6-2024

CLONALITY AND EPIDEMIOLOGICAL STUDIES OF E. COLI STRAIN COLLECTED FROM ABU DHABI EMIRATE USING NGS AND BIOINFORMATIC TOOLS

Ahmed Abdelrahman Mohamed

جامعة الإمارات العربية المتحدة
United Arab Emirates University

MASTER THESIS NO. 2024: 70

College of Medicine and Health Sciences

Department of Medical Microbiology and Immunology

CLONALITY AND EPIDEMIOLOGICAL STUDIES OF *E. COLI* STRAIN COLLECTED FROM ABU DHABI EMIRATE USING NGS AND BIOINFORMATIC TOOLS

Ahmed Abdelrahman Mohamed Abdelfattah

June 2024

United Arab Emirates University

College of Medicine and Health Sciences

Department of Medical Microbiology and Immunology

CLONALITY AND EPIDEMIOLOGICAL STUDIES OF *E. COLI* STRAIN COLLECTED FROM ABU DHABI EMIRATE USING NGS AND BIOINFORMATIC TOOLS

Ahmed Abdelrahman Mohamed Abdelfattah

This thesis is submitted in partial fulfilment of the requirements for the degree of Master of Medical Sciences (Microbiology and Immunology)

June 2024

United Arab Emirates University Master Thesis
2024: 70

Cover: Unraveling the Mysteries of the Bacterial Genome: A Bioinformatics Perspective
(Photo: By Ahmed Abdelrahman Mohamed Abdelfattah)

Print: University Print Service, UAEU 2024

Declaration of Original Work

I, Ahmed Abdelrahman Mohamed Abdelfattah. the undersigned, a graduate student at the United Arab Emirates University (UAEU), and the author of this thesis entitled "*Clonality and Epidemiological Studies of E. Coli Strain Collected From Abu Dhabi Emirate Using NGS and Bioinformatic Tools*", hereby, solemnly declare that this is the original research work done by me under the supervision of Dr. Mushtaq Khan, in the College of Medicine and Health Sciences at UAEU. This work has not previously formed the basis for the award of any academic degree, diploma or a similar title at this or any other university. Any materials borrowed from other sources (whether published or unpublished) and relied upon or included in my thesis have been properly cited and acknowledged in accordance with appropriate academic conventions. I further declare that there is no potential conflict of interest with respect to the research, data collection, authorship, presentation and/or publication of this thesis.

Student's Signature: 

Date: _____10-8-2024_____

Advisory Committee

1) Advisor: Mushtaq Khan

Title: Associate Professor

Department of Medical Microbiology and Immunology

College of Medicine and Health Sciences

2) Co-advisor: Mushal Allam

Title: Instructor

Department of Genetics and Genomics

College of Medicine and Health Sciences

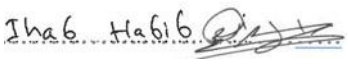
Approval of the Master Thesis

This Master Thesis is approved by the following Examining Committee Members:

1) Advisor (Committee Chair): Dr. Mushtaq Ahmad Khan
   Title: Associate Professor
   Department of Medical Microbiology and Immunology
   College of Medicine and Health Sciences, UAEU

   Signature  Date 13.08.2024

2) Member: Dr. Ihab Habib
   Title: Associate Professor
   Department of Veterinary Medicine
   College of Agriculture and Veterinary Medicine, UAEU

   Signature  Date 13.08.2024

3) Member (External Examiner): Dr Hafiz Ahmad
   Title: Associate Professor
   Department of Microbiology
   RAK College of Medical Sciences, RAK Medical and Health Sciences University, UAE

   Signature  Date 13.08.2024

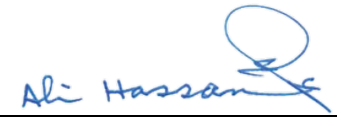
This Master Thesis is accepted by:

Acting Dean of the College of Medicine and Health Sciences: Professor Ahmed Alsuwaidi

Signature 

Date 22.08.2024

Dean of the College of Graduate Studies: Professor Ali Al-Marzouqi

Signature 

Date 23/08/204

Abstract

In the current landscape, where antibiotic resistance in bacteria is steadily rising, the role of bacterial genomics has assumed increasing significance. Therefore, it is necessary for microbiologists to acquire proficiency in the utilization of these methods and tools to harness their potential in gaining valuable insights into bacterial genomes. This can facilitate the development of innovative strategies for managing and treating bacterial infections. Here we present the power of the whole-genome sequencing technology in characterization of multidrug resistant *Escherichia coli* isolates. Fourteen *Escherichia coli* isolates were selected for this study. Following the standard microbiological techniques to confirm the sample identification, the genomic DNA from each isolate was extracted and sequenced using Illumina NovaSeq machine. The sequencing of the 14 isolates yielded an average of 100,079,168 high-quality reads which assembled to an average of 188 contiguous sequences longer than 200 bp. The draft genomes were then OH serotyped and MLST sequence typed. The common serotype and sequence type was O102:H6 (4/16, 42.8%) and ST648 (4/14, 28.6%), respectively. Antimicrobial resistant analysis characterized bla_{NDM} gene in all isolates (13 carried bla_{NDM-5} and 1 carried bla_{NDM-1}) which comfort resistance to carbapenem antibiotics. Of these, 8 isolates also carrying $bla_{CTX-M-15}$ and 7 carrying bla_{CMY-42} making them multidrug-resistant *E. coli* producing extended-spectrum β-lactamase (ESBL) enzymes and carbapenemases. Furthermore, the isolates contained a range of resistance genes related to aminoglycosides, tetracyclines, sulphonamides and macrolides. Noteworthy, all isolates retain their susceptibility to colistin and tigecycline. Whole-genome sequencing serves as a sophisticated and cutting-edge approach for precisely predicting antibiotic resistance patterns in complex and highly resistant *E. coli* isolates, particularly those exhibiting carbapenem resistance and ESBL production. It empowers researchers and clinicians with invaluable insights into the genetic basis of resistance, ultimately aiding in the development of more tailored and efficacious therapeutic interventions.

Keywords: Bacterial Genomics, Whole-Genome Sequencing, Antibiotic Resistance, *Escherichia Coli*, Multidrug Resistance, Carbapenem Resistance, Extended-Spectrum B-Lactamase (ESBL).

Title and Abstract (in Arabic)

دراسات التناسخية والوبائية لسلالة الإشريكية القولونية التي تم جمعها من إمارة أبوظبي باستخدام أدوات التسلسل الجيني والبَيُوْمعلوماتية

الملخص

الهدف: في المشهد الحالي، حيث تتزايد مقاومة المضادات الحيوية في البكتيريا بشكل مطرد، اكتسب دور علم الجينوم البكتيري أهمية متزايدة. لذلك، من الضروري أن يكتسب علماء الأحياء الدقيقة الكفاءة في استخدام هذه الأساليب والأدوات لتسخير إمكاناتها في اكتساب رؤى قيمة في الجينومات البكتيرية. يمكن أن يسهل هذا تطوير استراتيجيات مبتكرة لإدارة وعلاج الالتهابات البكتيرية. هنا نقدم قوة تقنية تسلسل الجينوم الكامل في توصيف عزلات الإشريكية القولونية المقاومة للأدوية المتعددة.تم اختيار أربعة عشر عزلة من الإشريكية القولونية لهذه الدراسة. باتباع التقنيات الميكروبيولوجية القياسية لتأكيد تحديد العينة، تم استخراج الحمض النووي الجينومي من كل عزلة وتسلسله باستخدام آلة .Illumina NovaSeq أسفر تسلسل الـ 14 عينة عن متوسط 100,079,168 قراءة عالية الجودة والتي تم تجميعها في متوسط 188 تسلسل متجاور أطول من 200 زوج أساسي. ثم تم تصنيف الجينومات حسب النمط المصلي OH ونمط التسلسل .MLST كان النمط المصلي المشترك ونوع التسلسل هما (4/16, 42.8%) O102:H6 و (4/14, 28.6%) ST648 على التوالي. وصف تحليل مقاومة مضادات الميكروبات جين blaNDM في جميع العينات: 13 عينة تحمل blaNDM-5 وعينة واحدة تحمل blaNDM-1 مما يعزز من المقاومة للمضادات الحيوية وخاصة الكاربابينيم. من بين هذه العينات، تحمل 8 عينات أيضًا blaCTX-M-15 و 7 عينات تحمل blaCMY-42 مما يجعلها مقاومة للأدوية المتعددة وتنتج إنزيمات بيتا لاكتاماز (ESBL) ذات الطيف الموسع وكاربابينيماز. علاوة على ذلك، احتوت العزلات على مجموعة من جينات المقاومة المرتبطة بالأمينوغليكوزيدات والتتراسيكلينات والسلفونواميدات والماكروليدات. والجدير بالذكر أن جميع العزلات تحتفظ بحساسيتها للكوليستين والتيجيسيكلين .الخلاصة: يعمل تسلسل الجينوم الكامل كنهج متطور للتنبؤ بدقة بأنماط مقاومة المضادات الحيوية في عزلات الإشريكية القولونية المعقدة والمقاومة للغاية، وخاصة تلك التي تظهر مقاومة الكاربابينيم وإنتاج .ESBL يمكّن للباحثين والأطباء من الحصول على رؤى لا تقدر بثمن حول الأساس الجيني للمقاومة، مما يساعد في نهاية المطاف في تطوير تدخلات علاجية أكثر ملاءمة وفعالية.

مفاهيم البحث الرئيسية: علم الجينوم البكتيري، تسلسل الجينوم الكامل، مقاومة المضادات الحيوية، الإشريكية القولونية، مقاومة الأدوية المتعددة، مقاومة الكاربابينيم، بيتا لاكتاماز الطيف الموسع (ESBL).

Acknowledgements

I extend my heartfelt thanks to my committee for their invaluable guidance, support, and assistance throughout the preparation of my thesis. I am particularly grateful to my advisor, Dr. Mushtaq Khan and Dr. Mushal Allam, for their unwavering support, profound knowledge, innovative ideas, and creative thinking, which have been instrumental in my master's study and research.

My deepest appreciation and special thanks are reserved for Dr. Akela Ghazawi, who, despite her busy schedule, has helped me refine my skills and served as a source of inspiration. I aspire to emulate her passion for knowledge.

I also wish to express my gratitude to all members of the Medical Microbiology and Immunology Department at the United Arab Emirates University Specially Dr. Ashrat for their kindness and support.

Lastly, I owe a debt of gratitude to my father, mother, brother and sister for their spiritual support, for helping me navigate through challenging times, and for their unwavering care. My expressions of thanks will always fall short of what they have given me, and for that, I am eternally grateful to my friends for their unstoppable support and all the good memories that we had during this journey which served as the fuel to be here today.

Dedication

To my beloved parents, family and friends

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AAC(3)-IId	Aminoglycoside 6'-N-acetyltransferase 3-IId
Aada2	Aminoglycoside Resistance Protein A2
Aada5	Aminoglycoside Resistance Protein A5
Acrb	Multidrug Transporter B
AcrD	Multidrug Transporter D
AcrE	Multidrug Transporter E
AcrS	Multidrug Efflux Pump Subunit AcrS
BAC	Bacterial Artificial Chromosome
Baca	BacA Metallo-beta-lactamase
Baer	Two-component System Response Regulator BaeR
BRP(MBL)	Bacteriophage-Resisting Protein (MBL)
catB3	Chloramphenicol Acetyltransferase B3
CATI	Chloramphenicol Acetyltransferase I
CHUA	Colloidal Hematogenous Utilization Associated Protein
CIA	Colanic Acid Synthesis Protein A
CMY-42	AmpC beta-lactamase 42
CpxA	Two-component System Sensor Protein CpxA
Dfra12	Trimethoprim Resistance Protein 12
EmrB	Multidrug Transporter B
EmrR	Multidrug Resistance Regulator EmrR
Escherichia coli acrA	Multidrug Transporter A of *Escherichia coli*
EvgA	Envelope Stress Response Regulator EvgA
FimH	Fimbrial Adhesion Protein H
FosA3	Fosfomycin Resistance Protein A3
FYUA	Ferric Yersiniabactin Uptake Protein A
GADW	Glutamate Decarboxylase W
HGP	Human Genome Project
H-NS	Histone-like Nucleoid Structuring Protein

IDBA-UD	Iterative De Bruijn Graph Assembler - Ultrafast De Bruijn Graph Assembler
IHA	Iron-regulated Heptose Adhesin
IRP2	Insulin Regulated Protein 2
ISS	Increased Serum Survival Protein
IUCC	Iron Uptake and Chaperone Protein C
IUTA	Iron uptake Transporter A
KdpE	Potassium-transporting ATPase E
KpsE	Capsular Polysaccharide Transport Protein E
kpsMII_K5	Capsule Polysaccharide Synthesis Protein K5
LeuO	Leucyl-tRNA Synthetase
LpfA	Long Polar Fimbriae Protein A
LTS	Long Term Support
MarA	Multiple Antibiotic Resistance Regulator A
McbA	Microcin Biosynthesis Protein A
MDR	Multidrug Resistance
MdtE	Multidrug Transporter E
MdtG	Multidrug Transporter G
MdtH	Multidrug Transporter H
MEGAHIT	Metagenome Scaler
MLST	Multilocus Sequence Typing
MphA	Macrolide Efflux Pump A
MRSA	Methicillin-resistant Staphylococcus Aureus
Mrx	Multidrug Resistance Transporter
MsbA	Multidrug Transporter A
NDM-5	New Delhi Metallo-beta-lactamase 5
NeuC	Neuraminidase C
NGS	Next-Generation Sequencing
OMPT	Outer Membrane Protein T
PCR	Polymerase Chain Reaction
PmrF	PhoP/PhoQ-regulated Membrane Protein F

QacEdelta1	Quaternary Ammonium Compound Resistance Protein Edelta 1
QC	Quality Control
rmtB	16S rRNA Methyltransferase B
RNA	Ribonucleic Acid
Sat	Secreted Autotransporter Toxin
SBS	Sequencing by Synthesis
SenB	Sensor Histidine kinase Component of a two-component Regulatory System
SitA	Sitron-dependent Transporter A
Sul1	Sulfonamide Resistance Protein 1
TEM-1	beta-lactamase TEM-1
TEM-57	beta-lactamase TEM-57
TerC	Tellurium Resistance Protein C
TolC	Outer Membrane Protein TolC
TraT	Transporter T
TRNA	Transfer RNA
YfcV	Uncharacterized Fimbrial-like Protein
YojI	AcrAB-TolC Resistance Protein YojI

Chapter 1: Introduction

1.1 Overview

Over the past five decades, the scientific community has made monumental strides in the field of molecular biology, particularly in determining the sequence order of nucleic acids within biomolecules. This journey from sequencing single genes to the advanced capability of Whole Genome Sequencing (WGS) represents a significant technological evolution. WGS, a method capable of sequencing millions of DNA fragments simultaneously, allows researchers to rapidly analyze large datasets. This technique is especially vital in studying complex biological phenomena such as antibiotic resistance, a major concern in global public health. A prime example of this concern is Carbapenem-resistant Enterobacteriaceae (CRE), a group of bacteria highly resistant to carbapenem antibiotics, typically used as a last-resort treatment for severe infections. CRE infections, including those caused by strains of *Escherichia coli*, pose a significant threat due to their high resistance to multiple antibiotics. In this study, we aim to leverage the versatility of WGS and bioinformatics tools to gain a deeper understanding of the genetic mechanisms behind carbapenem resistance in *Escherichia coli*. This approach not only aids in identifying specific resistance genes but also contributes to tracking the transmission and evolution of these resistant strains, ultimately enhancing our strategies to combat this critical public health issue.

1.2 Research Objectives

This study is designed to systematically evaluate the capabilities of bacterial bioinformatics tools in analyzing Next-Generation Sequencing (NGS) data. A key goal is to uncover the genetic diversity and antimicrobial resistance (AMR) patterns of *E. coli* strains within the study population from Abu Dhabi Emirate, providing critical data to support public health initiatives and infection control strategies. Using NGS, we will sequence 14 *E. coli* strains collected from patients with confirmed *E. coli* infections, aiming to understand their clonality and epidemiological characteristics. The NGS data will serve to ascertain the genetic relationships among these strains and identify clonal lineages prevalent in the region. Advanced bioinformatics analyses will be employed to detect genes linked to antibiotic resistance, enhancing our understanding of the

resistance mechanisms in these strains. We anticipate the study will reveal distinct clonal lineages of *E. coli* in Abu Dhabi Emirate and identify key genes associated with antibiotic resistance. These findings are expected to inform the development of targeted approaches for preventing and controlling *E. coli* infections in the UAE.

1.3 History of DNA Sequencing

The journey of DNA sequencing technology began with pioneering work in understanding the chemical structure of biopolymers. Fred Sanger, a key figure in this field, dedicated his career to elucidating these structures, believing that comprehending the sequence of living molecules was crucial (Sancar & Sancar, 1988). Ironically, despite DNA's central role in genetics, protein and RNA sequencing preceded DNA sequencing. Sanger's groundbreaking work in the early 1950s on insulin marked the first successful protein sequencing, showcasing the distinct amino acid sequences in proteins (Sanger & Tuppy, 1951). The subsequent development of the Edman degradation technique further accelerated protein sequencing (Takahashi et al., 1950). In the 1960s, RNA sequencing emerged, employing techniques like RNase digestion, chromatography, and electrophoresis, followed by exonuclease sequencing. This laborious process culminated in the sequencing of alanine tRNA from yeast, a significant milestone in molecular biology (Holley et al., 1965). Despite these advancements in protein and RNA sequencing, it wasn't until later that DNA sequencing gained prominence. The transition from understanding proteins and RNA to focusing on DNA marked a pivotal shift in molecular biology, leading to the development of techniques that laid the foundation for modern sequencing technologies. It's crucial to recognize these early achievements, as they set the stage for the rapid, high-throughput DNA sequencing methods we use today (Sanger et al., 1965).

1.3.1 First Generation Sequencing

The early stages of DNA sequencing were notably challenging. In 1968, Wu's pioneering work revealed 12 bases at the ends of bacteriophage lambda using primer extension methods (Wu & Kaiser, 1968). This was followed by Gilbert and Maxam's achievement in 1973, who decoded 24 bases of the lactose-repressor binding site, a process that impressively took two years per base (Gilbert & Maxam, 1973). The

landscape of sequencing was significantly changed in 1976 with the introduction of two groundbreaking techniques capable of decoding hundreds of bases daily: the Sanger-Coulson chain termination method and the Maxam-Gilbert chemical cleavage method, both of which employed radioactive labeling to determine nucleotide sequences.

In the Sanger-Coulson method, DNA polymerase was used to extend a primer, adding a dideoxynucleotide that lacks a 3' hydroxyl group, effectively terminating the DNA synthesis. The DNA fragments produced were then analyzed through gel electrophoresis to determine the sequence. On the other hand, the Maxam-Gilbert method chemically cleaved DNA at specific sites, with different reactions in separate tubes for each base. These fragments were also separated by gel electrophoresis. While they are now considered outdated, the Sanger-Coulson and Maxam-Gilbert methods were pivotal in the development of DNA sequencing technology. Sanger's approach involved the use of DNA polymerase to extend a primer with chain-terminating nucleotides, resulting in fragments of various sizes (Sanger et al., 1977). Gilbert's technique used chemical reactions to partially cleave a DNA-restriction fragment at specific bases (Maxam & Gilbert, 1977). Both methods utilized polyacrylamide slab gel electrophoresis, a technique introduced by Maniatis et al. in 1975, for fragment separation. The fragments, visualized on X-ray film, formed a ladder pattern that was used to read the base sequence.

These advancements led to further innovations. In 1979, Staden introduced shotgun sequencing, a method that involved sequencing random clones and assembling the sequences by identifying overlaps (Staden, 1979). The single-stranded M13 phage cloning vector, developed by Messing around 1980, greatly simplified RNA sequencing (Messing et al., 1981). This method was later applied to sequence the genomes of viruses, including bacteriophage lambda, by sequencing RNA molecules and using the overlaps between fragments to deduce the entire genome sequence (Sanger et al., 1982). In 1987, Smith, Hood, and Applied Biosystems introduced automated, fluorescence-based Sanger sequencing machines (Smith et al., 1986), boosting the daily sequencing capacity to about 1,000 nucleotides. The exponential increase in sequence data led to the establishment of databases like GenBank and search tools like BLAST (Altschul et al., 1990), enhancing the value of each sequence and encouraging a culture of data sharing.

By 1982, GenBank had amassed approximately 500,000 bases, and this number grew to around 10 million by 1986 (GenBank and WGS Statistics; https://www.ncbi.nlm.nih.gov/genbank/statistics/).

Table 1: The key differences between Sanger and Maxam methods

Method	**Principle**	**Advantages**	**Disadvantages**	**References**
Sanger-Coulson	Uses chain termination	Simple, relatively inexpensive	Low throughput, error-prone	(Sanger et al., 1977).
Maxam-Gilbert	Uses chemical cleavage	High throughput, more accurate	More complex, toxic chemicals	(Maxam & Gilbert, 1977).

1.3.2 Next Generation Sequencing

In the 1980s and 1990s, several research teams investigated alternatives to electrophoretic sequencing. Although the results of these efforts did not become apparent until after the Human Genome Project (HGP), a new method called \"massively parallel\" or \"next generation\" DNA sequencing (NGS) emerged, nearly surpassing Sanger sequencing within a decade of its completion replaced. The main difference between NGS technology and electrophoretic sequencing is multiplexing. In NGS, a complex library of DNA templates is securely fixed on a two-dimensional surface rather than using separate tubes for each reaction. This arrangement makes all templates accessible from a single reagent volume. Additionally, in vitro amplification creates duplicates of each template for sequencing, eliminating the need for bacterial cloning. Finally, the sequencing process, often referred to as “sequencing by synthesis,” involves biochemical cycles such as: For example, polymerase-mediated incorporation of fluorescently labeled nucleotides and imaging, rather than measurement of fragment

lengths (SBS) (Harris et al., 2008), in vitro clonal amplification was largely responsible for dense multiplexing of NGS with millions to Billions of immobilized templates were made possible. Primers immobilized on a surface are used to amplify a complicated template library such that copies of each template remain in close proximity to one another. This method is called polonies or bridge reinforcement (Adessi et al., 2000). In clonal emulsion PCR, multiple copies of a template are immobilized on beads before they are arranged on a surface and sequenced (Dressman et al., 2003; Margulies et al., 2005). Clonal "nanoballs" can be generated, assembled, and sequenced in solution using a third method called rolling circle amplification (Drmanac et al., 2010). The SBS used a total of three different primary tactics. Pyrosequencing is a technique developed by Ronaghi and Nyrèn. It involves the sequential addition of each deoxynucleotide in different steps (dNTP). When the firefly luciferase enzyme incorporates deoxyribonucleotide triphosphates (dNTPs) into a substrate, it releases pyrophosphate as a byproduct. Pyrophosphate is then used by firefly luciferase to produce light (Ronaghi et al., 1996). An ion-sensitive field effect transistor can be used in a similar manner as previously described to detect natural dNTP internals. Sequence-specific binding of fluorescent oligonucleotides to templates can also be achieved by using a second method that takes advantage of the specificity of DNA ligases (Brenner et al., 2000; Drmanac et al., 2010; McKernan et al., 2009; Shendure et al., 2005). Stepwise integration of fluorescently labeled deoxynucleotides via the action of polymerase is a third method and has been shown to be the most reliable (Braslavsky et al., 2003; Mitra et al., 2003). The success of single-stranded DNA synthesis using polymerase was largely achieved through the creation of dNTPs that could terminate and reversibly fluoresce. In addition, a specially designed polymerase was developed to ensure that each template only incorporates one dNTP per cycle. After imaging to identify the color integrated by each template on the surface, both blocking and fluorescent groups were removed to prepare for the next expansion step (Ruparel et al., 2005). This basic technique was used by Solexa, developed by Balasubramanian and Klenerman in 1998. In 2005, the first fully integrated NGS platforms were introduced with the resequencing of the *Escherichia coli* genome by Shendure, Porreca, Mitra and Church (Shendure et al., 2005), the de novo assembly of the Mycoplasma genitalium genome by Margulies, Rothberg and 454.

Solexa54's resequencing of PhiX174 and a human BAC showed that even very short reads can be helpful, as long as there is a reference genome to assign them to. Re-sequencing of the human genome on the Solexa platform using paired 35 base pair reads could be possible within the next three years (Bentley et al., 2008).

The first commercial NGS instrument was made available for purchase by 454, Inc. in 2005. With the completion of the HGP, only a select number of genomic centers were able to perform large-scale sequencing. Individual labs were able to immediately access the power of an entire HGP-era genome center once they had access to 454 and other competing devices that came soon after. The development of new DNA sequencing technologies has led to a democratization of sequencing functions, meaning more people and organizations have access to these technologies. This has had a dramatic impact on the field of genomics, leading to the emergence of new methods, insights, genomes, and other breakthroughs (Brenner et al., 2000). Historically, Applied Biosystems has had a monopoly on DNA sequencing technologies. However, with the development of new technologies, a number of companies have entered the market, including 454, Solexa, Agencourt, Helicos, Complete Genomics and Ion Torrent (Braslavsky et al., 2003; McKernan et al., 2009). These companies face fierce competition, which has driven down the cost of DNA sequencing.

The cost per base of sequencing DNA decreased by four orders of magnitude between 2007 and 2012 (Lindor et al., 2017). This makes DNA sequencing much more affordable, opening up new opportunities for research and applications. The 454, SOLiD and Helicos platforms are no longer being developed and the Illumina platform is the dominant one. However, Complete Genomics is still a possible rival (Drmanac et al., 2010). Since 2012, the pace of innovation has slowed, resulting in less competition. Nevertheless, it is noteworthy to consider the significant advances that have been made since the introduction of next-generation sequencing (NGS) in 2005. Although NGS read lengths are still shorter than those of Sanger sequencing and are typically a few hundred bases, they are still very precise, often exceeding 99.9 percent accuracy. Remarkably, a single graduate student using just one instrument, such as an Illumina NovaSeq, and investing a few thousand dollars can generate over a billion independent reads in just two days. This accumulation amounts to one terabase of sequence data and can be

achieved at relatively low cost. To put this into perspective, this production is almost 40 times larger than the estimated 23 gigabases produced by the Human Genome Project (HGP) to draft the human genome.

1.4 Application of Next Generation Sequencing in Clinical Microbiology

Public health clinical and microbiology laboratories are increasingly integrating next-generation sequencing (NGS) technology into their workflows and current diagnostic cycles due to rapid advances in NGS technologies. Microbiology laboratories in clinical and public health settings play a critical role in mitigating the impact of infectious diseases. Their main function is to identify and classify pathogens present in infected individuals or in the community. The use of next-generation sequencing has the potential to support clinical and public health decision-making by identifying the etiological agent of infectious diseases as well as the epidemiology and evolution of various pathogens in both hospitals and the community (Sintchenko & Holmes, 2015). Next-generation sequencing (NGS) is increasingly recognized as a preferred method in bacteriology due to its numerous advantages in identifying and classifying pathogens present in infected individuals, multilocus sequence typing (MLST), serotyping and virulence factor profiling as well Genes for antimicrobial resistance and outbreaks provides investigation.

1.4.1 Multilocus Sequence Typing (MLST)

The field of molecular epidemiology uses sequence typing of bacterial infections to track the spread of bacteria and identify outbreaks. Sanger sequencing, a method of DNA sequencing, was used in early gene-based approaches to typing bacterial isolates to sequence amplicons from a small set of loci or genetic regions. These approaches, including multilocus sequence typing (MLST), can be used to identify the genetic makeup of a bacterial isolate and compare it to other isolates. This information can be used to track the spread of bacteria and identify outbreaks (Maiden et al., 1998). Next-generation sequencing (NGS) has made it more practical to sequence complete bacterial genomes for typing purposes rather than relying on small numbers of amplicons (Hyytiä-Trees et al., 2007; Jackson et al., 2015; Maiden et al., 2013). Even higher resolution typing is possible with whole genome-based typing approaches such as Core Genome

MLST (cgMLST), Whole Genome MLST (wgMLST) and Average Nucleotide Identity (ANI) (Cody et al., 2013; Jolley et al., 2012). Although next-generation sequencing (NGS) is now available, determining sequence type (ST) is often still the first step in the analysis of bacterial isolates. Researchers and epidemiologists continue to rely on traditional MLST systems due to the large amount of legacy ST information collected from numerous surveys over the years (Goris et al., 2007; Katz et al., 2009).

It requires significant time and computational power to type bacterial isolates based on their genes, using current approaches to examine complete genome sequence data. The use of stringMLST was introduced, a method for rapid characterization of bacterial isolates from genome sequence data using k-mers. The advantages of stringMLST lie in the fact that it requires no assembly or alignment, requires very little memory, has a small code base, and can be installed quickly and easily (Katz et al., 2009). It can be used with either pre-existing MLST schemes or user-created, customized typing schemes, including those with scores (rMLST) or hundreds (cgMLST) of loci for larger scale analysis. The performance and correctness of stringMLST was tested by running it on a huge collection of sequence reads from entire bacterial genomes along with the corresponding ST information (Desoubeaux et al., 2016).

1.4.2 Serotyping

For more than 50 years, serotyping has been the backbone of *E. coli* infection surveillance in the public health system. Researchers can now identify epidemics and further subdivide each serotype using DNA testing. The latest generation of sequencing technology is an advance as it allows the laboratory to determine the bacterial species, serovar and subtype with a single analysis (Fratamico et al., 2016).

1.4.3 Antimicrobial Resistance and Virulence Factors

The identification and characterization of virulence factors, especially toxins, and antibiotic resistance markers of pathogens is crucial for the study of bacterial pathogenesis and their interactions with the host, as well as for the development of new treatments, vaccines and molecular diagnostic tools (Klein & Hultgren, 2020)

Furthermore, knowing which markers indicate virulence or resistance can improve outbreak control and therapeutic intervention. Multi-resistant (MDR) bacteria have become increasingly common in recent years and pose a threat to public health worldwide. All common medications are ineffective against multi-resistant bacteria (MDR). These include “last resort” antibiotics such as imipenem and colistin (Exner et al., 2017). Infections can be caused by a variety of MDR bacteria, including methicillin-resistant Staphylococcus aureus (MRSA), Enterobacteriaceae, Acinetobacter, and Pseudomonas species (Reffuveille et al., 2014). Deciphering the resistome of a bacterium has become necessary to understand resistance pathways, predict resistance phenotype, enable effective infection control, improve antibiotic therapy and patient care, develop polymerase chain reaction (PCR) tests to diagnose resistance-inducing genes or mutations, and to identify targets for novel drugs and prior to the use of bacteria that can be used as probiotics (Sintchenko & Holmes, 2015).

Bacteria can develop resistance to antibiotics through various mechanisms, including production of antibiotic-modifying enzymes, structural changes to the antibiotic target that prevent drug binding, membrane impermeability, and overexpression of efflux systems (Darby et al., 2022). Genome sequencing can be used to determine which MGEs propagate acquired resistance mechanisms such as plasmids, transposons and integrons (Fratamico et al., 2016).

Genome sequencing, or the systematic molecular sampling of a set of genes, led to the discovery that molecular cloning and/or mutagenesis play a role in disease (Waterston et al., 2002), but previously the detection of virulence factors relied on biochemical approaches. Pathogenic effects of extracted toxins or other virulence factors from bacteria are studied in vivo or in vitro (Devi et al., 2016). Molecular methods include the study of virulence genes through mutagenesis, cloning and expression in non-pathogenic strains (often *E. coli*) (Hendriksen et al., 2019). Over the past two decades, genomics and functional analyzes (transcriptomics and proteomics) have significantly increased the frequency with which virulence factors are identified. Homology searches for known virulence genes, matching strains with differential virulence, and analyzes of horizontally acquired genes can all be used to categorize bacterial virulence components in genomes (Bakour et al., 2016).

1.4.4 Outbreaks Investigation

Next-generation sequencing (NGS) technologies have revolutionized our approach to understand the infectious disease epidemics, which allows a comprehensive genetic investigation (Berry et al., 2020). The transition from pathogen-specific genome sequencing methods to more universal protocols represent a significant advancement in the field (Frey & Bishop-Lilly, 2015). This evolution has remarkably enhanced our understanding of the disease transmission dynamics, and extending it from hospital settings to broader communities. However, the integration of NGS into the routine clinical microbiology practice has several challenges, which include the absence of FDA-approved protocols, that requires systematic laboratory validation processes, financial limitations linked to non-reimbursable tests, and the need for specialized training in complex data analysis techniques (Cepas & Soto, 2020).

1.5 Bioinformatics Tools

Next Generation Sequencing (NGS) technologies provide rapid and reliable high-throughput methods for detecting the precise order of nucleotides contained in DNA and RNA molecules (Liu et al., 2012). al Since the development of more advanced sequencing methods, the identification of nucleic acid sequences has become a tool that is indispensable in almost all areas of biological science. As a direct result, the branch of science known as bioinformatics is at the forefront of both the interpretation and use of these biological data sets. Biological information at the molecular, cellular and genomic levels can be organized, analyzed and interpreted using tools developed in the field of bioinformatics. These tools use mathematical and statistical approaches implemented by a variety of programming languages. In diagnosis, medical treatment and epidemiological research (Blekherman et al., 2011). In this project, we relied on three main steps for the molecular characterization of the clinical *E. coli* samples: quality control, assembly and profiling.

1.5.1 Quality Control

One of the challenges of NGS is that the data can be noisy and difficult to interpret. This is because NGS generates millions of short reads and each read can

contain errors. Therefore, it is important to perform quality control (QC) on NGS data before it can be analyzed (Zhou et al., 2014). QC is the process of assessing the quality of NGS data and identifying potential errors. This information can then be used to filter out low-quality data and improve the accuracy of downstream analysis. A number of QC tools are available for bacterial NGS analysis. These tools typically assess a variety of parameters, including read length, base quality, GC content, and overrepresented sequences (Guo et al., 2014). Some tools also offer advanced features, such as: B. the ability to identify and remove contaminants, like Cutadapt. QC for NGS data typically involves a number of steps, including:

- Assessing the quality of the raw reads
- Trimming low-quality reads
- Removing contaminants
- Aligning the reads to a reference genome
- Calculating metrics to assess the quality of the assembly

A number of bioinformatics tools are available to perform quality control on NGS data. These tools vary in their capabilities and performance, and the best tool for a particular project depends on the researcher's specific needs. The best way to choose a QC bioinformatics tool is to consider the specific needs of the project. If the research is only interested in assessing the quality of the raw data, a tool like FastQC might be a good choice. If interested in reducing low-quality readings and removing contaminants, a tool like MultiQC may be a better choice. If it is needed to align reads to a reference genome and calculate metrics to assess the quality of the assembly, then a tool like Pilon may be a good choice. It is also important to consider the features of different tools. Some tools have features not available in others, such as the ability to align reads to a reference genome or calculate metrics to assess the quality of the assembly.

Table 2: A Comparative Analysis of Quality Control Bioinformatics Tools for NGS Data

Tool	Capabilities	Performance	Features	References
FastQC	Can assess the quality of raw reads, trim low-quality reads, and remove contaminants.	Fast	Can be used with a variety of NGS platforms	(Guo et al., 2014).
MultiQC	Can provide a graphical overview of the results of multiple QC tools	Fast	Can be used with a variety of NGS platforms	(Ewels et al., 2016)
Pilon	Can align reads to a reference genome and calculate metrics to assess the quality of the assembly	Slow	Can be used with a variety of NGS platforms	(Walker et al., 2014)
BWA	Can align reads to a reference genome	Fast	Can be used with a variety of NGS platforms	(Li & Durbin, 2009)
LAST	Can align reads to a reference genome	Fast	Can be used with a variety of NGS platforms	(Altschul et al., 1990)

1.5.2 Genome Assembly

Assembly bioinformatics tools play a critical role in analyzing NGS data, enabling the reconstruction of complete or near-complete genomes from short sequence reads (Wee et al., 2019). In this section, there will be evaluation of several widely used assembly bioinformatics tools, including SPAdes, Skesa, Velvet, IDBA-UD, and MEGAHIT, highlighting their key features, algorithms, and performance metrics. The underlying methods, advantages, limitations and recommended use cases will be addressed.

The assembly tools are instrumental in converting short sequence reads obtained from NGS platforms into longer contiguous sequences, known as contigs, which can then be further analyzed for genome annotation, comparative genomics, and other downstream analyses (Wee et al., 2019).
These tools use a variety of algorithms to assemble short DNA reads into a complete genome. Some of the most common algorithms used for assembly include:

- De Bruijn graph assembly
- Iterative assembly
- Hierarchical assembly

De Bruijn graph assembly is an assembly method that uses a graph to represent the relationships between short reads. The graph is created by connecting reads that share common sequences. Once the graph is created, it can be traversed to find the complete genome (Compeau et al., 2011).

Iterative assembly is an assembly method that uses a series of steps to assemble the genome. In the first step, the reads are aligned to a reference genome. In the second step, the reads that do not match the reference genome are merged. In the third step, the assembled sequences are aligned again with the reference genome. This process is repeated until the genome is complete (Hitch & Creevey, 2018).

Hierarchical assembly is an assembly method in which the genome is assembled at multiple levels. At the first level, the reads are assembled into contigs, which are short sequences that are likely part of the same gene. At the second level, the contigs are

assembled into scaffolds, which are longer sequences that are likely part of the same chromosome. At the third level, the scaffolds are assembled into the complete genome (Miller et al., 2010).

Table 3: compares some of the most popular assembly bioinformatics tools. The table provides information about the tools' features, performance, and capabilities.

Table 3: Evaluating the Effectiveness of Different Genome Assembly Software Tools

Tool	**Capabilities**	**Performance**	**Features**	**References**
ABySS	Can assemble DNA sequences from short reads, even with high error rates	Fast	Can assemble large genomes	(Simpson et al., 2009)
CLC Genomics Workbench	Can assemble DNA sequences from short reads, and can also perform other bioinformatics tasks such as annotation and visualization	Slow	Can assemble large genomes, and has a user-friendly interface	(Huszar et al., 2021)
Flye	Can assemble DNA sequences from short reads, and is particularly good at assembling large genomes	Fast	Can assemble large genomes, and is open source	(Wick & Holt, 2019)
IDBA-UD	Can assemble DNA sequences from short reads, and is particularly good at assembling genomes with repetitive sequences	Slow	Can assemble large genomes, and is open source	(Peng et al., 2012)

Table 3: Evaluating the Effectiveness of Different Genome Assembly Software Tools (continued)

Tool	Capabilities	Performance	Features	References
BLAST	Can assemble DNA sequences from short reads, and is particularly good at assembling genomes with low error rates	Fast	Can assemble small genomes, and is open source	(Altschul et al., 1990)
MEGAHIT	Can assemble DNA sequences from short reads, and is particularly good at assembling genomes with high error rates	Fast	Can assemble large genomes, and is open source	(Li et al., 2015)
Mr. Bayes	Can assemble DNA sequences from short reads, and uses a Bayesian approach to assembly	Slow	Can assemble large genomes, but is not as fast as some other tools	(Huelsenbeck & Ronquist, 2001)
Oases	Can assemble DNA sequences from short reads, and uses a de Bruijn graph approach to assembly	Slow	Can assemble large genomes, but is not as fast as some other tools	(Schulz et al., 2012)

Table 3: Evaluating the Effectiveness of Different Genome Assembly Software Tools (continued)

Tool	Capabilities	Performance	Features	References
PEAR	Can assemble DNA sequences from short reads, and is particularly good at assembling genomes with repetitive sequences	Slow	Can assemble large genomes, and is open source	(Zhang et al., 2014)
SPAdes	Can assemble DNA sequences from short reads, and is particularly good at assembling genomes with low error rates	Fast	Can assemble large genomes, and is open source	(Bankevich et al., 2012)

There are a number of factors to consider when selecting an assembly bioinformatics tool for NGS bacterial analysis (Almeida & De Martinis, 2019). These factors include:

The size of the genome: The genome size significantly impacts the choice of the assembly tools. The tools differ in their ability and suitability to assemble small versus large genomes. In general, small genomes, having fewer repetitive sequences, are easier to assemble when compared to large genomes where these repeats may have greater assembly challenges.

The quality of the reads: The read quality is a crucial factor in selecting an assembly tool. The assembly tools differ in their sensitivity to read errors, which can lead to a complication during the genome assembly. Errors may cause mis-assembly, leading to inaccurate genome assembly outcomes.

The desired output: When selecting an assembly tool, it is important to know the desired output. Although some tools can generate a single assembly, others can produce

multiple assemblies. In general, a single assembly is usually ideal for a comprehensive genome overview. Nevertheless, multiple assemblies can be as well beneficial in identifying regions of the genome that are challenging to assemble.

The computational resources available: When choosing an assembly tool, a researched needs to consider the computational resources available in a particular setting. Some of the tools require more computational power than others, making it essential to select a tool that matches with your specific needs.

1.5.3 Species Identification and Closest Reference Detection

There are several bioinformatics tools available for NGS bacterial identification and closest reference detection. Some of the most popular tools include (Ma & Zhang, 2010):

The popular web-based tool Bactinspector: It utilizes methods like BLAST, Kraken, and Centrifuge for bacterial identification from NGS data. While it is very effective, this web-based tool may execute very slowly with large datasets.

The algorithm of Bactinspector works by dividing the NGS data into overlapping windows. Then, it generates a profile for each window, which represents the nucleotide distribution. Afterwards, these profiles are compared against known organisms in a database. The identification is based on the organism with the closest matching profile.

Bactinspector is highly effective and more accurate than other tools which relies only on BLAST or Kraken for identifying bacteria across various organisms. However, it is slower in detecting contamination, and it can take several hours to analyze large datasets. Regardless, it still remains a user-friendly and accessible tool, where the upload of NGS data and the result retrieval processes are very easy. In addition, Bactinspector is free to use for contamination detection.

Bactinspector uses different tools for bacterial identification and reference genome detection:

BLAST: A tool for DNA sequence comparison, useful to identify the closest reference genome (Ma & Zhang, 2010).

Centrifuge: This tool uses different algorithms for bacterial identification at the species level and able to find the nearest reference genome (Li et al., 2018).

Kraken: It applies a probabilistic method for bacterial species-level identification (Lu et al., 2022).

MetagenomicsDB: A database of NGS bacterial genomes which helps in reference genome identification (Bragg & Tyson, 2014).

SILVA: it is very similar to MetagenomicsDB, a database for identifying the closest reference genome (Quast et al., 2013).

All these tools work by comparing the DNA sequence of a bacterium to those available in all databases, in order to identify the most similar sequence. Every tool has its own advantages: BLAST is fast but may be less accurate, Centrifuge is accurate but slower, and Kraken combines both speed and accuracy but it requires large genome database. MetagenomicsDB and SILVA are valuable databases for reference genome identification.

1.5.4 Contamination Detector Tools

NGS data are prone to contamination with DNA from other organisms, potentially can lead to false positives in the identification and reference detection. To reduce and eliminate this possibility, it is crucial to use contamination detection bioinformatics tools to identify and eliminate such possible contamination. There are several popular tools for this purpose which include Confinder, Kraken 1, and Kraken 2, each one has its unique strengths and weaknesses. It is really important to select the most suitable tool which match the specific needs (Lu et al., 2022).

1.5.4.1 Confinder

It is a contamination detection tool for NGS data, it uses a probabilistic approach which is suitable for a wide range of organisms. The algorithm works by dividing the NGS data into overlapping windows, then creating a nucleotide distribution profile for each, followed by comparing these profiles with the known organisms in a database. The organism whose profile closely matches is identified as the probable contaminant.

1.5.4.2 Kraken 1

It is another contamination detection tool for NGS data, which utilize a probabilistic method known to be effective against a large database of known organisms. The process starts with segmenting the NGS data into overlapping windows, then creating a k-mer signature for each window, followed by comparing these signatures with those of known organisms in a database. The organism with the most closely matching k-mer signatures is identified as the likely contaminant.

1.5.4.3 Kraken 2

Kraken 2 is an updated version of the Kraken 1, which offers an enhanced accuracy and capabilities in detecting contamination from a wider range of organisms in NGS data. The work process is similar to that of the Kraken 1. This updated version makes Kraken 2 a more effective tool to ensure the purity and reliability of NGS data analysis.

Overall, the three tools are considered potential in detecting the contamination from the NGS data. While, Confinder excels with the accuracy across a diverse range of organisms. Both the Kraken 1 and Kraken 2 are reliable since they are using an extensive database of known organisms. In addition, Kraken 2 exceeds Kraken 1 in its accuracy and it can identify contaminants from an even wider variety of organisms. The choice between these tools heavily depends on the user's specific requirements.

1.6 The Global Crisis of Antibiotic Resistance

The discovery of antibiotics in the 20th century served as a medical revolution, which significantly reduce the rates of infectious disease and saved lives (Mohr, 2016). These drugs allowed the effective treatment of previously deadly diseases, which lead to longer lifespans and better public health. However, the years of overuse and misuse of antibiotics have led to an increased resistance in all the bacterial species, in this part, we will focus particularly on *E. coli* (English & Gaur, 2010). *E. coli*, for instance, has developed resistance to carbapenem antibiotics, which is often used as a last resort, limiting the possible treatment options and increasing the risk of severe infections (Liang et al., 2018). This alarming trend highlights the urgency for responsible antibiotic usage

and strong research into finding new drugs and treatment alternatives, to prevent a potential relapse to an era where even minor infections posed significant threats.

The problem of antibiotic resistance is international and the global use and misuse of antibiotics has reduced the number of effective medicines. More than 50% of bacteria in southern Europe are resistant to antibiotics; For the USA this figure is 25% and for Kenya it is 60%, although regional comparisons such as this are difficult. The problem has become so serious that resistance to major pathogens in the hospitals of many developed countries is so high that effective treatment is virtually impossible. In many developing countries, the cost of antibiotics means that the cheapest drug is usually the first choice. This often leads to the choice of a broad-spectrum antibiotic, and uncontrolled use of these antibiotics is common. Impacts in individually wealthy countries are reflected in global consumption, and a dramatic example of a lower middle income country that has seen a dramatic rise in resistance is India (Mancuso et al., 2021; Tiri et al., 2020; Urban-Chmiel et al., 2022).

1.7 Enterobacteriaceae

Enterobacteriaceae, belonging to the order of Enterobacter, is a large and diverse group of rod-shaped, Gram-negative, facultative anaerobic bacteria. This family includes different genera such as *Escherichia*, *Citrobacter*, *Enterobacter*, *Klebsiella*, *Salmonella*, and *Shigella*. They are all characterized by being catalase-positive and oxidase-negative, these bacteria are abundant, found in the intestinal flora of most animals, including humans, and are also present in water, soil, and plants (Janda & Abbott, 2021). While some members play a crucial role in gut health by aiding digestion and nutrient absorption, others can become opportunistic pathogens causing infections ranging from urinary tract issues to severe bloodstream infections. The adaptability of Enterobacteriaceae, especially in acquiring antibiotic resistance, poses significant challenges in healthcare settings, emphasizing the importance of responsible antibiotic use and continuous research to manage these ever-evolving threats effectively. Studying this family is important for understanding their complex biology and developing strategies to balance their beneficial and harmful impacts (Bennett et al., 2010; Hong Nhung et al., 2007).

1.8 Pathogenic Strains of *Escherichia Coli*

Escherichia coli was discovered by Theodor Escherich in 1885, *E. coli* is a Gram-negative, rod-shaped, facultative anaerobic bacterium, predominantly found in the human and animal intestinal tract (Escherich, 1988). Despite the low abundance in the healthy human microbiome (around 0.1%), *E. coli* is one of the first species to colonize infants' digestive systems and is present in over of 90% of the adult population (Huttenhower et al., 2012); Tenaillon et al., 2010; (Secher et al., 2016). While most strains of *E. coli* are harmless, contributing to gut health by producing B-complex vitamins and vitamin K2, nine pathogenic strains have been documented to cause illnesses, particularly in individuals with weakened immune systems or compromised gastrointestinal barriers (Mellies et al., 2001; Bentley & Meganathan, 1982).

These pathogenic strains, can cause diseases which range from diarrhea to extraintestinal illnesses like hemolytic uremic syndrome, include Enteroaggregative *E. coli* (EAEC), Enterohaemorrhagic *E. coli* (EHEC), Enteroinvasive *E. coli* (EIEC), Enteropathogenic *E. coli* (EPEC), Diffusely adherent *E. coli* (DAEC), Enterotoxigenic *E. coli* (ETEC) and Adherent-Invasive *E. coli* (AIEC). In particular, EHEC has been linked to severe outbreaks as a foodborne pathogen (Yang et al., 2017). The pathogenicity of these *E. coli* strains is because of a range of virulence factors encoded by genetic variants, enabling them to exhibit diverse behaviors such as adhesion, invasion, and toxin production (Wu et al., 2008). Each *E. coli* pathotype possesses distinct pathogenicity mechanisms, as well as a distinct profile of virulence components that are encoded by distinct gene clusters. An array of behaviors, including adhesion, invasion, attachment, iron uptake, motility, and toxin activity, may be encoded by genes linked to pathogenicity. These virulence factors, usually encoded on chromosomes or mobile genetic elements like plasmids, which makes *E. coli* a significant concern in public health, underlining the need for continuous monitoring and research into this adaptable and potentially dangerous bacterium.

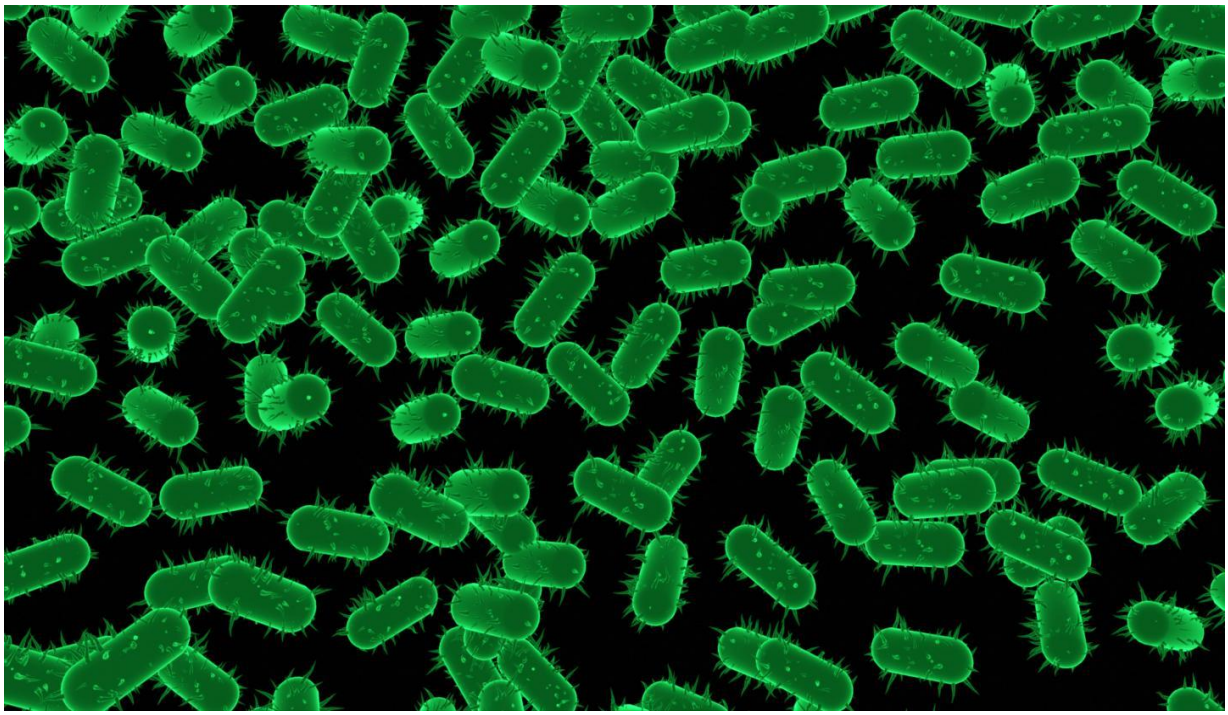


Figure 1: Pathogenic *E. coli* Strains Emerge by Acquiring Genetic material.

The pathogenicity mechanisms of *E. coli* have been extensively studied due to their role in intestinal and extraintestinal diseases. For example, enteroaggregative *E. coli* (EAEC) is known to cause diarrhea and is common in AIDS patients (Peng et al., 2024). Its pathogenesis involves attachment to intestinal cells, formation of biofilms and secretion of toxins, regulated by AggR factor. Likewise, enterohemorrhagic *E. coli* (EHEC) cause severe symptoms such as hemorrhagic colitis and have been linked to various outbreaks (Sharma et al., 2020). Its virulence is largely due to Shiga-like toxins (SLTs), which cause cellular damage and can lead to diseases such as hemolytic uremic syndrome (Sharma et al., 2020). Enterotoxin *E. coli* (ETEC), a major diarrhea pathogen in developing countries, attaches to intestinal cells and releases toxins that lead to fluid loss and diarrhea. ETEC's enterotoxins, both heat-labile and heat-stable, disrupt normal cellular functions and result in secretory diarrhea (Stephen, 2001). Enteroinvasive *E. coli* (EIEC), similar to Shigella, cause dysentery-like symptoms. It invades and multiplies in intestinal cells, which is facilitated by a large virulence plasmid encoding various proteins essential for its pathogenicity (Ud-Din & Wahid, 2014). Diffuse adherent *E. coli* (DAEC) is associated with diarrhea in young children and various intestinal diseases in adults. Its pathogenesis involves adhesion to host cells mediated by specific adhesins,

and it can secrete toxins that contribute to inflammation and tissue damage (Segura & Garcia, 2016). Adherent invasive *E. coli* (AIEC), often associated with inflammatory bowel diseases such as Crohn's disease, attaches to and invades intestinal epithelial cells. It lacks specific virulence factors found in other *E. coli* pathotypes but can still cause significant inflammation and damage to the intestinal mucosa. In this chapter, the discussion will be exclusively about the beta-lactam antibiotics (Palmela et al., 2018).

1.9 Antibiotics: Mechanism and Resistance

Antibiotic resistance in bacteria can be understood in two ways: firstly, through laboratory analysis known as "microbiological resistance" and secondly through observing how bacteria respond to treatment in actual patients, referred to as "clinical resistance." The laboratory method identifies specific resistance genes, classifying bacteria as resistant or not based on specific criteria. On the other hand, the clinical approach focuses on the effectiveness of antibiotics in treating infections in patients, considering the real-world outcomes rather than just relying on lab findings (MacGowan & Macnaughton, 2017).

The discovery of antibiotics in the 20th century revolutionized medicine by drastically reducing deaths from bacterial infections. These drugs, once hailed as "miracle drugs," have saved countless lives. However, over time, some bacteria have evolved into forms highly resistant to most antibiotics, known as multi-drug resistant (MDR) and extensively drug-resistant (XDR) Enterobacteriaceae, or “superbugs” (Venezia et al., 2017). These resilient bacteria present a significant challenge in treatment, highlighting the urgent need for new strategies in battling resistant infections.

The rise of MDR and XDR Enterobacteriaceae poses a severe threat to public health, often leading to more prolonged hospital stays, increased medical costs, and higher mortality rates. Key factors contributing to this problem include the overuse and misuse of antibiotics, which accelerates the development of resistance (English & Gaur, 2010), the transmission of resistant bacteria, particularly in healthcare environments (Rao, 1998), and a slowdown in the development of new antibiotics, complicating the treatment of infections caused by these resistant bacteria (Falagas et al., 2008).

1.10 Classes of Antibiotics and Their Mechanism of Action

Antibiotics, derived from microorganisms, function to hinder the growth of competing microbes in the same environment. These substances have been adapted for medical use, often undergoing modifications or even total synthesis, as seen with drugs like chloramphenicol. Depending on factors like the type of bacteria, the specific antibiotic, and sometimes its concentration, antibacterial medications can act as either bactericidal (killing bacteria) or bacteriostatic (inhibiting bacterial growth) (Aminov, 2010). Additionally, antibiotics are categorized based on their mechanism of action, as detailed in Table 4 (Murray et al., 2018; Kaufman, 2011).

Table 4: Main mode of action of antibiotics (Kaufman, 2011)

Mode of action	Examples
Cell wall synthesis inhibition	ß-Lactams, Glycopeptides, Fosfomycin
Protein synthesis inhibition	Aminoglycosides, Macrolides, Tetracyclines, Chloramphenicol
Interference with nucleic acid synthesis	Quinolones, Fluoroquinolones, Rifampin
Disruption of Cell membrane	Polymyxins
Antimetabolites	Sulfonamides, Trimethoprim

1.11 Beta-lactam Antibiotics

Beta-lactam antibiotics are categorized into four main groups: penicillins, cephalosporins, monobactams, and carbapenems. These antibiotics are named for their defining chemical feature, the beta-lactam ring, as depicted in Figure 2.

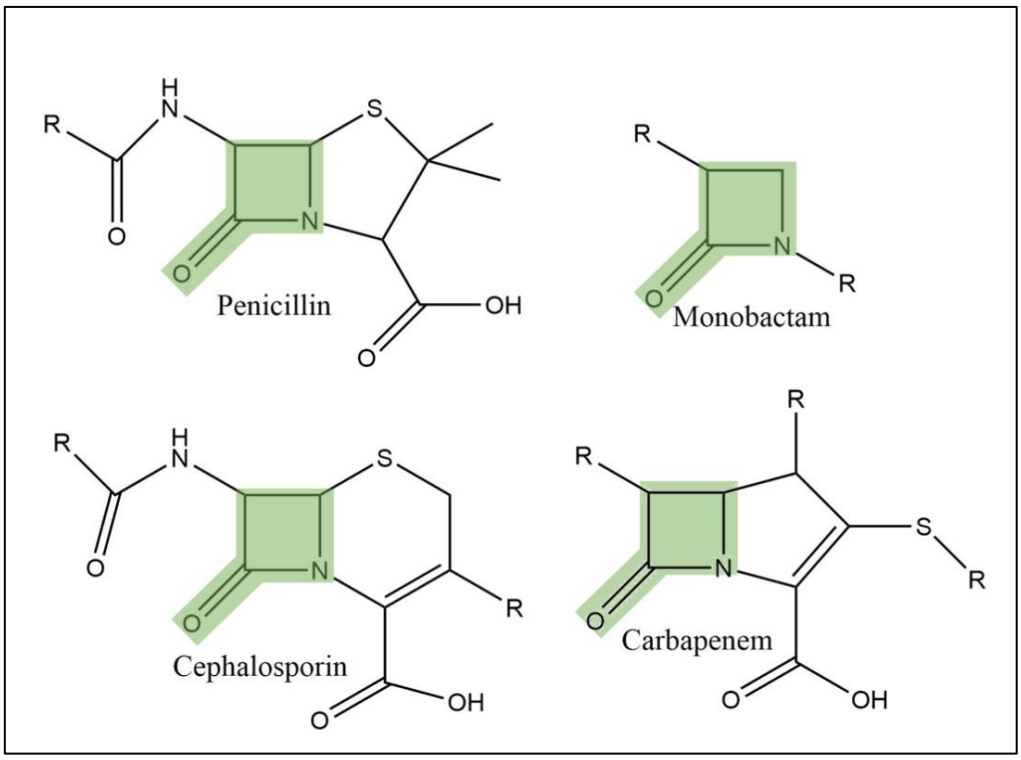


Figure 2: The Major Groups of Beta-lactam Antibiotics, The green Shade Shows the Beta-lactam Structure

The initial beta-lactam antibiotic, later termed penicillin G, was identified by Alexander Fleming in 1928, though its widespread clinical application began in the 1940s (Banerjee et al., 2010). Penicillins are categorized into several types: i) natural penicillins, like penicillin G; ii) penicillinase-resistant penicillins, such as methicillin; iii) aminopenicillins, for instance, ampicillin; iv) carboxypenicillins, including ticarcillin; and v) ureidopenicillins, like piperacillin, with the last two types also being effective against Pseudomonas aeruginosa (Banerjee et al., 2010).

Cephalosporins, broad-spectrum antibiotics, were initially found as a compound from the fungus Cephalosporium acremonium. They are grouped into five generations based on their antimicrobial properties, with newer generations generally offering better efficacy against Gram-negative bacteria but sometimes reduced activity against Gram-positive organisms. The third and fourth generations are known as extended-spectrum cephalosporins, and those from the fifth generation are effective against methicillin-resistant Staphylococcus aureus (MRSA) (Craig et al., 2010).

Aztreonam, the sole monobactam in clinical use, specifically targets aerobic Gram-negative bacteria. While susceptible to hydrolysis by various beta-lactamases, it notably resists hydrolysis by metallo-beta-lactamases (MBL) (Craig et al., 2010).

Carbapenems are known for their resistance to many hydrolytic enzymes and have the widest spectrum of antibacterial activity among beta-lactam drugs, effective against almost all Gram-negative and Gram-positive bacteria. Common carbapenems include imipenem, meropenem, doripenem, and ertapenem (Wallace et al., 2011).

Beta-lactam antibiotics disrupt bacterial cell wall peptidoglycan synthesis. They bind to and inhibit the function of penicillin-binding proteins (PBPs), which include transpeptidases, transglycosylases, and carboxypeptidases. PBPs are crucial for cross-linking peptides between polymer chains of N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM), forming a robust peptidoglycan mesh. The interruption of this process by beta-lactam antibiotics triggers cell wall autolytic hydrolases, leading to bacterial cell death (Craig et al., 2010, Nordmann et al., 2012).

1.12 Mechanism of Antibiotic Resistance in *E. coli*

1.12.1 Reducing Antibiotic Permeability

Pathogenic *E. coli* strains uses different strategies to diminish their permeability to antibiotics, essentially building barriers to keep these drugs out (Delcour, 2009). These strategies include:

1.12.1.1 Lipopolysaccharide (LPS) Modification

The outermost layer of the *E. coli* cell wall, LPS, provides structural integrity and protection (Maldonado et al., 2016). *E. coli* modifies its LPS by adding sugars to the O-antigen, making the layer denser and reducing antibiotic entry points (Maldonado et al., 2016). This alteration lessens the permeability to various antibiotics, including fluoroquinolones and β-lactams (Maldonado et al., 2016).

1.12.1.2 Porin Mutations

Porins are protein channels that regulate molecule passage, including antibiotics (Schmid et al., 1998). *E. coli* mutations can narrow these pores or reduce their abundance,

hindering antibiotic entry (Schmid et al., 1998). Such mutations in porins like OmpF and OmpC can lead to resistance, especially against cephalosporins (Schmid et al., 1998).

1.12.1.3 Capsule Formation

Some *E. coli* can form a polysaccharide capsule, adding another barrier (Cross et al., 1988). This capsule limits the diffusion of antibiotics towards the cell (Cross et al., 1988). Capsules contribute to resistance, particularly against cephalosporins and fluoroquinolones.

1.12.2 Antibiotic Inactivation

E. coli can deactivate antibiotics before they cause harm, using several mechanisms.

1.12.2.1 Enzymatic Breakdown

The bacteria produce enzymes to dismantle antibiotic structures (Sykes & Matthew, 1976). β-lactamases target β-lactams by breaking down their essential structure, conferring wide resistance (Wong-Beringer, 2001). Beta-lactamase classes A, B, and D include enzymes known as carbapenemases, which possess significant carbapenamase activity. Generally, these enzymes are effective against a broad range of beta-lactam antibiotics, with a few exceptions (Queenan & Bush, 2007). The characteristics of these enzymes are concisely outlined in Table 5, adapted from (Logan & Weinstein, 2017).

Table 5: The main features of the different β-lactamases (Logan & Weinstein, 2017)

Molecular class	Active site	Enzyme type	Examples	Inhibitors
A	Serine	Penicillinases Broad Spectrum	TEM1 SHV1	Generally susceptible to most beta-lactamase inhibitors, making combinations with these inhibitors a potential therapeutic approach.
		Extended Spectrum ß-lactamases (ESBL)	TEM-derived, CTX- derived	
		Carbapenemases	KPC-2 KPC-3	
B	Zn^{2+}	Metallo-β-lactamases	IMP, VIM, GIM, NDM	Their activity can be inhibited by chelating agents like EDTA.
C	Serine	Cephalosporinases	AmpC-type	Their activity can be inhibited by cloxacillin and monobactam
D	Serine	Oxacillinases Broad spectrum	OXA-family in *P. aeruginosa*	Their activity can be influenced by the concentration of sodium chloride (NaCl)
		Extended spectrum	OXA-derived in *P. aeruginosa*	
		Carbapenemases	OXA-derived in *Acinetobacter*	

Metallo-β-lactamases (MβLs, Ambler class B), are enzymes which depend on zinc ions, and play a critical role in antibiotic resistance. These enzymes hydrolyze the β-lactam ring in antibiotics, making them inactive (Nordmann & Poirel, 2002; Bush & Johnson, 1998). One of the key MBL types in *E. coli* include New Delhi metallo-β-lactamase (NDM), these enzymes can hydrolyze most β-lactams except aztreonam and they are not inhibited by commercial β-lactamase inhibitors, instead they are susceptible to EDTA and dipicolinic acid. Their spread represents a global healthcare crisis, with pan-resistant bacteria emerging (Nordmann & Poirel, 2002; Maltezou, 2009). Since 2009, NDM-type MBLs have gained international attention, when the bla_{NDM-1} was discovered in a patient treated in India. They are characterized by rapid gene transfer and widespread dissemination, today there are more than 40 NDM enzyme variants in (Figure 3) mostly originating in Asia. This type of bacteria harboring NDM, usually

carry multiple resistance mechanisms, maintaining susceptibility to only a few antibiotics like colistin and Fosfomycin (Maltezou, 2009; Nordmann & Poirel, 2014; Poirel et al., 2010; Rana et al., 2023; Yong et al., 2009).

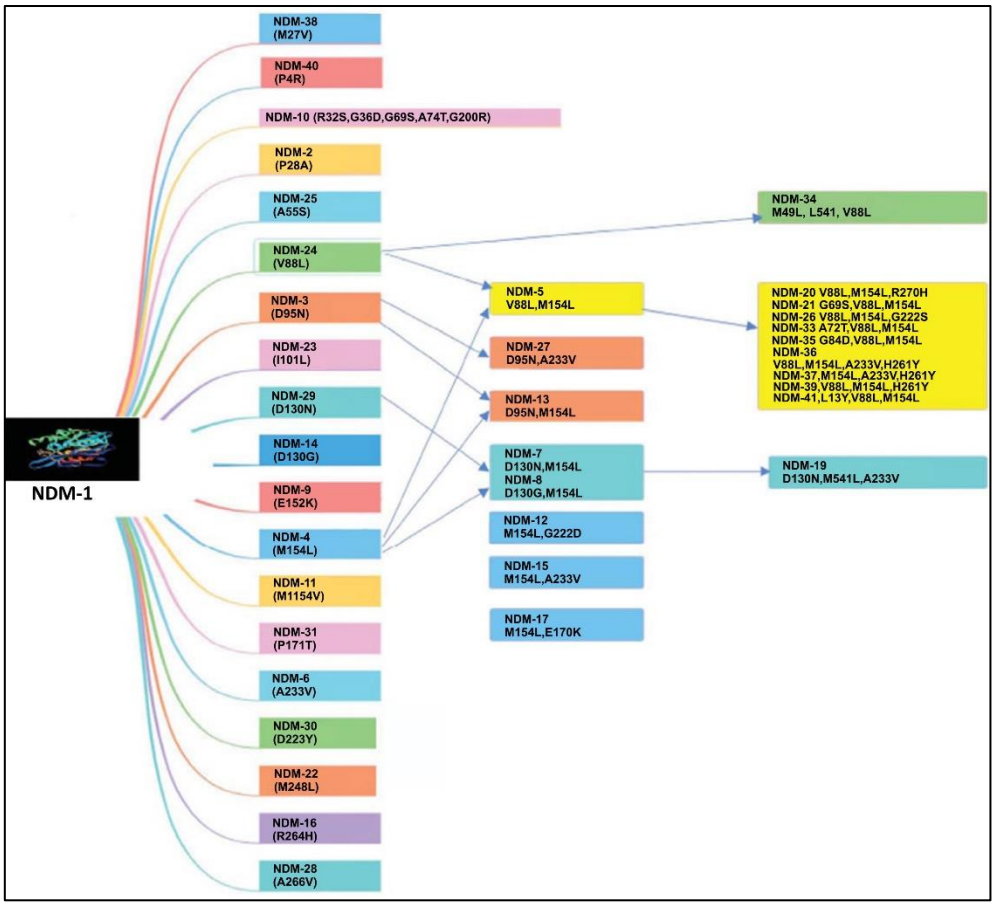


Figure 3: New Delhi metallo-beta-lactamase (NDM) variants.

The spread of NDM-producing *E. coli* is not clone-specific but is usually mediated by various plasmids. There are several epidemic clones which have been identified carrying NDM-type MBL genes along with other resistance determinants. NDM producers are now commonly found in Europe, USA and Canada. NDM-producing *E. coli* are becoming more prevalent in Africa and they are endemic in the Middle East contributing significantly to carbapenem resistance (Jamal et al., 2016; Sonnevend et al., 2015; Yezli et al., 2015)

1.12.2.2 Efflux Pumps

Efflux pumps actively remove antibiotics from the cell, lowering their intracellular concentration (Webber & Piddock, 2003). These pumps can expel a variety of antibiotics, contributing to multidrug resistance (Webber & Piddock, 2003).

1.12.3 Target Modification

E. coli can alter its antibiotic target sites, effectively making the drugs ineffective.

1.12.3.1 Modifying Key Proteins

Changes in crucial protein targets, like penicillin-binding proteins and ribosomes, impede antibiotic binding (Schaenzer & Wright, 2020). Gene mutations lead to structural changes in these proteins, preventing effective antibiotic action (Schaenzer & Wright, 2020).

1.12.3.2 Mimicry and Decoys

E. coli produces decoy molecules, resembling target sites, to bind antibiotics and block their action (Kronenberger et al., 2019). Certain strains produce enzymes that mimic PBPs, latching onto β-lactams and protecting actual target proteins (Wong-Beringer, 2001). These sophisticated strategies allow *E. coli* to resist various antibiotics, presenting a challenging scenario in medical treatment and public health.

1.13 *E. coli* Lineages Associated with Carbapenem Resistance

Carbapenem-resistant *Escherichia coli* (*E. coli*) lineages have been the focus of several studies utilizing whole genome sequencing to identify the most common lineages and characterize their virulence factors (Devi et al., 2018). These studies have revealed a diverse range of *E. coli* lineages, with 24 distinct sequence types (STs) identified, including clinically important STs such as ST131, ST69, ST95, and ST73 (Devi et al., 2018). The genomic characterization of *E. coli* lineages has also highlighted shared genetic traits that may be linked to the emergence of carbapenem resistance plasmids (Meunier et al., 2017). Specific evolutionary events have been identified in the emergence of globally distributed carbapenem-resistant *E. coli* lineages, particularly within phylogroups A and B1, with the *Escherichia coli* ST410 lineage being a key carbapenem-resistant clone (Jaramago et al., 2017). The ST131 lineage of *E. coli* has

been recognized as a high-risk pandemic multidrug-resistant strain, indicating its successful emergence and spread (Mohsin et al., 2018). Additionally, the population dynamics of the ST131 lineage have been studied in relation to carbapenem antibiotic treatment, showing potential impacts on the prevalence of symptomatic urinary tract infections (Meunier et al., 2017). The evolution of pathogenic *E. coli* lineages, including carbapenem-resistant clones, is influenced by genetic mutations that contribute to their fitness and spread (Liu et al., 2018). The spread of resistance to carbapenems in *E. coli* lineages has been linked to the presence of specific resistance genes, such as New Delhi Metallo-β-Lactamase-5 (bla_{NDM-5}), which can be found in multiple-antibiotic-resistant strains (Paveenkittiporn et al., 2021). Reports of carbapenem-resistant *E. coli* in companion animals further highlight the cross-border emergence of clonal lineages, emphasizing the importance of surveillance and control measures to prevent the spread of carbapenem resistance (Jain et al., 2021; Jing et al., 2022).

Studies revealed that *E. coli* strains carrying the New Delhi metallo-β-lactamase (NDM) gene are often linked to specific sequence types (STs). Among these, ST101, ST167, ST131, ST405, ST410, and ST648 are the most commonly reported across various countries. ST101 has been identified in 13 different countries, while ST167, ST405, and ST410 have been reported in 8 countries each. Other notable STs include ST648, present in 5 countries; ST131, found in 4 countries; and ST156, ST448, ST361, and ST617, each found in 3 countries. Additionally, ST10, ST38, ST43, ST2659, ST354, and ST744 have each been reported in two countries (Li et al., 2023) (Figure 4).

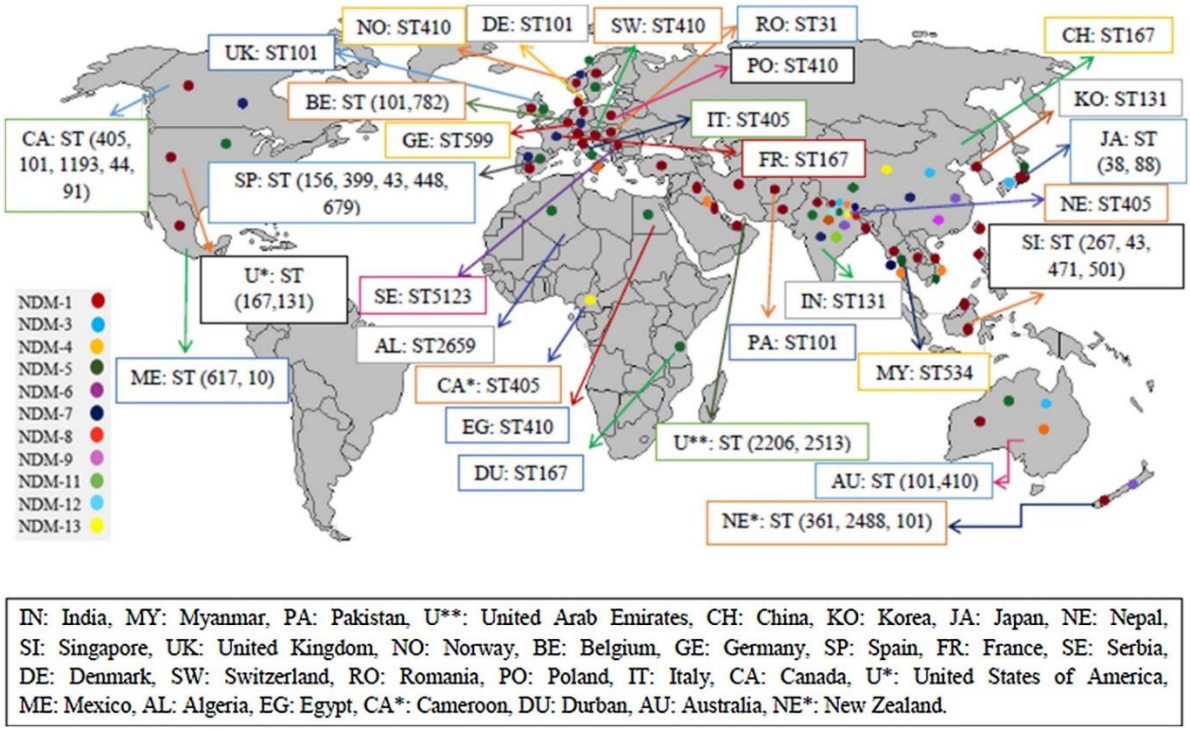


Figure 4: Most Common Sequence Types (STs) of New Delhi metallo-β-lactamase (NDM)-producing *Escherichia coli*

In addition, it has been found that NDM gene can recruits mobile genetic elements, such as plasmids belonging to different replicon or Inc types (IncFII, IncHI2, IncN, and IncX3), insertion sequences (ISAba125, ISCR1), and transposons (Tn125) (Figure 5) (Dong et al., 2022).

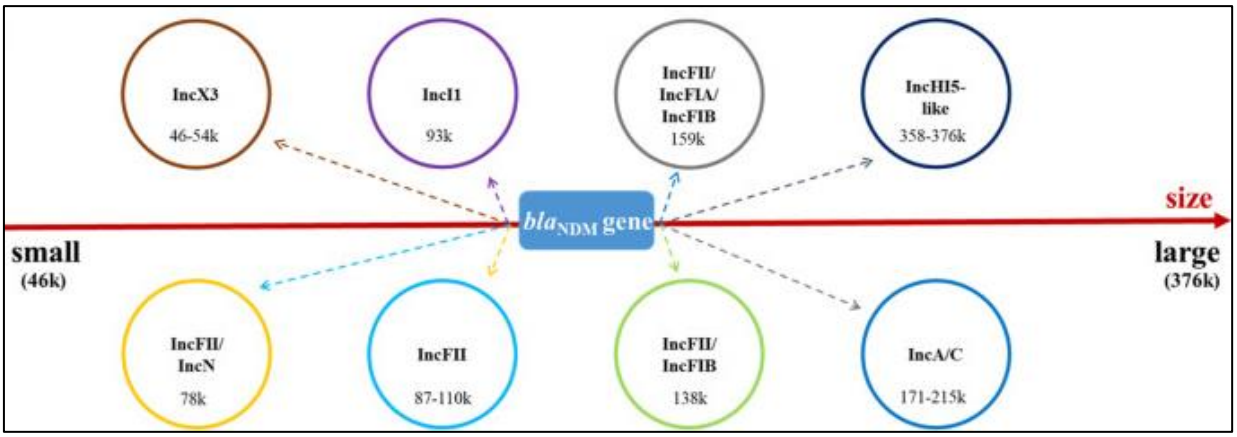


Figure 5: Diversity of NDM-bearing Plasmids in Terms of Replicon Types and Sizes

1.14 Using Whole Genome Sequencing for Detecting and Identifying CR-*E. coli*

The rise of Carbapenem-resistant *E. coli* (CR-*E. coli*) presents a formidable challenge in healthcare due to their resistance to key antibiotics. Whole Genome Sequencing (WGS) has emerged as a pivotal tool in detecting and understanding these pathogens. For instance, a study by (Mate et al., 2014) reported a 30% carbapenem resistance rate among Gram-negative bacteria in a North-East Indian hospital. (Meunier et al., 2017) utilized WGS to uncover a novel carbapenemase, FRI-2, in Enterobacter cloacae isolate. Similarly, (Paveenkittiporn et al., 2021) identified mcr gene-bearing CR-*E. coli* in Thailand, with a prevalence in phylogroup C, especially ST410, using WGS. This technology has proven crucial in revealing resistance mechanisms in clinical samples. For example, (Mohsin et al., 2018) detected the plasmid-mediated colistin resistance gene Mcr-1 in an *E. coli* bloodstream isolate from Oman. (Jain et al., 2021) revealed the presence of multiple antibiotic and virulence genes in an extensively drug-resistant *E. coli* isolate through WGS analysis. A new bla_{NDM} variant, $bla_{NDM\text{-}21}$, was identified by (Bin et al., 2018) in a clinical *E. coli* isolate. The implications of these findings extend beyond human health (Usman et al., 2023) highlighted the co-occurrence of Mcr-1 and carbapenem resistance in avian pathogenic *E. coli* from broiler chickens. Understanding the epidemiology and genetic characteristics of carbapenem-resistant Enterobacterales is vital for effective infection control, as evidenced by (Jing et al., 2022) in their multicentre study in Henan, China. In summary, Whole Genome Sequencing plays an instrumental role in not only detecting but also providing deeper insights into the resistance mechanisms and epidemiological patterns of CR-*E. coli*. Continuous surveillance and research leveraging WGS are essential to address the growing threat of carbapenem resistance in healthcare environments.

The primary aim of this thesis is to apply advanced bioinformatics tools and Whole Genome Sequencing to characterize the antibiotic resistance mechanisms in *E. coli* strains. This approach is intended to dissect the intricate genetic patterns and variations contributing to resistance, thereby enriching our understanding and enhancing the precision of our analyses. Through this focused application of bioinformatics and WGS, the goal is to develop more effective and targeted intervention strategies for the treatment and management of infections caused by antibiotic-resistant *E. coli*. This

research also aims to make a significant contribution to global efforts in antimicrobial resistance surveillance and research, fostering improved and timely clinical and public health responses to the challenges posed by antimicrobial resistance.

Chapter 2: Methods

2.1 Bacterial Collection

This retrospective study involved 14 *E. coli* isolates, collected between January 2017 and August 2018 from four hospitals in the Abu Dhabi, UAE. These isolates were identified using standard microbiological techniques. For preservation, each isolate was stored in duplicate in Tryptic Soy Broth with 10% glycerol at -80°C. Prior to experimentation, strains were revitalized by 24-hour incubation on Tryptic Soy Agar at 37°C.

2.2 Antibiotic Susceptibility Profile

Antibiotic susceptibility was evaluated using the VITEK 2 System (bioMerieux), following the manufacturer's instructions. Bacterial suspensions were adjusted to a 0.5 McFarland standard in 0.45% saline, with further dilution achieved by adding 145 µl of this suspension to 3 ml of saline. The VITEK 2 system, specifically the AST-N419 card, assessed susceptibility to a range of antibiotics, including Cefotaxime, Amikaci, Imipenem, Meropenem, Ciprofloxacin, Piperacillin/tazobactam, Ampicillin/ Sulbactam, Tigecycline, Gentamicin, Trimethoprim/sulfamethoxazole, Ceftolozane/tazobactam, Ceftazidime, Ceftazidime/Avibactam and Cefepime. Susceptibility was determined based on Clinical and Laboratory Standards Institute (CLSI) breakpoints.

Colistin susceptibility was determined using the broth microdilution method. This involved serial dilutions of colistin sulphate (Sigma, USA) in Muller Hinton Broth (Oxoid, UK) in 96-well microtiter plates (Sarstedt, Nümbrecht, Germany), with bacterial inoculation at 10^5 CFU/ml. After incubating for 18 hours at 37°C, the Minimum Inhibitory Concentration (MIC) was visually determined as the lowest concentration preventing visible growth. Duplicate assays were conducted for all tests, using *Escherichia coli* ATCC25922 as a quality control.

The classification of the isolates was based on their resistance levels as follows:

- Extensively Drug-Resistant (XDR): Isolates are classified as XDR if they are only susceptible to two or fewer classes of the tested antibiotics.

- Multi-Drug Resistant (MDR): An isolate is considered MDR if it shows resistance (non-susceptibility) to at least one agent in three or more of the tested antibiotic classes.
- Pan-Drug Resistant (PDR): Isolates are deemed PDR if they exhibit resistance to all agents in all antibiotic classes tested.

2.3 Whole Genome Sequencing

2.3.1 DNA Extraction for Whole Genome Sequencing

For the extraction of genomic DNA, the Wizard Genomic DNA Purification Kit from Promega, USA, was employed as per the manufacturer's guidelines. Initially, an overnight culture in Tryptic Soy Broth was centrifuged at 14,800 rpm for 3 minutes. The resultant pellet was then gently resuspended in 600 µl of Nuclei Lysis Solution and heated at 80°C for 5 minutes. After cooling to room temperature, 3 µl of RNase Solution was added, followed by an incubation at 37°C for 45 minutes. The next step involved adding 200 µl of Protein Precipitation Solution, vortexing the mixture, and then incubating it on ice for 5 minutes. The DNA-containing supernatant was carefully transferred into a clean Eppendorf tube, mixed with 600 µl of isopropanol, and centrifuged for another 3 minutes at 14,800 rpm. The DNA pellet obtained was washed with 600 µl of 70% ethanol and centrifuged for 3 minutes at room temperature at 14,800 rpm. Once the ethanol was discarded, the pellet was left to air dry before being redissolved in nuclease-free water. The final step involved assessing the extracted DNA's quantity and quality using the ND-1000 Spectrophotometer from NanoDrop Technologies, USA.

2.3.2 Whole Genome Sequencing of the isolates

The extracted genomic DNA was then sequenced on an Illumina NovaSeq platform (150 bp paired-end) using a commercial send-out service provided by Novogene in the United Kingdom. The JEKESA pipeline (https://github.com/stanikae/jekesa) was employed for whole-genome sequencing. To analyze our sequencing data, we utilized the Jekesa pipeline. To ensure seamless installation and management of dependencies, we created a dedicated conda environment within our Ubuntu operating system. This environment effectively isolated Jekesa and its

required packages, preventing potential conflicts with other software on the system. By leveraging the flexibility of conda, we were able to install Jekesa and its dependencies with ease, streamlining our bioinformatics analysis workflow. Briefly, Trim Galore v0.6.2 (https://github.com/FelixKrueger/TrimGalore) was used to filter the sequence reads This is usually based on Phred quality scores, de novo assembly was performed using SPAdes v3.13.2 (https://github.com/ablab/spades),the assemblies were polished and optimized using Shovill v1.1.0 (https://github.com/tseemann/shovill), and sequence typing was done using the multilocus sequence typing (MLST) tool v2.16.4 (https://github.com/tseemann/mlst). QUAST v5.0.2 (http://quast.sourceforge.net/quast) was used to determine assembly metrics such as GC content and number of contigs and N50. CARD (Comprehensive Antibiotic Resistance Database) was used for antimicrobial resistant genes analysis (https://card.mcmaster.ca/), the plasmid incompatibility type was done by plasmid finder (PlasmidFinder 2.1 (dtu.dk)), Virulence factor analysis was done by Virulence finder (VirulenceFinder 2.0 (dtu.dk)), eaphylogenetic tree was done using Galaxy tool and visualized Evolview (Evolview:Home (evolgenius.info)) and Serotyping was done by serotype finder (CGE Server (dtu.dk)).

Chapter 3: Results and Discussions

3.1 Collection and Distribution of *E. coli* Isolates

Between January 2017 and August 2018, a total of 14 *E. coli* isolates were collected across four major hospitals in the Abu Dhabi Emirate. These hospitals, identified as SKMC (Sheikh Khalifa Medical City), TAWAM Hospital, MZ Hospital (Mohammed bin Zayed City Hospital), and MAFRAQ Hospital.

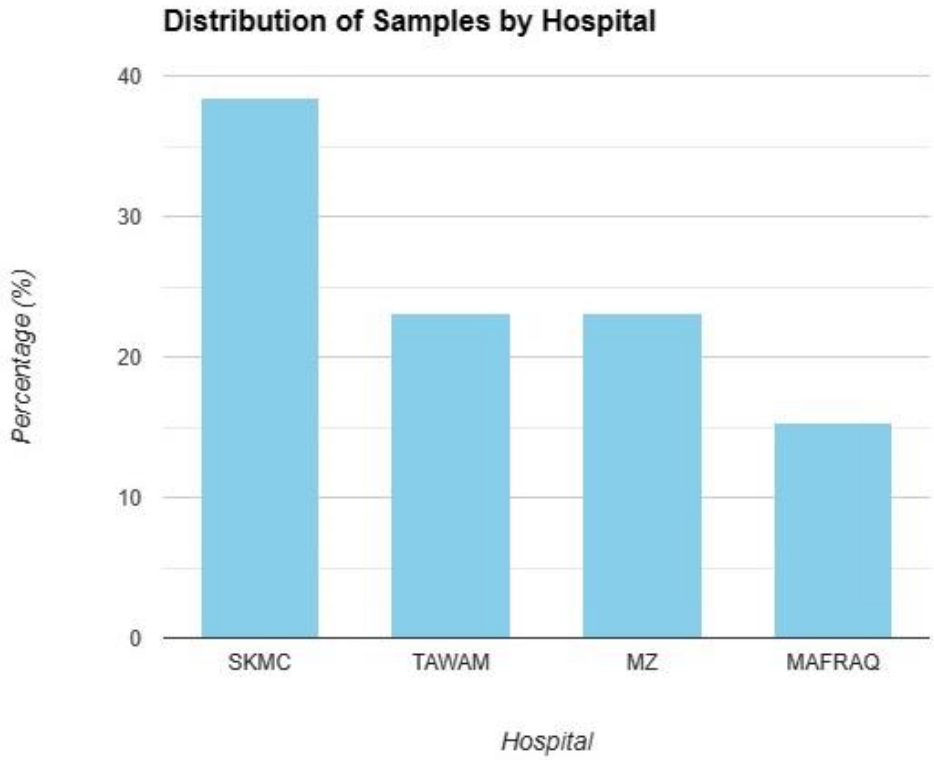


Figure 6: Distribution of *E. coli* Samples Across the Hospitals in Abu Dhabi

SKMC hospital has the largest share, with approximately 36% of the total *E. coli* isolates. Followed by SKMC, TAWAM Hospital contributed 29% of the isolates. For MZ and MAFRAQ Hospitals, each of these hospitals contributed around 17% of the samples.

All the 14 *E. coli* isolates were identified by sample type. Figure 7 provides a detailed overview of the various clinical sources from which *E. coli* samples were collected.

Rectal swabs account for approximately 40% of the *E. coli* isolates. Urine samples constitute about 30% of the isolates, marking it as the second most common source.

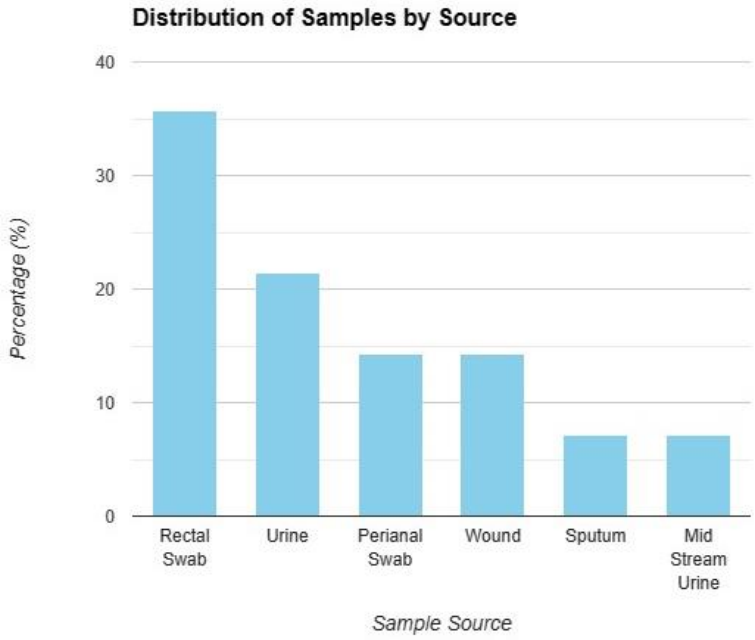


Figure 7: Distribution of *E. coli* Samples by Clinical Source

3.2 Antibiotic Susceptibility Profile of The Isolates

Table 6 provides the antibiotic susceptibility and resistance profiles for the 14 *E. coli* isolates. Notably, the isolates exhibited a high level of resistance to a majority of the antibiotics tested. Ampicillin/sulbactam, piperacillin/tazobactam, cefotaxime, ceftazidime, ceftazidime/avibactam, ceftolozane/tazobactam, cefepime, imipenem, and ciprofloxacin showed a 100% resistance rate among the isolates, indicating no susceptibility. Similarly, high resistance levels were observed with co-trimoxazole and meropenem, with 92.8% of isolates resistant to these antibiotics. In contrast, tigecycline and colistin were entirely effective, with a 100% susceptibility rate, highlighting their potential as effective treatments against these *E. coli* strains. Amikacin and gentamicin displayed mixed responses, with 85.8% and 50% of the isolates being susceptible, respectively.

Table 6: Antibiotic Susceptibility Profile of the isolates

Antibiotics	**Resistant**	**Susceptible**
	%	**%**
Ampicillin/sulbactam	100	0
Piperacillin/tazobactam	100	0
Cefotaxime	100	0
Ceftazidime	100	0
Ceftazidime/avibactam	100	0
Ceftolozane/tazobactam	100	0
Cefepime	100	0
Imipenem	100	0
Meropenem	92.8	7.2
Amikacin	14.2	85.8
Gentamicin	50	50
Ciprofloxacin	100	0
Tigecycline	0	100
Colistin	0	100
co-trimoxazole	92.8	7.2

3.3 Whole Genome Sequencing Analysis of the Isolates

3.3.1 FastQC Report

The FastQC analysis of the 14 *E. coli* samples confirmed that the sequencing data was of good quality. The sequencing resulted in an average of 100,079,168 high-quality reads per isolate. These reads successfully assembled into approximately 188 contiguous sequences, each exceeding 200 base pairs in length. The quantity and quality of these reads indicate robust data sufficient to create reliable and comprehensive genome assemblies. This level of sequencing detail is critical for subsequent analyses, including detailed genomic characterization and identification of genetic variations among isolates, to ensure that the data serve as a solid foundation for further bioinformatics investigations and antimicrobial resistance studies.

3.3.2 Sample Confirmation

In this study, all 14 isolates were conclusively identified as *Escherichia coli* using the BactInspector species identification tool, supplemented by whole genome sequencing

(WGS)-based contamination tests. The confirmation process, illustrated in Table 7, used advanced genomic tools to ensure accuracy in species determination.

Table 7: Bactinspector Analysis

SampleID	**Species_Identification**
CRE-92	*Escherichia coli* (100%)
CRE-115	*Escherichia coli* (100%)
CRE-67	*Escherichia coli* (100%)
CRE-65	*Escherichia coli* (100%)
CRE-114	*Escherichia coli* (100%)
CRE-3	*Escherichia coli* (100%)
CRE-91	*Escherichia coli* (100%)
CRE-72	*Escherichia coli* (100%)
CRE-74	*Escherichia coli* (100%)
CRE-90	*Escherichia coli* (100%)
CRE-116	*Escherichia coli* (100%)
CRE-102	*Escherichia coli* (100%)
CRE-94	*Escherichia coli* (100%)
CRE-71	*Escherichia coli* (100%)

3.3.3 Assessing Contamination Level

To evaluate the potential contamination in our whole genome sequencing (WGS) studies, a taxonomic classification of sequencing reads from the 14 *Escherichia coli* WGS samples was conducted using Kraken 2. This tool is renowned for its ability to rapidly classify reads into viral, bacterial, and archaeal domains based on a k-mer-based approach. As illustrated in Table 8, the alignment score for the primary match (Matching 1) was significantly higher than those for the secondary and tertiary matches (Matching 2 and Matching 3). This substantial difference confirms that the primary query sequence aligns with *Escherichia coli*, confirming the purity of our isolates. The high alignment score for Matching 1, relative to the others, effectively rules out contamination, allowing

us to assert with confidence that our samples are not contaminated and correctly identified as *Escherichia coli*.

Table 8: contamination conformation using kraken2 tool

Assembly	**Kraken_Match #1**	**Kraken_Match_#2**	**Kraken_Match_#3**
CRE-92	*Escherichia coli* (17.11%)	*Klebsiella variicola* (0.36%)	*Enterobacter asburiae* (0.18%)
CRE-115	*Escherichia coli* (14.93%)	*Escherichia coli* SE11 (0.10%)	*Escherichia fergusonii* (0.07%)
CRE-67	*Escherichia coli* (46.43%)	*unclassified* (0.08%)	*Shigella flexneri* (0.04%)
CRE-65	*Escherichia coli* (17.28%)	*unclassified* (0.08%)	*Escherichia fergusonii* (0.06%)
CRE-114	*Escherichia coli* (41.81%)	*unclassified* (0.05%)	*Escherichia fergusonii* (0.04%)
CRE-3	*Escherichia coli* (40.75%)	*unclassified* (0.08%)	*Shigella flexneri* (0.05%)
CRE-91	*Escherichia coli* (15.31%)	*Klebsiella pneumoniae* (0.06%)	*Shigella dysenteriae* (0.05%)
CRE-72	*Escherichia coli* (46.07%)	*unclassified* (0.07%)	*Shigella flexneri* (0.04%)
CRE-74	*Escherichia coli* (43.04%)	*unclassified* (0.09%)	*Klebsiella pneumoniae* (0.07%)
CRE-90	*Escherichia coli* (11.61%)	*Klebsiella pneumoniae* (0.38%)	*unclassified* (0.32%)
CRE-116	*Escherichia coli* (41.64%)	*unclassified* (0.05%)	*Shigella flexneri* (0.04%)
CRE-102	*Escherichia coli* (44.54%)	*unclassified* (0.22%)	*Klebsiella pneumoniae* (0.15%)
CRE-94	*Escherichia coli* (17.06%)	*Shigella dysenteriae* (0.07%)	*Shigella flexneri* (0.06%)
CRE-71	*Escherichia coli* (46.37%)	*unclassified* (0.05%)	*Escherichia fergusonii* (0.04%)

3.3.4 Genome Assembly Quality

The genome assemblies of the 14 isolates comprised an average of 187.8 contigs, each longer than 200 base pairs. These assemblies included an average maximum contig length of approximately 483,809 base pairs. In total, the genome assemblies covered an

average of 5,216,671 base pairs per isolate. All the genome has a consistent average GC content of 50.6%. Additionally, the assemblies demonstrated a robust average N50 value of 152,331 base pairs, which indicates the quality of the assembly in terms of contig length—half of the entire genome is contained in contigs of this length or longer.

3.3.5 Multilocus Sequence Typing (MLST)

The Multilocus Sequence Typing (MLST) analysis revealed a diverse distribution of *E. coli* sequence types (STs) among the isolates. The most prevalent sequence type identified was ST648, accounting for 28.6% of the isolates. This was followed by ST405, which comprised 21.4%. Other sequence types, including ST46, ST448, ST2083, ST410, ST131, ST167, and ST617, were each represented by 7% of the isolates (Table 9).

Table 9: Multilocus Sequence Typing of the isolates

Sample	SCHEM	ST	adk(allell)	fumC (allell)	gyrB(allell)	icd(allell)	mdh(allell)	purA(allell)	recA(allell)
CRE-92	ecoli	46	adk(8)	fumC (7)	gyrB (1)	icd (8)	mdh(8)	purA(8)	recA(6)
CRE-115	ecoli	448	adk (6)	fumC (6)	gyrB (5)	icd (16)	mdh(11)	purA(8)	recA(7)
CRE-67	ecoli	648	adk (92)	fumC (4)	gyrB (87)	icd (96)	mdh(70)	purA(58)	recA(2)
CRE-65	ecoli	2083	adk (6)	fumC (322)	gyrB (5)	icd (16)	mdh(11)	purA(8)	recA(7)
CRE-114	ecoli	405	adk (35)	fumC (37)	gyrB (29)	icd (25)	mdh(4)	purA(5)	recA(73)
CRE-3	ecoli	405	adk (35)	fumC (37)	gyrB (29)	icd (25)	mdh(4)	purA(5)	recA(73)
CRE-91	ecoli	410	adk (6)	fumC (4)	gyrB (12)	icd (1)	mdh(20)	purA(18)	recA(7)
CRE-72	ecoli	648	adk (92)	fumC (4)	gyrB (87)	icd (96)	mdh(70)	purA(58)	recA(2)
CRE-74	ecoli	131	adk (53)	fumC (40)	gyrB (47)	icd (13)	mdh(36)	purA(28)	recA(29)
CRE-90	ecoli	167	adk (10)	fumC (11)	gyrB (4)	icd (8)	mdh(8)	purA(13)	recA(2)
CRE-116	ecoli	405	adk (35)	fumC (37)	gyrB (29)	icd (25)	mdh(4)	purA(5)	recA(73)
CRE-102	ecoli	648	adk (92)	fumC (4)	gyrB (87)	icd (96)	mdh(70)	purA(58)	recA(2)
CRE-94	ecoli	617	adk (10)	fumC (11)	gyrB (4)	icd (8)	mdh(8)	purA(13)	recA(73)
CRE-71	ecoli	648	adk (92)	fumC (4)	gyrB (87)	icd (96)	mdh(70)	purA(58)	recA(2)

3.3.6 E. coli Serotyping

The serotyping analysis of the *E. coli* isolates in our study highlights the prevalence of specific O and H serotypes. The O serotypes O102 and O45 were identified as the most frequently occurring among the isolates. Similarly, H6 emerged as the predominant H serotype. Notably, the OH serotype O102:H6 was the most prevalent combination, found in 6 of the 14 samples (42.8%). Other significant serotypes identified include O45:H6, present in 3 samples (21.4%), O25:H4, and O101:H10, each found in 2 samples (14.3%) (Table 10).

Table 10: Distribution of O and H Serotypes Across *E. Coli* Isolates

Sample	O serotype	H serotype
CRE-3	O102	H6
CRE-65	O131	H6
CRE-67	O45	H6
CRE-71	O45	H6
CRE-72	O45	H6
CRE-74	O25	H4
CRE-90	O101	H5
CRE-91	-	H21
CRE-92	O9	H10
CRE-94	O101	H10
CRE-102	O8	H4
CRE-114	O102	H6
CRE-115	O188	H19
CRE-116	O102	H6

3.3.7 Antimicrobial Resistance Genes

Antimicrobial resistance analysis revealed that the bla_{NDM} gene was present in all isolates: 13 isolates possessed the bla_{NDM-5} variant, and one isolate had the *blaNDM-1* variant. This gene confers resistance to carbapenems. In addition, out of the total, 8 isolates were also found to carry the $bla_{CTX-M-15}$ gene, and 7 carried the bla_{CMY-42} gene, classifying them as multidrug-resistant *E. coli*. The resistance profile was broad, with various genes providing resistance to aminoglycosides, tetracyclines, sulphonamides, and macrolides. Importantly, all isolates remained susceptible to colistin and tigecycline.

Table 11: Antimicrobial Resistance Genes

Sample	Carbapenem Resistance Gene	Aminoglycoside Resistance Genes	Beta-Lactams Resistance Genes
CRE-91	*bla*$_{NDM-5}$	*aadA5*	*bla*$_{CTX-M-15}$, *bla*$_{CMY-42}$, *bla*$_{OXA-1}$
CRE-92	*bla*$_{NDM-5}$	*aadA2, rmtB*	*bla*$_{CTX-M-15}$, *bla*$_{TEM-1}$, *bla*$_{OXA-1}$
CRE-65	*bla*$_{NDM-5}$	*aadA5, aac(3)-Iid*	*bla*$_{CMY-42}$, *bla*$_{TEM-1}$
CRE-67	*bla*$_{NDM-5}$	*aadA2*	*bla*$_{CMY-42}$
CRE-71	*bla*$_{NDM-5}$	*aadA2*	*bla*$_{CMY-42}$
CRE-72	*bla*$_{NDM-5}$	*aadA2*	*bla*$_{CMY-42}$
CRE-74	blaNDM-1	-	*bla*$_{CTX-M-15}$
CRE-90	*bla*$_{NDM-5}$	*aadA2*	*bla*$_{CTX-M-15}$, *bla*$_{OXA-1}$
CRE-94	*bla*$_{NDM-5}$	*aadA2*	*bla*$_{CMY-42}$
CRE-102	*bla*$_{NDM-5}$	*aadA5*	*bla*$_{CTX-M-15}$, *bla*$_{TEM-1}$, *bla*$_{OXA-1}$
CRE-3	*bla*$_{NDM-5}$	aadA2	-
CRE-114	*bla*$_{NDM-5}$	aadA2, aadA5, aac(3)-Iid	*bla*$_{CTX-M-15}$, *bla*$_{TEM-1}$, *bla*$_{OXA-1}$
CRE-115	*bla*$_{NDM-5}$	aadA5, aac(3)-Iid	*bla*$_{CTX-M-15}$, *bla*$_{CMY-42}$, *bla*$_{OXA-1}$
CRE-116	*bla*$_{NDM-5}$	aadA5, aac(3)-Iid	*bla*$_{CTX-M-15}$, *bla*$_{OXA-1}$

3.3.8 Plasmid Incompatibility Type

Whole genome sequencing analysis revealed different distribution of Inc types across different strains (Table 12). For example, Cre-3 hosts a mixture of P0111, IncFIA and IncFIB. The isolates (Cre-65 to Cre-74) exhibit a complex array of Inc types, including combinations such as IncFIB, IncI1-I, IncX3, and IncFIA. Furthermore, cre-65 and cre-74 have greater diversity in their plasmid types compared to others. The strains (cre-67 to cre-72) are characterized by a consistent presence of IncFIB, IncY and IncFII, suggesting a specific plasmid profile pattern. Isolates such as cre-90, cre-92 and cre-102 also contain the rare type of Col (BS512) paired with different IncF types and IncI1-I. In contrast, strains such as cre-94 and cre-116 are simpler and mainly include Inc types such as IncFIA and IncFII.

Table 12: Plasmid Incompatibility Types Across *E. Coli* Bacterial isolates

Strain name	Inc Type
CRE-3	P0111, IncFIA, IncFIB
CRE-65	IncFIB, Incl1-I, IncX3, IncFIA, IncFII
CRE-67	IncFIB, IncY, IncFII
CRE-71	IncFIB, IncY, IncFII
CRE-72	IncFIB, IncY, IncFII
CRE-74	IncI1-I, IncFII, IncFIB, IncFIA
CRE-90	Col(BS512), IncFIA, IncI1-I
CRE-91	IncX3, IncX4, IncFIB, IncFIA
CRE-92	Col(BS512), IncFIA, IncFII, IncFIB
CRE-94	IncFIA
CRE-102	IncX3, IncFIA, IncFIB, IncFII, Col(BS512)
CRE-114	IncFIA, IncFII, IncFIB
CRE-115	IncFIB, IncX3, IncFIA, IncFII
CRE-116	IncFIA, IncFII, IncFIB

3.3.9 Virulence Factor

Our results revealed a variety of virulence factors in the bacterial isolates, which differ by their sequence types. The results represented a spectrum of genes related to adhesins, invasion, toxins, bacteriocins, mechanisms of iron uptake, serum resistance and capsule production, among others. All genes such as afaA, chuA, fyuA and irp2 were examined for their presence (shown in green) or absence (shown in red) in each strain (Table 13). The isolate CRE-74 with sequence type 131 showed the most comprehensive virulence profile with 14 unique factors identified, indicating highly virulent potential. This strain had genes related to iron absorption, serum resistance, and a variety of adhesins and invasion. In contrast, CRE-71 with sequence type 648 had a more limited virulence gene set, with only two or three factors detected, possibly suggesting a lower virulence capacity. The column totals provided a quantifiable measure of the virulence genes present. The higher the number, the greater the number of virulence factors.

Table 13: Virulence genes carried by *Escherichia coli* isolates

Strains				CRE-92	CRE-115	CRE-67	CRE-65
ST				46	448	648	2083
Virulence factors and their genes	Adhesins, invasins		*afaA*	0	0	0	0
		Transcriptional	*air*	0	0	0	0
		Enteroaggregative	*eilA*	0	0	0	0
		Salmonella HilA	*hra*	0	0	0	0
		Heat-resistant	*iha*	0	0	0	0
		Adherence protein	*lpfA*	0	1	0	1
		Long polar fimbriae	*papA_F19*	0	0	0	0
		Major pilin subunit	*papC*	0	0	0	0
		Outer membrane	*tsh*	0	0	0	0
		Temperature-sensitive	*yfcV*	0	0	0	1
		Fimbrial protein	*ibeA*	0	0	0	0
		Invasin of brain	*astA*	0	0	0	0
	Toxins, bacteriocins	EAST-1 heat-stable	*cea*	0	0	0	0
		Colicin E1	*celb*	0	0	0	0
		Endonuclease	*cia*	0	0	0	1
		Colicin ia	*cma*	0	0	0	0
		Colicin M	*cvaC*	0	0	0	0
		Microcin C	*hlyF*	0	0	0	0
		Hemolysin F	*hlyE*	0	0	0	0
		Avian *E. coli*	*mchB*	0	0	0	0
		Microcin H47 part	*mchC*	0	0	0	0
		MchC protein	*mchF*	0	0	0	0
		ABC transporter	*mcmA*	0	0	0	0
		Microcin M part of	*chuA*	0	0	0	0
	Iron acquistion	Outer membrane	*fyuA*	1	0	0	0
		Siderophore	*irp2*	0	0	0	1
		High molecular	*iroN*	0	0	0	0
		Enterobactin	*ireA*	0	0	0	0
		Siderophore	*iucC*	0	0	0	0
		Aerobactin	*iutA*	0	0	0	0
		Ferric aerobactin	*sitA*	1	0	1	0
		Iron transport	*iss*	0	0	0	0
	Serum resistance	Increased serum survival	*traT*	0	1	1	0
		Outer membrane protein complement	*kpsE*	0	0	1	0
	Capsule production	Capsule	*kpsMII_K5*	0	0	0	0
		Polysialic acid	*kpsMII*	0	0	0	0
		Polysialic acid	*capU*	0	0	0	0
		Hexosyltransferase	*etsC*	0	0	0	0
	Others	Putative type I	*gad*	1	0	0	0
		Glutamate	*ompT*	0	0	0	0
		Outer membrane	*terC*	0	1	0	1
		Tellurium ion	*csgA*	0	0	0	0
		curlin major	*traJ*	0	0	0	0
		Positive regulator	*nlpI*	0	0	0	0
		lipoprotein NlpI					
	Total presence			3	3	3	4

Table 13: Virulence genes carried by *Escherichia coli* isolates (continued)

CRE-116	CRE-90	CRE-74	CRE-72	CRE-91	CRE-3	CRE-114
405	167	131	648	410	405	405
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	1	0	0	0	0	0
0	0	1	0	0	0	0
0	0	0	0	1	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	1	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	1	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
1	0	1	0	0	1	1
1	0	1	0	0	1	1
1	0	1	0	0	1	1
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	1	0	1	0	0
0	0	1	0	1	0	0
1	0	1	1	1	1	1
0	0	1	0	0	0	0
0	1	1	1	1	0	0
0	0	1	1	0	0	0
1	0	1	0	0	1	1
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	1	0	0
0	0	1	0	0	0	0
0	1	0	0	1	0	0
0	1	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
5	4	14	3	7	5	5

Table 13: Virulence genes carried by *Escherichia coli* isolates (continued)

CRE-102	648	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	7
CRE-94	617	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	3
CRE-71	648	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	2

3.3.10 Genetic Environment of bla_{NDM-5}

The genetic analysis of the bacterial isolates demonstrated that, with the exception of one, that all the isolates harbored the bla_{NDM-5} gene, a key determinant of carbapenem resistance. The gene was located within a gene cassette, suggesting a potential mechanism for its dissemination among bacterial populations. Downstream to bla_{NDM-5} , the gene cassettes also included a bleomycin resistance gene and an isomerase and reductase gene. Comparative genomic analysis across all isolates confirmed the conservation of this gene arrangement, highlighting its potential importance in the bacterial genome. Moreover, isolates CRE71, CRE72, and CRE90 were found to possess an identical 10 kb genomic fragment containing the bla_{NDM-5} gene cluster, reinforcing the concept of a conserved genetic transfer event or a shared ancestry among these specific isolates (Figure 9).

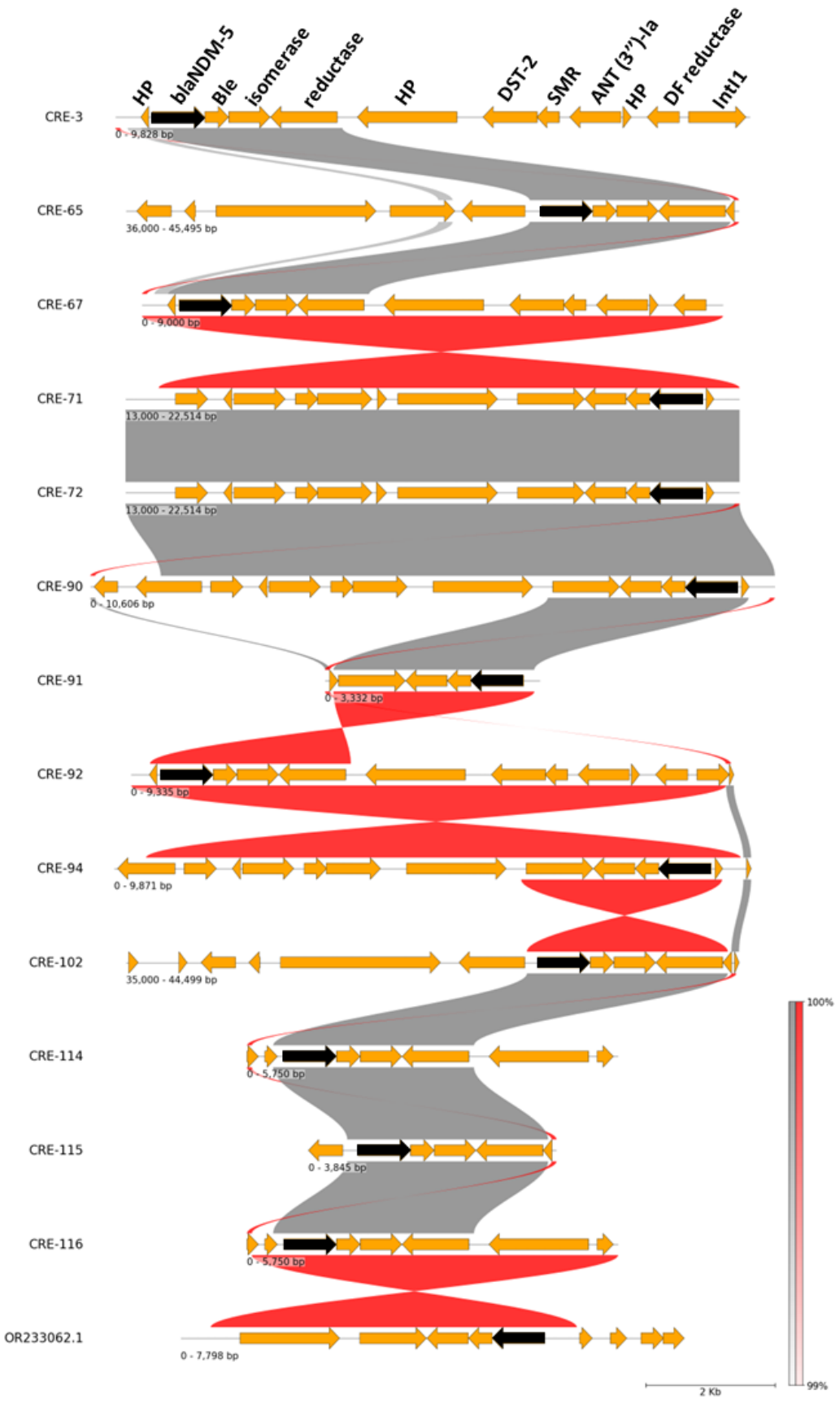


Figure 8: Genetic Environment of the blaNDM-5 Gene

3.3.11 Phylogenetic Analysis of the E. coli Isolates

Phylogenetic analysis of the *E. coli* isolates played a crucial role in our results and provided essential insights into the evolutionary relationships and genetic diversity of our bacterial samples. By constructing a phylogenetic tree, we were able to delineate the genetic lineage and possible common ancestors of the isolates. This analysis revealed distinct clades corresponding to different strains, highlighting the genetic heterogeneity within our sample population. Furthermore, the phylogenetic positioning of isolates carrying the *bla*$_{NDM-5}$ gene provided valuable insights into the evolutionary emergence and distribution of this crucial determinant of antibiotic resistance.

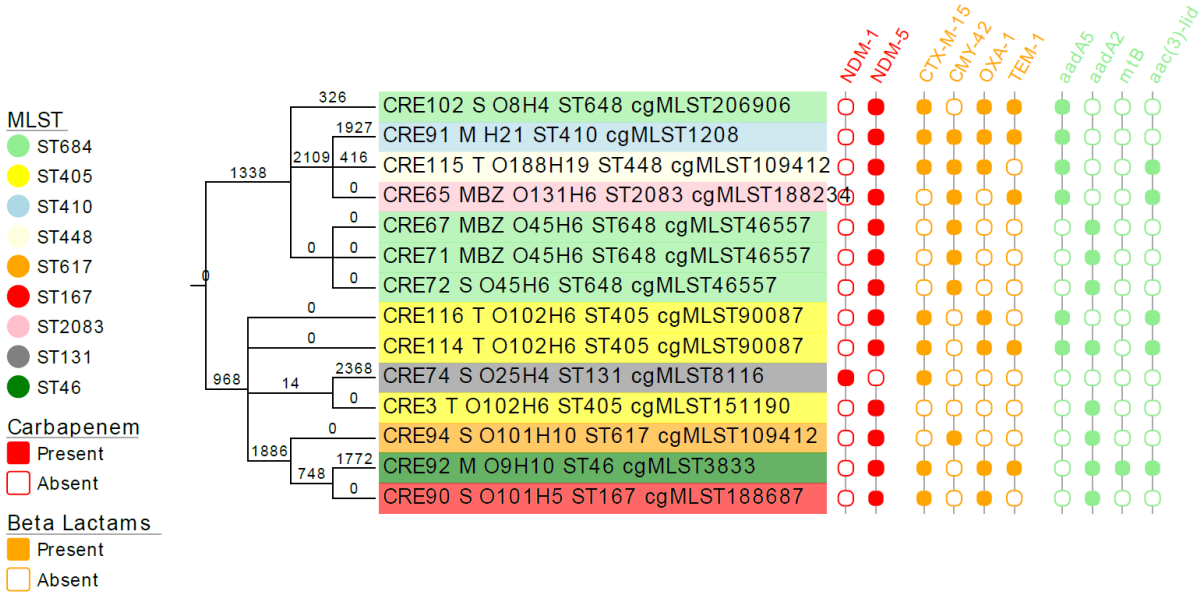


Figure 9: Phylogenetic Tree of the *E. coli* Isolates

3.3.12 Core Genome MLST (cgMLST) Analysis of E. coli Isolates

The cgMLST (core genome multilocus sequence typing) analysis conducted on the *E. coli* isolates revealed distinct clustering patterns. Specifically, Cluster 1 was composed of three isolates that shared identical ST which is ST648 and originated from two different hospitals. On the other hand, Cluster 2 included two isolates with the same ST which is ST405, both originating from a single hospital, which might indicate a localized transmission within that facility. Interestingly, despite having the same ST (ST648 and ST405), two other isolates (CRE3, CRE102) did not cluster with their similar STs, highlighting a divergence in their genetic profiles.

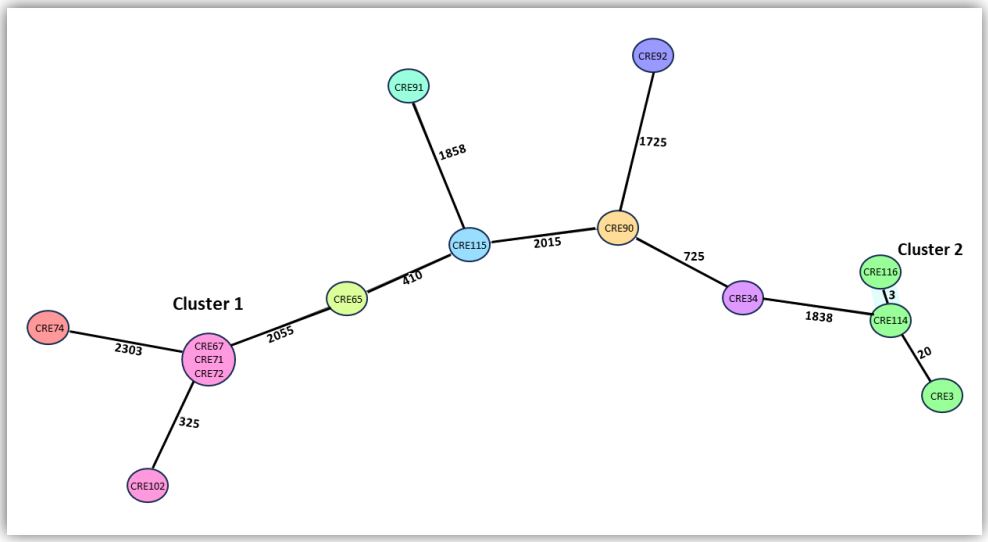


Figure 10: Core Genome MLST (cgMLST)

Chapter 4: Discussion

E. coli is a bacterium typically found in the human gut. While most strains are harmless, some can cause serious infections like urinary tract infections (UTIs), bloodstream infections, and pneumonia. The emergence of carbapenem-resistant *E. coli* (CRE), resistant to the last-resort carbapenem antibiotics, is particularly alarming in the fight against antibiotic resistance.

Carbapenem-resistant *E. coli* (CRE) leaves healthcare providers with limited treatment options, leading to higher morbidity and mortality rates, reliance on less effective antibiotics with severe side effects, and the potential for spreading resistance genes to other bacteria. This requires stricter infection control measures, resulting in longer hospital stays and increased healthcare costs. Factors contributing to CRE's spread include antibiotic misuse in humans, animals, and agriculture, global travel and trade, and the selective pressures in hospital environments due to frequent antibiotic use. Challenges in combating CRE include the limited development of new antibiotics due to high costs and rapid resistance development, complex resistance mechanisms, and the need for strict infection control. Combating CRE requires responsible antibiotic stewardship, investment in novel antibiotics, exploration of alternative therapies like phage therapy, improved surveillance systems to monitor resistance, and global collaboration for information sharing and intervention strategies.

Carbapenem-resistant *Escherichia coli* (CR-*E. coli*) is a growing threat in healthcare settings worldwide, and the United Arab Emirates (UAE) faces this challenge as well. Understanding how CR-*E. coli* spreads and develops resistance is crucial for implementing effective control strategies. Whole-genome sequencing (WGS) is a revolutionary tool for studying CR-*E. coli*, providing detailed insights into its emergence, spread, and evolution. Traditional methods like antibiotic susceptibility testing (phenotypic) and pulsed-field gel electrophoresis (PFGE) (genotypic) offer some useful data but lack the precision of WGS, preventing full identification of resistance mechanisms. In contrast, WGS detects genes encoding carbapenemase enzymes that break down carbapenems, identifies mutations affecting porin channels (key for antibiotic entry), and distinguishes closely related isolates to determine whether cases are

part of an outbreak or are sporadic. It also reveals evolutionary history by analyzing single nucleotide polymorphisms (SNPs), reconstructing transmission chains, and providing data to create rapid diagnostic tests for early identification and appropriate treatment. This comprehensive information is crucial for combating CR-*E. coli*, enabling targeted control measures, effective outbreak monitoring, and improved patient care.

We're using whole-genome sequencing to trace the spread of carbapenem-resistant *E. coli* (CR-*E. coli*) across the Emirates. We collected 14 unique CR-*E. coli* samples from four major Abu Dhabi hospitals, with Sheikh Khalifa Medical City (SKMC) providing nearly 37% of the total. All samples came from clinical cases, indicating suspected infections. Our findings revealed 100% resistance to Ampicillin/sulbactam, Piperacillin/tazobactam, Cefotaxime, Ceftazidime, Ceftazidime/avibactam, Ceftolozane/tazobactam, Imipenem, and Ciprofloxacin. These results are consistent with previous studies (Duan et al., 2019; Galindo-Méndez & Galindo-Méndez, 2020; Poirel et al., 2018; Regassa et al., 2023). A 2022 study in China reported 100% resistance to Cefotaxime and 35.41% resistance to Ciprofloxacin in *E. coli* samples from neonates (Jing et al., 2022). A 2023 study in Ethiopia found more than 57% resistance to 17 antibiotics, including Ampicillin and Cefuroxime (Regassa et al., 2023). Further analysis showed 14.2% resistance to Amikacin, 50% to Gentamicin, 92.8% to Meropenem, and 92.8% to Co-trimoxazole. While a 2013 study showed Amikacin and Meropenem had the lowest resistance rates (Vading et al., 2011), recent research indicates rising resistance to these antibiotics as well. In 2023, a Chinese study found a 55.68% resistance rate to Trimethoprim/Sulfamethoxazole (Jing et al., 2022). Our findings showed 100% susceptibility to Tigecycline and Colistin, consistent with published research. These antibiotics are reserved for last-resort cases due to their potential side effects, reducing the selective pressure that could lead to resistance (Galindo, 2020; Regassa et al., 2023).

Multilocus sequence typing (MLST) analysis of the 14 CR-*E. coli* isolates revealed a notable predominance of ST648 (28.6% of isolates), followed by ST405 (21.4%). Other sequence types (ST46, ST448, ST2083, ST410, ST131, ST167, and ST617) were also present, each accounting for 7% of the isolates. This diversity highlights the need to understand the local epidemiology of multidrug resistance. ST648,

the most prevalent sequence type at 28.6%, is recognized as a high-risk clone linked to the spread of carbapenem-resistant *E. coli* (Furlan et al., 2020; Harada et al., 2021). It is often associated with the production of New Delhi metallo-beta-lactamase (NDM), which confers resistance to a broad range of beta-lactam antibiotics, including carbapenems (Findlay et al., 2021). Its dominance signals a shift in CR-*E. coli* epidemiology. ST405, found in 21.4% of isolates, is a well-established high-risk clone associated with global dissemination of carbapenem resistance (Linkevicius et al., 2023). Like ST648, it is frequently linked to NDM production, reinforcing its significant role in spreading multidrug-resistant *E. coli*. The remaining sequence types, including ST46, ST448, ST2083, ST410, ST131, ST167, and ST617, each accounted for 7% of our isolates, adding to the diversity of circulating CR-*E. coli* strains. Many of these STs have been documented globally, indicating widespread distribution (Dadashi et al., 2019; Kim et al., 2021; Luo et al., 2018; Peng et al., 2022). Since 2017, various studies using MLST have revealed regional variability in CR-*E. coli* distribution. For instance, ST167 is prevalent in India, Egypt, France, and the United States, while ST410 is found in Thailand, the UK, France, and the U.S. ST131 appears frequently in the UK, Italy, the U.S., and China (Huang et al., 2024). Although ST648 and ST405 are dominant globally, the specific mix of sequence types varies by region. Some regions report a higher prevalence of other high-risk lineages, such as ST131, emphasizing the importance of local surveillance to understand the epidemiology of CR-*E. coli*.

The identification of *E. coli* serotypes O102 and O45 as the most common, along with the predominance of the H6 flagellar type, reveals interesting findings. O102 is not among the classic "big six" Shiga toxin-producing *E. coli* (STEC) serogroups but has been linked to diarrhea and occasional outbreaks. Some O102 strains possess virulence factors similar to those found in STECs, contributing to their pathogenicity (Kocsis et al., 2022). Serogroup O45, considered a non-traditional STEC serogroup, is associated with sporadic cases and outbreaks of diarrhea and hemolytic uremic syndrome (HUS), indicating its pathogenic potential (Alharbi et al., 2022; Beutin & Martin, 2012; Cabal et al., 2016). The H6 flagellar type is commonly found in both pathogenic and non-pathogenic *E. coli* strains. Its prevalence reflects its widespread environmental presence, although some H6 strains are linked to specific pathotypes like enteropathogenic *E. coli*

(EPEC) (Trabulsi et al., 2002). The growing recognition of non-O157 STECs like O102 and O45 highlights the evolving pathogenic landscape of *E. coli*. Surveillance and characterization of these strains are vital for identifying potential public health threats (Bielaszewska et al., 2011).

All 14 carbapenem-resistant *E. coli* (CR-*E. coli*) isolates in our study harbored the bla_{NDM} gene, a critical public health concern. The bla_{NDM-5} variant dominated, with one isolate carrying bla_{NDM-1}, underscoring the urgent need for antimicrobial stewardship and infection control measures. Carbapenems like Imipenem and Meropenem are 'last resort' antibiotics for severe or multidrug-resistant infections. However, bla_{NDM} (New Delhi metallo-beta-lactamase) grants resistance to nearly all beta-lactams, including carbapenems, severely limiting treatment options. The prevalence of bla_{NDM-5} is particularly alarming due to its enhanced enzymatic activity and higher resistance compared to earlier variants (Hornsey et al., 2011). Although present in only one isolate, the detection of bla_{NDM-1} points to the diversity of NDM variants circulating in our setting. Continuous surveillance is crucial to monitor the emergence and spread of these variants. NDM genes, often located on highly mobile plasmids, enable rapid dissemination within *E. coli* populations and other bacterial species, posing a significant global threat (Kumarasamy et al., 2010). NDM-producing bacteria were first reported in India and are now a worldwide problem, with NDM-positive Enterobacteriaceae, including *E. coli*, reported in various countries, underscoring the endemic nature of this resistance mechanism (Nordmann et al., 2011). The spread of NDM-5-producing *E. coli* is a significant concern in the field of infectious diseases. Studies have shown the dissemination of bla_{NDM-5} gene via various plasmids among different strains of *Escherichia coli* and other Enterobacteriaceae. For instance, (X. Li et al., 2018) identified eleven NDM-5-producing strains, including *Escherichia coli*, Klebsiella pneumoniae, and Citrobacter freundii. The study by (Xu et al., 2019) highlighted the dissemination of mcr-1 in an epidemic NDM-5-producing *E. coli* clone, which could contribute to the spread of colistin resistance. (Pérez-Vázquez et al., 2019) found that the spread of NDM-producing *E. coli* was polyclonal, while (Zhu et al., 2020) demonstrated the dissemination of bla_{NDM-5} carrying plasmids among multiclonal Klebsiella pneumoniae strains. Furthermore, (Zheng et al., 2022) emphasized the importance of

closely monitoring the clonal spread of NDM-5-producing *E. coli*, particularly the ST167 clone. (Li et al., 2020) reported the emergence of an NDM-5-producing *E. coli* sequence type 410 clone in infants in a children's hospital in China, highlighting the vulnerability of certain populations to these infections. Additionally, (Dong et al., 2022) identified a clinical *Escherichia coli* isolate co-harboring a novel bla_{NDM-5}-harboring IncI1-I plasmid and an mcr-1.1-harboring IncHI2 plasmid, further complicating the landscape of antibiotic resistance. Overall, these studies underscore the urgent need for continued surveillance and control measures to prevent the further spread of NDM-5-producing *E. coli* strains, which pose a significant threat to public health. Additionally, the integration of WGS within national surveillance programs has been crucial in identifying the transmission of NDM-producing *E. coli* outbreaks associated with specific exposures, such as endoscopes (Dong et al., 2022).

The UAE recognizes the importance of whole-genome sequencing (WGS) in fighting antimicrobial resistance (AMR). Several initiatives can be adopted to use WGS to study carbapenem-resistant *E. coli* (CR-*E. coli*):

- National Surveillance Programs: The Ministry of Health and Prevention (MoHAP) collects data on antibiotic resistance patterns across hospitals in the UAE. WGS can analyze bacterial isolates from these programs, providing crucial insights into the prevalence and distribution of CR-*E. coli*.
- Outbreak Investigations: During CR-*E. coli* outbreaks in healthcare facilities, WGS can rapidly compares bacterial genomes from patients and staff, helping officials assess the outbreak's scope and implement measures to prevent further transmission.
- Research Studies: UAE universities and research institutions use WGS to study CR-*E. coli* epidemiology and transmission mechanisms, revealing key factors that contribute to the bacterium's emergence and spread, which can be always linked and delivered to the concerned health facilities.

Significant challenges persist in incorporating genomics into routine surveillance in different settings. These include supply chain issues and the procurement of whole-

genome sequencing (WGS) reagents and equipment, substantial cost differences between high- and low-income regions, and a lack of skilled local bioinformaticians due to limited training opportunities and challenges in retaining staff. Additionally, platforms to deliver actionable genomic data to sentinel sites are scarce. Moreover, the timeline for implementing disruptive technologies like WGS to reach full adoption by all relevant stakeholders is expected to exceed the duration of a single academically funded project.

Chapter 5: Conclusion

Critical Challenge of CRE:

- Carbapenem-resistant *E. coli* (CRE) poses a severe health risk due to limited treatment options and increased mortality rates.
- The widespread dissemination of resistance genes contributes to the rapid spread of multidrug resistance.

Whole-Genome Sequencing (WGS) Insights:

- WGS offers valuable data on CRE transmission and epidemiology, surpassing traditional methods.
- Our research found high resistance to key antibiotics, and the presence of high-risk clones like ST648 and ST405.
- The bla_{NDM} gene, particularly the $bla_{NDM\text{-}5}$ variant, was detected across all isolates, emphasizing the urgent need for stewardship and infection control.

Importance of UAE Surveillance Programs:

- WGS should be integrated into national surveillance programs, to reveal insights into CR-*E. coli* distribution and prevalence.
- The Ministry of Health and Prevention (MoHAP) collects data on antibiotic resistance, which can be used to adopt WGS, enabling targeted outbreak management and research.

Challenges in Genomics Integration:

- Supply chain issues, high equipment costs, and a lack of skilled bioinformaticians hamper the integration of genomics into routine surveillance.
- The absence of platforms to deliver actionable genomic data complicates effective response strategies.
- Implementing disruptive technologies like WGS requires a timeline beyond a single project funding period.

Future Directions:

- Investment in national surveillance programs, research, and international collaboration is crucial for addressing these challenges.
- Improving stewardship practices and infection control can help mitigate the spread of CRE and other drug-resistant pathogens.

References

Adessi, C., Matton, G., Ayala, G., Turcatti, G., Mermod, J. J., Mayer, P., & Kawashima, E. (2000). Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. *Nucleic Acids Research*, *28*(20), e87. https://doi.org/10.1093/NAR/28.20.E87

Agarwala, V., Flannick, J., Sunyaev, S., Vinetta, A., & Altshuler, D. (2013). Evaluating empirical bounds on complex disease genetic architecture. Nature Genetics, 45(12), 1418-1429.

Akawi, N. A. (December 2013). Characterizing the genetic bases of autosomal recessive disorders. PhD, The UAE University. Al-Ain, UAE.

Alharbi, M. G., Al-Hindi, R. R., Esmael, A., Alotibi, I. A., Azhari, S. A., Alseghayer, M. S., & Teklemariam, A. D. (2022). The "Big Six": Hidden Emerging Foodborne Bacterial Pathogens. Tropical Medicine and Infectious Disease, 7(11). https://doi.org/10.3390/TROPICALMED7110356

Almeida, O. G. G., & De Martinis, E. C. P. (2019). Bioinformatics tools to assess metagenomic data for applied microbiology. Applied Microbiology and Biotechnology, 103(1), 69–82. https://doi.org/10.1007/S00253-018-9464-9/METRICS

Al-Suwaid, E. K. (July 2005). Obesity in the United Arab Emirates [e-thesis]. PhD, University of Aberdeen, UK. Retrieved September 22, 2013, from UAEU Library databases, ProQuest Dissertations & Theses Global

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aminov, R. I. (2010). A brief history of the antibiotic era: Lessons learned and challenges for the future. *Frontiers in Microbiology*, *1*(DEC), 8040. https://doi.org/10.3389/FMICB.2010.00134/BIBTEX

Bakour, S., Sankar, S. A., Rathored, J., Biagini, P., Raoult, D., & Fournier, P. E. (2016). Identification of virulence factors and antibiotic resistance markers using bacterial genomics. *Future Microbiology*, *11*(3), 455–466. https://doi.org/10.2217/FMB.15.149

Banerjee, R., Gretes, M., Harlem, C., Basuino, L., & Chambers, H. F. (2010). A mecA-negative strain of methicillin-resistant Staphylococcus aureus with high-level β-lactam resistance contains mutations in three genes. *Antimicrobial Agents and Chemotherapy*, *54*(11), 4900–4902. https://doi.org/10.1128/AAC.00594-10/Asset/C7ddc93d-4f51-48ee-B159-C37bafa0fb12/Assets/Graphic/Zac9991094030001.Jpeg

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455. https://doi.org/10.1089/CMB.2012.0021

Barcellos, S. (2000). "Understanding intermediation in a digital environment: an exploratory case study.". Virtual Reference Desk Second Annual Digital Reference Conference, Vol.2 [Softcopy].

Bennett, J. W., Mende, K., Herrera, M. L., Yu, X., Lewis, J. S., Wickes, B. L., Jorgensen, J. H., & Murray, C. K. (2010). Mechanisms of carbapenem resistance among a collection of Enterobacteriaceae clinical isolates in a Texas city. *Diagnostic Microbiology and Infectious Disease*, *66*(4), 445–448. https://doi.org/10.1016/J.DIAGMICROBIO.2009.11.013

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray, S. J., … Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature 2008 456:7218*, *456*(7218), 53–59. https://doi.org/10.1038/nature07517

Bentley, R., & Meganathan, R. (1982). Biosynthesis of vitamin K (menaquinone) in bacteria. *Microbiological Reviews*, *46*(3), 241. https://doi.org/10.1128/MR.46.3.241-280.1982

Beutin, L., & Martin, A. (2012). Outbreak of Shiga Toxin–Producing Escherichia coli (STEC) O104:H4 Infection in Germany Causes a Paradigm Shift with Regard to Human Pathogenicity of STEC Strains. *Journal of Food Protection*, *75*(2), 408–418. https://doi.org/10.4315/0362-028X.JFP-11-452

Bielaszewska, M., Mellmann, A., Zhang, W., Köck, R., Fruth, A., Bauwens, A., Peters, G., & Karch, H. (2011). Characterisation of the Escherichia coli strain associated with an outbreak of haemolytic uraemic syndrome in Germany, 2011: a microbiological study. *The Lancet. Infectious Diseases*, *11*(9), 671–676. https://doi.org/10.1016/S1473-3099(11)70165-7

Bin, P., Tang, Z., Liu, S., Chen, S., Xia, Y., Liu, J., Wu, H., & Zhu, G. (2018). Intestinal microbiota mediates Enterotoxigenic Escherichia coli-induced diarrhea in piglets 11 Medical and Health Sciences 1107 Immunology. *BMC Veterinary Research*, *14*(1), 1–13. https://doi.org/10.1186/S12917-018-1704-9/FIGURES/3

Blekherman, G., Laubenbacher, R., Cortes, D. F., Mendes, P., Torti, F. M., Akman, S., Torti, S. V., & Shulaev, V. (2011). Bioinformatics tools for cancer metabolomics. *Metabolomics 2011 7:3*, *7*(3), 329–343. https://doi.org/10.1007/S11306-010-0270-3

Bragg, L., & Tyson, G. W. (2014). Metagenomics using next-generation sequencing. *Methods in Molecular Biology*, *1096*, 183–201. https://doi.org/10.1007/978-1-62703-712-9_15/COVER

Braslavsky, I., Hebert, B., Kartalov, E., & Quake, S. R. (2003). Sequence information can be obtained from single DNA molecules. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 3960–3964. https://doi.org/10.1073/PNAS.0230489100/ASSET/E7EF6A08-040C-4798-B154-4E9B448613E3/ASSETS/GRAPHIC/PQ0630489005.JPEG

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S. R., Moon, K., Burcham, T., Pallas, M., … Corcoran, K. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology*, *18*(6), 630–634. https://doi.org/10.1038/76469

Bush, K., & Johnson, R. W. (1998). Metallo-β-Lactamases: A Class Apart. *Clinical Infectious Diseases*, *27*(Supplement_1), S48–S53. https://doi.org/10.1086/514922

Cabal, A., Porrero, M. C., De La Cruz, M. L., Saez, J. L., Barcena, C., Lopez, G., Gortazar, C., Dominguez, L., & Alvarez, J. (2016). Molecular characterization and antimicrobial resistance of STEC strains isolated from healthy cattle in 2011 and 2013 in Spain. *Epidemiology and Infection*, *144*(14), 2956. https://doi.org/10.1017/S0950268816001370

Cajal, S. R. (1928). *Degeneration and Regeneration of the Nervous System.* London: Oxford University Press.

Cepas, V., & Soto, S. M. (2020). Relationship between Virulence and Resistance among Gram-Negative Bacteria. *Antibiotics 2020, Vol. 9, Page 719*, *9*(10), 719. https://doi.org/10.3390/ANTIBIOTICS9100719

Cody, A. J., McCarthy, N. D., Van Rensburg, M. J., Isinkaye, T., Bentley, S. D., Parkhill, J., Dingle, K. E., Bowler, I. C. J. W., Jolley, K. A., & Maiden, M. C. J. (2013). Real-time genomic epidemiological evaluation of human Campylobacter isolates by use of whole-genome multilocus sequence typing. *Journal of Clinical Microbiology*, *51*(8), 2526–2534. https://doi.org/10.1128/JCM.00066-13/SUPPL_FILE/ZJM999092676SO1.PDF

Compeau, P. E. C., Pevzner, P. A., & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology 2011 29:11*, *29*(11), 987–991. https://doi.org/10.1038/nbt.2023

Craig, W. A., Andes, D. R., & Stamstad, T. (2010). In vivo pharmacodynamics of new lipopeptide MX-2401. *Antimicrobial Agents and Chemotherapy*, *54*(12), 5092–5098. https://doi.org/10.1128/AAC.00238-10/ASSET/30B9E540-4DE1-491D-94A9-2F2116FB28DB/ASSETS/GRAPHIC/ZAC9991094430005.JPEG

Cross, A., Sadoff, J., Gemski, P., & Kim, K. S. (1988). The Relative Role of Lipopolysaccharide and Capsule in the Virulence of E. Coli. *Bacteria, Complement and the Phagocytic Cell*, 319–334. https://doi.org/10.1007/978-3-642-85718-8_25

Dadashi, M., Yaslianifard, S., Hajikhani, B., Kabir, K., Owlia, P., Goudarzi, M., Hakemivala, M., & Darban-Sarokhalil, D. (2019). Frequency distribution, genotypes and prevalent sequence types of New Delhi metallo-β-lactamase-producing Escherichia coli among clinical isolates around the world: A review. *Journal of Global Antimicrobial Resistance*, *19*, 284–293. https://doi.org/10.1016/j.jgar.2019.06.008

Darby, E. M., Trampari, E., Siasat, P., Gaya, M. S., Alav, I., Webber, M. A., & Blair, J. M. A. (2022). Molecular mechanisms of antibiotic resistance revisited. *Nature Reviews Microbiology 2022 21:5*, *21*(5), 280–295. https://doi.org/10.1038/s41579-022-00820-y

Delcour, A. H. (2009). Outer membrane permeability and antibiotic resistance. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1794*(5), 808–816. https://doi.org/10.1016/J.BBAPAP.2008.11.005

Desoubeaux, G., Dominique, M., Morio, F., Thepault, R. A., Franck-Martel, C., Tellier, A. C., Ferrandière, M., Hennequin, C., Bernard, L., Salamé, E., Bailly, É., & Chandenier, J. (2016). Epidemiological outbreaks of Pneumocystis jirovecii pneumonia are not limited to kidney transplant recipients: Genotyping confirms common source of transmission in a liver transplantation unit. *Journal of Clinical Microbiology*, *54*(5), 1314–1320. https://doi.org/10.1128/JCM.00133-16/ASSET/01E8B590-9651-4E31-84AC-33EBBE6BB0F4/ASSETS/GRAPHIC/ZJM0051649420002.JPEG

Devi, U., Bora, R., Das, J. K., & Mahanta, J. (2018). Extended-spectrum β-lactamase & carbapenemase-producing gram-negative bacilli in neonates from a tertiary care centre in Dibrugarh, Assam, India. *Indian Journal of Medical Research*, *147*(January), 110–114. https://doi.org/10.4103/IJMR.IJMR_1288_16

Dong, H., Li, Y., Cheng, J., Xia, Z., Liu, W., Yan, T., Chen, F., Wang, Z., Li, R., Shi, J., & Qin, S. (2022). Genomic Epidemiology Insights on NDM-Producing Pathogens Revealed the Pivotal Role of Plasmids on bla_{NDM} Transmission. *Microbiology Spectrum*, *10*(2). https://doi.org/10.1128/SPECTRUM.02156-21/SUPPL_FILE/SPECTRUM02156-21_SUPP_1_SEQ11.PDF

Dressman, D., Yan, H., Traverso, G., Kinzler, K. W., & Vogelstein, B. (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 8817–8822. https://doi.org/10.1073/PNAS.1133470100

Drmanac, R., Sparks, A. B., Callow, M. J., Halpern, A. L., Burns, N. L., Kermani, B. G., Carnevali, P., Nazarenko, I., Nilsen, G. B., Yeung, G., Dahl, F., Fernandez, A., Staker, B., Pant, K. P., Baccash, J., Borcherding, A. P., Brownley, A., Cedeno, R., Chen, L., … Reid, C. A. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science (New York, N.Y.)*, *327*(5961), 78–81. https://doi.org/10.1126/SCIENCE.1181498

Duan, Q., Xia, P., Nandre, R., Zhang, W., & Zhu, G. (2019). Review of Newly Identified Functions Associated With the Heat-Labile Toxin of Enterotoxigenic Escherichia coli. *Frontiers in Cellular and Infection Microbiology*, *9*, 292. https://doi.org/10.3389/FCIMB.2019.00292/BIBTEX

English, B. K., & Gaur, A. H. (2010). The use and abuse of antibiotics and the development of antibiotic resistance. *Advances in Experimental Medicine and Biology*, *659*, 73–82. https://doi.org/10.1007/978-1-4419-0981-7_6/FIGURES/6_1_189198_1_EN

Escherich, T. (1988). The Intestinal Bacteria of the Neonate and Breast-Fed Infant. *Reviews of Infectious Diseases*, *10*(6), 1220–1225. https://doi.org/10.1093/CLINIDS/10.6.1220

Evans, E., Whitney, Aviva, & Must. (2014). "Recent Developments in the Epidemiology of Obesity". In *Treatment of the Obese Patient* (pp. 151-166). Springer New York.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047. https://doi.org/10.1093/BIOINFORMATICS/BTW354

Exner, M., Bhattacharya, S., Christiansen, B., Gebel, J., Goroncy-Bermes, P., Hartemann, P., Heeg, P., Ilschner, C., Kramer, A., Larson, E., Merkens, W., Mielke, M., Oltmanns, P., Ross, B., Rotter, M., Schmithausen, R. M., Sonntag, H.-G., & Trautmann, M. (2017). Antibiotic resistance: What is so special about multidrug-resistant Gram-negative bacteria? *GMS Hygiene and Infection Control*, *12*, Doc05. https://doi.org/10.3205/DGKH000290

Falagas, M. E., Grammatikos, A. P., & Michalopoulos, A. (2008). Potential of old-generation antibiotics to address current need for new antibiotics. *Expert Review of Anti-Infective Therapy*, *6*(5), 593–600. https://doi.org/10.1586/14787210.6.5.593

Findlay, J., Poirel, L., Kessler, J., Kronenberg, A., & Nordmann, P. (2021). New Delhi Metallo-β-Lactamase–Producing Enterobacterales Bacteria, Switzerland, 2019–2020. *Emerging Infectious Diseases*, *27*(10), 2628. https://doi.org/10.3201/EID2710.211265

Fratamico, P. M., DebRoy, C., Liu, Y., Needleman, D. S., Baranzoni, G. M., & Feng, P. (2016). Advances in molecular serotyping and subtyping of Escherichia coli. *Frontiers in Microbiology*, *7*(MAY), 644. https://doi.org/10.3389/FMICB.2016.00644/BIBTEX

Frey, K. G., & Bishop-Lilly, K. A. (2015). Next-Generation Sequencing for Pathogen Detection and Identification. *Methods in Microbiology*, *42*, 525–554. https://doi.org/10.1016/BS.MIM.2015.06.004

Furlan, J. P. R., Savazzi, E. A., & Stehling, E. G. (2020). Widespread high-risk clones of multidrug-resistant extended-spectrum β-lactamase-producing Escherichia coli B2-ST131 and F-ST648 in public aquatic environments. *International Journal of Antimicrobial Agents*, *56*(1). https://doi.org/10.1016/J.IJANTIMICAG.2020.106040

Galindo-Méndez, M., & Galindo-Méndez, M. (2020). Antimicrobial Resistance in Escherichia coli. *E. Coli Infections - Importance of Early Diagnosis and Efficient Treatment*. https://doi.org/10.5772/INTECHOPEN.93115

Gilbert, W., & Maxam, A. (1973). The Nucleotide Sequence of the lac Operator. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(12 Pt 1-2), 3581. https://doi.org/10.1073/PNAS.70.12.3581

Gopal Rao, G. (1998). Risk factors for the spread of antibiotic-resistant bacteria. *Drugs*, *55*(3), 323–330. https://doi.org/10.2165/00003495-199855030-00001/METRICS

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, *57*(1), 81–91. https://doi.org/10.1099/IJS.0.64483-0/CITE/REFWORKS

Guo, Y., Zhao, S., Sheng, Q., Ye, F., Li, J., Lehmann, B., Pietenpol, J., Samuels, D. C., & Shyr, Y. (2014). Multi-perspective quality control of Illumina exome sequencing data using QC3. *Genomics*, *103*(5–6), 323–328. https://doi.org/10.1016/J.YGENO.2014.03.006

Harada, S., Suzuki, M., Sasaki, T., Sakurai, A., Inaba, M., Takuya, H., Wakuda, M., & Doi, Y. (2021). Transmission of NDM-5-Producing and OXA-48-Producing Escherichia coli Sequence Type 648 by International Visitors without Previous Medical Exposure. *Microbiology Spectrum*, *9*(3). https://doi.org/10.1128/SPECTRUM.01827-21

Harris, T. D., Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Causey, M., Colonell, J., DiMeo, J., Efcavitch, J. W., Giladi, E., Gill, J., Healy, J., Jarosz, M., Lapen, D., Moulton, K., Quake, S. R., Steinmann, K., Thayer, E., … Xie, Z. (2008). Single-molecule DNA sequencing of a viral genome. *Science (New York, N.Y.)*, *320*(5872), 106–109. https://doi.org/10.1126/SCIENCE.1150427

Hendriksen, R. S., Bortolaia, V., Tate, H., Tyson, G. H., Aarestrup, F. M., & McDermott, P. F. (2019). Using Genomics to Track Global Antimicrobial Resistance. *Frontiers in Public Health*, *7*, 242. https://doi.org/10.3389/FPUBH.2019.00242

Hitch, T. C. A., & Creevey, C. J. (2018). Spherical: An iterative workflow for assembling metagenomic datasets. *BMC Bioinformatics*, *19*(1), 1–8. https://doi.org/10.1186/S12859-018-2028-2/FIGURES/2

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., & Zamir, A. (1965). Structure of a Ribonucleic Acid. *Science*, *147*(3664), 1462–1465. https://doi.org/10.1126/SCIENCE.147.3664.1462

Hong Nhung, P., Ohkusu, K., Mishima, N., Noda, M., Monir Shah, M., Sun, X., Hayashi, M., & Ezaki, T. (2007). Phylogeny and species identification of the family Enterobacteriaceae based on dnaJ sequences. *Diagnostic Microbiology and Infectious Disease*, *58*(2), 153–161. https://doi.org/10.1016/J.DIAGMICROBIO.2006.12.019

Hornsey, M., Phee, L., & Wareham, D. W. (2011). A novel variant, NDM-5, of the New Delhi metallo-β-lactamase in a multidrug-resistant Escherichia coli ST648 isolate recovered from a patient in the United Kingdom. *Antimicrobial Agents and Chemotherapy*, *55*(12), 5952–5954. https://doi.org/10.1128/AAC.05108-11

Huang, J., Lv, C., Li, M., Rahman, T., Chang, Y. F., Guo, X., Song, Z., Zhao, Y., Li, Q., Ni, P., & Zhu, Y. (2024). Carbapenem-resistant Escherichia coli exhibit diverse spatiotemporal epidemiological characteristics across the globe. *Communications Biology 2024 7:1*, *7*(1), 1–13. https://doi.org/10.1038/s42003-023-05745-7

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754–755. https://doi.org/10.1093/BIOINFORMATICS/17.8.754

Huszar, T. I., Gettings, K. B., & Vallone, P. M. (2021). An Introductory Overview of Open-Source and Commercial Software Options for the Analysis of Forensic Sequencing Data. *Genes*, *12*(11). https://doi.org/10.3390/GENES12111739

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., Fitzgerald, M. G., Fulton, R. S., Giglio, M. G., Hallsworth-Pepin, K., Lobos, E. A., Madupu, R., Magrini, V., Martin, J. C., Mitreva, M., Muzny, D. M., Sodergren, E. J., … White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature 2012 486:7402*, *486*(7402), 207–214. https://doi.org/10.1038/nature11234

Hyytiä-Trees, E. K., Cooper, K., Ribot, E. M., & Gerner-Smidt, P. (2007). Recent developments and future prospects in subtyping of foodborne bacterial pathogens. *Http://Dx.Doi.Org/10.2217/17460913.2.2.175*, *2*(2), 175–185. https://doi.org/10.2217/17460913.2.2.175

Jackson, B. R., Salter, M., Tarr, C., Conrad, A., Harvey, E., Steinbock, L., Saupe, A., Sorenson, A., Katz, L., Stroika, S., Jackson, K. A., Carleton, H., Kucerova, Z., Melka, D., Strain, E., Parish, M., Mody, R. K., & Centers for Disease Control and Prevention (CDC). (2015). Listeriosis Associated with Stone Fruit — United States, 2014. *Morbidity and Mortality Weekly Report*, *64*(10), 282. /pmc/articles/PMC4584806/

Jain, P., Bepari, A. K., Sen, P. K., Rafe, T., Imtiaz, R., Hossain, M., & Reza, H. M. (2021). High prevalence of multiple antibiotic resistance in clinical E. coli isolates from Bangladesh and prediction of molecular resistance determinants using WGS of an XDR isolate. *Scientific Reports 2021 11:1*, *11*(1), 1–13. https://doi.org/10.1038/s41598-021-02251-w

Jamal, W. Y., Albert, M. J., & Rotimi, V. O. (2016). High Prevalence of New Delhi Metallo-β-Lactamase-1 (NDM-1) Producers among Carbapenem-Resistant Enterobacteriaceae in Kuwait. *PLOS ONE*, *11*(3), e0152638. https://doi.org/10.1371/JOURNAL.PONE.0152638

Janda, J. M., & Abbott, S. L. (2021). The changing face of the family enterobacteriaceae (Order: Enterobacterales): New members, taxonomic issues, geographic expansion, and new diseases and disease syndromes. *Clinical Microbiology Reviews*, *34*(2), 1–45. https://doi.org/10.1128/CMR.00174-20/ASSET/A4E357C9-3F13-4400-8086-D62BAC2E7455/ASSETS/IMAGES/LARGE/CMR.00174-20-F0002.JPG

Jaramago, J. M., Ibáñez, R. A., Miñana, J. J., & Morales Suárez-Varela, M. (2017). Resistance profiles and risk factors of resistant microorganisms in bacteraemia of abdominal origin. *Revista Española de Anestesiología y Reanimación (English Edition)*, *64*(9), 490–498. https://doi.org/10.1016/J.REDARE.2017.03.007

Jing, N., Yan, W., Zhang, Q., Yuan, Y., Wei, X., Zhao, W., Guo, S., Guo, L., Gao, Y., Zhao, L., Shi, C., & Li, Y. (2022). Epidemiology and genotypic characteristics of carbapenem resistant Enterobacterales in Henan, China: a multicentre study. *Journal of Global Antimicrobial Resistance*, *29*, 68–73. https://doi.org/10.1016/J.JGAR.2022.01.029

Jolley, K. A., Hill, D. M. C., Bratcher, H. B., Harrison, O. B., Feavers, I. M., Parkhill, J., & Maiden, M. C. J. (2012). Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid web-based analysis methods. *Journal of Clinical Microbiology*, *50*(9), 3046–3053. https://doi.org/10.1128/JCM.01312-12/SUPPL_FILE/ZJM999091927SO4.PDF

Katz, L. S., Bolen, C. R., Harcourt, B. H., Schmink, S., Wang, X., Kislyuk, A., Taylor, R. T., Mayer, L. W., & Jordan, I. K. (2009). Meningococcus genome informatics platform: a system for analyzing multilocus sequence typing data. *Nucleic Acids Research*, *37*(suppl_2), W606–W611. https://doi.org/10.1093/NAR/GKP288

Kaufman, G. (2011). Antibiotics: mode of action and mechanisms of resistance. *Nursing Standard (Royal College of Nursing (Great Britain) : 1987)*, *25*(42), 49. https://doi.org/10.7748/NS.25.42.49.S52

Kim, J. S., Yu, J. K., Jeon, S. J., Park, S. H., Han, S., Park, S. H., Kang, M., Jang, J. I., Park, J., Shin, E. K., Kim, J., Hong, C. K., Lee, J. H., Hwang, Y. O., & Oh, Y. H. (2021). Dissemination of an international high-risk clone of Escherichia coli ST410 co-producing NDM-5 and OXA-181 carbapenemases in Seoul, Republic of Korea. *International Journal of Antimicrobial Agents*, *58*(6). https://doi.org/10.1016/j.ijantimicag.2021.106448

Klein, R. D., & Hultgren, S. J. (2020). Urinary tract infections: microbial pathogenesis, host–pathogen interactions and new treatment strategies. *Nature Reviews Microbiology 2020 18:4*, *18*(4), 211–226. https://doi.org/10.1038/s41579-020-0324-0

Kocsis, B., Gulyás, D., & Szabó, D. (2022). Emergence and Dissemination of Extraintestinal Pathogenic High-Risk International Clones of Escherichia coli. *Life 2022, Vol. 12, Page 2077*, *12*(12), 2077. https://doi.org/10.3390/LIFE12122077

Kronenberger, T., de Oliveira Fernades, P., Drumond Franco, I., Poso, A., & Gonçalves Maltarollo, V. (2019). Ligand- and Structure-Based Approaches of Escherichia coli FabI Inhibition by Triclosan Derivatives: From Chemical Similarity to Protein Dynamics Influence. *ChemMedChem*, *14*(23), 1995–2004. https://doi.org/10.1002/CMDC.201900415

Kumarasamy, K. K., Toleman, M. A., Walsh, T. R., Bagaria, J., Butt, F., Balakrishnan, R., Chaudhary, U., Doumith, M., Giske, C. G., Irfan, S., Krishnan, P., Kumar, A. V., Maharjan, S., Mushtaq, S., Noorie, T., Paterson, D. L., Pearson, A., Perry, C., Pike, R., … Woodford, N. (2010). Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *The Lancet. Infectious Diseases*, *10*(9), 597–602. https://doi.org/10.1016/S1473-3099(10)70143-2

Law, D. (2009). Academic digital libraries of the future: An environment scan. *New Review of Academic Librarianship, 15*(1), 53-67.

Li, D., Liu, C. M., Luo, R., Sadakane, K., & Lam, T. W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)*, *31*(10), 1674–1676. https://doi.org/10.1093/BIOINFORMATICS/BTV033

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/BIOINFORMATICS/BTP324

Li, J., Yu, T., Tao, X. Y., Hu, Y. M., Wang, H. C., Liu, J. L., Zhou, H. J., & Zou, M. X. (2020). Emergence of an NDM-5-Producing Escherichia coli Sequence Type 410 Clone in Infants in a Children's Hospital in China. *Infection and Drug Resistance*, *13*, 703–710. https://doi.org/10.2147/IDR.S244874

Li, X., Fu, Y., Shen, M., Huang, D., Du, X., Hu, Q., Zhou, Y., Wang, D., & Yu, Y. (2018). Dissemination of *blaNDM-5* gene via an IncX3-type plasmid among non-clonal Escherichia coli in China. *Antimicrobial Resistance and Infection Control*, *7*(1). https://doi.org/10.1186/S13756-018-0349-6

Li, Z., Breitwieser, F. P., Lu, J., Jun, A. S., Asnaghi, L., Salzberg, S. L., & Eberhart, C. G. (2018). Identifying Corneal Infections in Formalin-Fixed Specimens Using Next Generation Sequencing. *Investigative Ophthalmology & Visual Science*, *59*(1), 280–288. https://doi.org/10.1167/IOVS.17-21617

Liang, C. A., Lin, Y. C., Lu, P. L., Chen, H. C., Chang, H. L., & Sheu, C. C. (2018). Antibiotic strategies and clinical outcomes in critically ill patients with pneumonia caused by carbapenem-resistant Acinetobacter baumannii. *Clinical Microbiology and Infection*, *24*(8), 908.e1-908.e7. https://doi.org/10.1016/J.CMI.2017.10.033

Lindor, N. M., Thibodeau, S. N., & Burke, W. (2017). Whole-Genome Sequencing in Healthy People. *Mayo Clinic Proceedings*, *92*(1), 159–172. https://doi.org/10.1016/j.mayocp.2016.10.019

Linkevicius, M., Bonnin, R. A., Alm, E., Svartstrom, O., Apfalter, P., Hartl, R., Hasman, H., Roer, L., Raisanen, K., Dortet, L., Pfennigwerth, N., Hans, J. B., Toth, A., Buzgo, L., Cormican, M., Delappe, N., Monaco, M., Giufre, M., Hendrickx, A. P. A., … Kohlenberg, A. (2023). Rapid cross-border emergence of NDM-5-producing Escherichia coli in the European Union/European Economic Area, 2012 to June 2022. *Eurosurveillance*, *28*(19), 1. https://doi.org/10.2807/1560-7917.ES.2023.28.19.2300209

Liu, L., Feng, Y., McNally, A., & Zong, Z. (2018). blaNDM-21, a new variant of blaNDM in an Escherichia coli clinical isolate carrying blaCTX-M-55 and rmtB. *The Journal of Antimicrobial Chemotherapy*, *73*(9), 2336–2339. https://doi.org/10.1093/JAC/DKY226

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, *2012*. https://doi.org/10.1155/2012/251364

Logan, L. K., & Weinstein, R. A. (2017). The Epidemiology of Carbapenem-Resistant Enterobacteriaceae: The Impact and Evolution of a Global Menace. *The Journal of Infectious Diseases*, *215*(suppl_1), S28–S36. https://doi.org/10.1093/INFDIS/JIW282

Lu, J., Rincon, N., Wood, D. E., Breitwieser, F. P., Pockrandt, C., Langmead, B., Salzberg, S. L., & Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols 2022 17:12*, *17*(12), 2815–2839. https://doi.org/10.1038/s41596-022-00738-y

Luo, Y., Luo, R., Ding, H., Ren, X., Luo, H., Zhang, Y., Ye, L., & Cui, S. (2018). Characterization of Carbapenem-Resistant Escherichia coli Isolates Through the Whole-Genome Sequencing Analysis. *Microbial Drug Resistance*, *24*(2), 175–180. https://doi.org/10.1089/mdr.2017.0079

Ma, J., & Zhang, L. (2010). Modern BLAST Programs. *Problem Solving Handbook in Computational Biology and Bioinformatics*, 3–19. https://doi.org/10.1007/978-0-387-09760-2_1

MacGowan, A., & Macnaughton, E. (2017). Antibiotic resistance. *Medicine*, *45*(10), 622–628. https://doi.org/10.1016/J.MPMED.2017.07.006

Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D. A., Feavers, I. M., Achtman, M., & Spratt, B. G. (1998). Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*, *95*(6), 3140–3145. https://doi.org/10.1073/PNAS.95.6.3140

Maiden, M. C. J., Van Rensburg, M. J. J., Bray, J. E., Earle, S. G., Ford, S. A., Jolley, K. A., & McCarthy, N. D. (2013). MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology 2013 11:10*, *11*(10), 728–736. https://doi.org/10.1038/nrmicro3093

Maldonado, R. F., Sá-Correia, I., & Valvano, M. A. (2016). Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. *FEMS Microbiology Reviews*, *40*(4), 480–493. https://doi.org/10.1093/FEMSRE/FUW007

Maljkovic Berry, I., Melendrez, M. C., Bishop-Lilly, K. A., Rutvisuttinunt, W., Pollett, S., Talundzic, E., Morton, L., & Jarman, R. G. (2020). Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity. *The Journal of Infectious Diseases*, *221*(Supplement_3), S292–S307. https://doi.org/10.1093/INFDIS/JIZ286

Maltezou, H. C. (2009). Metallo-β-lactamases in Gram-negative bacteria: introducing the era of pan-resistance? *International Journal of Antimicrobial Agents*, *33*(5), 405.e1-405.e7. https://doi.org/10.1016/J.IJANTIMICAG.2008.09.003

Mancuso, G., Midiri, A., Gerace, E., & Biondo, C. (2021). Bacterial antibiotic resistance: the most critical pathogens. *Pathogens*, *10*(10), 1310. https://doi.org/10.3390/PATHOGENS10101310/S1

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., … Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature 2005 437:7057*, *437*(7057), 376–380. https://doi.org/10.1038/nature03959

Mate, Dr. Ph. H., Sulochana Devi, Dr. K., Mamta Devi, Dr. K., Damrolien, Dr. S., Lilavati Devi, D. N., & Devi, P. P. (2014). Prevalence of Carbapenem Resistance among Gram-Negative Bacteria in a Tertiary Care Hospital in North-East India. *IOSR Journal of Dental and Medical Sciences*, *13*(12), 56–60. https://doi.org/10.9790/0853-131235660

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(2), 560–564. https://doi.org/10.1073/PNAS.74.2.560

McKernan, K. J., Peckham, H. E., Costa, G. L., McLaughlin, S. F., Fu, Y., Tsung, E. F., Clouser, C. R., Duncan, C., Ichikawa, J. K., Lee, C. C., Zhang, Z., Ranade, S. S., Dimalanta, E. T., Hyland, F. C., Sokolsky, T. D., Zhang, L., Sheridan, A., Fu, H., Hendrickson, C. L., … Blanchard, A. P. (2009). Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. *Genome Research*, *19*(9), 1527–1541. https://doi.org/10.1101/GR.091868.109

Mellies, J. L., Navarro-Garcia, F., Okeke, I., Frederickson, J., Nataro, J. P., & Kaper, J. B. (2001). espC Pathogenicity Island of Enteropathogenic Escherichia coli Encodes an Enterotoxin. *Infection and Immunity*, *69*(1), 315–324. https://www.academia.edu/4997033/espC_Pathogenicity_Island_of_Enteropathogenic_Escherichia_coli_Encodes_an_Enterotoxin

Messing, J., Crea, R., & Seeburg, P. H. (1981). A system for shotgun DNA sequencing. *Nucleic Acids Research*, *9*(2), 309–321. https://doi.org/10.1093/NAR/9.2.309

Meunier, D., Findlay, J., Doumith, M., Godoy, D., Perry, C., Pike, R., Gronthoud, F., Shryane, T., Poirel, L., Welfare, W., Woodford, N., & Hopkins, K. L. (2017). FRI-2 carbapenemase-producing Enterobacter cloacae complex in the UK. *Journal of Antimicrobial Chemotherapy*, *72*(9), 2478–2482. https://doi.org/10.1093/JAC/DKX173

Meza-Segura, M., & Estrada-Garcia, T. (2016). Diffusely adherent Escherichia coli. *Escherichia Coli in the Americas*, 125–147. https://doi.org/10.1007/978-3-319-45092-6_6/COVER

Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, *95*(6), 315–327. https://doi.org/10.1016/J.YGENO.2010.03.001

Mitra, R. D., Shendure, J., Olejnik, J., Krzymanska-Olejnik, E., & Church, G. M. (2003). Fluorescent in situ sequencing on polymerase colonies. *Analytical Biochemistry*, *320*(1), 55–65. https://doi.org/10.1016/S0003-2697(03)00291-4

Mohr, K. I. (2016). History of antibiotics research. *Current Topics in Microbiology and Immunology*, *398*, 237–272. https://doi.org/10.1007/82_2016_499/FIGURES/15

Mohsin, J., Pál, T., Petersen, J. E., Darwish, D., Ghazawi, A., Ashraf, T., & Sonnevend, A. (2018). Plasmid-Mediated Colistin Resistance Gene mcr-1 in an Escherichia coli ST10 Bloodstream Isolate in the Sultanate of Oman. *Microbial Drug Resistance (Larchmont, N.Y.)*, *24*(3), 278–282. https://doi.org/10.1089/MDR.2017.0131

Murray, A. K., Zhang, L., Yin, X., Zhang, T., Buckling, A., Snape, J., & Gaze, W. H. (2018). Novel insights into selection for antibiotic resistance in complex microbial communities. *MBio*, *9*(4). https://doi.org/10.1128/MBIO.00969-18/ASSET/A545C951-AFC1-4804-B3DD-0F2FB339CCBB/ASSETS/GRAPHIC/MBO0041839730005.JPEG

Navon-Venezia, S., Kondratyeva, K., & Carattoli, A. (2017). Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance. *FEMS Microbiology Reviews*, *41*(3), 252–275. https://doi.org/10.1093/FEMSRE/FUX013

Nordmann, P., & Poirel, L. (2002). Emerging carbapenemases in Gram-negative aerobes. *Clinical Microbiology and Infection*, *8*(6), 321–331. https://doi.org/10.1046/J.1469-0691.2002.00401.X

Nordmann, P., & Poirel, L. (2014). The difficult-to-control spread of carbapenemase producers among Enterobacteriaceae worldwide. *Clinical Microbiology and Infection*, *20*(9), 821–830. https://doi.org/10.1111/1469-0691.12719

Nordmann, P., Dortet, L., & Poirel, L. (2012). Carbapenem resistance in Enterobacteriaceae: Here is the storm! *Trends in Molecular Medicine*, *18*(5), 263–272. https://doi.org/10.1016/j.molmed.2012.03.003

Nordmann, P., Poirel, L., Walsh, T. R., & Livermore, D. M. (2011). The emerging NDM carbapenemases. *Trends in Microbiology*, *19*(12), 588–595. https://doi.org/10.1016/j.tim.2011.09.005

Palmela, C., Chevarin, C., Xu, Z., Torres, J., Sevrin, G., Hirten, R., Barnich, N., Ng, S. C., & Colombel, J. F. (2018). Adherent-invasive Escherichia coli in inflammatory bowel disease. *Gut*, *67*(3), 574–587. https://doi.org/10.1136/GUTJNL-2017-314903

Papp-Wallace, K. M., Endimiani, A., Taracila, M. A., & Bonomo, R. A. (2011). Carbapenems: Past, present, and future. *Antimicrobial Agents and Chemotherapy*, *55*(11), 4943–4960. https://doi.org/10.1128/AAC.00296-11/ASSET/96F96B1C-D18C-46D2-A73C-AC22762B3899/ASSETS/GRAPHIC/ZAC9991003070009.JPEG

Paveenkittiporn, W., Kamjumphol, W., Ungcharoen, R., & Kerdsin, A. (2021). Whole-Genome Sequencing of Clinically Isolated Carbapenem-Resistant Enterobacterales Harboring mcr Genes in Thailand, 2016-2019. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/FMICB.2020.586368

Peng, C., Feng, D. H., Zhan, Y., Wang, Q., Chen, D. Q., Xu, Z., & Yang, L. (2022). Molecular Epidemiology, Microbial Virulence, and Resistance of Carbapenem-Resistant Enterobacterales Isolates in a Teaching Hospital in Guangzhou, China. *Microbial Drug Resistance*, *28*(6), 698–709. https://doi.org/10.1089/mdr.2021.0156

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, *28*(11), 1420–1428. https://doi.org/10.1093/BIOINFORMATICS/BTS174

Peng, Z., Wang, X., Huang, J., & Li, B. (2024). Pathogenic Escherichia coli. *Molecular Medical Microbiology, Third Edition*, 1065–1096. https://doi.org/10.1016/B978-0-12-818619-0.00069-1

Pérez-Vázquez, M., Sola Campoy, P. J., Ortega, A., Bautista, V., Monzón, S., Ruiz-Carrascoso, G., Mingorance, J., González-Barberá, E. M., Gimeno, C., Aracil, B., Sáez, D., Lara, N., Fernández, S., González-López, J. J., Campos, J., Kingsley, R. A., Dougan, G., & Oteo-Iglesias, J. (2019). Emergence of NDM-producing Klebsiella pneumoniae and Escherichia coli in Spain: phylogeny, resistome, virulence and plasmids encoding blaNDM-like genes as determined by WGS. *The Journal of Antimicrobial Chemotherapy*, *74*(12), 3489–3496. https://doi.org/10.1093/JAC/DKZ366

Poirel, L., Madec, J.-Y., Lupo, A., Schink, A.-K., Kieffer, N., Nordmann, P., & Schwarz, S. (2018). Antimicrobial Resistance in Escherichia coli. *Microbiology Spectrum*, *6*(4). https://doi.org/10.1128/MICROBIOLSPEC.ARBA-0026-2017

Poirel, L., Naas, T., & Nordmann, P. (2010). Diversity, epidemiology, and genetics of class D β-lactamases. *Antimicrobial Agents and Chemotherapy*, *54*(1), 24–38. https://doi.org/10.1128/AAC.01512-08/ASSET/E38344FD-EB00-405A-BDF0-070F29DF27DA/ASSETS/GRAPHIC/ZAC0011086500002.JPEG

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/NAR/GKS1219

Queenan, A. M., & Bush, K. (2007). Carbapenemases: the Versatile β-Lactamases. *Clinical Microbiology Reviews*, *20*(3), 440–458. https://doi.org/10.1128/CMR.00001-07

Rama Devi, K., Srinivasan, R., Kannappan, A., Santhakumari, S., Bhuvaneswari, M., Rajasekar, P., Prabhu, N. M., & Veera Ravi, A. (2016). In vitro and in vivo efficacy of rosmarinic acid on quorum sensing mediated biofilm formation and virulence factor production in Aeromonas hydrophila. *Http://Dx.Doi.Org/10.1080/08927014.2016.1237220*, *32*(10), 1171–1183. https://doi.org/10.1080/08927014.2016.1237220

Rambo, P. (2010). Changes in research libraries as the result of the e-science initiatives; A snapshot. In D. Marcum, & G. George (Eds), *The data deluge: Can libraries cope with e-science?* (pp. 43-600). Santa Barbara, CA: Libraries Unlimited.

Rana, T., Farooq, U., Kaur, N., Khan, A., & Khan, A. (2023). Plants Derived Efflux Pump Inhibitors: An approach against Multidrug-Resistant Gram-negative bacteria Klebsiella pneumoniae. *Pharmacophore*, *14*(1–2023), 71–79. https://doi.org/10.51847/7FQXFNYNT5

Redman, P., & Maples, W. (2011). *Good essay writing: a social sciences guide* (Third ed.). London: Sage Publications.

Reffuveille, F., De La Fuente-Nũnez, C., Mansour, S., & Hancock, R. E. W. (2014). A Broad-Spectrum Antibiofilm Peptide Enhances Antibiotic Action against Bacterial Biofilms. *Antimicrobial Agents and Chemotherapy*, *58*(9), 5363–5371. https://doi.org/10.1128/AAC.03163-14

Regassa, B. T., Tosisa, W., Eshetu, D., Beyene, D., Abdeta, A., Assefa, A., Shiferew, D., Tasew, G., Tulu, B., & Awoke, T. (2023). Antimicrobial resistance profiles of bacterial isolates from clinical specimens referred to Ethiopian Public Health Institute: analysis of 5-year data. *BMC Infectious Diseases*, *23*(1). https://doi.org/10.1186/S12879-023-08803-X

Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M., & Nyrén, P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry*, *242*(1), 84–89. https://doi.org/10.1006/ABIO.1996.0432

Ruparel, H., Bi, L., Li, Z., Bai, X., Kim, D. H., Turro, N. J., & Ju, J. (2005). Design and synthesis of a 3'-O-allyl photocleavable fluorescent nucleotide as a reversible terminator for DNA sequencing by synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(17), 5932–5937. https://doi.org/10.1073/PNAS.0501962102

Sancar, A., & Sancar, G. B. (1988). DNA repair enzymes. *Annual Review of Biochemistry*, *57*, 29–67. https://doi.org/10.1146/ANNUREV.BI.57.070188.000333

Sanger, F., & Tuppy, H. (1951). The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochemical Journal*, *49*(4), 463. https://doi.org/10.1042/BJ0490463

Sanger, F., Brownlee, G. G., & Barrell, B. G. (1965). A two-dimensional fractionation procedure for radioactive nucleotides. *Journal of Molecular Biology*, *13*(2), 373-IN4. https://doi.org/10.1016/S0022-2836(65)80104-8

Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F., & Petersen, G. B. (1982). Nucleotide sequence of bacteriophage λ DNA. *Journal of Molecular Biology*, *162*(4), 729–773. https://doi.org/10.1016/0022-2836(82)90546-0

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–5467. https://doi.org/10.1073/PNAS.74.12.5463

Schaenzer, A. J., & Wright, G. D. (2020). Antibiotic Resistance by Enzymatic Modification of Antibiotic Targets. *Trends in Molecular Medicine*, *26*(8), 768–782. https://doi.org/10.1016/J.MOLMED.2020.05.001

Schmid, B., Maveyraud, L., Krömer, M., & Schulz, G. E. (1998). Porin mutants with new channel properties. *Protein Science*, *7*(7), 1603–1611. https://doi.org/10.1002/PRO.5560070714

Schulz, M. H., Zerbino, D. R., Vingron, M., & Birney, E. (2012). Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, *28*(8), 1086–1092. https://doi.org/10.1093/BIOINFORMATICS/BTS094

Secher, T., Brehin, C., & Oswald, E. (2016). Early settlers: which E. coli strains do you not want at birth? *American Journal of Physiology. Gastrointestinal and Liver Physiology*, *311*(1), G123–G129. https://doi.org/10.1152/AJPGI.00091.2016

Sharma, I., Yaiphathoi, S., & Hazarika, P. (2020). Pathogenic Escherichia coli: Virulence Factors and Their Antimicrobial Resistance. *Model Organisms for Microbial Pathogenesis, Biofilm Formation and Antimicrobial Drug Discovery*, 159–173. https://doi.org/10.1007/978-981-15-1695-5_10/COVER

Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., Wang, M. D., Zhang, K., Mitra, R. D., & Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science (New York, N.Y.)*, *309*(5741), 1728–1732. https://doi.org/10.1126/SCIENCE.1117389

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, *19*(6), 1117. https://doi.org/10.1101/GR.089532.108

Sintchenko, V., & Holmes, E. C. (2015). The role of pathogen genomics in assessing disease transmission. *BMJ*, *350*. https://doi.org/10.1136/BMJ.H1314

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. H., & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature 1986 321:6071*, *321*(6071), 674–679. https://doi.org/10.1038/321674a0

Song, Y., Manson, J. E., Lee, I. M., & Zhang, S. M. (2012). Effect of combined folic acid, vitamin B6, and vitamin B12 on colorectal adenoma [online]. *Journal of the National Cancer Institute*, *104*(20), 1562-1575. Retrieved September 18, 2014, from http://jnci.oxfordjournals.org/content/104/20/1562.full.pdf

Sonnevend, Á., Ghazawi, A. A., Hashmey, R., Jamal, W., Rotimi, V. O., Shibl, A. M., Al-Jardani, A., Al-Abri, S. S., Tariq, W. U. Z., Weber, S., & Pál, T. (2015). Characterization of Carbapenem-Resistant Enterobacteriaceae with High Rate of Autochthonous Transmission in the Arabian Peninsula. *PLOS ONE*, *10*(6), e0131372. https://doi.org/10.1371/JOURNAL.PONE.0131372

Sprain, L., Endres, D., & Rai-Peterson, T. (2010). Research as a transdisciplinary network process. *Communication Monograph*, *77*(4), 441-444.

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research*, *6*(7), 2601–2610. https://doi.org/10.1093/NAR/6.7.2601

Stephen, J. (2001). Pathogenesis of Infectious Diarrhea. *Canadian Journal of Gastroenterology and Hepatology*, *15*(10), 669–683. https://doi.org/10.1155/2001/264096

Sykes, R. B., & Matthew, M. (1976). The beta-lactamases of gram-negative bacteria and their role in resistance to beta-lactam antibiotics. *The Journal of Antimicrobial Chemotherapy*, *2*(2), 115–157. https://doi.org/10.1093/JAC/2.2.115

Takahashi, A., Matsuoka, H., & Uda, Y. (1950). Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand.*, *4*(1), 283–293. https://doi.org/10.3123/JEMS.26.1

Tenaillon, O., Skurnik, D., Picard, B., & Denamur, E. (2010). The population genetics of commensal Escherichia coli. *Nature Reviews Microbiology 2010 8:3*, *8*(3), 207–217. https://doi.org/10.1038/nrmicro2298

Tiri, B., Bruzzone, P., Priante, G., Sensi, E., Costantini, M., Vernelli, C., Assunta Martella, L., Francucci, M., Andreani, P., Mariottini, A., Capotorti, A., D'Andrea, V., Francisci, D., Cirocchi, R., & Cappanera, S. (2020). Impact of Antimicrobial Stewardship Interventions on Appropriateness of Surgical Antibiotic Prophylaxis: How to Improve. *Antibiotics 2020, Vol. 9, Page 168*, *9*(4), 168. https://doi.org/10.3390/ANTIBIOTICS9040168

Trabulsi, L. R., Keller, R., & Tardelli Gomes, T. A. (2002). Typical and atypical enteropathogenic Escherichia coli. *Emerging Infectious Diseases*, *8*(5), 508–513. https://doi.org/10.3201/EID0805.010385

Ud-Din, A., & Wahid, S. (2014). Relationship among Shigella spp. and enteroinvasive Escherichia coli (EIEC) and their differentiation. *Brazilian Journal of Microbiology*, *45*(4), 1131–1138. https://doi.org/10.1590/S1517-83822014000400002

Urban-Chmiel, R., Marek, A., Stępień-Pyśniak, D., Wieczorek, K., Dec, M., Nowaczek, A., & Osek, J. (2022). Antibiotic Resistance in Bacteria—A Review. *Antibiotics*, *11*(8), 1079. https://doi.org/10.3390/ANTIBIOTICS11081079/S1

Usman, M., Rasool, M. H., Khurshid, M., Aslam, B., & Baloch, Z. (2023). Co-Occurrence of mcr-1 and Carbapenem Resistance in Avian Pathogenic E. coli Serogroup O78 ST95 from Colibacillosis-Infected Broiler Chickens. *Antibiotics 2023, Vol. 12, Page 812*, *12*(5), 812. https://doi.org/10.3390/ANTIBIOTICS12050812

Vading, M., Samuelsen, Haldorsen, B., Sundsfjord, A. S., & Giske, C. G. (2011). Comparison of disk diffusion, Etest and VITEK2 for detection of carbapenemase-producing Klebsiella pneumoniae with the EUCAST and CLSI breakpoint systems. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, *17*(5), 668–674. https://doi.org/10.1111/J.1469-0691.2010.03299.X

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLoS ONE*, *9*(11), 112963. https://doi.org/10.1371/JOURNAL.PONE.0112963

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., … Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature 2003 420:6915*, *420*(6915), 520–562. https://doi.org/10.1038/nature01262

Webber, M. A., & Piddock, L. J. V. (2003). The importance of efflux pumps in bacterial antibiotic resistance. *Journal of Antimicrobial Chemotherapy*, *51*(1), 9–11. https://doi.org/10.1093/JAC/DKG050

Wee, Y., Bhyan, S. B., Liu, Y., Lu, J., Li, X., & Zhao, M. (2019). The bioinformatics tools for the genome assembly and analysis based on third-generation sequencing. *Briefings in Functional Genomics*, *18*(1), 1–12. https://doi.org/10.1093/BFGP/ELY037

Wick, R. R., & Holt, K. E. (2019). Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Research*, *8*, 2138. https://doi.org/10.12688/F1000RESEARCH.21782.4

Wong-Beringer, A. (2001). Therapeutic Challenges Associated with Extended-Spectrum, β-Lactamase-Producing Escherichia coli and Klebsiella pneumoniae. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*, *21*(5), 583–592. https://doi.org/10.1592/PHCO.21.6.583.34537

Wu, H. J., Wang, A. H. J., & Jennings, M. P. (2008). Discovery of virulence factors of pathogenic bacteria. *Current Opinion in Chemical Biology*, *12*(1), 93–101. https://doi.org/10.1016/J.CBPA.2008.01.023

Wu, R., & Kaiser, A. D. (1968). Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *Journal of Molecular Biology*, *35*(3), 523–537. https://doi.org/10.1016/S0022-2836(68)80012-9

Xu, L., Wang, P., Cheng, J., Qin, S., & Xie, W. (2019). Characterization of a novel bla NDM-5-harboring IncFII plasmid and an mcr-1-bearing IncI2 plasmid in a single Escherichia coli ST167 clinical isolate. *Infection and Drug Resistance*, *12*, 511–519. https://doi.org/10.2147/IDR.S192998

Yang, S. C., Lin, C. H., Aljuffali, I. A., & Fang, J. Y. (2017). Current pathogenic Escherichia coli foodborne outbreak cases and therapy development. *Archives of Microbiology*, *199*(6), 811–825. https://doi.org/10.1007/S00203-017-1393-Y

Yezli, S., Shibl, A. M., & Memish, Z. A. (2015). The molecular basis of β-lactamase production in Gram-negative bacteria from Saudi Arabia. *Journal of Medical Microbiology*, *64*(2), 127–136. https://doi.org/10.1099/JMM.0.077834-0/CITE/REFWORKS

Yong, D., Toleman, M. A., Giske, C. G., Cho, H. S., Sundman, K., Lee, K., & Walsh, T. R. (2009). Characterization of a new metallo-β-lactamase gene, bla NDM-1, and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. *Antimicrobial Agents and Chemotherapy*, *53*(12), 5046–5054. https://doi.org/10.1128/AAC.00774-09/ASSET/A59978DC-CBC4-4167-B1B4-16F74620E900/ASSETS/GRAPHIC/ZAC0120985730006.JPEG

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614. https://doi.org/10.1093/BIOINFORMATICS/BTT593

Zheng, B., Xu, H., Guo, L., Yu, X., Ji, J., Ying, C., Chen, Y., Shen, P., Han, H., Huang, C., Zhang, S., Lv, T., & Xiao, Y. (2022). Genomic and Phenotypic Diversity of Carbapenemase-Producing Enterobacteriaceae Isolates from Bacteremia in China: A Multicenter Epidemiological, Microbiological, and Genetic Study. *Engineering*, *12*, 90–100. https://doi.org/10.1016/J.ENG.2020.10.015

Zhou, Q., Su, X., & Ning, K. (2014). Assessment of quality control approaches for metagenomic data analysis. *Scientific Reports 2014 4:1*, *4*(1), 1–11. https://doi.org/10.1038/srep06957

Zhu, W., Wang, X., Qin, J., Liang, W., Shen, Z., & Gales, A. C. (2020). Dissemination and Stability of the *blaNDM-5* -Carrying IncX3-Type Plasmid among Multiclonal Klebsiella pneumoniae Isolates. *MSphere*, *5*(6). https://doi.org/10.1128/MSPHERE.00917-20

UAEU
جامعة الإمارات العربية المتحدة
United Arab Emirates University

UAEU MASTER THESIS NO. 2024:70

The increasing threat of antibiotic-resistant bacteria highlights the crucial role of bacterial genomics in combating infections. Microbiologists need to master genomic techniques to understand bacteria better and develop new ways to treat infections. This study introduces the use of whole-genome sequencing to study multidrug-resistant E. coli.

Ahmed Abdelfattah received his Master of Medical Sciences from the Department of Medical Microbiology and Immunology, College of Medicine and Health Sciences at the United Arab Emirates University, UAE. he received his Bachelor of Science in Biology from the College of Science, United Arab Emirates University, UAE.

www.uaeu.ac.ae