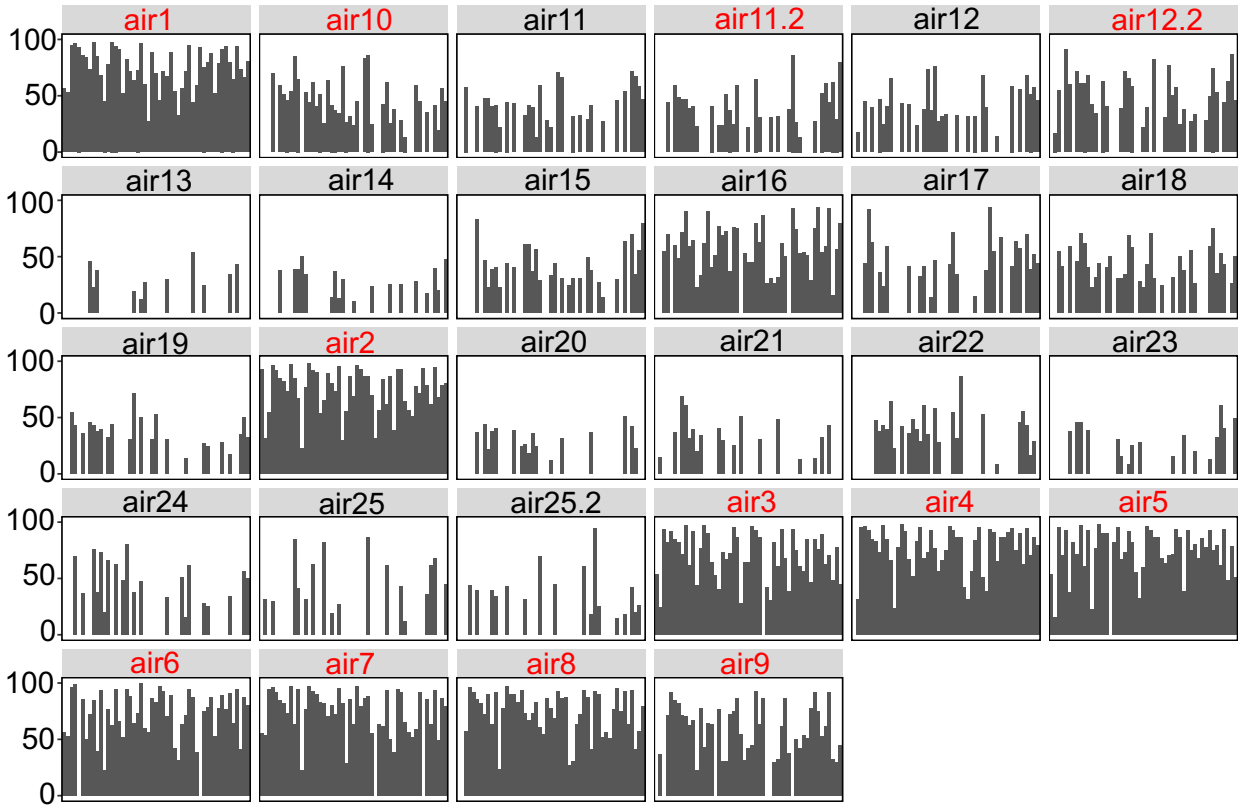
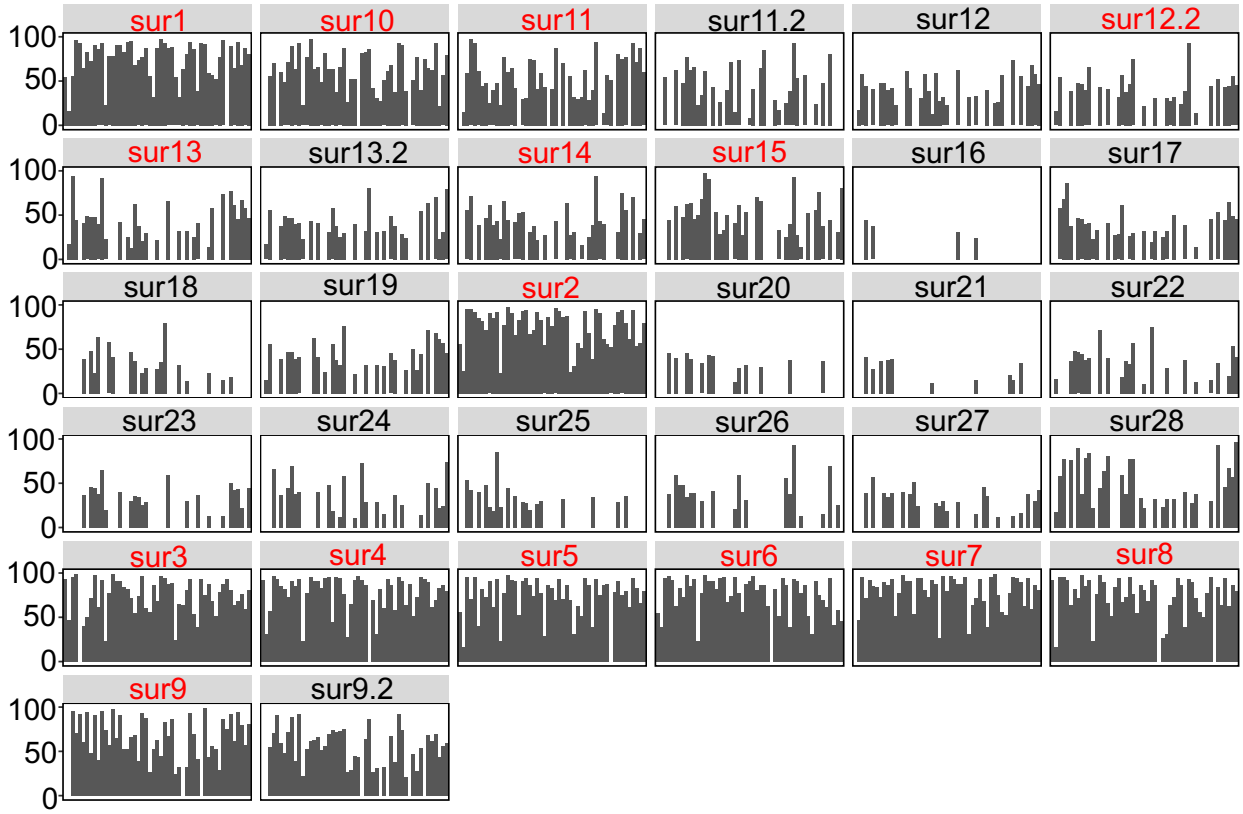


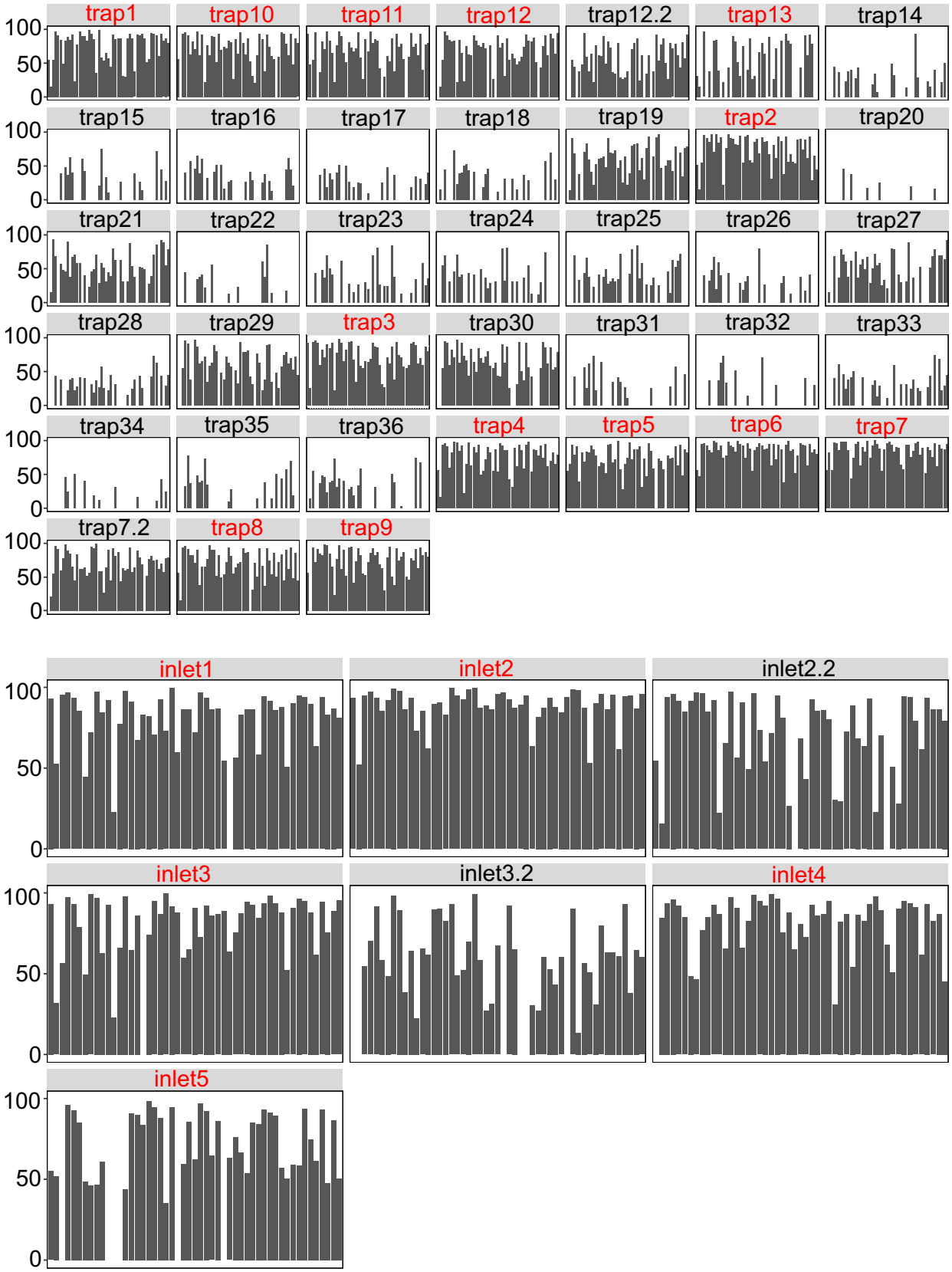
Figure S5.2 Distribution of data size of seven-run MinION sequencing of 145 amplicon sequencing samples.


pos1
pos2
pos3
pos4
pos5
pos6
100
50
0


tap1
tap10
tap10.2
tap11
tap12
tap13
tap14
tap15
tap16
tap17
tap18
tap19
tap2
tap20
tap21
tap22
tap23
tap24
tap25
tap3
tap4
tap5
tap6
tap6.2
tap7
tap8
tap8.2
tap9
tap9.2
100
50
0


air1
air10
air11
air11.2
air12
air12.2
air13
air14
air15
air16
air17
air18
air19
air2
air20
air21
air22
air23
air24
air25
air25.2
air3
air4
air5
air6
air7
air8
air9
100
50
0


sur1
sur10
sur11
sur11.2
sur12
sur12.2
sur13
sur13.2
sur14
sur15
sur16
sur17
sur18
sur19
sur2
sur20
sur21
sur22
sur23
sur24
sur25
sur26
sur27
sur28
sur3
sur4
sur5
sur6
sur7
sur8
sur9
sur9.2
100
50
0


trap1
trap10
trap11
trap12
trap12.2
trap13
trap14
trap15
trap16
trap17
trap18
trap19
trap2
trap20
trap21
trap22
trap23
trap24
trap25
trap26
trap27
trap28
trap29
trap3
trap30
trap31
trap32
trap33
trap34
trap35
trap36
trap4
trap5
trap6
trap7
trap7.2
trap8
trap9
inlet1
inlet2
inlet2.2
inlet3
inlet3.2
inlet4
inlet5
100
50
0

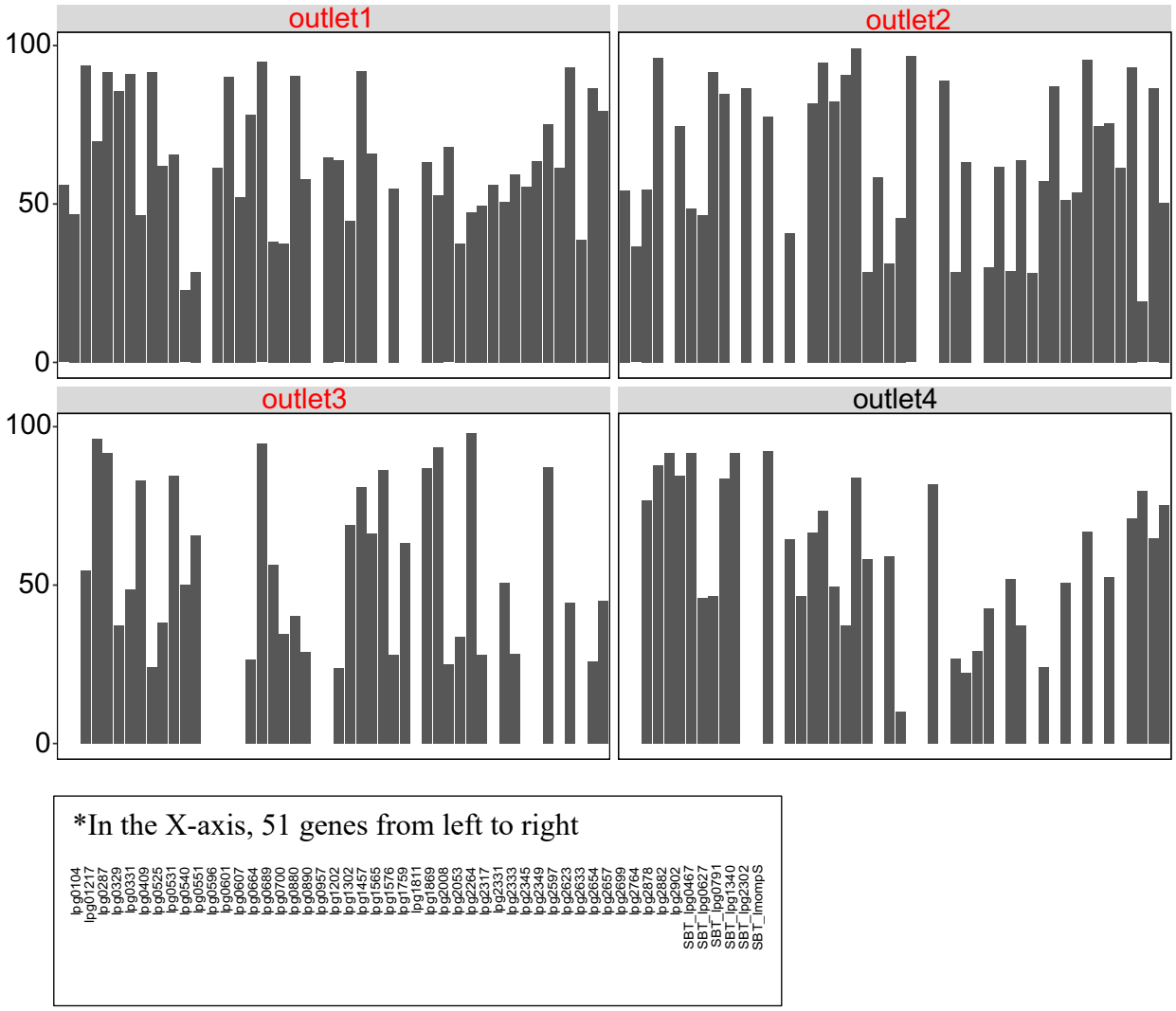
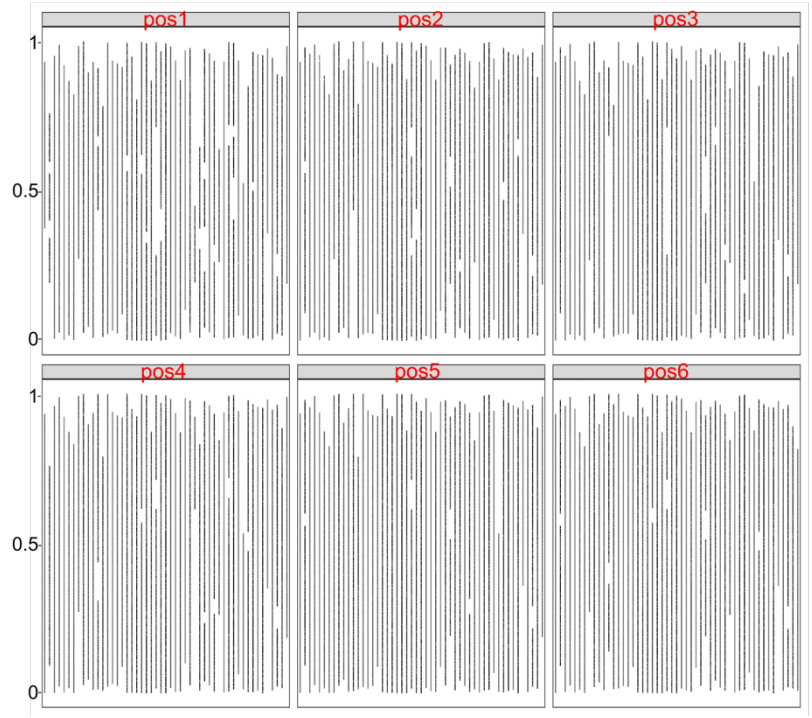
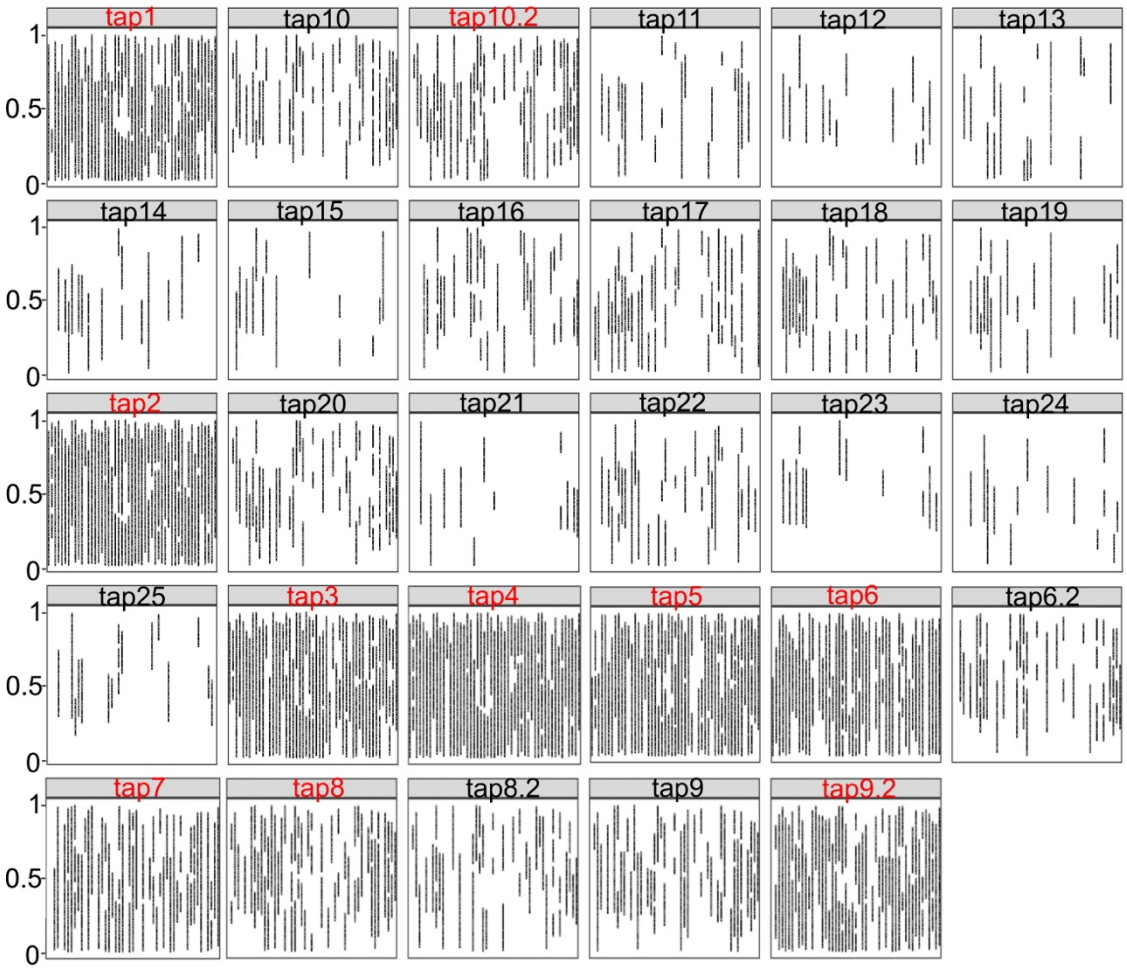
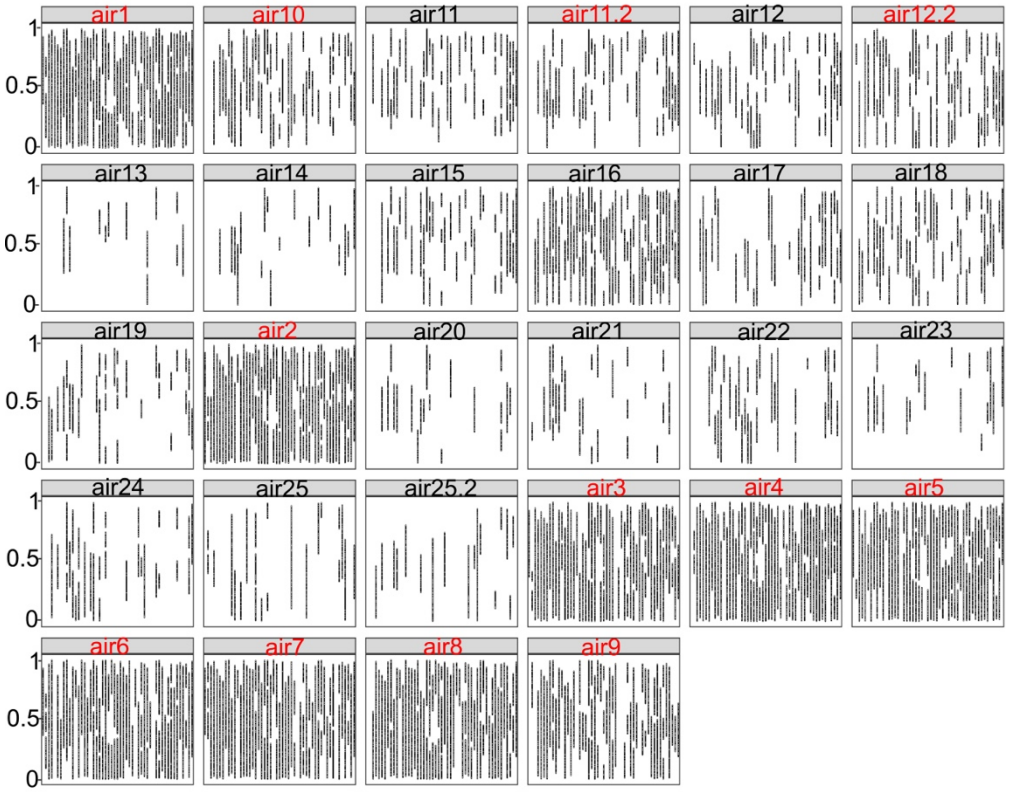
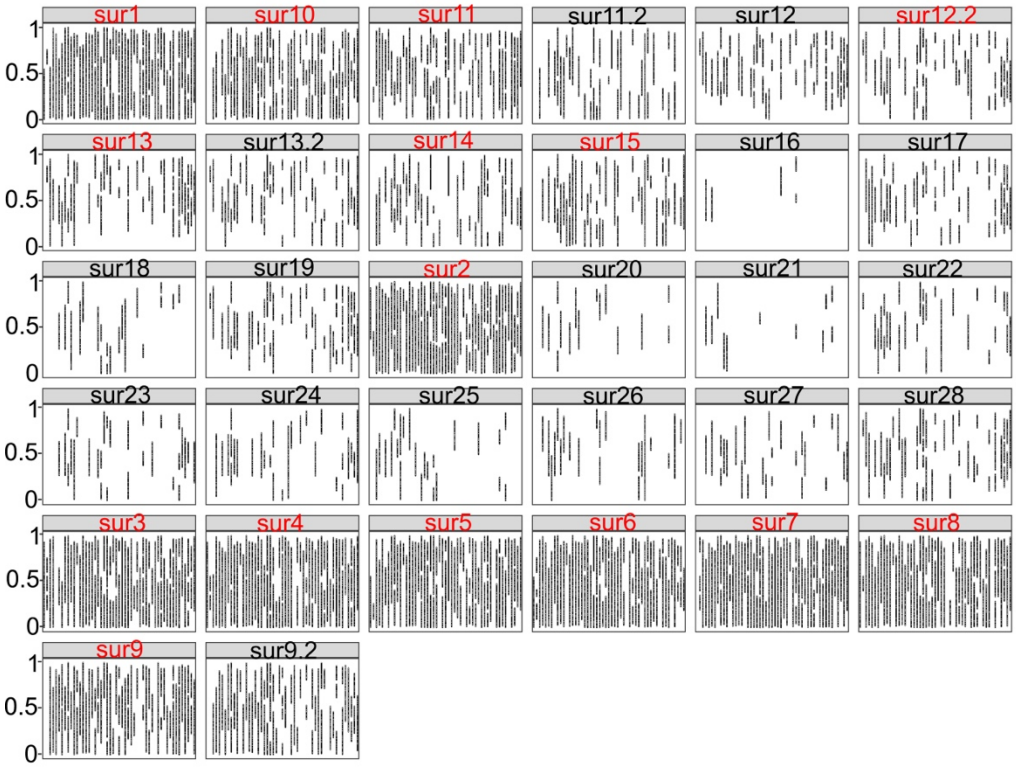


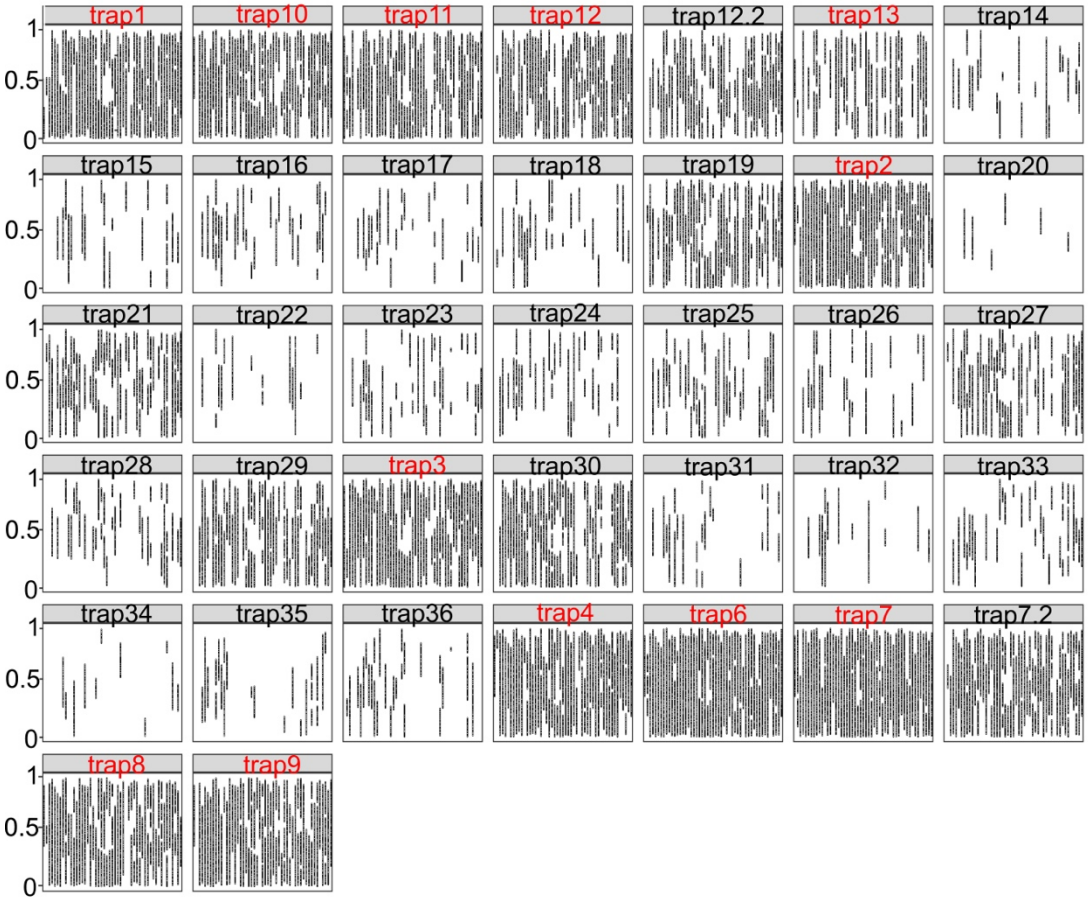
Figure S5.3 Distribution of base coverage against the reference of 51 core genes across all 145 amplicon sequencing samples. The red marked samples were 64 samples used for building the phylogenetic tree.

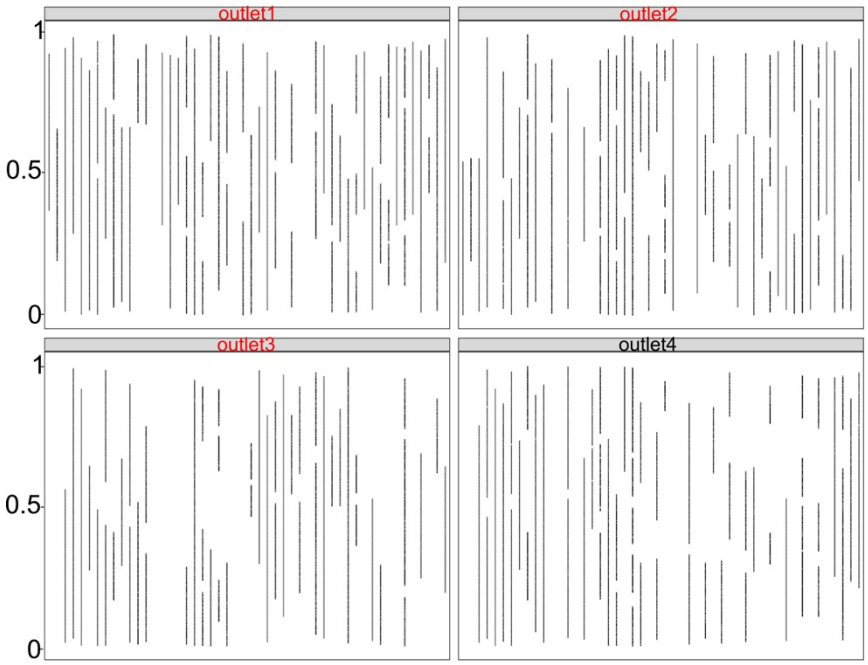

pos1
pos2
pos3
pos4
pos5
pos6
1
0.5
0


tap1
tap10
tap10.2
tap11
tap12
tap13
tap14
tap15
tap16
tap17
tap18
tap19
tap2
tap20
tap21
tap22
tap23
tap24
tap25
tap3
tap4
tap5
tap6
tap6.2
tap7
tap8
tap8.2
tap9
tap9.2
1
0.5
0


air1
air10
air11
air11.2
air12
air12.2
air13
air14
air15
air16
air17
air18
air19
air2
air20
air21
air22
air23
air24
air25
air25.2
air3
air4
air5
air6
air7
air8
air9
1
0.5
0


sur1
sur10
sur11
sur11.2
sur12
sur12.2
sur13
sur13.2
sur14
sur15
sur16
sur17
sur18
sur19
sur2
sur20
sur21
sur22
sur23
sur24
sur25
sur26
sur27
sur28
sur3
sur4
sur5
sur6
sur7
sur8
sur9
sur9.2
1
0.5
0


trap1
trap10
trap11
trap12
trap12.2
trap13
trap14
trap15
trap16
trap17
trap18
trap19
trap2
trap20
trap21
trap22
trap23
trap24
trap25
trap26
trap27
trap28
trap29
trap3
trap30
trap31
trap32
trap33
trap34
trap35
trap36
trap4
trap6
trap7
trap7.2
trap8
trap9
1
0.5
0


outlet1
outlet2
outlet3
outlet4
1
0.5
0

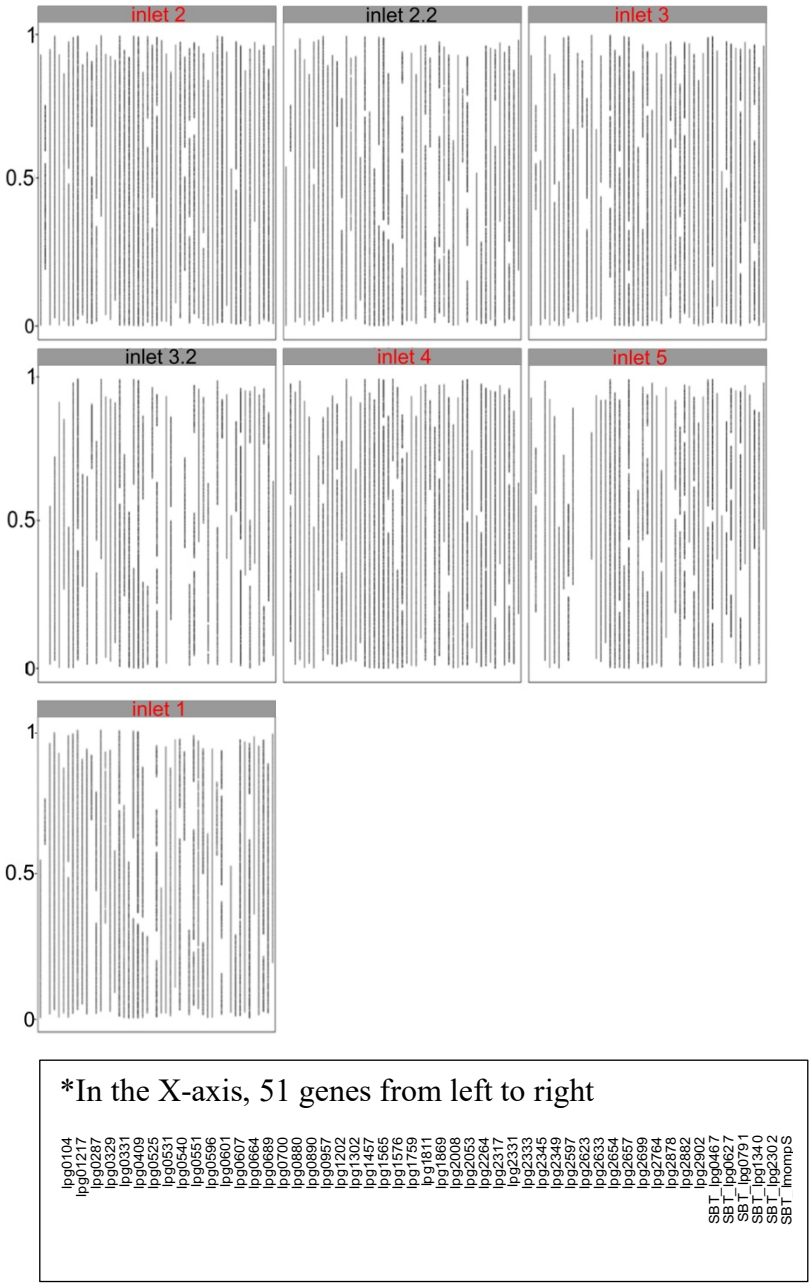


Figure S5.4 Distribution of relative base position (1-base coordinates) where the depth over 20 within each gene (amplicon) across all 145 amplicon sequencing samples.The red marked samples were 64 samples used for building the phylogenetic tree.

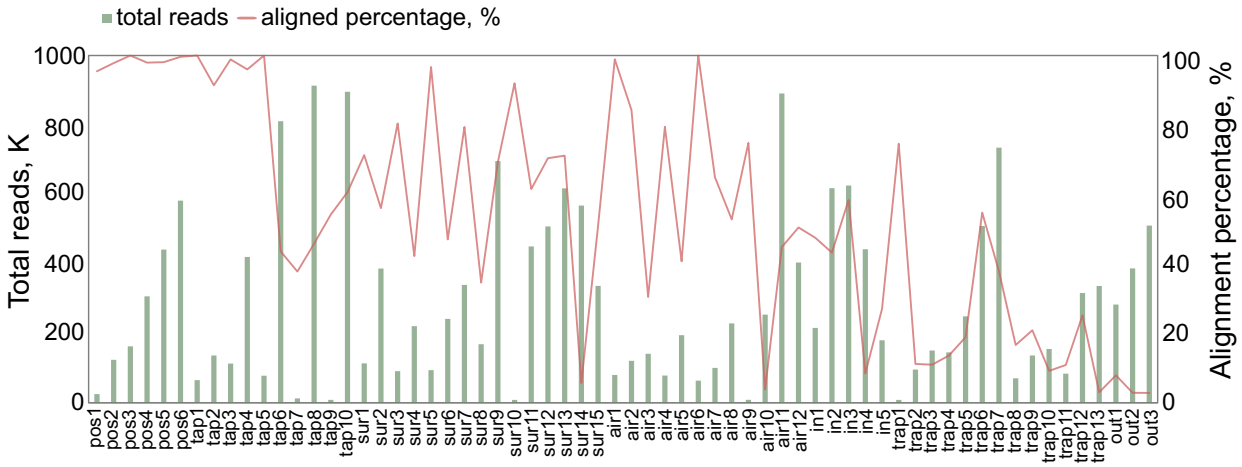


Figure S5.5 Distribution of raw Fastq reads and alignment percentage against the reference of 51 core genes across 64 environmental samples.

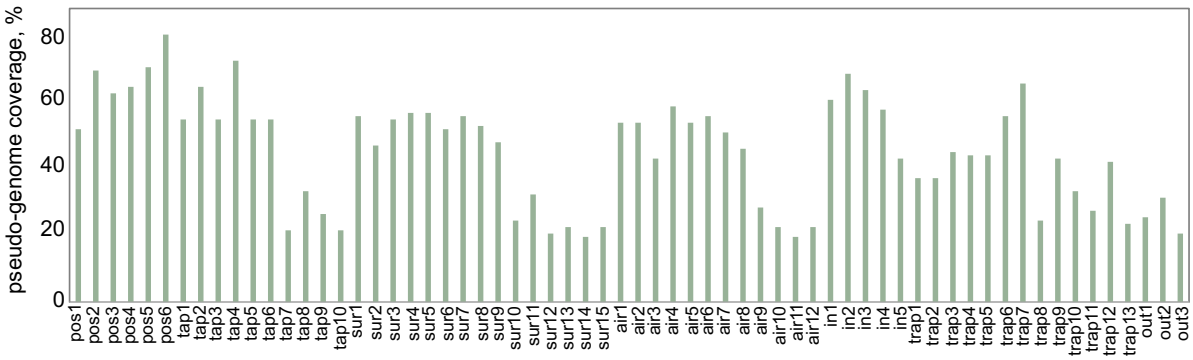


Figure S5.6 Distribution of pseudo-genome coverage across 64 environmental samples.

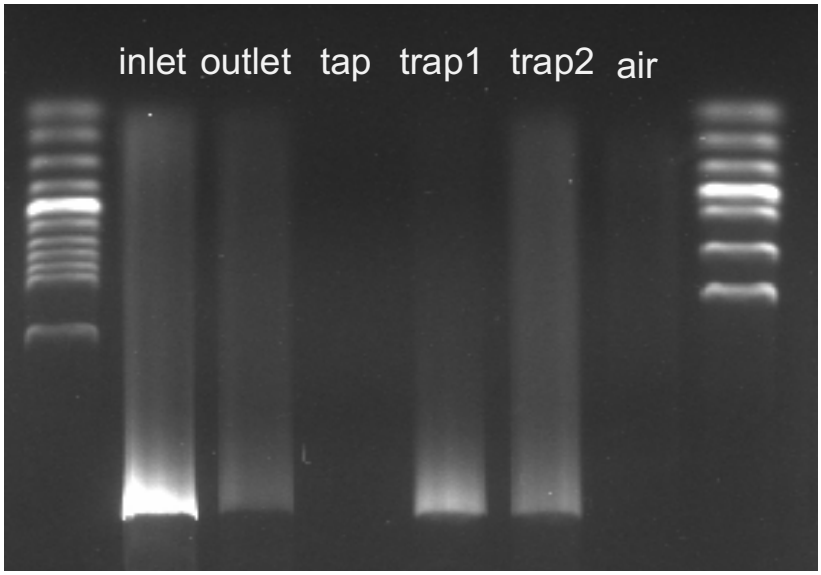


Figure S5.7 The electrophoresis of DNA samples extracted from sewage inlet, sewage outlet, tap water, p-trap water and air samples. The left marker is 100bp ladder, and the right marker is 1000bp ladder.

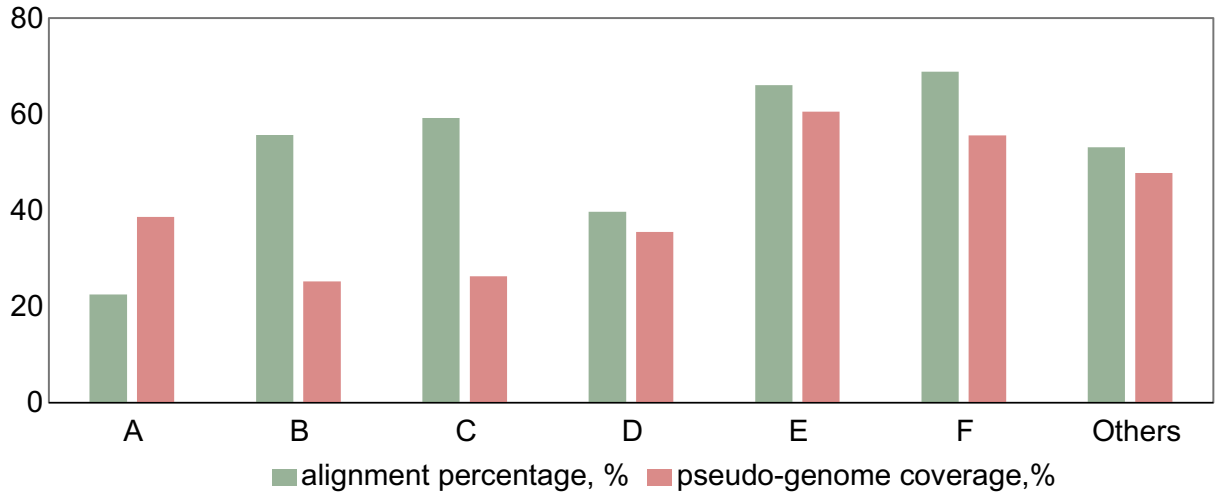


Figure S5.8 Distribution of average mapping ratio and average pseudo-genome coverage of each clade in the phylogenetic tree in Figure 5.5.

Appendix 3 R scripts

1 Network analysis

```
library(psych)

library(graphics)

###gynecology
otu_gynecology <- read_excel("arg_otu_network_forthesis.xlsx", sheet
= "gynecology_otu")

arg_mge_gynecology<-read_excel("arg_otu_network_forthesis.xlsx",
sheet = "gynecology_arg_mge")

df1= otu_gynecology

df2= arg_mge_gynecology

rownames(df1)<-otu_gynecology$sample

rownames(df2)<-arg_mge_gynecology$sample

df1 = df1[,-1]

df2 = df2[,-1]

result_pair_gynecology=data.frame(print(corr.test(df1, df2,
use="pairwise", method="spearman", adjust="fdr", alpha=.05, ci=TRUE,
minlength=10), short=FALSE))
```

```
result_pair_gynecology1 <-
subset(result_pair_gynecology,result_pair_gynecology$raw.p <=0.05)

result_pair_gynecology2 <-
subset(result_pair_gynecology1,abs(result_pair_gynecology1$raw.r) >=0
.6)

result_pair_gynecology3 <-
subset(result_pair_gynecology2,abs(result_pair_gynecology2$raw.r) >=0
.8)

write.csv(result_pair_gynecology2,file =
"corr.gynecology.arg.mge.otu.0.6.csv")

write.csv(result_pair_gynecology3,file =
"corr.gynecology.arg.mge.otu.0.8.csv")

###for gephi network

n<-corr.test(df1, df2, use="pairwise", method="spearman",
adjust="fdr", alpha=.05, ci=TRUE, minlength=10)

occor_r <- n$r

occor_p <-n$p

occor_r_test <- occor_r
occor_r_test[occor_p > 0.01 | abs(occor_r) < 0.8] = 0

diag(occor_r_test) <- 0
```

```
write.csv(occor_r_test,file = "corr.forarg_mge_otu_network.csv")
```

2 ANOSIM, LDA and PCA analysis

```
pca_arg_mge <- read_excel("arg_forthesis.xlsx", sheet =
"arg_MGE_pca")

meta <- read_excel("arg_forthesis.xlsx", sheet = "annotation_sample")

##ANOSIM arg_mge

arg_mge = pca_arg_mge[2:109]

rownames(arg_mge)<-pca_arg_mge$sample

dist<-vegdist(arg_mge, method = "bray")

PCoA1<-cmdscale(dist,k=nrow(arg_mge)-1, eig=T, add = F)

PCoA1<-add.spec.scores(PCoA1,arg_mge,method="wa.scores", Rscale=T,
scaling=1, multi=1)

class(PCoA1)<-c("cmdscale")

PCoA1.anosim.1<-anosim(dist, meta$season, permutations = 9999,
distance = "bray", strata = NULL)

PCoA1.anosim.1

### PCA
```

```
df1<-arrange(arg_only_pca,sample)

meta<-arrange(meta, sample)

t<-df1[,2:ncol(df1)]

t<-t+1e-20 # avoid 0 in origianl data

t<-log2(t) #logscale transform

t[,93]=df1$sample

colnames(t)[93]<-"sample"

t<-merge(t,meta,by="sample",all = TRUE)

rownames(t)<-df1$sample

t= t[,-1]

PCA1<-prcomp(t[,c(1:92)],center = TRUE,scale = TRUE)

summary(PCA1)

plot(PCA1, type="l")

pc1<-PCA1$x[,1]

pc2<-PCA1$x[,2]

p0<-ggplot(t,aes(x=pc1,y=pc2)) + geom_point()
```

```
p0
```

Random Forest between early-summer and late-summer

```
df1<-arrange(pca_summer,sample)

meta<-arrange(meta_summer, sample)

tmp<-merge(df1,meta,by="sample")

# remove unnecessary columns to keep only categories of temperature

tmp<-tmp[,-which(colnames(tmp) %in%
c("sample","type","feature","season","medium","medium2","date"))]

tmp$season2<-as.factor(tmp$season2) # the categories must be factors

tmp2<-data.frame(tmp)

library(randomForest)

df1.rf<-randomForest(season2~.,data=tmp2,importance=T)

df1.imp<-data.frame(importance(df1.rf))

df1.imp$GH<-rownames(df1.imp)

df1.imp<-arrange(df1.imp, MeanDecreaseAccuracy, decreasing=T)
```

```
# then the top 10 most discriminating genus between early-summer and
late-summer are:

df1.imp$GH[1:10]

##Plot loadings values for the most discriminating variables

load<-as.data.frame(PCA1$rotation)

load<-data.frame(genus=rownames(load),PC1=load$PC1,PC2=load$PC2)

df1.imp [3,5] ="Candidatus Alysiosphaera"

load<-load[which(load$genus %in% df1.imp$GH[1:5]),]

g1<-ggplot(load,aes(x=factor(genus),y=PC1))+
  geom_bar(stat = "identity", width =
0.5,fill=alpha("darkblue",0.7))+
  ggtitle("loading on PC1")+
  xlab("Top 5 most discriminating variables") +theme(axis.text.x =
element_text(angle = 90))

g1

g2<-ggplot(load,aes(x=factor(genus),y=PC2))+
  geom_bar(stat = "identity", width = 0.5,
fill=alpha("darkred",0.7))+
  ggtitle("loading on PC2")+
  xlab("Top 5 most discriminating variables") + theme(axis.text.x =
element_text(angle = 90))
```

```
g2

library(cowplot)

plot_grid(g1,g2)

##LDA

t<-df1[,2:ncol(df1)]

t<- t+1e-20 # avoid 0 in origianl data

t<-log2(t)*100 #logscale transform

t[,109]=df1$sample

colnames(t)[109]<-"sample"

t<-merge(t,meta,by="sample")

t[2:108] <- scale(t[2:108], center = TRUE, scale = TRUE)

apply(t[2:108], 2, mean)

apply(t[2:108], 2, sd)

set.seed(123)

training <- sample(rownames(t), nrow(t)*0.8)

train.data <- subset(t, rownames(t) %in% training)
```

```
test.data <- subset(t, ! rownames(t) %in% training)

train.data.1 = train.data[,c(-1,-5,-111,-112,-113,-115)]

test.data.1 = test.data[,c(-1,-5,-111,-112,-113,-115)]

model <- lda(season2~., data = train.data.1)

plot(model)

model

library(ggplot2)

ggplot(cbind(train.data.1, predict(model)$x), aes(LD1,
LD2,color=season2)) + geom_point() + stat_ellipse(level = 0.95,
show.legend = FALSE)

predictions <- predict(model, train.data.1)

mean(predictions$class == train.data.1$season2)

predictions <- predict(model,test.data.1)

mean(predictions$class == test.data.1$season2)
```

3. PCoA analysis

```
dist_arg_mge <-vegdist(arg_mge, method = "bray")
```

```
PCoA1<-cmdscale(dist_arg_mge,k=nrow(arg_mge)-1, eig=T, add = F)

PCoA1<-add.spec.scores(PCoA1,arg_mge,method="wa.scores", Rscale=T,
scaling=1, multi=1)

class(PCoA1)<-c("cmdscale")

dist_otu <- vegdist(otu, method = "bray")

PCoA2<-cmdscale(dist_otu,k=nrow(otu)-1, eig=T, add = F)

PCoA2<-add.spec.scores(PCoA2,otu,method="wa.scores", Rscale=T,
scaling=1, multi=1)

class(PCoA2)<-c("cmdscale")

comp.ord<-procrustes(PCoA1$points,PCoA2$points)

c1<-protest(PCoA1$points[,1:2], comp.ord$Yrot[,1:2], permutations =
10000)

c1

plot(c1)

plot(c1,kind = 2)

residuals(c1)

res_arg_mge_otu <- data.frame(residuals(c1))
```

```
res_arg_mge_otu$sample <- rownames(res_arg_mge_otu)

meta_corr = meta[-14,]

res_arg_mge_otu <-merge(res_arg_mge_otu,meta,by="sample")

write.csv(res_arg_mge_otu,file = "arg_mge_otu_residuals.csv")

library(shape)

plot(0,0,
     xlim=extendrange(r = range(c(c1$X[,1], c1$Yrot[,1])), f = 0.15),
     ylim=extendrange(r = range(c(c1$X[,2], c1$Yrot[,2])), f = 0.15),
     type="n",xlab = "PC1",ylab="PC2")

points(c1$X, pch=19)#arg circle

points(c1$Yrot, pch=17,col="red")#col=as.factor(meta$medium2))#genus
trangle

legend("bottomleft",
legend=levels(as.factor(meta$medium2)),col=unique(as.factor(meta$medi
um2)))

with(c1,Arrows(X[,1], X[,2],Yrot[,1], Yrot[,2],
               code=2,
               arr.adj = 0.6,
               arr.length = 0.05))

ordiellipse(c1$X,meta$medium2,conf=0.68)
```

4. Mantel test

```
otu_pcoa <- read_excel("arg_otu_network_forthesis.xlsx", sheet =
"otu_pcoa")

arg_mge_pcoa<-read_excel("arg_otu_network_forthesis.xlsx", sheet =
"arg_MGE_pcoa")

meta<-read_excel("16s_genus_forthesis.xlsx", sheet =
"sample_annot_pca_meta")

otu <-merge(meta,otu_pcoa,by="sample")

arg_mge <-merge(meta,arg_mge_pcoa,by='sample')

otu_gynecology <- subset(otu,otu$type=="gynecology")

arg_mge_gynecology <-subset(arg_mge,arg_mge$type=="gynecology")

meta_gynecology<-subset(meta,meta$type=="gynecology")

dist_arg_mge_gynecology <-vegdist(df2, method = "bray")

dist_otu_gynecology <-vegdist(df1, method = "bray")

mantel(dist_arg_mge_gynecology,dist_otu_gynecology,method =
"spearman")
```

4. Shannon diversity and Venn analysis

```
### Shannon
matrix_shannon = matrix(data=NA, nrow=64, ncol=1)

print(matrix_shannon)

i=1
 while (i<65) {
  i=i+1
  di = diversity(shannon_genus[,i],index="shannon")
  print (di)
  matrix_shannon[i-1,1]=di
}

print(matrix_shannon)

write.csv(data.frame(matrix_shannon),file="genus_shannondiversity.csv
")

###genus venn

genus_venn <- read_excel("16s_genus_forthesis.xlsx",
                         sheet = "genus_venn")

###create the genus marix for venn

matrix_genus = matrix(data=NA, nrow=13760, ncol=3)
print(matrix_genus)

i=0
```

```
j=0
a=-1
while (j<64) {
  j=j+1
  a=a+1
  c=0
  i=215*a
  while (i<215*(a+1)){
    i=i+1
    c=c+1
    matrix_genus[i,1]=as.matrix(genus_venn[c,1])
    matrix_genus[i,2]=as.matrix(genus_venn[c,j+1])
    matrix_genus[i,3]=colnames(genus_venn[j+1])
 }
}

matrix_genus<-data.frame(matrix_genus)

print(matrix_genus)

write.csv(matrix_genus,file = "genus_forvenn.csv")

###draw genus venn

color_v <- c("dodgerblue", "goldenrod1", "darkorange1", "seagreen3",
"orchid3","red","black","gray")

library(VennDiagram)

v2<-venn.diagram( x =
list(genus_water_venn,genus_air_venn,genus_winter_venn,genus_summer_v
```

```
enn),category.names =
c("water","air","winter","summer"),filename=NULL, fill=color_v[1:4])

dev.off()

grid.draw(v2)
```

###genus_department

```
list4=list(blood=genus_d1_venn,emergency=genus_d2_venn,gynecology=gen
us_d3_venn,ophthalmology=genus_d5_venn,
otolaryngology=genus_d6_venn,pneumology=genus_d7_venn,surgery=genus_d
8_venn)

list4 <- fromList(list4)

color_v <- c("dodgerblue", "goldenrod1", "darkorange1",
"orchid3","red","black","gray")

##upset(list4,nsets = length(list4),sets.bar.color
=color_v,matrix.color = "blue",
      keep.order = TRUE,order.by = "freq")

v3 <- venn.diagram ( x = list4, filename = NULL, fill= color_v)
```

5. SNP-distance calculation

```
#seqinr::dist.alignment is the square root version of ape::dist.gene,
not ape::dist.dna.
#ape::dist.dna claim to calculate the number of sites that differ
between each pair of
```

```
#sequences, whereas ape::dist.gene calculate the distance between each
pair of sequences
#through the number of different sites It does look similar in a glimpse
when you look at
the documentation, but it does not.  Imagine two sequences,

#Sequence 1: CCTGCA

#Sequence 2: TTCXXG

#The total number difference is 6 (dist.gene), but the type of
different sites are 4 (dist.dna).
#That is why the values between dist.dna and dist.alignment different.
They are calculating
#different things.

library(ape)
install.packages("seqinr")
library(seqinr)

##cladef
seq <-read.dna('cladef.mafft.fa',format = "fasta")
#seq_bin <-as.DNAbin(seq)
myseq<-read.alignment('cladef.mafft.fa',format = "fasta")
#dist <- as.data.frame(dist.dna(seq, model="N",as.matrix = T))
#dist_GraphSNP <- cbind('rowCol' = rownames(dist), dist)
dist <- as.matrix(dist.alignment(myseq, matrix = "identity" ),gap)
write.csv(dist,file="snp_distance_matrix.csv")
#dist_tmp <- as.matrix(dist.gene(seq))

library(tidyverse)
```

```
library(readxl)
dist2 <- read_excel("snpdistance.xlsx")
##wide matrix to long matrix
dist3 <- dist2 %>% pivot_longer(cols = pos6:trap11,names_to =
"sample2",values_to = "distance")
write.csv (dist3,file="snp_distance_long_format.csv")
library("ggplot2")
#library("GGally")
p<-ggplot(dist3, aes(sample, sample2)) + geom_tile(aes(fill =
distance)) + scale_fill_gradient(low = "white", high = "coral4") +
  theme(axis.text.x = element_text(angle = 90))
p

##violin plot
dist3$clade ="cladef"
dist3$distance2 <-dist3$distance^2 ##distance2 is the real pairwise
distance
v<-ggplot(dist3, aes(x=clade,y=distance2)) +
  geom_violin() + geom_boxplot(width=0.1)
v
mean(dist3$distance2) #[1] 0.1938347
median(dist3$distance2)#[1] 0.1750417
##range (0-0.5401423)


##Clade D
cladeD<-read.alignment('cladeD.mafft.fa',format = "fasta")
dist_D <- as.matrix(dist.alignment(cladeD, matrix = "identity" ),gap)
write.csv(dist_D,file="snp_distance_matrix.csv")

distD2 <- read_excel("snpdistance.xlsx", sheet = "cladD_squareroot")
```

```
distD3 <- distD2 %>% pivot_longer(cols = trap5:trap2,names_to = "sample2",values_to = "distance")
write.csv(distD3,file="snp_distance_long_format.csv")

distD4 <- read_excel("snpdistance.xlsx", sheet = "cladD1D2_squaroot")
distD4$distance2 <-distD4$distance^2 ##distance2 is the real pairwise distance

v_D<-ggplot(distD4, aes(x=clade,y=distance2)) +
  geom_violin() + geom_boxplot(width=0.1)
v_D

mean(distD4$distance2) ##[1] 0.1736177
aggregate(distD4$distance2,by=list(distD4$clade),mean)
##D1 0.1664589
##D2 0.1807765
```

6. R markdown of RF-distance

```
---
title: "Compare phylogenetic tree between 51-core-gene-based alignment and whole-genome-based alignment of L. pneumophila"
author: "Qing Yang"
date: "4/16/2024"
output: html_document
---

```{css, echo = TRUE}
pre {
  max-height: 300px;
  overflow-y: auto;
```
```

````
}
pre[class] {
  max-height: 100px;
}
```
```{r setup, include=FALSE}
knitr::opts_chunk$set(
        echo = TRUE,
        message = FALSE,
        warning = FALSE
)
```

## R Markdown

This is an R Markdown document to compare phylogenetic tree between 51-core-gene-based alignment and whole-genome-based alignment of L. pneumophila. The similarty between two trees are calculated by generalized RF distance. Generalized RF distances work by finding a matching that pairs splits from one tree with splits in the other. Each pairing is scored according to the similarity of the paired splits; the sum of these scores is the score of the matching. The tree distance is given by the score of the optimal matching distance value of 1 (= similarity of 0)

```{r import tree, fig.height=30, fig.width=20}
#step 1: import 51-core-gene (cg) tree and whole-genome (wg) tree
library(ape)
cg_import <- "51cg.annot.nex.tree"
cg <-ape::read.nexus (cg_import,force.multi = TRUE)
cg_tree<-cg[[8]] #cg
````

```
plot(cg_tree)

wg_import <- "wg.annot.nex.tree"
wg<-ape::read.nexus (wg_import,force.multi = TRUE)
wg_tree <-wg[[8]] #wg
plot(wg_tree)

#step 2:compare tree using generalized RF distance
library(TreeDist)

##step 2.1: We can view the splits in each tree, named according to
the number of their associated node:
summary(TreeTools::as.Splits(cg_tree))
summary(TreeTools::as.Splits(wg_tree))

```
```{r compare tree, fig.height=30, fig.width=20}
##step 2.2:tell if two trees are equal, and calculate generalized RF
distance value or similarity value
all.equal(cg_tree,wg_tree)
TreeDistance(cg_tree,wg_tree)
SharedPhylogeneticInfo(cg_tree,wg_tree,normalize                    =
TRUE,reportMatching = FALSE,diag = TRUE)

##step 2.3: Find matching splits and generate splits matrix with pair
score between two trees
attri<-attributes(SharedPhylogeneticInfo(cg_tree,           wg_tree,
reportMatching = TRUE))

###step 2.3.1: view splits matrix with pair score,where the row name
represents cg, and the column name represents wg.
```

```
pairscore<-data.frame(attri[["pairScores"]])
rownames(pairscore) = paste("cgsplit", 1:67, sep = "")
colnames(pairscore) = paste("wgsplit", 1:88, sep = "")
library(pheatmap)
pheatmap(pairscore,border=F)

###step 2.3.2: view matching splits between two trees
attri[["matchedSplits"]]
### step 2.3.3 : visualize the matching splits with pair score between
two trees
VisualizeMatching(SharedPhylogeneticInfo, cg_tree, wg_tree,matchZeros
= FALSE)

###step 2.3.4: select some matching splits exhibited on two trees with
ggtree package

library(readxl)
library(ggtree)
cg_annot <- read_excel("matchsplits.xlsx",
                       sheet = "forggtree")
cg_annot2<-data.frame(node=cg_tree$tip.label)
cg_annot3<-merge(cg_annot2,cg_annot,by="node")
row.names(cg_annot3)<-cg_annot3$node
p_cg<-ggtree(cg_tree) + geom_tiplab(align = T)
p_cg2<-gheatmap(p_cg,  cg_annot3["Split"],offset  =  0,legend_title  =
"cg_split")
wg_annot = cg_annot3
p_wg <-ggtree(wg_tree) + geom_tiplab(align = T)
p_wg2<-gheatmap(p_wg,  wg_annot["Split"],offset  =  0,legend_title  =
"wg_split")
#show cg-tree with 8 matching splits
```

```
p_cg2
#show wg-tree with 8 mathcing splits
p_wg2

```
```
```