Kingdom of Morocco

Ministry of National Education, Vocational Training, Higher Education and Scientific Research

MOHAMMED V UNIVERSITY of RABAT

FACULTY OF MEDICINE AND PHARMACY OF RABAT

MASTER OF MEDICAL BIOTECHNOLOGY

Major: BIOMEDICAL

MM0492020

MASTER'S THESIS

Comparative genomic analysis of four Moroccan clinical strains of *Klebsiella Pneumoniae* displaying different antibiotic resistance profiles

by

Zineb KABBAJ

Graduation Committee

Pr. Azeddine IBRAHIMI	Faculty of Medicine and Pharmacy, Rabat	**Committee Chair**
Mr. Tarek ALOUANE	Faculty of Medicine and Pharmacy, Rabat	**Internal Supervisor**
Pr. Latefa BISKRI	Mohammed VI Polytechnic University, Benguerir	**External Supervisor**
Pr. Mouna OUADGHIRI	Faculty of Medicine and Pharmacy, Rabat	**Examiner**
Mme. Souad KARTTI	Faculty of Medicine and Pharmacy, Rabat	**Examiner**

Academic Year 2019-2020

DEDICATION

All the words in the world can't be enough to express my feelings towards you.

To my dear and respectful parents,

No dedication could express the depth of my feelings for you, your countless sacrifices and dedication were encouragement for me.
So many sacrifices, so much affection, an unrestrained commitment, an unconditional love
This work is the result of the spirit of sacrifice that you have shown me, the encouragement and support that you never cease to show, I hope that you will find there the fruits of your seed and the testimony of my great pride to have you as my parents.

I love you.

To my brothers Mohammed, Hassan and Houssine,

Thank you for your advices and the support you show me every day. May this work be the expression of my esteem for you.

To my dearest friends Chaimae, Maryam and Badiaa,

For all the happy moments spent together, I wish you success, happiness, health and prosperity.

AKNOWLEDGEMENTS

Expressing gratitude to the people who have surrounded and supported is natural. This is by no means a mere formality, but sincere feelings that come from the bottom of my heart.

First of all, I would like to thank **Pr. Azeddine IBRAHIMI,** Head of the Medical Biotechnology Laboratory and director of the Bio-Inova Research Center at the Faculty of Medicine and Pharmacy of Rabat, Mohammed V University. I am extremely proud to be part of the MEDBIOTECH family. His enthusiasm and passion for science were a source of motivation and perseverance to all of us students. I am extremely grateful for his guidance and support throughout these past two years. His human and professional qualities are known and will always have my admiration.

I express my sincere thanks to **Pr. Latefa BISKRI,** Professor Researcher and Responsible at Centre Interface Et Application Médicale (CIAM), for her confidence in me from the start and accepting me as an intern, her supervision, and her time and patience to help me with this project. I also thank her for her availability and her encouragements all along.

I would like to warmly thank **Mr. Tarek Alouane**, without whom this project would probably not have been possible. My thanks also for the quality of his supervision, his permanent availability, his kindness, his wise advices and without forgetting his scientific qualities that enlighten my vision of scientific research that has gone beyond the simple stage of interest and curiosity. Thank you for everything you have taught me.

I am very honored that **Pr. Mouna OUADGHIRI,** coordinator of the MEDICAL BIOTECHNOLOGY master, has agreed to carefully examine my manuscript. Here is the expression of my deepest consideration and my sincere gratitude.

I would like to sincerely thank **Mme Souad KARTTI**, professor at the MEDICAL BIOTECHNOLOGY master, who did me the great honor of evaluating this work.

My warmest thanks go to all members of the Centre Interface Et Application Médicale (CIAM) and the medical biotechnology laboratory (MEDBIOTECH) for everything.

ABSTRACT

Klebsiella Pneumoniae is now recognized as a major threat to human health globally because of the emergence of multidrug-resistant strains associated with hospital outbreaks and hypervirulent strains associated with severe community-acquired infections. However, very little research work has been carried out concerning the sequencing of the complete genome of *K. pneumoniae.*

The objective of this study is to carry out a genomic analysis of four strains of *K. Pneumoniae* with different AMR profiles from different Moroccan patients in order to identify the resistance genes and perform a genomic comparison.

Four strains of *K. Pneumoniae* were subjects of a de Novo assembly followed by an annotation in order to identify the genetic elements (resistance genes, plasmids...) of the strains. This genomic analysis pipeline made it possible to identify 31 resistance genes from different antimicrobial classes (β-lactam, aminoglycosides, Fosfomycin, fluoroquinolone…), 8 plasmids of different types (IncF, IncX, IncR, IncC, Col), 4 sequence types using MLST (ST664, ST13, ST2695, ST35) and a pan-genome of 4299 core genes and 1946 unique genes.

Our study provided a global view of the information carried in the genome of *K. Pneumoniae*. There are multiple resistance genes across diverse *Kp* genomes and plasmids in circulation that are capable of carrying this resistance. Unless appropriate interventions are rapidly placed, these may lead to a massive problem of untreatable infection in vulnerable populations.

Keywords: *Klebsiella pneumoniae,* outbreak, comparative genomic, antimicrobial resistance, MLST, Plasmids, Pan-genome.

RÉSUMÉ

Klebsiella Pneumoniae est désormais reconnue comme une menace majeure pour la santé humaine dans le monde en raison de l'émergence de souches multirésistantes associées à des épidémies hospitalières et de souches hypervirulentes associées à des infections communautaires sévères. Cependant, les travaux de recherche axés sur les souches de *K. pneumoniae* présentant une différence de profil de résistance restent relativement limités.

L'objectif de cette étude est de réaliser une analyse génomique de quatre souches cliniques marocaines de *K. Pneumoniae* avec différents profils de résistance aux antimicrobiens afin d'identifier les déterminants génétique en effectuant une analyse génomique comparative.

Quatre souches de *K. Pneumoniae* ont fait l'objet d'un assemblage de Novo suivi d'une annotation afin d'identifier les éléments génétiques (gènes de résistance, plasmides ...) de chacune des quatres souches. Ce pipeline d'analyse génomique a permis d'identifier 31 gènes de résistance de différentes classes d'antimicrobiens (β-lactame, aminoglycosides, fosfomycine, fluoroquinolone…), 8 plasmides de différents types (IncF, IncX, IncR, IncC, Col), 4 types de sequences-types en utilisant MLST (ST664, ST13, ST2695, ST35) et un pan-génome de 4299 de core protéines et 1946 de protéines uniques.

Notre étude a fourni une vue globale des informations disponibles dans le génome de *K. Pneumoniae*. Il existe plusieurs gènes de résistance à travers divers génomes et plasmides de *K. Pneumoniae* en circulation qui sont capables de porter cette résistance. À moins que des interventions appropriées ne soient mises en place rapidement, ceux-ci peuvent conduire à un problème massif d'infection incurable dans les populations vulnérables.

Mots clés: *Klebsiella pneumoniae*, épidémie, génomique comparative, résistance aux antimicrobiens, MLST, Plasmides, Pan-génome.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AMR : Antimicrobial Resistance
ARGs : Antibiotic Resistance Genes
ATP : Adenosine Triphosphate
bp : Base Pair
C3G : 3rd Generation Cephalosporins
CCD : Charge Coupled Device
CDS : Coding Sequences
CGH : Comparative Genomic Hybridization
CRKp : Carbapenem-resistant *K. pneumoniae*
DBG : De Bruijn Graph
DNA : Deoxyribonucleic Acid
ESBL : Extended Spectrum β-lactam
FC : Flow cell
FRET : Fluorescence Resonance Emission Transfer
HGAP : Hierarchical Genome Assembly Process
ICEs : Integrated Conjugative Elements
IHGSC: International Human Genome Sequencing Consortium
K. Pneumoniae: *Klebsiella pneumoniae*
LPS : Lipopolysaccharides
MDR : Multidrug-resistant
MGEs : Mobile Genetic Elements
MLST : Multi-Locus Sequence Type
NGS : Next Generation Sequencing
OLC : Overlap-Layout-Consensus
PBPs : Penicillin-binding proteins
PCR : Polymerase Chain Reaction
PGM : Personal Genome Machine
PTP : Pico Titer Plate
QRDR: Quinolone Resistance Determining Region
RAST : Rapid Annotations using Subsystems Technology
RNA : Ribonucleic Acid
tRNA : Transfer Ribonucleic Acid
SMRT : Single Molecule Real-Time Analysis
SOliD : Sequencing by Oligonucleotide Ligation and Detection
WGS : Whole Genome Sequencing
ZMW : Zero-Mode waveguide

INTRODUCTION

In medical pathologies, microbial infection actually holds the first place. It has been known for a long time, with the use of antibiotics, it has progressively changed its face and clinicians have been faced with infections from germs previously considered non-pathogenic or saprophytic. *Klebsiella pneumoniae* (*K. pneumoniae*) is one of the most remarkable examples of this. [1]

Antimicrobial treatment of *Klebsiella pneumoniae* contaminations has gotten progressively troublesome as an outcome of the rise and spread of strains that are resistant to various antimicrobials. *K. pneumoniae* has for a long time been perceived as a factor of illness (first portrayed as a reason for pneumonia via Carl Friedländer in 1882), and stays among the world's most common nosocomial microorganisms [2]. It is additionally a significant reason for neonatal sepsis, positioning in the main three causative factors in most settings [3] [4].

The World Health Organization perceives extended spectrum β-lactam (ESBL)-producing and carbapenem-resistant *K. pneumoniae* (CRKp) as a critical general health danger [5].
In Europe alone, such strains purportedly represent >90,000 contaminations, >7,000 deaths every year and 25% of the overall *disability-adjusted life years* lost to multidrug-resistant (MDR) bacterial contaminations [6]. In Morocco, unlike most developed countries, we do not yet have national surveillance programs for controlling antimicrobial resistance. However, bacterial resistance surveillance studies are one of the most effective measures in terms of monitoring the spread of resistant bacteria. Still, antimicrobial surveillance in Morocco is restricted to only few large hospitals [7].

To control and treat these infections caused by *K. pneumoniae,* a variety of antibiotics such as β-lactams and aminoglycosides are effective [8] [9]. However, antimicrobial resistance (AMR) caused by the abuse of antibiotics considerably decreases the effectiveness of these antibiotics, resulting in increasing difficulties in the treatment of *K. pneumoniae*.

Whole genome sequencing (WGS) has become a fundamental tool for understanding the power of pathogenic bacteria. Major advances have been made in the field of sequencing over the past decade with the development of next-generation sequencing platforms,

characterized by high throughput and the ability to simultaneously analyze many samples. Still, researchers find many difficulties with the enormous mass of data generated by the Next Generation Sequencing (NGS). To address this issue, several computer and statistical tools are developed to store, analyze and organize the data.

On the other hand, research work focusing on strains of *K. pneumoniae* with a difference in resistance profile remains relatively limited. The availability of whole genomes of *K. pneumoniae* makes it promising to explore the genomic characteristics (resistance to antibiotics, virulence factors and specific genes) having a role in the rise of the multi-resistant phenotype, and to understand the history of its evolution, adaptation and dispersal around the world.

The objectives of this study are to carry out a comparative genomic analysis of four Moroccan clinical strains of *K. Pneumoniae* showing a difference in resistance profiles in order to:

- Identify the genetic determinant potentially involved in resistance
- Determine the typical sequences via the MLST analysis of strains circulating in Morocco
- Estimate the complete repertoire of *K. pneumoniae* genes through the pangenome analysis.

This work is made up of four chapters:

- The first chapter corresponds to a bibliographical synthesis aiming to set the work in its scientific context. It includes general information on *K. pneumoniae* and the comparative genomic analysis.
- The second chapter is devoted to the experimental study, the use of various bioinformatic tools and software in order to characterize the common molecular and genetic characteristics (Multi-Locus Sequence Type (MLST), AMR genes, plasmids and pan-genome).
- The third chapter includes results of the work of analysis and comparison between the four strains of *K. pneumoniae*.
- The fourth chapter is dedicated to a general discussion of the results and to the open perspectives of this work.

BIBLIOGRAPHY

Chapter one

I. Klebsiella pneumoniae

1. History

The genus *Klebsiella* was first discovered by the German microbiologist Edwin Klebs in 1834 and named after his name by Trevisan in 1887 to honor him.

The type species is *Klebsiella pneumoniae*, known again under the name Friedlander's *pneumobacillu*s. [10]

2. Taxonomic position of *Klebsiella pneumoniae*

According to the classification of the 2nd edition of Bergy's manual [11], it belongs to:

Domain: *Bacteria*
Phylum: *Proteobacteria*
Class: *Gamma Proteobacteria*
Order: *Enterobacteriales*
Family: *Enterobacteriaceae*
Genus: *Klebsiella*
Species: *Klebsiella pneumoniae*

The *K. pneumoniae* species is subdivided into 3 subspecies: *K. pneumoniae Subsp. Pneumoniae*, *K. pneumoniae Subsp. Ozenae* and *K. pneumoniae Subsp. Rhinoscleromatis*.

3. Epidemiology

Klebsiella is an opportunistic pathogen and natural intestinal flora enteric bacteria. Enteric bacteria generally do not cause illness and may also support proper function and nutrition in the intestine. The bacteria become pathogenic only when it reaches tissues outside of their normal intestinal or other less common normal flora sites [12]. In the respiratory tract and feces of about 5 % of normal individuals *K. pneumoniae* is present and responsible for a small proportion (approximately 1 %) of bacterial pneumonias.

K. pneumoniae is the most medically important species of the Family Enterobacteriaceae after *E. coli* [13]. *K. pneumoniae* is a significant hospital-acquired pathogen that is a frequent cause of septicemia, wound and blood infection, infection of the urinary system, intra-abdominal infections and pneumonia in individuals affected by immunodeficiency [14] [15].

K. pneumoniae strains holding extended spectrum beta-lactamases (ESBL) and metallocarbapenemase, presenting resistance to many of accessible antibiotics that making treatment possibilities limited, have been reported in many parts of the world [16]. These resistant pathogens are considered clinically significant because they cause nosocomial

infections that frequently occur in outbreaks and associated with treatment failure, extended hospital stay, higher insurance costs and a potential rise in mortality [17].

4. Transmission

K. pneumoniae is responsible for public infections in 25% of cases, but in particular for serious nosocomial infections that are difficult to treat [18]. The transmission of these bacteria from one patient to another is simply achieved through by the patient's hand, nursing staff or the work equipments of medical or paramedical staff (catheter, oxygen mask, etc.) [19], or to a reduced degree by environment contamination. The transmission of *K. pneumoniae* is very fast but does not spread through the air according to Boston Medical Research Occupational Health Program.

5. Risk factors

Risk factors for developing nosocomial infection with *K. pneumoniae* include: cancer, chronic liver disease, diabetes, solid organ transplantation and dialysis, treatment with corticosteroids, chemotherapy. Newborns and the elderly are also at risk [20].

6. Virulence factors

Klebsiella virulence factors differ depending on the sites of infection because the host defense mechanisms differ from site to other. Its pathogenicity has been attributed to many virulence factors that facilitate infection and survival in the host. Among all the virulence factors it is important to point out the capsule, the exopolysaccharide associated with cystic fibrosis, lipopolysaccharide (LPS), adhesins and iron absorption systems [21] and the biofilm [22]. (Fig.1)

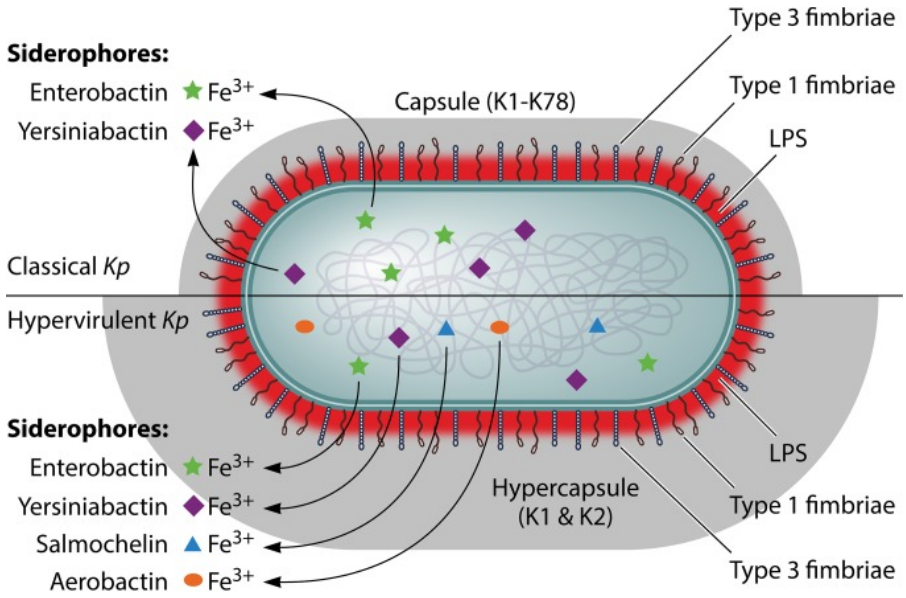


Figure 1:Four well-characterized virulence factors in classical and hypervirulent *K. pneumoniae* strains. *[23]*

Table 1 shows the factors of virulence involved in the pathogenesis of *K. pneumoniae* and their role.

Table 1: Summary of virulence factors involved in the pathogenesis of *K. pneumoniae*. *[24]*

Virulence factor	Role of the factor in pathogenesis
Capsule	Inhibits and prevents phagocytosis by host cells, neutralizes antibacterial peptide and complement activity
LPS	O antigen provides resistance to serum
Siderophore	Recovers iron essential for survival, hypermucoviscous phenotypes have been linked to increased iron binding activity
Type 1 fimbriae	They have the greatest capacity for membership. They are involved in the colonization of the respiratory and urinary tract
Type 3 fimbriae	Important for the formation of biofilm on biotic and abiotic surfaces
Biofilm formation	Its formation promotes resistance to host defense mechanisms and antibiotics

II. Antibiotics

1. Definition

The term antibiotic was derived from the word "antibiosis" which literally means "against life" [25]. These are agents whose selective toxicity results from a specific mechanism of action. They work in low doses to inhibit the growth of microorganisms or to destroy them [26], those that kill bacteria are called bactericides while those that inhibit bacterial growth are called bacteriostats [27]. They can be produced naturally by fungi and bacteria or obtained by synthesis and hemisynthesis.

2. Classification and mechanism of action

Antibiotics classification can be done essentially on the basis of chemical nature, site of action or mechanism of action [28].

The antimicrobial potency of most classes of antibiotics is directed to an exclusive feature of bacterial structure or their metabolic processes [29], the target sites for antibiotics are as follows:

- Inhibition of cell wall synthesis
- Disruption of the structure or function of the cell membrane
- Inhibition of the structure and function of nucleic acids
- Inhibition of protein synthesis
- Blockage of the main metabolic pathways, [30]

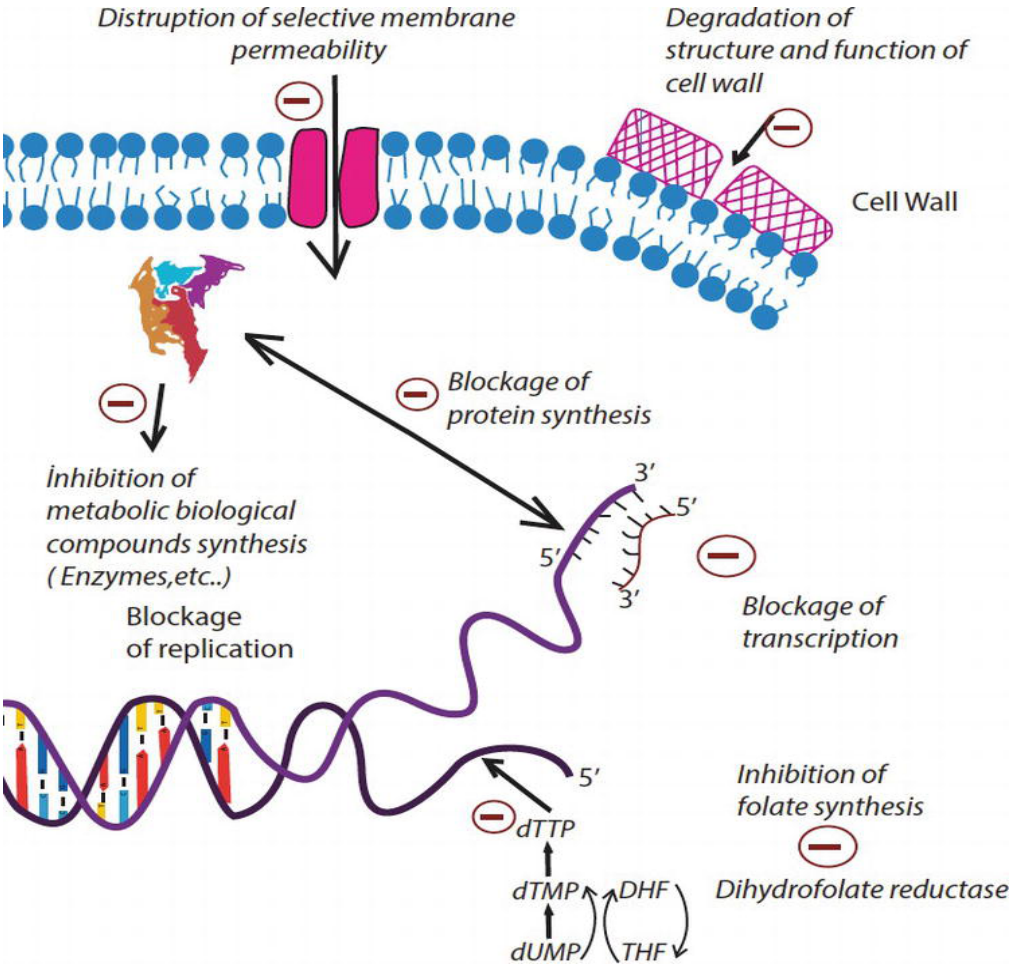


Figure 2: Mechanism of action of Antimicrobials *[31]*

III. Antimicrobial Resistance (AMR)

Antibiotic resistance is the ability of one bacterium to survive and multiply in the presence of drugs that inhibit or destroy them. Initially, resistance was thought to be a kind of defense mechanism formed by pathogenic bacteria to survive the pressure of selection of antibiotics [32]. Over the decades, several studies have focused only on clinical bacteria and their resistant phenotypes. However, metagenome and resistance studies have shown that a wide variety of antibiotic resistance genes (ARGs) exist; their expression and interaction are essential for drug resistance [33] [34]. A broad variety of ARGs are involved in other functional bacterial pathways, such as metabolism and resistance-related functions. The mutations in the DNA molecule configure one of the most powerful driving forces in evolution. The presence of antibiotics in the environment imposes selective pressure on bacteria with higher mutation levels. Increases in mutations may contribute to more regular genetic recombination, a process that supports a variety of antibiotic resistance mechanisms. [35]

1. Types of antibiotic resistance

Antibiotic effectiveness depends on at least three factors: the amount of antibiotic in contact with the target, the affinity of the antibiotic for the target, and the production of enzymes that inactivate the antibiotic. These factors are responsible either for natural resistance, and therefore present in all strains of the species, or for resistance acquired by certain strains, due to the appearance of chromosomal mutations or the acquisition of genetic material such as plasmids, transposons or integrons [36].

2. Mechanisms of antibiotic resistance

In bacteria, the three mechanisms underlying antibiotic resistance are natural, acquired and adaptive:

2.1.Natural resistance

Natural resistance is associated with the innate ability of all or almost all prokaryotes to resist specific drugs; independently of adverse conditions such as drug selective pressure.
K. pneumoniae naturally has a gene encoding a chromosomal penicillinase (with several SHV, LEN, OKP variants) which gives it low-level resistance to penicillin. This penicillinase is sensitive to the action of inhibitors (clavulanic acid, tazobactam). Thus, the wild-type *K. pneumoniae* phenotype is sensitive to combinations of amoxicillin (or ticarcillin) + clavulanic acid and piperacillin + tazobactam as well as to all cephalosporins. It is also sensitive to cephamycin, aztreonam and carbapenems [37]. *K. pneumoniae* is sensitive to other antibiotics like aminoglycosides, fluoroquinolones, fosfomycin and cotrimoxazole.

2.2.Acquired resistance

Acquired resistance is a consequence of spontaneous chromosomal mutations or gene gain through mobile genetic elements (MGEs - plasmids, integrons, transposons, and genomic islands) [38].

K. pneumoniae contributed amply to the hospital dissemination of extended spectrum β-lactamases (ESBLs) which give them resistance to 3rd generation cephalosporins (C3G). *K. pneumoniae* has always played a major role in the emergence and/or dissemination of resistance mechanisms, in particular towards β-lactams. [39] [40].

More recently, this species occupies a significant position in the emergence and dissemination of certain enzymes conferring resistance to carbapenems (KPC, OXA-48).

K. pneumoniae is in 80% of cases the host in which these new enzymes (of the TEM, SHV or class C type) were determined [39]. When it is not the first host, this species is very frequently found to be responsible for the epidemic diffusion of new enzymes such as the metalloenzymes of the VIM and NDM-1 type [41] [42]. Most resistance observed in *K. pneumoniae* is associated with horizontally acquired accessory AMR genes rather than mutations in chromosomal genes [43].

2.3.Adaptive resistance

Adaptive resistance is under investigation. Gradual changes in antibiotic concentrations impact the populations of bacteria and, in response to such environmental selective pressure, bacteria grow reversible drug resistance profiles [44] [45]. Typically, bacteria may develop normally at non-lethal antibiotic concentrations, but this can lead to proven resistance mechanisms against both the drug in use and related drugs. In the absence of an antibiotic inducer, the phenotype is then reversed to its prior state [44] [46] [47]. In clinical settings, antibiotics can play a concentration-dependent role in inducing higher levels of selective pressure. Antibiotics tend to be used in high doses, bacteria respond by enabling and controlling those ARGs via complex mechanisms, but only during drug exposures [48].

Despite the apparent three pathways used to acquire drug resistance, the major modifications or adaptations observed in microorganisms are often the same [49] [50] [51]: decrease of bacterial envelope permeability, increase in both the expression and function of efflux pump systems, synthesis of enzymes capable of destroying or modifying the drug, modification, substitution or disruption of antibacterial targets, and formation biofilms.

In order to provide examples of the mechanisms cited above, Fig. (3) diagrams the main classes of antibiotics used in the therapy, and the most possible microorganism resistance mechanisms to develop when exposed to such antibiotics.

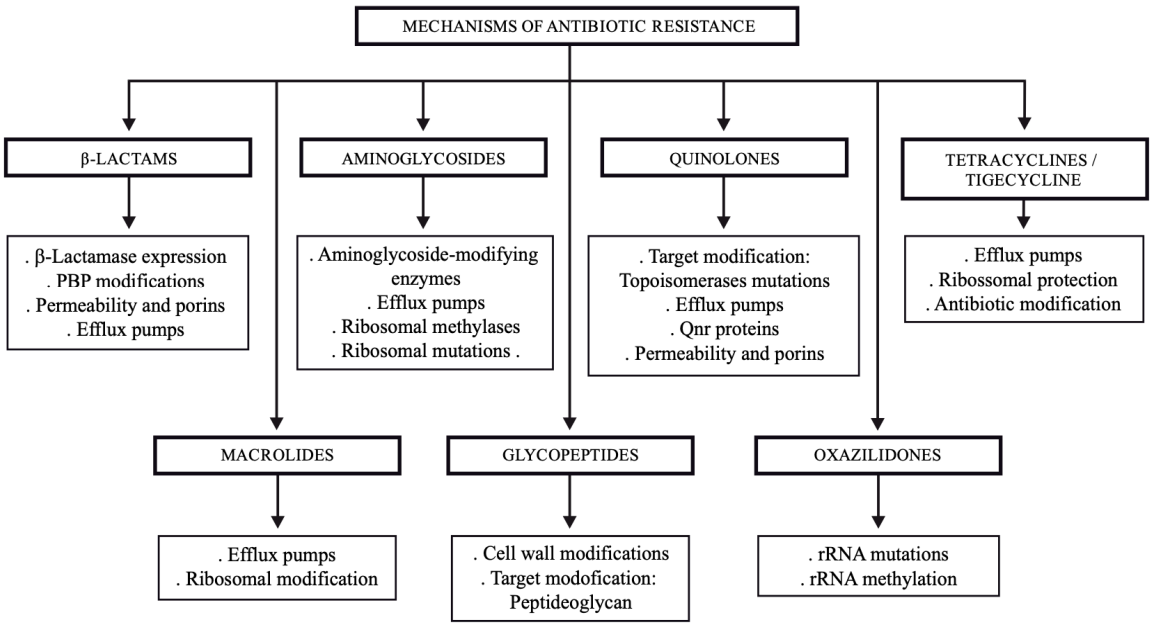


Figure 3: Diagram representing the main classes of antibiotics and their respective resistance mechanism. *[52]*

3. Resistance of *K. pneumoniae* to β-lactams

Beta-lactams are among the most common antibiotics used in treating infections caused by *K. pneumoniae* due to their diversity, low toxicity, bactericidal activity and broad spectrum of action. They include penicillin derivatives, cephalosporins, monobactams and carbapenems, and all share the β-lactam ring [37].

3.1.Mechanism of resistance

Natural or acquired resistance to β-lactams is characterized by at least four mechanisms which can also be combined: the impermeability of the outer membrane, the expression of an active efflux pump, the modification of the affinity of the target (PBPs) and finally enzymatic inactivation by β-lactamases, constituting the principal mechanism of the natural and acquired resistance of Gram-negative bacilli to β-lactams [37].

3.2.Impermeability

K. pneumoniae outer membrane of is composed of LPS, the structure of which is hydrophilic due to its surface electrical charges and very compact in depth due to its unsaturated fatty acids. This organization is responsible of the natural resistance to hydrophobic and / or high molecular weight antibiotics (penicillin G, V and M, macrolides, rifampicin, fusidic acid and glycopeptides) [39]. Porins facilitate the transmission between the periplasm and the external environment through passive diffusion of nutrients and other substances. the outer membrane can be crossed by hydrophilic β-lactams by borrowing the porins OmpK36, OmpD and Omp36 in *K. pneumoniae* [37]. However, resistance acquired by decrease in wall

permeability has been reported in this species following a quantitative or qualitative alteration of the porins. This resistance mechanism is usually expressed at low level, can affect multiple families of antibiotics and is frequently associated with other resistance mechanisms: efflux and / or production of β-lactamase.

3.3.Excretion by efflux systems

K. pneumoniae efflux systems are made of three proteins [39]:

- One, inserted in the cytoplasmic membrane acting as a pump,
- The second is implanted in the outer membrane guaranteeing the outer membrane crossing,
- The third, periplasmic, that connect the pump to the protein of outer membrane.

These systems are in fact metabolic pumps that guarantee successful expulsion of metabolic products or toxicants, such as antibiotics, using a protonmotive force.

Efflux resistance is also associated to a decrease in permeability. The combination of these two mechanisms can lead to high-level and simultaneous resistance to structurally unrelated antibiotics, thus constituting true multi-resistance systems.

3.4.Modification of PBP

Several factors can lead to resistance by alteration of the target: loss of affinity of PBPs for β-lactams by mutation, acquisition of genes or fragments of genes encoding PBPs of reduced affinity or hyperproduction of normal PBPs. This type of resistance mechanism remains very rare in *K. pneumoniae* [39].

4. Resistance of *K. pneumoniae* to quinolones

In both hospitals and community medicine, resistance of bacteria, particularly *K. pneumoniae*, to fluoroquinolones has become a concern. Quinolones are synthetic antibacterial compounds, the leader of which, nalidixic acid, was described by Lesher et *al* in 1962 [53]. All these molecules have a pyridine ring in which the nitrogen can be substituted and have a ketone function in 4 and a carboxylic group in 3. This ring is connected to another variable aromatic ring: benzene, pyridine, pyrimidine [37]. Fluoroquinolones are characterized by the presence of a fluorine atom in position 6 and a nitrogenous ring, most often a piperazine, in position 7 [37].

4.1.Mechanism of resistance

In *K. pneumoniae*, two main mechanisms of quinolone resistance occur separately or in combination and confer varying levels of resistance.

4.1.1. Resistance by chromosomal mutation

Resistance by chromosomal mutation which is due either to the decrease in the affinity of the intracellular targets which are the DNA-DNA gyrase and DNA-DNA topoisomerase IV complexes, or to the decrease in intracellular accumulation of the antibiotic, by default of passive penetration and / or active excretion [54]. The loss of affinity for the target is due to a structural change in a region called the Quinolone Resistance Determining Region (QRDR), where most of the mutations responsible for resistance to fluoroquinolones are found.

a. Mutations in the target site gene

The most common mechanism of high-level resistance is due to mutations in the quinolone resistance-determining regions (QRDR) of at least one of the genes that encode the targets of these drugs, the type II topoisomerases. and IV (gyrA, parC), result in amino acid substitutions that structurally alter the target protein and subsequently the drug-binding affinity of the enzyme [55] [56].

b. Mutations leading to reduced drug accumulation

Mutations cause a decrease in intracellular drug concentration, either through reduced absorption or increased efflux or a combination of both, may confer resistance to quinolone [55] [56] [57].

4.1.2. Plasmid resistance

Plasmid resistance is due to the protection of DNA gyrase from the binding of quinolones. This resistance was described for the first time in 1998 in a strain of *K. pneumoniae* harboring a plasmid carrying the gene qnrA which codes for a qnrA protein [58] which protects the DNA-gyrase complex from inhibition by quinolones [59]. A second mechanism was subsequently discovered and consisted of an aminoglycoside modifying acetyltransferase, aac (6 ') - Ib-cr. This variant is able to acetylate the unsubstituted nitrogen of the C7 piperazine ring found in quinolones such as ciprofloxacin, thus conferring resistance to quinolones by decreasing the activity of the drug [60]. The third mechanism is the non-specific oqxAB efflux pump [61], which confers a level of resistance to fluoroquinolones in *K. pneumoniae* [62].

5. Resistance of *K. pneumoniae* to aminoglycosides

Aminoglycosides have a wide antibacterial spectrum that includes Gram negative and positive bacteria. The treatment of serious nosocomial infections caused by Gram-negative bacilli requires their use most often in combination with a β-lactam or a fluoroquinolone [40].

5.1.Mechanism of resistance

The general mechanisms of resistance to aminoglycosides are:

- Altering the target:

The mode of action of aminoglycosides suggests the mutation of 16S RNA as a means of resistance. Three 16S RNA methylation activities modify the A site at positions G1405 (N7), A1408 and C1407 (N5) [63].

- Enzymatic modification of the antibiotic:

When an aminoglycoside is modified by bacterial enzymes its binding to 16S RNA can be affected and result in loss of its activity. Aminoglycoside modifying enzymes are the most common resistance mechanism in Gram negative and positive bacteria. They have been grouped according to the reaction they catalyze [42]:

- Acetylation of an amino group (N-acetyltransferase).
- Phosphorylation of a hydroxyl group (O-phosphotransferase).
- Nucleotidylation of a hydroxyl group (O-nucleotidyltransferase).

Pathogenic bacteria have acquired these enzymes via transferable plasmids carried by bacteria which have a specific DNA recombination system such as transposons and integrons [64] [65].

6. Resistance of *K. pneumoniae* to fosfomycins

Fos A is a Mn^{2+} and K^{+} dependent glutathione dimeric S-transferase, which catalyzes the addition of glutathione to fosfomycin, rendering the antibiotic inactive [66]. Plasmids or chromosomes may be coding this enzyme. The gene *Fos A* is plasmid-mediated, such as *fosA5* and *fosA6*, whose origin appears to be the chromosome of *K. pneumoniae* [67] [68].

6.1.Mechanism of action of Fosfomycin

In bacterial cell wall biosynthesis, Fosfomycin interferes with the first cytoplasmic step [69]. Blocks the production of N-acetylmuramic acid and prevents the synthesis of peptidoglycan [70]. Fosfomycin binds coevally in the active site of *Mur A* (enzyme initiates the synthesis of peptidoglycan) and thereby inactivates it; [69].

IV. State of Art: Genomic analyses

1. Genomic

In just 50 years, biology has shifted from the study of a single gene to the study of genomes, transcriptomes or proteomes of entire organisms, with granularity of analysis ranging from the single molecule to the organization as a whole. This has been made possible by the technological advances made jointly in biology and bioinformatics which have been motivated by the desire of scientists to extract ever more information from our genome.

1.1.Biological revolution

The biological revolution was initiated by four major events: the discovery of DNA as a carrier of genetic information [71] and of its double helix structure [72], thus as the establishment of the central dogma of molecular biology and the deciphering of the genetic code [73]. These events form the foundations of genomics.

The central dogma is the simplified modeling of the flow of genetic information through different molecules of the cell and can be summed up in three processes (Fig. 4).

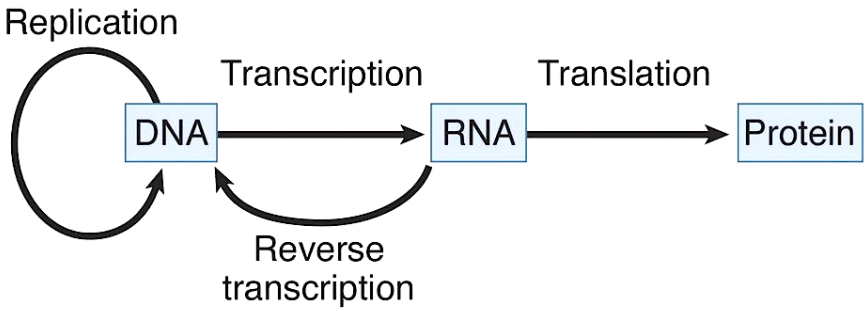


Figure 4: Central dogma of molecular biology (http://www.dnalife.org/).

Thanks to these principles, new molecular biology techniques have been developed, and in synergy with the growing power and availability of computers, a new discipline has emerged: bioinformatics.

The emergence of two new approaches to DNA sequencing [74] [75] has allowed bioinformatics to really take off. The production of sequences by these methods is an opportunity to create the new EMBL [76] and GenBank [77] databases, in order to list these nucleic acid sequences, and to develop new algorithms for processing biological data. These have led to the major tools of bioinformatics that are FASTA [78], CLUSTALW [79], and BLAST [80].

With the advent of automatic sequencers and new molecular biology tools, the production of sequences is accelerating and we see the emergence of whole genome sequencing projects that lead to the end of the century.

After ten years of efforts, the arrival of the first preliminary sequences of the Human genome [81] marks the end of the genomic era and the entry into the post-era era. genomics. However, it was not until 2004 that a version was obtained from the International Human Genome Sequencing Consortium (IHGSC) that can be considered finalized [82].

Currently, thanks to high-throughput sequencing techniques, sequencing projects have multiplied (around 176 000) so that the scientific community has access to 33 943 complete and published genomes (Fig. 5).

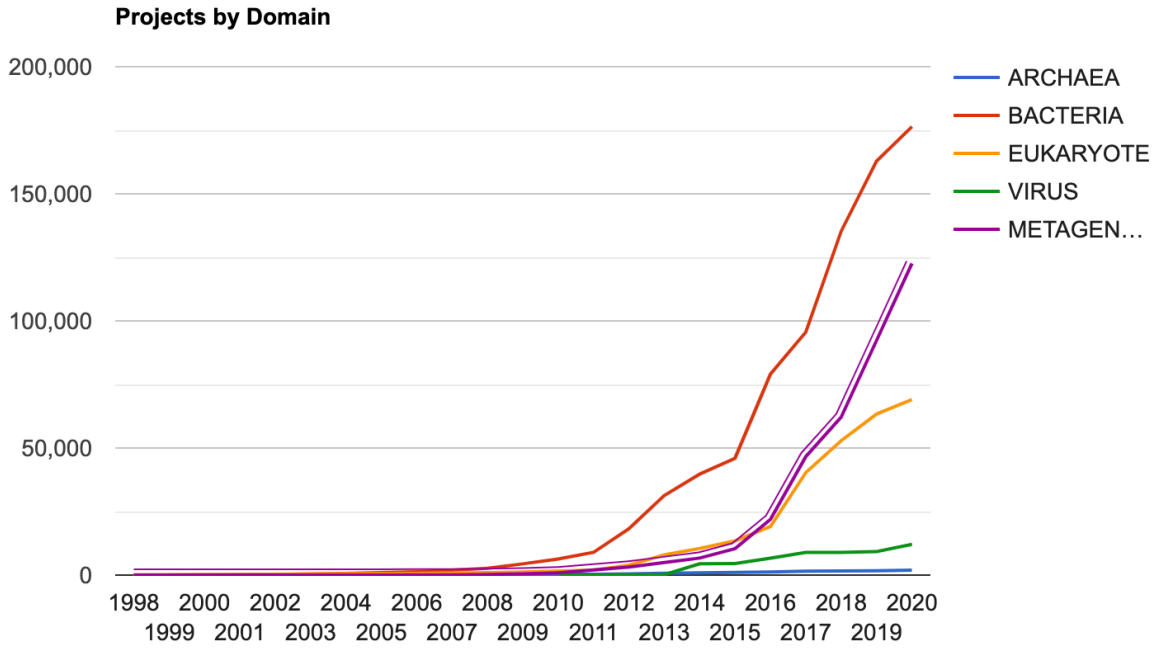


Figure 5: Evolution of the number of complete genomes available (https://gold.jgi.doe.gov/statistics).

1.2.Genomic sequencing

DNA sequencing is a method whose goal is to determine the linear succession of bases A, C, G and T taking part in the structure of DNA. Reading this sequence makes it possible to study the biological information contained by it. Given the uniqueness and specificity of the DNA structure in each individual, the DNA sequence allows many applications in the field of medicine, such as, for example, diagnostics, genetic studies, paternity study, criminology, understanding of physiopathological mechanisms, drug synthesis, epidemiological investigations. The objective of the section below is to describe the evolution of manual sequencing to the high-speed sequencers which are the most used today.

1.2.1. History

In 1965, Holley and his colleagues sequenced the first two nucleic acids in history, the tRNA of alanine from the bacterium Escherichia coli, then that of the yeast Saccharomyces cerevisiae. It was thanks to the ability to purify particular tRNAs and to the knowledge of RNAses, whose specificity was known, that these first sequencing could take place. In addition, it was possible to determine the secondary structure of tRNA, since hybridization between bases was known at the time. It was in 1971 that the first DNA molecule was sequenced. This molecule consisted of a sequence of 12 nucleotides, the sequence of the cohesive ends of phage lambda [83]. These early sequences were obtained using specific chemical reactions, such as depurination. These methods made it possible to obtain long sequences of 10 to 20 nucleotides.

In 1975, Sanger and Coulson introduced the method of chain termination for DNA sequencing. In 1977, Maxam and Gilbert devised a method similar to Sanger's, but instead used nucleotides that did not allow chain elongation. That same year, Sanger introduced the dideoxynucleotide method, a method that allowed up to 100 nucleotides to be sequenced. This technique allowed the sequencing of the genome of the phage PhiX [75].

The next big innovation in the history of sequencing has been the automation of protocols and analysis (First Generation) [84]. This important advance made it possible to democratize sequencing, to the point of allowing the sequencing of complete genomes (95%), including the human genome in February 2001 [81] and the genome of TriTryp in 2005.

The second generation of sequencing tools appeared in 2005 in response to the high price and low throughput of first-generation sequencing. Here, tens of thousands of sequences are processed together and in parallel. This is the emergence of high-throughput sequencing ("next-generation sequencing").

While the human genome sequencing project cost $ 3 billion and lasted 13 years (completed in 2006), that of James Watson (79, co-discoverer of DNA structure) cost a million dollars and was completed in two months. It was performed on an FLX sequencer from Roche (company 454 Life Sciences, Baylor College of Medicine, Houston, Texas, United States,). Four months later, the Craig Venter Institute published Craig Venter's complete genome [85]. Unlike that of James Watson, this one was sequenced according to the classic Sanger technique. In 2009, one of the co-founders of Helicos Bionsciences, Stephen R. Quake, sequences his genome [86] with a depth of 28x and a genome coverage of 90% at a cost of 48,000 dollars. In the same year (2009), four other human genomes were described: those of a Yoruba man from Nigeria [87] sequenced at a depth of 30x, of 2 Koreans [88] [89] at a depth of 28 and 29x and a Han Chinese [90] at a depth of 36x. These individual sequencing are a major step towards personal medicine.

The NGS platforms currently available in the market use second generation high throughput sequencing technologies offered by Roche 454 Life Sciences, Illumina, Solid and Ion Torrent and the third generation (“next-next-generation sequencing”) offered by Pacific Biosciences (PacBio RS) [91] [92].

1.2.2. Constitution of DNA bank

Any sequencing project begins with the constitution of one or more DNA banks. This library is a collection of DNA fragments to be sequenced that have been integrated into the genome of host cells (usually microorganisms) for storage and replication. Integration is carried out by means of a DNA molecule, called the cloning vector, inside which has been placed a fragment of the DNA that we want to sequence (called in this case ' insert ').

There are two types of DNA banks: genomic DNA banks whose inserts are derived from the fragmentation of the initial genetic material to be sequenced, and cDNA banks whose inserts are mRNAs which have been "copied" into. DNA under the effect of a retrovirus enzyme, reverse transcriptase. Genomic libraries are used in genome sequencing, while cDNA libraries are used in gene expression studies.

Aside from a different first step between building a genomic library and building a cDNA library, the other steps are common to both (Fig. 6).

⇒ **First step:**

- **Genomic bank:** consists in fractionating genomic DNA by partial digestion with an endonuclease, or a restriction enzyme, but physical methods are preferred because they are more reproducible and the fragmentations are more random. These physical methods can involve sonication, high pressure nebulization, or shear force [93].
- **cDNA bank**: no fragmentation is necessary. On the contrary, special care is taken to preserve RNA molecules which are more fragile than DNA, since they consist of only one strand. In this first step, mRNA is reverse transcribed into DNA by the action of reverse transcriptase, a retrovirus enzyme.

⇒ **The other steps:**

They consist of a separation by electrophoresis on agarose gel of the DNA or cDNA fragments in order to select and extract from the gel the fragments of the desired size, then to integrate them into the selected cloning vector. Host cells are then transformed by inserting a vector into their genetic material. Finally, the cells which have been transformed are cultured and isolated in very distinct colonies. Cells that have integrated a vector are selected using one of the vector's markers, usually a gene for resistance to an antibiotic present in the culture medium, and which prevents the multiplication of untransformed cells. Each colony is then sub cultured, preserved and labeled with a unique identifier with a view to its sequencing.

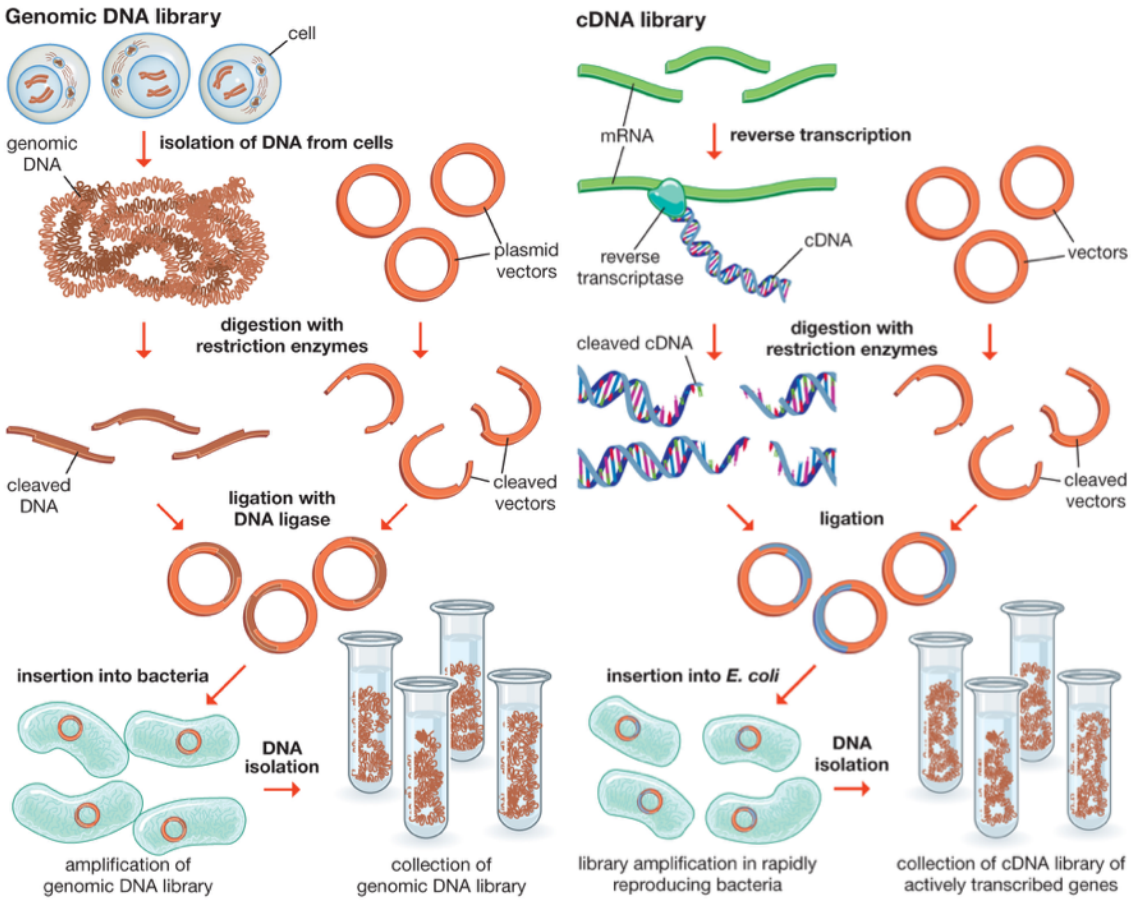


Figure 6: Diagram for the preparation of genomic DNA and cDNA libraries (Encyclopædia Britannica, Inc.)

1.2.3. 1st generation sequencing

a. Sequencing methods

There are two so-called "classic" sequencing methods: the Maxam and Gilbert method, by selective chemical degradation, and the Sanger method by enzymatic synthesis. While the use of the former has remained confidential, the latter has been widely developed and is now the benchmark technique.

- Chemical method [74]:

This method is based on a chemical degradation of DNA and uses the different reactivities of the four bases A, T, G and C, to achieve selective cleavages. By reconstructing the order of the cuts, we can go back to the nucleotide sequence of the corresponding DNA. We can break down this chemical sequencing into six successive steps:

- **Labeling**: The ends of the two DNA strands to be sequenced are labeled with a radioactive label (32P). This reaction is usually carried out by means of radioactive ATP and polynucleotide kinase.

- **Isolation of the DNA fragment** to be sequenced: This is separated by means of electrophoresis on a polyacrylamide gel. The DNA fragment is cut from the gel and recovered by diffusion.
- **Separation of strands**: The two strands of each DNA fragment are separated by thermal denaturation, then purified by a new electrophoresis.
- **Specific chemical modifications**: Single-stranded DNAs are subjected to specific chemical reactions of the different basic types. Walter Gilbert developed several types of specific reactions, carried out in parallel on a fraction of each strand of labeled DNA: for example, a reaction for Gs (alkylation by dimethyl sulfate), a reaction for Gs and Gs. A (depurination), a reaction for C, as well as a reaction for C and T (alkaline hydrolysis).
- **Clipping**: After these reactions, the DNA is cleaved at the modification level by reaction with a base, piperidine.
- **Analysis**: For each fragment, the products of the different reactions are separated by electrophoresis under denaturing conditions and analyzed to reconstitute the DNA sequence. This analysis is analogous to that carried out for the Sanger method (Fig.7).

The chemicals used in the reaction media during specific cuts being excessively dangerous for health, this method was abandoned in favor of the method by enzymatic synthesis.

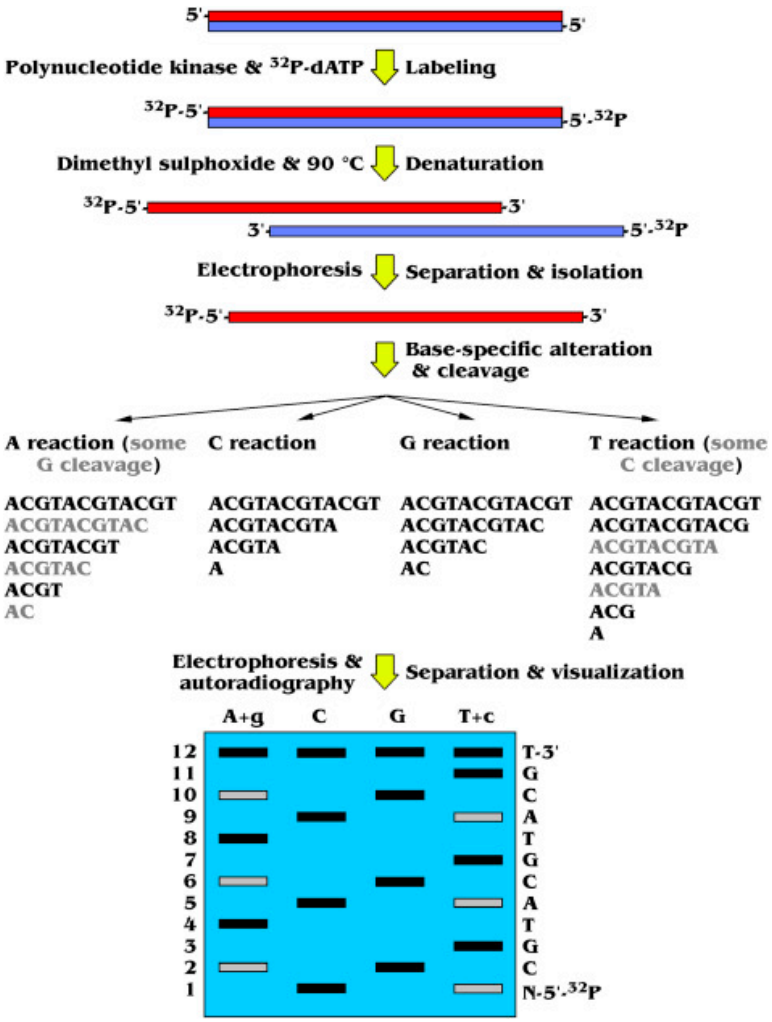


Figure 7: Maxam-Gilbert sequencing *[94]*

- Enzymatic method [75]:

This method, also called the Sanger method because of its inventor, is based on the activity of DNA polymerase, which makes it possible to polymerize a strand of DNA complementary to a template strand, from an oligonucleotide, called a primer. This capacity is used to synthesize a complementary strand, but in an incomplete manner, by stopping the reaction randomly so as to statistically obtain products resulting from a reaction interrupted at each of the bases of the fragment to be sequenced.

The reaction mix consists of the cloning vector containing the fragment to be cloned, the polymerase, the primers and the dNTPs. To randomly stop the reaction, a low concentration of ddNTP is added. Since these ddNTPs do not contain a 3'-OH group, they act as terminators of the polymerization reaction by preventing the completion of a subsequent 5'-3 'phosphodiester bond.

b. Automated Sanger method

Manual reading of sequencing gels is tedious and prone to multiple misinterpretations, especially after reading the first thousand bases. This is why this task has been automated with the help of sequencers. The step of sequencing by enzyme synthesis is still necessary. It is delegated to robots which carry it out in microplates of 96 or 384 wells, corresponding to as many sequenced clones. These plates are then transferred into a sequencer, which will perform electrophoresis while recording on a computer the light intensity profile of the fluorochromes. These profiles are called chromatograms.

There are two types of Sanger technology automatic sequencers:

⇒ Flat gel sequencers have a gel sandwiched between two glass plates top of which are arranged wells.

⇒ Capillary sequencers such as ABI 3730 DNA Analyzer from Life technologies.

This technology has enabled progress in genomics since its commercialization in 2002. In 2004, using the ABI3700 automatic capillary sequencer, as part of a collaborative project, the *P. atrosepticum* SCRI1043 genome was sequenced, which allowed the characterization of the virulence factors of this species by comparison with pathogens of plants or animals of the *Enterobacteriaceae* family [95].

1.2.4. Next Generation Sequencing (NGS)

New sequencing technologies known as NGS (Next Generation Sequencing), have made it possible to sequence DNA faster with better quality while reducing the cost which has fallen from more than 1000 dollars for a million bases in 2001 to less than $ 100 in 2020 (http://www.genome.gov/sequencingcosts/).

Benchtop sequencers have been developed to allow access to NGS technologies to most laboratories, including medical microbiology laboratories, and no longer only to specialized laboratories. They are characterized by lower cost, small size and limited performance but sufficient for most NGS applications in standard research laboratory [96] [97].

1.2.5. 2nd generation sequencing

The Sanger sequencing technique has reached its limits in terms of yield. It is only possible to speed up sequencing by increasing the number of gels, capillaries or machines, and the speed of electrophoresis can hardly be accelerated without reducing the quality of the sequencing. This is why new technologies have emerged to result in massively parallelized real-time sequencers (base-by-base sequencing), using a wide variety of techniques such as:

- High-throughput pyrosequencing (Technology 454)
- Reversible cyclic termination techniques (Solexa / Illumina and Helicos)
- Ligation sequencing (SOliD)

- Sequencing by nano-measurement of PH (PGM)

Despite their methodological differences, the high-throughput sequencing techniques currently available follow a protocol with similar broad lines [87]. Although the methods used at each of the steps are different, the order in which they are performed is invariable:

- Preparation of a library of sequences marked by adapters;
- Amplification of the marked sequences so that they are separated spatially;
- Sequencing by cyclic enzymatic reactions measured in real time.

a. Pyrosequencing (454 technology)

Pyrosequencing is the technique which is currently the most successful competing with the Sanger method. This DNA sequencing technique, introduced by Hyman since 1988, has been improved by the introduction of a PCR step. It is a sequencing by synthesis which is characterized by the real-time detection of DNA polymerase activity and which adds only one non-fluorescent nucleotide at a time.

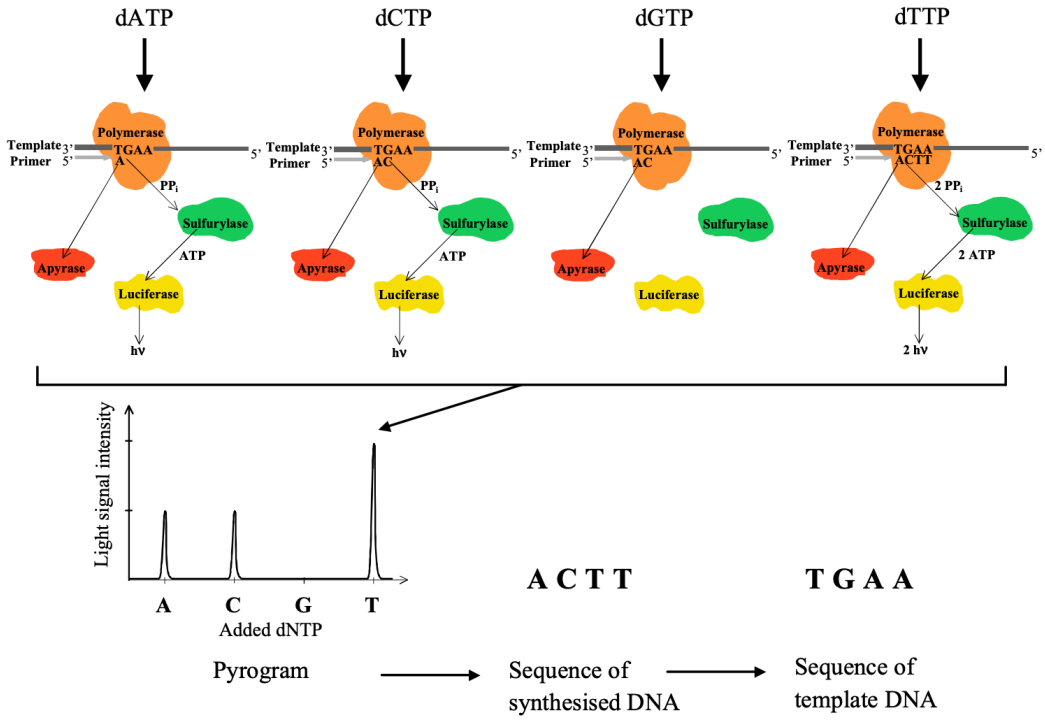


Figure 8: Schematic representation of the pyrosequencing enzyme system. *[98]*

The 454 Life Sciences technology designed by Rothberg [99] is based on the integration of several techniques:

- Pyrosequencing,
- Pico-titrated optical fiber plate technologies (PTP),
- PCR in emulsion "emPCR" in microreactors,
- Advanced computer technologies for the acquisition, processing and image analysis.

b. Reversible cyclic termination techniques (Illumina / Solexa)

Initially developed by Solexa, the Illumina Genome Analyzer platform (Sandiego, USA) has implemented a DNA chip type slide sequencing technology, based on the integration of several techniques: DNA biochips, nanotechnology, a variant of the Sanger technique called reversible cyclic termination (RCT), as well as advanced computer technologies for image acquisition, processing and analysis [100].

The specificity of this technology is based on a bridge amplification (PCR bridge) of the fragments to be sequenced. It takes place on a glass surface called a flow cell (FC), similar to a microscope slide, divided into eight lines (originally one line per sample). The fragments of the library to be sequenced have adapters at their ends. These will allow them to bind randomly to the FC, by hybridization to the primers which cover its surface (Fig. 12). A new strand is then synthesized by a polymerase (a): it is covalently attached to the FC. The original strand is then removed by denaturation (b) and the free end of the remaining strand hybridizes to an adjacent primer to form a bridge (c). The polymerase synthesizes the complementary strand again to form a double-stranded DNA bridge (d) and then the two copies are released by denaturation (e). The bridging amplification cycle (steps c to e) begins again to eventually form a cluster of clonal DNA in an area called cluster (f). The antisense strands (corresponding to the green primers) are then cleaved (g): this is linearization.

The free 3 'end of the DNA fragments is blocked and the sequencing primer hybridizes to it (Fig. 9). Sequencing is carried out on hundreds of millions of clusters simultaneously, thanks to a chemistry of reversible terminators: blocked nucleotides labeled by fluorescence are added, one of them is incorporated, the fluorescence emitted is recorded then the fluorophore and the blocker are cleaved allowing the addition of a new nucleotide. At each cycle of incorporation, a base can be determined. [101].

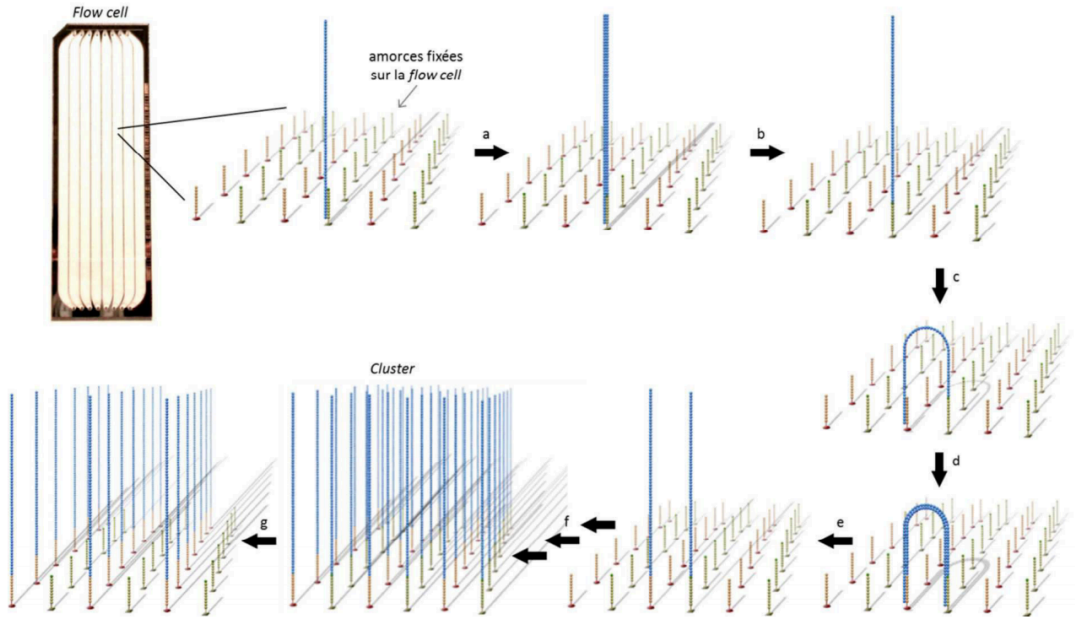


Figure 9: NGS Illumina technology by bridge amplification (Images taken from Illumina documents (http://www.illumina.com/)

Today the HiSeq 2000 system has a yield of 6 billion sequences for a total of about 600 Gb in 11 days (with an average sequence size of 100 bp). This technology is widely used today (sequencing, resequencing, Chip-Seq, RNA-Seq).

c. Ligation sequencing (SOliD)

The SOLiD (Sequencing by Oligonucleotide Ligation and Detection) platform of the system described by Shendure et al. [102] employs sample preparation and amplification processes similar to those developed for the 454 technology. But the sequencing process differs considerably, in that it relies on "ligation" sequencing, where fluorescently labeled 8 base oligonucleotides are sequentially linked to the sequencing primer. Each of the four types of probe carries a separate 3′ fluorophore, representing the sequence of the sample complementary to the fourth and fifth bases of the oligonucleotide probe. The fluorescent signal generated by the hybridization of the probe is then detected and recorded by a laser scanner and a data acquisition process similar to the Solexa system. The bound probe is then cleaved between the 5th and 6th base, removing the fluorescent half, and repeated rounds of probe ligation [91]. Using this method, it is possible to read up to 1.5 billion parallel sequences 75 bases long (short fragments). The built-in error correction system associated with the use of the ligase makes this technology very reliable and usable in genome and transcriptome resequencing projects.

d. Sequencing by nano-measurement of PH (PGM)

In 2007, Jonathan M. Rothberg founded the company Ion Torrent, bought in 2010 by Life Technologies, which will market in December 2010 its first sequencer, the PGM (Personal

Genome Machine) for $ 50,000. The technology uses a synthetic sequencing method, similar to pyrosequencing, based on the natural release of an H + ion upon the incorporation of a nucleotide into a fragment of DNA during production. The sequence is determined by measuring the variations in pH following the incorporation of the different nucleotides, within a semiconductor chip. This "simple" detection system does not require neither optical system nor fluorescent marker, make Ion Torrent technology a fast and inexpensive sequencing method compared to other NGS platforms.

The latest version of this sequencer (chip 318) now allows it to read 1 Gb with great precision in just two hours and the read length will reach 400 bp in 2012. The PGM was thus used to decode the genome of the strain d 'Escherichia coli which wreaked havoc in Europe and caused a health scandal in May 2011 [103].

1.2.6. 3rd generation sequencing

The third generation of mass sequencing is symbolized by the sequencing of a single DNA molecule "SMS" [104]. Unlike the second generation, no DNA (or RNA) amplification is required to perform the measurements. Only one molecule is "read".

These innovations would reduce time and considerably simplify the preparation steps for DNA banks. In addition, the elimination of these steps would also make it possible to reduce the possible sources of bias that may be introduced by the PCR phases (polymerization errors or recombinations by PCR). Another advantage of this technique lies in the fact that weakly concentrated or even degraded DNAs could still become usable [105].

a. HeliScope technology

Helicos Biosciences has developed the first single molecule sequencer: the HeliScope Single Molecule Sequencer. Unlike the chemistry used on the Illumina platform, here the nucleotides are labeled with the same fluorophore and this technology appears perfectly suited to quantitative transcriptomic studies. Helicos no longer sells a device and offers a sequencing service.

b. SMRT (Single Molecule Real-Time Analysis) technology

PacBio sequencing technology, developed by the company Pacific Biosciences (California, USA), makes it possible to sequence a DNA molecule in real time (SMRT: single molecule real time) using an instrument, the PacBio RS. The sequencing process is carried out on a support called “SMRT Cell” composed of tens of thousands of wells in the form of a detector structure called ZMW (zero-mode waveguide). At the bottom of each well, with a diameter of 100 nm, is fixed a single polymerase. These ZMW structures allow to record in real time the incorporated dNTPs where each of the 4 nucleotides is linked to a different flurochrome.

The signals obtained are recorded and analyzed by advanced computer methods [101] [106].

For the assembly of reads from the PacBio sequencer, the company has developed several assembly algorithms adapted according to the size of the reads and the sequencing coverage. In the case of bacterial genomes, it is HGAP (Hierarchical Genome Assembly Process) which includes several steps in the assembly process [107].

The sequencing speed of this technology is 10 bases per second with an average read length of over 1000 bp (up to 3000 bp). Despite its success, PacBio technology also generates sequencing errors linked to the quantification of homopolymers.

c. Nanopore technologies

Several platforms allowing the sequencing of a DNA molecule by passage through a nanopore of a biological or synthetic membrane are currently in development. The different nucleotides A, C, G and T are successively detected during their passage through the pore by optical systems (labeling of nucleotides) or electrical (modification of the current according to the degree of pore occupation and the transit time, specific to each nucleotide).

The nanopores used can be solid (graphene) or biological (α-hemolysin, Mycobacterium smegmatis porin A)

1.2.7. Emerging technologies

A fourth wave of sequencers will emerge, highlighting the qualities of each of the platforms mentioned above.

a. FRET technology (Fluorescence Resonance Emission Transfer)

This sequencing technology is based on the distance-dependent energy transfer between two molecules: a donor and an acceptor. Sequencing is performed synthetically using DNA polymerase (donor) and nucleotides labeled with a fluorophore (acceptors).

The DNA molecule to be sequenced is deposited on a surface in the presence of a mobile DNA polymerase: the labeled nucleotides are then added to the medium. When the polymerase incorporates a nucleotide on a DNA fragment being synthesized, there is a transfer of energy between the donor group and the acceptor group by proximity. Fluorescence is emitted, and detected.

A sequencing platform using this technology is currently under development: **Starlight (Life Technologies)**.

b. In situ sequencing technology

A team recently developed a technique for sequencing short fragments of RNA within tissues to study the gene expression profile in situ. The 4th generation of sequencers may also allow the sequencing of DNA or RNA molecules directly from a cell or tissue [97] [108].

1.2.8. High throughput sequencing applications

The applications available with high throughput sequencing can broadly be grouped into three broad categories.

a. De novo sequencing

This is the sequencing of a genome for which there is no reference sequence, therefore determining an unknown sequence. In order to successfully obtain good quality genome versions, it is often necessary to combine several methods. Pyrosequencing allows, thanks to long reads, to build a first version of the genome skeleton when the methods by reversible terminators will correct the errors present in this first reconstruction to produce a draft of the genome of quality. In the medical field, these tools are used for the discovery of genomes of unknown pathogens or new viruses [109].

b. Resequencing

In genome studies, the term resequencing is used instead of sequencing. This name, mainly used in genetics, designates the sequencing of a DNA segment followed by the comparison of the result obtained with that of a known reference sequence. High-throughput sequencing is then used to find out what the genomic variations are in the sample being studied compared to the one taken as a reference. These approaches are certainly among the most widely used in the medical field today. These tools are typically intended to replace traditional methods of comparative genomic hybridization (CGH) and to make it possible to specify diagnoses either in a preventive manner or to characterize a pathology already declared.

2. Comparative genomics and use of genomes

2.1. Genome assembly

Sequence assembly is the process of aligning, orienting and fusing the DNA fragments obtained during the sequencing phase. This step is necessary because DNA sequencing cannot output the entire genome in one read. Instead, small sequences of 20bp to 40,000bp (depending on the technology used) are obtained, often using fragments from shotgun sequencing. From the reads from the sequencing phase, the assembler (a computer program) will try to put the readings together based on their overlap. Assembly is generally done in 2 steps:

- First, the fragments are put together using overlapping reads to build larger complete sequences called contigs. The information used is thus generally limited to single readings (without the pair information).
- Then comes the phase of scaffolding (scaffolding): this step aims to orient and place the contigs with respect to each other to obtain a unique sequence, although having gaps

(i.e. holes in the final sequence). These shortcomings will be corrected during the genome finishing step, which aims to fill the remaining holes in the assembly by experimental methods, such as PCR or primer extension [110].

The assembly of complete genomes for comparison requires a verification phase. The quality criteria for a complete genomic sequence are sufficient coverage, a small number of contigs and a minimum of uncertainty in the order of the contigs. The stages of finishing and validating a sequencing project are generally long and difficult, depending both on the sequencing method used, on the assembly strategy and on the richness of repeated elements of the genomes. In addition, the construction of genomic DNA libraries or the use of assembly programs (and their parameterization) can generate errors detrimental to the subsequent analysis of sequences such as chimeric contigs and reading artifactual phase shifts.

In general, there are two types of assemblies, the assembly with genomic reference and de novo assembly.

2.1.1. Assembly by reference - Mapping

Assembly with reference makes it possible to align the results from the sequencer with a reference genome existing in international biological databases. The assembly by reference (or mapping) is a process which makes it possible to compare the sequences obtained by the sequencer with a known genomic sequence (reference) in order to identify common areas, to determine functional sites (catalytic site, interaction zone, ...) and predict the function(s) of a protein (if detected). This process allows a number of predictions.
The available alignment tools are based on two different algorithms: *hash table method* [111] and *Burow Wheler Method* [112].

2.1.2. De Novo assembly

Even if a reference genome is available, de novo assembly [113] must be performed, as it allows the recovery of transcripts transcribed from genome sequences missing in the genome assembly. This type of genome assembly involves analyzing the reads so as to reconstruct long sequences via possible overlaps between reads and combining the sequences obtained from the sequencer into contiguous sequences, called contigs.

New generation sequencers generate two types and for each type of read an assembly algorithm is dedicated: Overlap-Layout-Consensus (OLC) type algorithms are suitable for assembling long reads [114], These algorithms start by calculating all possible overlaps between all supplied pairs of reads [115]. This method is one of the first successfully used to assemble a genomic sequence [113]. The general principle is to merge overlapping sequences,

starting with those with the most significant overlaps, until a single sequence is formed. Therefore, most of the available assemblers capable of handling typical data generated by Illumina use a De Bruijn graph-based and k-mer-based approach. The principle of this algorithm aims to simplify the search for a Hamiltonian path in the grahe by transforming it into a search for a Eulerian path [116]. This method begins at first, to fragment the sequences which are to be assembled into subsequences of the same size (k), called k-mer. Two k-mer will be able to be assembled if their sequences diverge only by the first nucleotide of one and the last nucleotide of the other. In this way it is possible to establish a link between each of the k-mer [116] [117]. Then the overlapping sequences are cut into 4-mer. The nodes of the graph represent the common sequences of length k-1 = 3. Two subsequences are linked if there is a sequence match between the first 3 nucleotides of one and the last 3 of the other. It is then possible to provide the consensus sequence. One of the parameters essential to a good use of DBG, the size of k-mer, is not so easy to handle. The size of k-mer chosen greatly changes the appearance of the graph produced. There is an optimal size of k-mer which we will describe as "k" which is an odd number [118] necessarily smaller than the size of reads, the choice of the value of k-mer could become detrimental to assembly.

2.2. Genome annotation

Genomic annotation is the identification of all the significant elements of genetic material and their role, it consists in analyzing the nucleotide sequence which constitutes the raw information in order to extract the biological information necessary for understanding the functioning of the cell and the relationships between genes [119].

The annotation can be automatic, that is to say rely solely on algorithms looking for similarities (in sequence, structure, patterns, etc.), making it possible to predict the function of a gene. It results in the "automatic" transfer of the information contained in the label of a gene "similar" from a genome already annotated to the genome being annotated [120].

The initial automatic annotation is sometimes supplemented by a manual annotation by experts who validate or invalidate the prediction according to their knowledge or experimental results. This can thus avoid the automatic transfer of errors and therefore their propagation, which can become the great problem that genomics will have to face, given the massive influx of data resulting in particular from new sequencing techniques [120].

Classically, there are three main stages in the genomic annotation process: syntactic annotation, functional annotation and relational annotation [119]. These three stages are initially based on the use of sophisticated algorithmic tools, the development of which constitutes one of the fields of bioinformatics.

2.2.1. Syntax annotation

Syntactic annotation, or structural annotation, it is the step which makes it possible to identify the genetic objects having a biological relevance, the main part of these objects are the sequences encoding proteins, RNA molecules (tRNA, rRNA, SsRNA...) [119]. But, also pseudo-genes, retrotransposons, certain repeated sequences and regulatory sequences of gene expression can also be detected thanks to their particular structural motifs. The detection of these regions of interest is done using ab initio predictions and by comparison of sequences or motifs [120].

2.2.2. Functional annotation

Functional annotation is the step that makes it possible to predict the potential functions of previously identified genetic objects and to collect any experimental information (literature, large-scale datasets) [119]. It is based on the search for homology between the unknown sequence considered and a set of annotated sequences. All of this research (homology or intrinsic analysis of CDS) thus makes it possible to identify several pieces of information useful for understanding the function of the predicted gene and its product. However, the prediction will have to be validated experimentally to guarantee its annotation quality [120].

2.2.3. Relational annotation

Relational, or contextual, annotation uses more complex information than information attached to sequences. It determines the relationships likely to exist between the elements predicted and characterized previously. It appeared about ten years ago [121] [122] when the first fully sequenced genomes had sufficient data to establish relationships between the elements of the genome. These relationships are of various kinds [120]:

- Homology: proteins can be grouped into families of homologs;
- Physical interaction: the elements interact physically with each other: protein / acids nucleics, protein / protein and nucleic acids / nucleic acids;
- Common involvement in a biological process: participation in the same path metabolic, same transport route or same regulatory network.

2.3. Contribution of comparative genomics

Interest in sequencing was stimulated by a need for knowledge. In fact, sequencing the genome of a micro-organism gives access to all of its genetic information. However, this knowledge is more valuable if it is put into perspective and compared with other genomic or phenotypic information.

The sequencing and comparison of several genomes within a species or a bacterial genus makes it possible to define the central genome (or “core-genome”) and the total genome (or “pan-genome”) of the species. The **pan-genome** describes the total number of genes found at

least once in a species or genus. Indeed, bacteria can have large variations in gene content between closely related strains. The analysis of these genes therefore makes it possible to understand the evolution of a bacterial group, which is particularly relevant in the study of metagenomes. However, it can also be used in a more restricted genomic context. The pan-genome includes the core-genome containing the genes present in all the organisms in question, the accessory genome containing the genes present in two or more strains or species, and the unique genes, specific to a single strain or species. The core genome generally includes all the genes responsible for major aspects of the physiology and major phenotypic characters of a species or genus. In contrast, accessory and unique genes contribute to strain or species diversity and may encode additional metabolic pathways and functions which are not essential for growth but which may confer an adaptive advantage to the organism, such as by example adaptation to an ecological niche, resistance to antibiotics or the colonization of a particular host. Such genes are generally found in genomic islands [123] [124].

The analysis of accessory genes therefore makes it possible to understand the specificities of each organism sequenced. Note here that these distinctions are not strictly biological because they depend in part on the strains or species included in the analysis as well as the parameters used for the comparisons and grouping of genes.

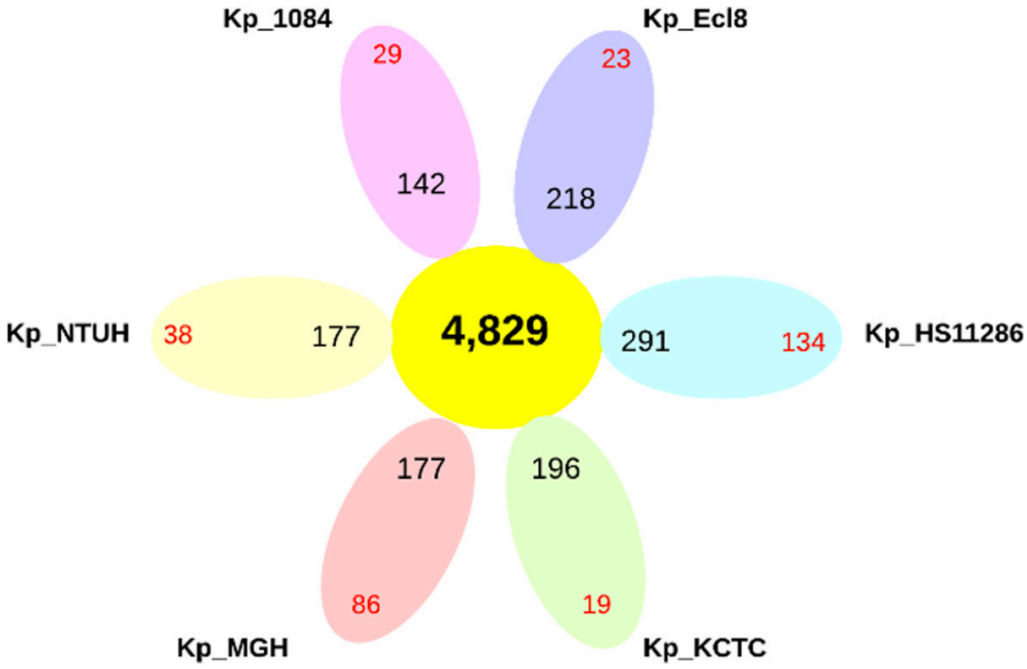


Figure 10: Pan-genome representation for 6 studied strains of *Klebsiella pneumoniae*. The number of core genes is shown in the yellow circle. For each strain, the number of accessory genes is shown in black and the number of unique genes is shown in red. *[125]*

The in silico comparative studies carried out on the sequences of microorganisms have made it possible to progress in many fields concerning in particular the structure and diversity of genomes and in particular the identification of virulence genes and pathogenicity determinants for pathogenic species. These genes can be many and varied. These are mainly genes encoding toxins, macromolecule export systems (or secretion systems) and genes encoding functions related to interactions with the host such as genes encoding adhesion functions, genes for the synthesis of surface polysaccharides and molecules of interaction with the host's immune system and iron uptake systems (siderophores) which are sought as determinants of virulence.

The gene content of an organism reflects its possibilities for interactions with its environment. For example, the presence of genes encoding enzymes of a given catabolic pathway in a genome indicates that the microorganism is probably capable of using certain resources present in its environment. Also, comparing the contents of several microorganisms in genes encoding systems for importing and exporting macromolecules such as polysaccharides or polyamino acids can provide information on the ability of these microorganisms to use resources and nutrients. present in their living environments under the physico-chemical conditions necessary for their growth. The gene content of a microorganism genome therefore reflects its ecological niche [103] [126] [127].

Comparison of genomic data from multiple isolates within a species identifies genes ubiquitous in the species. The contribution of genomes from other closely related species can then make it possible to identify the specific genes of this species. These specific and ubiquitous genes can for example be used as targets for the development of detection means or vaccines.

MATERIALS & METHODS

Chapter two

The main objective of this study is to contribute to improving and deepening the knowledge of the complete genome of the K. Pneumoniae species of Moroccan origin, and to characterize the resistance for each strain and establish a comparative genomic analysis**.**

Genome assembly by SPAdes

After checking the quality of the raw data (fastq) the assembly was done using the SPAdes software [128] which is an assembler created for small genomes and especially for the genomes of bacteria. SPAdes builds a DBG from k-mer of different sizes. Thus, a smaller value of k-mer in regions with low coverage minimizes fragmentation, while a larger value of k-mer in regions with high coverage reduces the number of chimeric reads. Contigs are then deduced from the graph. It is recommended for SPAdes to increase the length of k-mer from 21 until the length k-mer reaches 127.

⇒ **Quality control of the assembly:**

The evaluation of the quality of the result of the assembly in "*. Fasta" format obtained by the software SPAdes, was carried out by the software Quast [129], which allows to calculate several parameters (number of contigs / scaffolds, size of the largest contig, N50, GC rate ...) in order to define good de novo assembly quality.

The quality criteria for a complete genomic sequence are:

- Number of contigs: The total number of contigs obtained after assembly.
- Size of the assembled genome: the number of bases assembled.
- N50: means that half of the genome sequence is covered by contigs larger than or equal the N50 contig size. [130].

Annotation by RAST **(Rapid Annotations using Subsystems Technology)**

The annotation of the four *K. Pneumoniae* strains genomes in this study were generated using the RAST server pipeline.

RAST is a web server dedicated for annotation of prokaryotic genomes especially for bacteria and archaea. This tool is based on annotation using the 2.0 subsystem. It can identify the coding sequences of proteins (CDS), ribosomal RNA and transfer RNA using FIGfams and Glimmer which makes it possible to predict the positions of genes in microbial DNA, in particular the genomes of bacteria and viruses. It provides high quality genomic annotations. In addition, the annotated genome can be browsed in an environment that supports comparative analysis with the annotated genomes maintained in the SEED environment. [131]

The FASTA file containing the contigs from our genome assembly is uploaded to the server which will allow us to specify a set of gene calls like the annotation scheme where we used the

ClassicRAST for all strains. Normally, the annotated genomes are available within 12 to 24 hours of submission.

Multi-locus sequence typing (MLST)

The sequence types of the four *K. Pneumoniae* strains "KP02-M, KP03-M, KP04-M, KP06-M" were determined through multilocus sequence typing (MLST).

MLST for *K. pneumoniae* was developed by Diancourt et al., and is currently used in the world over to determine the ST of isolates [132]. This is done by obtaining the DNA sequences of seven genes and by plotting the sequences into a database for comparison using (https://pubmlst.org/bigsdb?db=pubmlst_mlst_seqdef&page=sequenceQuery).

The database will report allele numbers for each gene, and the combination of alleles of housekeeping genes in *K. pneumoniae* [rpoB (β-subunit of RNA polymerase B), gapA (glyceraldehyde-3-phosphate dehydrogenase), mdh (malate dehydrogenase), pgi (phosphoglucose isomerase), phoE (phosphoporine E), infB (translation initiation factor 2), and tonB (periplasmic energy transducer)] gives the sequence type.

Antimicrobial resistance genes

ResFinder 4.0 includes the AMR genes database (ResFinder). The ResFinder entries have been associated with an AMR phenotype both at the antimicrobial class and at the antimicrobial compound level. A selection of antimicrobial compounds within each class was made to include antimicrobial agents important for clinical and surveillance purposes for the different bacterial species included. [133]

To identify acquired resistance genes, we first have to specify a bacterial species (in our case: *Klebsiella Pneumoniae*) in order to define the specific antimicrobial panel for the *in silico* antibiogram [134].

ResFinder is freely available at (https://cge.cbs.dtu.dk/services/ResFinder/). It is based on a database of more than 2,000 resistance genes covering 12 types of antimicrobial resistance agents (aminoglycoside, betalactamase, fluoroquinolone, fosfomycin, fusidic acid, glycopeptide, macrolide-lincosamide-streptograminB, phenicol, rifampicin, sulphoamide, tetracycline, and trimethophorim), which is searched using BLAST.39 The threshold for reporting a match between a gene in the ResFinder database and the input phage genome was set to be 90% identity over at least 60% of the length of the resistance gene for all four strains.

Plasmids

For the PlasmidFinder database, a total of 745 sequences related to nonredundant, complete sequences of plasmids identified in bacterial species belonging to the family Enterobacteriaceae were collected from the NCBI nucleotide database (http://www.ncbi.nlm.nih.gov/nuccore/). This database was used to build a Web tool freely available at (http://cge.cbs.dtu.dk/services/PlasmidFinder/) utilizing the BLASTn algorithm to look for DNA homologies in both raw and assembled sequencing data from four different sequencing platforms [135].

In our case, the four replicon sequences from the PlasmidFinder database were BLASTed against the assembled genomes, and the best-matching hits in each genome for each replicon sequence were given as output, using a selected % ID threshold value of 95% and a selected minimum length of 60%.

Pangenome

We constructed the pan-genome from the proteins identified in the complete genomes of the four strains of this study. The orthoMCL tool was used to define the orthologous proteins shared between the 4 proteomes of *K. pneumoniae*. Thus, we can predict the size of the pan-genome, the core-genome (the proteins shared by the 4 strains), the accessory genome (shared by at least 2 strains) and specific (the proteins found only in a single strain) [125], using the best reciprocal success strategy in a command line with 70% for both identity and coverage.

RESULTS

Chapter three

Annotation and analysis of the *K. pneumoniae* genome

In this study, we used genomes associated with four *K. Pneumoniae* strains isolated from clinical samples of hospitalized patients in Morocco.

1. Genome sequencing

Genomes sequencing was done in the Medbiotech Laboratory by Illumina MiSeq using the Nextera XT DNA Library Prep V500 kit.

The MiSeq platform is the first system to offer a fully integrated sequencing ecosystem.

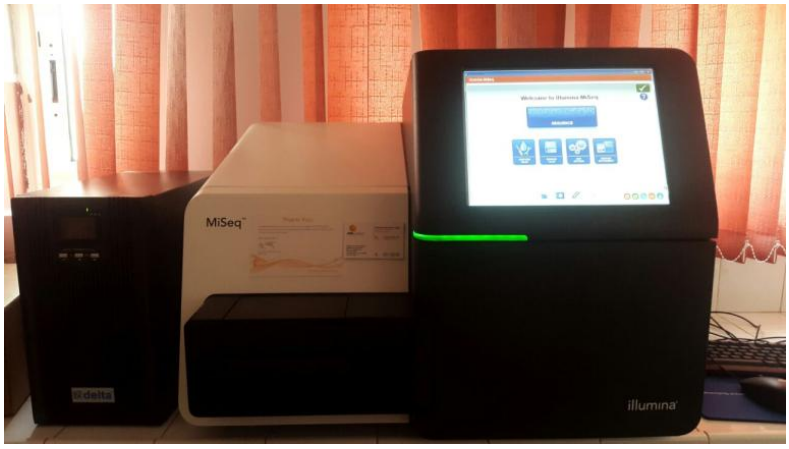


Figure 11: Illumina Miseq Sequencer

2. Genome assembly by SPAdes

De novo assembly for K. Pneumonia strains was done by the SPAdes 3.7.1 assembler. The results obtained with SPAdes are detailed in Table 2.

All the studied *K. pneumoniae* genomes had an average length of 5.53 Mb. The number of contigs varies from 81 for KP06-M to 138 for KP04-M including contigs with a greater size than 1000 bp varying from 59 to 106 for KP06-M and KP04-M respectively. The results also show that KP06-M has the highest N50 (285494 bp), meaning that half of the final SPAdes assembly is found in contigs larger than 285494 bp length.

Table 2: Results of the de Novo assembly with SPAdes.

	Strains			
	KP02-M	**KP03-M**	**KP04-M**	**KP06-M**
Contigs	83	98	138	81
Contigs (>= 1000 bp)	66	70	106	59
Total length	5441045	5461352	5672726	5580918
GC (%)	57.25	57.15	57.00	57.18
N50	207845	165526	152985	285494

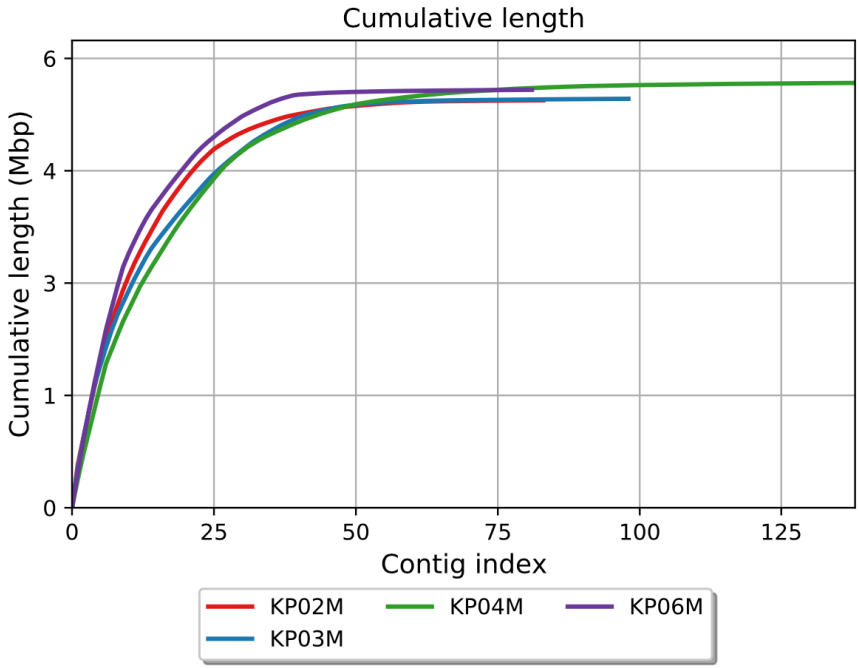


Figure 12: Cumulative Length Graph (Quast)

Quast makes it possible to compare all the strains in a graphical representation of cumulative length which shows the growth of the contig lengths (Fig.12). On the x axis, contigs are ordered from smallest to largest. The y axis gives the size of the largest x contigs in the assembly.

3. Annotation by RAST

K. pneumoniae HS11286 was used as a reference.

The general characteristics of the genome of the four studied strains are presented in Table 3.

Table 3: Annotation results.

Characteristics	*Strains*			
	KP02-M	**KP03-M**	**KP04-M**	**KP06-M**
Genome size (Mb)	5.44	5.46	5.67	5.58
GC Content (%)	57.3	57.2	57.0	57.2
Number of Contigs (with PEGs)	83	98	138	81
Number of Subsystems	597	602	589	597
Number of Coding Sequences	5188	5139	5401	5362
Number of RNAs	99	104	107	97

This annotation results predicted the number of subsystems, coding sequences (CDS) and RNAs of each strain with their genome size and GC content.

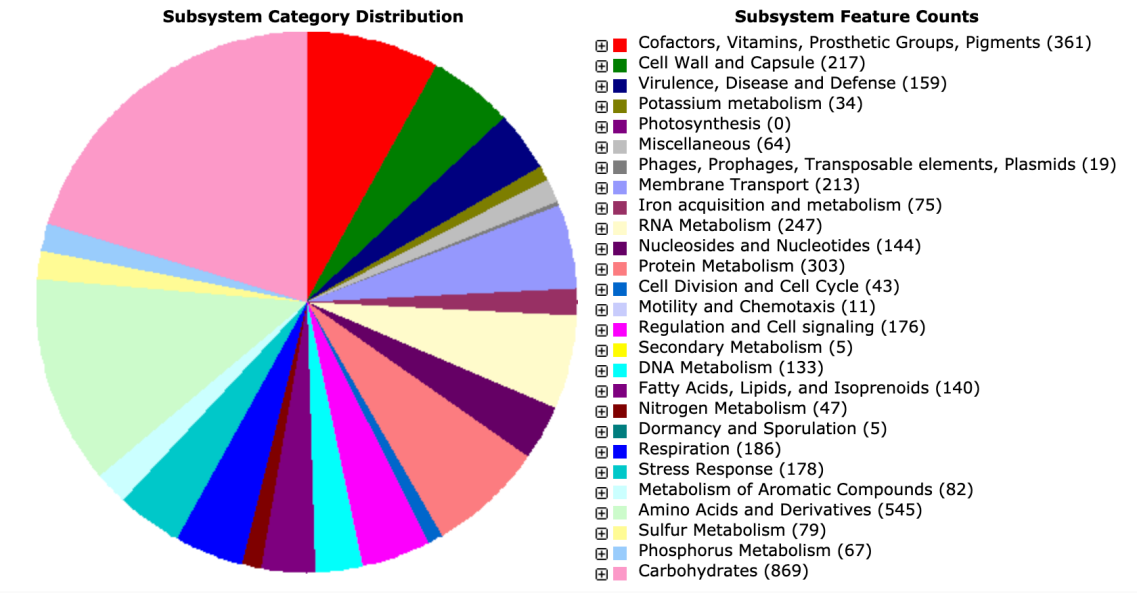


Figure 13: Annotated genes distribution of KP02-M strain.

Fig. 13 shows that most of the annotated genes have been involved in the metabolism of carbohydrates (869), amino acids and their derivatives (545), cofactors, vitamins, prosthetic groups and pigment formation (361), Protein metabolism (247), RNA metabolism (247), cell

wall and capsule development (217). This bacterial genome also carries 159 virulence, disease and defense genes.

4. Multi-Locus Sequence Type (MLST)

The determination of the Multi-Locus Sequence Type (MLST) is essential to classify the four strains of our study in a comprehensible and comparable global context. It revealed a novel combination of alleles known to household genes of *K. pneumoniae* MLST (gapA, infB, mdh, pgi, phoE, rpoB and tonB).

Based on this combination using the PubMLST system, we found four different type sequences for each of the four *Klebsiella pneumoniae* strains; ST664, ST13, ST2695, and ST35 belonged to KP02-M, KP03-M, KP04-M, and KP06-M, respectively

Table 4: A combination of the alleles of seven household genes of *K. pneumoniae*.

Strain	***Locus***							***Sequence Type***
	gapA	*infB*	*mdh*	*pgi*	*phoE*	*rpoB*	*tonB*	
KP02M	gapA_52	infB_1	mdh_20	pgi_2	phoE_1_508	rpoB_1	tonB_47	664
KP03M	gapA_2	infB_3	mdh_1	pgi_1	phoE_10	rpoB_1	tonB_19	13
KP04M	gapA_2	infB_1	mdh_1	pgi_1	phoE_4	rpoB_4	tonB_373	2695
KP06M	gapA_2	infB_1	mdh_2	pgi_1	phoE_10	rpoB_1	tonB_19	35

5. Antimicrobial resistance genes

As shown in Table 5 below, the four strains of *K. pneumoniae* had multiple resistance genes detected with significant identity.

In total 31 genes were identified and distributed with a different resistance profile in each of the four strains, of which three antibiotic resistance genes were present in both the four strains, namely, oqxA, oqxB, and fosA, belonging to two antimicrobial classes: fluoroquinolone and fosfomycin respectively. Resistance to beta-lactams was also observed in all strains but across different resistance genes.

Remarkably, we noticed that the KP04-M strain having the most genes compared to other *K. pneumoniae* strains, including the tet (A), catB3, and ARR-3 genes are respectively associated with resistance to tetracycline, phenicol, and rifamycin was found only in this strain (KP04-M).

In addition, the folate pathway antagonist and aminoglycoside resistance genes were present in only KP04-M and KP03-M. While aminocyclitol aadA1 resistance genes were found only in strain KP03-M.

Table 5: Antimicrobial resistance genes identified in 4 strains of *Klebsiella Pneumoniae.*

Antimicrobial class	***Resistance genes***	***Strains***			
		KP02-M	**KP03-M**	**KP04-M**	**KP06-M**
Aminoglycoside	aadA1	-	+	-	-
	aph(3")-Ib	-	-	+	-
	aph(6)-Id	-	-	+	-
	aac(3)-IId	-	-	+	-
	aac(6')-Ib-cr	-	-	+	-
Aminocyclitol	aadA1	-	+	-	-
β-lactam	blaSHV-12	-	-	+	-
	blaSHV-33	-	-	-	+
	blaSHV-40	+	-	-	-
	blaSHV-56	+	-	-	-
	blaSHV-79	+	-	-	-
	blaSHV-85	+	-	-	-
	blaSHV-89	+	-	-	-
	blaOXA-1	-	+	+	-
	blaCMY-4	-	-	+	-
	blaTEM-1B	-	-	+	+
	blaTEM-106	-	-	-	+
	blaTEM-126	-	-	-	+
	blaTEM-135	-	-	-	+
Fluoroquinolone	oqxA	+	+	+	+
	oqxB	+	+	+	+
	qnrB9	-	-	+	-
	qnrB32	-	-	+	-
	aac(6')-Ib-cr	-	-	+	-
Fosfomycin	fosA	+	+	+	+
Folate pathway antagonist	dfrA14	-	-	+	-
	sul1	-	+	+	-
	sul2	-	-	+	-
Phenicol	catB3	-	-	+	-
Rifamycin	ARR-3	-	-	+	-
Tetracycline	tet(A)	-	-	+	-

Note: (+): present, (-): absent.

6. Plasmids

PlasmidFinder showed that all strains contained at least one plasmid from the incompatibility group F (IncF), namely IncFIB(K). With the exception of the KP04-M strain where its plasmid (pCAV1099-114) exhibits significant homology of 99.64% with the *K. Oxytoca* plasmid.

Our plasmid analysis also revealed that the IncF plasmid detected in KP03-M, KP04-M, and KP06-M was associated with the same plasmid replicons of *K. pneumoniae* ST258 (pKPN-IT). The IncFII(K) plasmid was detected only in the "KP03M" strain and also represented a significant 97.97% homology with the *K. pneumoniae* plasmid MGH78578.

In addition, two plasmids detected Col(RNAI) and Col(BS512) belonged to strains KP03M and KP06M respectively. These two plasmids showed significant homology with the plasmids of other strains, in particular the plasmid Col (BS512) which represented 100% homology with the plasmid *Shigella boydii* CDC 3083-94. While plasmid types such as IncR, IncC, IncX have also been predicted in KP02-M, KP04-M, and KP06-M showing significant homology to K. *pneumoniae* NK245, *K. pneumoniae* Kp7, and *Escherichia coli* respectively.

Table 6: Plasmids predicted for 4 strains of *Klebsiella Pneumoniae.*

Plasmid	***Strains***			
	KP02-M	**KP03-M**	**KP04-M**	**KP06-M**
ColRNAI	-	+	-	-
Col(BS512)	-	-	-	+
IncC	-	-	+	-
IncFIB(K)	+	+	-	+
IncFIB(K)(pCAV1099-114)	-	-	+	-
IncFII(K)	-	+	-	-
IncR	+	-	-	-
IncX1	-	-	-	+

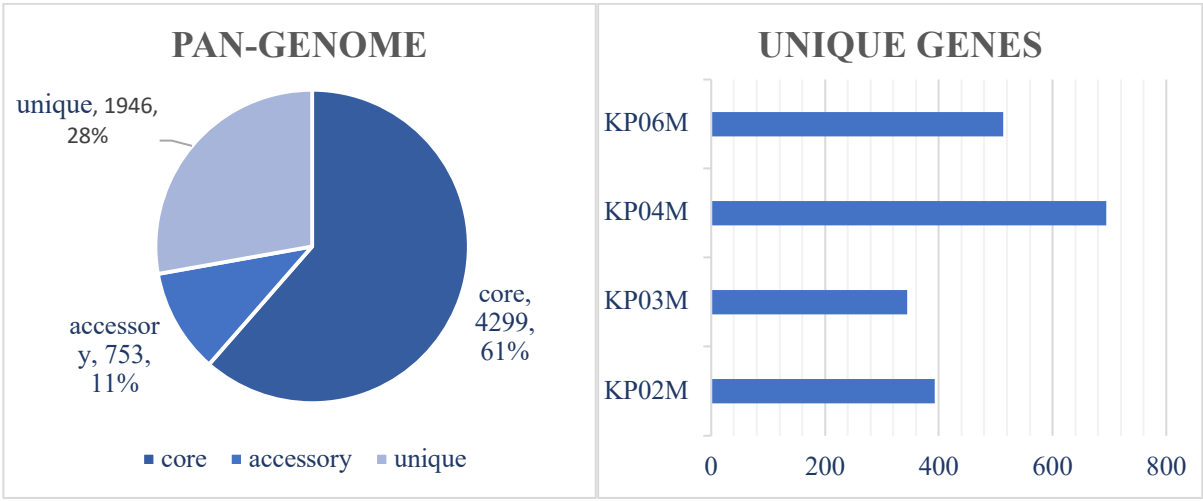
Note: (+): present, (-): absent.

7. Pan-Genome

In order to study the genomic variability between *K. pneumoniae* strains, we constructed a pan-genome by grouping together the sets of annotated proteins from each of the 4 strains.

Our results showed that the pan-genome size of the four strains in this study was 6998 protein clusters, of which 4299 proteins belonged to the core (found in all four strains) and the core / pan-genome ratio was 61% (Fig. 14a), which could indicate a medium conservation rate among these strains. While the accessory protein (the variable part between strains) was only 753 (11%) proteins shared by at least two but not all strains.

It is also interesting to note that strain KP04-M contained the highest number of unique proteins (found in a single strain) with 695 proteins (Fig. 14b) followed by KP06-M, KP02-M, and KP03-M with 513, 393 and 74 proteins, respectively.



a. *b.*

Figure 14: a. A pie chart showing the composition of the pan-genome. b. Unique genes of the four strains of *K. pneumoniae.*

DISCUSSION

Chapter four

K. pneumoniae clinical isolates are progressing toward increasing levels of antimicrobial drug resistance, placing this species among the infectious bacterial pathogens that are most challenging to control [136]. Although bacterial genomics has provided a better understanding of the evolutionary mechanisms of bacterial genomes, only little is known about the genomics of this specific pathogen. Therefore, the availability of new genome sequences from *K. pneumoniae* strains with differential antibiotic resistance profiles furthers our understanding and appreciation of drug resistance mechanisms and their evolution.

De Novo assembly of the four *K. Pneumoniae* strains was performed by the SPAdes assembler based on the DBG method. It was done on the set of k-mers generated by SPAdes in order to obtain a partially complete *K. pneumonia* sequence. The genomic annotation of the four *K. pneumoniae* strains contigs was performed with the platform RAST. All the studied *K. pneumoniae* genomes had an average length of 5.53 Mb. The KP02-M strain genome was the smallest with only 5.44 Mb and KP04-M was the largest genome with 5.69 Mb. The GC content varied from 57.0 % for KP04-M to 57.3 % for KP02-M, with an overall average of 57.17 %. The number of coding sequences (CDS) varied from 5139 for KP03-M to 5401 for KP04-M. The number of RNAs also differed depending on the strains, ranging from 97 tRNA in KP06-M to 107 in KP04-M. Most of the annotated genes have been involved in the metabolism of carbohydrates, amino acids and their derivatives, cofactors, vitamins, prosthetic groups and pigment formation, the metabolism of RNAs, cell wall and capsule development. It may also carry genes related to virulence and defense. This annotation results are consistent with other *K. pneumoniae* strains already annotated and published [137] [138]. *K. pneumoniae* genomes genetic diversity has been previously demonstrated to be primarily due to elements that often migrate by horizontal gene transfer, including plasmids, phages, integrated conjugative elements (ICEs) and insertion sequences (ISs) [139].

MLST is one of the gold standards for determining the epidemiological link of organisms and for reliable classification, including determination of multilocus sequence type, based on a combination of alleles of seven housekeeping genes (rpoB, gapA, mdh, pgi, phoE, infB and tonB). In this study, the MLST analysis revealed four different sequence types (ST664, ST13, ST2695 and ST35) belonging to the four clinical strains KP02-M, KP03-M, KP04-M and KP06-M respectively.

These STs were observed in various countries. For example, the sequence-type ST13 was identified since the 1980s in numerous regions of the world as the widely spread *K. pneumoniae* genetic line [140]. In recent years, MDR strains of this genetic line carrying the TEM-3 and

DHA-1 beta-lactamase genes were identified in Spain, and OXA-48-producing strains in Finland, Ireland, and Algeria [141] [142] [143]. ST13 has also been reported in Curaçao during an ESBL-producing *K. pneumoniae* outbreak in 2005 [132] and in Spain during a DHA-1-producing *K. pneumoniae* outbreak in 2007 [144]. The sequence-type ST35 has also been reported in Tunisia (SHV-12 in 2002) [145], in Spain (CTX-M-15 in 2008) [146], and in Denmark (without ESBL in 2008) [147].The sequence-type ST664 was reported in a study in Malawi and Tunisia [148] [149] but no further studies to identify specifically this sequence-type were made. As for the sequence-type ST2695, that happens to be an original strain and no study was ever made upon it.

Combined with an accurate epidemiological information and the characterization of antibiotic resistance mechanisms, MLST analysis of greater sample sets can provide a much-improved explaining of the evolutionary origin and the spread of outbreak *K. pneumoniae* strains. [132]

Thirty-one antibiotic resistance genes were identified in total. 8 in the genome of KP02-M, 7 in that of KP03-M, 20 in that of the KP04-M and 8 in that of the KP06-M strain (Table 5). The KP02-M strain harbored multiple beta-lactamases: blaSHV-40, blaSHV-56, blaSHV-79, blaSHV-85 and blaSHV-89. It also acquired resistance genes able to impair effectiveness of fluoroquinolones (oqxA, oqxB), and fosfomycins (fosA); however, genes conferring resistance to aminoglycosides were not found.

The only β-lactam resistance gene in the KP03-M genome is blaOXA-1. It harbored other drug resistance genes, such as aadA1 encoding for aminoglycoside and aminocyclitol; oqxA and oqxB encode fluoroquinolone resistance; fosA encodes fosfomycin resistance and sul1 encodes folate pathway antagonist resistance.

KP04-M is highly resistant to a number of different antibiotics but particularly refractory to the fluoroquinolone class. Moreover, we have also found other drug resistance genes, such as blaSHV-12, blaOXA-1, blaCMY-4 and blaTEM-1B encoding β-lactam resistance; aph(3")-Ib, aph(6)-Id, aac(3)-IId and aac(6')-Ib-cr encoding aminoglycoside resistance; oqxA, oqxB, qnrB9, qnrB32 and aac(6')-Ib-cr encode fluoroquinolone resistance (It is acknowledged that aac(6′)-Ib-cr is a variant of the aac(6′)-Ib gene which can induce resistance against aminoglycoside and fluoroquinolone simultaneously. [150]); fosA encodes fosfomycin resistance; dfrA14, sul1, sul2 encode folate pathway antagonist resistance, catB3 encodes phenicol resistance; ARR-3 encodes rifamycin resistance and tet(A) encodes tetracycline resistance.

As for the KP06-M strain, it harbored blaSHV-33, blaTEM-1B, blaTEM-106, blaTEM-126, and blaTEM-135 encoding resistance to beta-lactams, oqxA and oqxB encoding resistance

to fluoroquinolones and fosA to fosfomycin; nevertheless, genes conferring resistance to aminoglycosides were not found. These findings demonstrate the multi-drug resistant phenotype of these *K. pneumoniae* isolates.

K. pneumoniae generally harbors more than one plasmid. In this study, we could identify eight plasmids from the total DNA of the four strains namely ColRNAI, Col(BS512), IncC, IncFIB(K), IncFIB(K)(pCAV1099-114), IncFII(K), IncR and IncX1. 2 in KP02-M, 3 in KP03-M, 2 in KP04-M and 3 in KP06-M strain (Table 6).

IncF plasmids were the commonest plasmids seen among *K. pneumoniae,* playing a significant role in the dissemination of antimicrobial resistance in *Enterobacteriaceae*. They contribute to the fitness of their bacterial hosts by providing virulence and antimicrobial resistance genes and displaying rapid evolution [151]. In this study, all isolates were identified with IncF type for instance in KP02-M with IncFIB(K), KP03-M with IncFIB(K) and IncFII(K), KP04-M with IncFIB(K)(pCAV1099-114) and KP06-M with IncFIB(K) as well.

The IncFII plasmid family exists in various Enterobacter species and plays an important role in the spread of antibacterial resistance genes [152]. It has also been reported in the United States, Israel, the United Kingdom, Italy and Colombia [139]. Other types were included as IncX, IncR, IncC plasmids in KP06-M (IncX1), KP02-M and KP04-M respectively. A Col replicon was also detected in two isolates (KP03-M and KP06-M) namely ColRNAI and Col(BS512) respectively. Col plasmids are related to colicin production by and against Escherichia coli and related bacteria [153].

The pan-genome for the four strains of *K. Pneumoniae* contained 4299 core genes (Fig. 14a) and the core/pan-genome ratio was 61 %. This percentage was indicative of a medium rate of conservation among these strains. Compared to Aurélia Caputo et al. study the core/pan-genome ratio (94 %) was greater. This high percentage (more than 90 %) was indicative of a high rate of conservation among their strains [125].And regarding unique genes, they are defined as those found solely in only one strain, which reflects differences between strains. In this study, pan-genome analysis revealed 1946 unique genes in the four strains, KP04-M had the highest number of unique genes (695) which happens to be the same strain with most AMR genes. We conclude that the emergence of a pan-genome would allow for the development of a more reliable approach to species classification, in which species are defined as distinct biological entities with large differences [125].

CONCLUSION AND PERSPECTIVES

Comparative analyses of the clinical *K. pneumoniae* strains have demonstrated that this species possesses an extremely plastic genome. Multiple resistances to various classes of antibiotics are also observed in hospital-adapted *K. Pneumoniae* due to the accumulation of antibiotic resistance genes that could be encoded on several plasmids [154].

Our results show the presence of an easily distinguishable *K. Pneumoniae* genomes and illustrate that MDR populations of this species are mainly non-intersecting. However, *K. pneumoniae* is beginning to evolve toward rising levels of antimicrobial drug resistance and may represent a new and significant challenge to public health.

Four strains of *K. Pneumoniae* were subjects of a de Novo assembly followed by an annotation in order to identify the genetic elements (resistance genes, plasmids...) of the strains. This genomic analysis pipeline made it possible to identify 31 resistance genes from different antimicrobial classes (β-lactam, aminoglycosides, Fosfomycin, fluoroquinolone…), 8 plasmids of different types (IncF, IncX, IncR, IncC, Col), 4 sequence types using MLST (ST664, ST13, ST2695, ST35) and a pan-genome of 4299 core genes and 1946 unique genes.

In conclusion, our study provided a global view of the information carried in the genome of *K. Pneumoniae*, hence the need of setting up a succession genomic data analysis therefore maintain a surveillance program for controlling antimicrobial resistance.

As perspectives, mining the totally sequenced *K. pneumoniae* genomes would be helpful to reveal the key roles of mobile genetic elements in the adaptive evolution and spread of antibiotic resistance. As well as a proper identification and analysis of annotated unique genes would be very ideal to better our understanding of factors that determine the specificity of resistance and virulence genes.

BIBLIOGRAPHY

[1] Kariuki S., Corkill J.E., Revathi G., Musoke R., Hart A., Keynan Y. and Rubinstein E., "The changing face of Klebsiella pneumoniae infections in the community.," *International journal of Antimicrobiol Agents.,* vol. 6, pp. 2474-2479, 2007.

[2] Jack N Pendleton, Sean P Gorman and Brendan F Gilmore, "Clinical relevance of the ESKAPE pathogens," *Expert Review of Anti-infective Therapy, Vol 11,* p. 297–308, 2013.

[3] Okomo U, Akpalu ENK, Le Doare K, Roca A, Cousens S, Jarde A, Sharland M, Kampmann B, Lawn JE., "Aetiology of invasive bacterial infection and antimicrobial resistance in neonates in sub-Saharan Africa: a systematic review and meta-analysis in line with the STROBE-NI reporting guidelines.," *Lancet Infect. Dis.,* vol. 19, p. 1219–1234, 2019.

[4] Zaidi AK, Huskins WC, Thaver D, Bhutta ZA, Abbas Z, Goldmann DA., "Hospital-acquired neonatal infections in developing countries.," *Lancet,* vol. 365, p. 1175–1188, 2005.

[5] World Health Organization., "Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics," 2017.

[6] Alessandro Cassini, Liselotte Diaz Högberg, Diamantis Plachouras, Annalisa Quattrocchi, Ana Hoxha, Gunnar Skov Simonsen, et al., "Attributable deaths and disability- adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis," *Lancet Infect. Dis.,* vol. 19, p. 56–66, 2019.

[7] M.C.El Bouamri, L.Arsalane, Y.El Kamouni, S.Zouhair, "Antimicrobial susceptibility of urinary Klebsiella pneumoniae and the emergence of carbapenem-resistant strains: A retrospective study from a university hospital in Morocco, North Africa," *African Journal of Urology,* vol. 21, no. 1, pp. 36-40 , 2015.

[8] Bush K, Jacoby GA., "Updated functional classification of betalactamases.," *Antimicrob. Agents. Ch,* vol. 54, p. 969–976, 2010.

[9] Krause KM, Serio AW, Kane TR, Connolly LE., "Aminoglycosides: an overview.," *Csh. Perspect. Med.,* vol. 6, p. a 027029, 2016.

[10] Renaud, Freney, Cédric Leclercq, Précis de bactériologie clinique., 2000.

[11] George M., Garrity Julia A., Bell and Timothy G.L., "Taxonomic Outline of the Procaryotes.," *In Bergey's Manual of Systematic Bacteriology.,* 2004.

[12] Brooks, G.; Carroll, K.; Butel, J. and Morse, S, Jawetz, Melnick, and Adelberg's Medical Microbiology., McGraw-Hill, 2007.

[13] Irving, P.; Ala'Aldeen, P. and Boswell, T., Medical Microbiology., New York.: Taylor and Francis e-Library, 2006.

[14] Brisse, S.; Fevre, C.; Passet, V.; Issenhuth, J.; Tournebize, R.; Diancourt, L. and Grimont, P., "Virulent Clones of Klebsiella pneumoniae:Identification and Evolutionary Scenario Based on Genomic and Phenotypic Characterization.," *J. PLoS ONE.,* vol. vol.4(3), p. 4982, 2009.

[15] J. Turton, C. Perry and H. C. Elgohari, "PCR characterization and typing of Klebsiella pneumoniae using capsular typespecific, variable number tandem repeat and virulence gene targets.," *J. Med. Microbiol.,* vol. 59, pp. 541-547, 2010.

[16] Keynan, Y. and Rubinstein, E., "The changing face of Klebsiella pneumoniae infections in the community.," *Int. J. Antimicrob. Agents.,* vol. 30, pp. 385-389, 2007.

[17] A. Leavitt, I. Chmelnitsky, R. Colodner, I. Ofek and S. Carmeli, "Ertapenem resistance among extended-spectrumβ-lactamase-P producing Klebsiella pneumoniae isolates.," *J. Clin. Microbiol,* vol. 47, pp. 969-974, 2009.

[18] Arafa N., Smati F., Scheftel J.M. and Meunier O., "Caractérisation phénotypique et génotypique de souche de Klebsiella pneumoniae subsp pneumoniae isolées a l'hôpital universitaire de Constantine, Algerie.," *Sciences and Technologie,* pp. 43-49., 2009.

[19] Raud P., "Etude de la diversité génétique des souches de Klebsiella pneumoniae productrices de beta-lactamases à spectre étendu (BLSE), isolées au CHU de Nantes, de 1990 à 2001.," pp. 39-41, 77-79, 2003.

[20] Paczosa MK, Mecsas J., "Klebsiella pneumoniae: going on the offense with a strong defense.," *Microbiol Mol Biol Rev.,* vol. 80, p. 629–661, 2016.

[21] Broberg C.A., Palacios M. and Miller V.L., "Klebsiella: a long way to go towards understanding this enigmatic jet-setter," *F1000Prime Reports,* vol. 6, p. 64, 2014.

[22] Bansal S., Harjai K. and Chhibber S., "Aeromonas punctata derived depolymerase improves susceptibility of Klebsiella pneumoniae biofilm to gentamicin.," *BMC.Microbiology.,* vol. 15, p. 119, 2015.

[23] C. N., Clegg S., Murphy., "Epidemiology and virulence of Klebsiella pneumoniae.," *Microbiol. Spectrum.,* vol. 4, pp. UTI-0005-2012, 2016.

[24] Russell A., Denyer S. P., Hodges N. A. and German S. P., Types of antibiotics and synthetic antimicrobial agents., Blackwell Science, UK: Hugo and Russells pharmaceutical microbiology, 2004, pp. 152-186.

[25] L., Mangin, "Antibiotiques et résistances : enquête sur les connaissances et les comportementsdugrandpublic.," Thése de doctorat, université de Lorraine, 2016.

[26] C., Walsh, Antibiotics: actions, origins, resistance., Washington, DC: ASM Press, 2003.

[27] B.Demoré, M.Grare and R.Duval, "Chapitre 40 : Généralités sur les antibiotiques par voie systémique et principes d'utilisation.," in *Pharmacie clinique et thérapeutique.*, Elsevier Masson (Eds), 2012, p. 198.

[28] A., Ebimieowei E. Ibemologi, "Antibiotics: Classification and mechanisms of action with emphasis on molecular perspectives.," *International Journal of Applied Microbiology and Biotechnology Research.,* vol. 4, pp. 90-101, 2016.

[29] K., Firoz., "Antibiotics Classification and Visual Target Sites for Bacterial Inhibition.," *Adv. Pharmacol. Clin. Trials.,* vol. 3, p. 000137, 2018.

[30] Sahra Kırmusaoğlu, Nesrin Gareayaghi and Bekir S. Kocazeybek, "Introductory Chapter: The Action Mechanisms of Antibiotics and Antibiotic Resistance," in *Antimicrobials, Antibiotic Resistance, Antibiofilm Strategies and Activity Methods*, 2019.

[31] J. Davies and D. Davies, "Origins and evolution of antibiotic resistance.," *Microbiol. Mol. Biol. Rev.,* vol. 74, pp. 417-433, 2010.

[32] V. D'Costa, E. Griffiths and G. Wright, "Expanding the soil antibiotic resistome: exploring environmental diversity.," *Curr. Opin. Microbiol.,* no. 10, pp. 481-489, 2007.

[33] Y. Hu, X. Yang, J. Li, N. Lv, F. Liu, J. Wu, I. Lin, N. Wu, B. Weimer, G. Gao, Y. Liu and B. Zhu, "The bacterial mobile resistome transfer network connecting the animal and human microbiomes.," *Appl. Environ. Microbiol.,* no. 82, pp. 6672-6681, 2016.

[34] M. Alekshun and S. Levy, "Molecular mechanisms of antibacterial multidrug resistance.," *Cell,* no. 128, pp. 1037-1050, 2007.

[35] M. Drancourt, C. Bollet, A. Carta, P. Rousselier, "Phylogenetic analyses of Klebsiella species delineate Klebsiella and Raoultella gen. nov., with description of Raoultella ornithinolytica comb. nov., Raoultella terrigena comb. nov. and Raoultella planticola comb. nov," *Int.J.Syst.Evol.Microbiol,* vol. 51, pp. 925-932, 2001.

[36] P. Courvalin et al., Antibiogramme., vol. 2, 2006, pp. 142-162;27-246;263-277.

[37] H. Boucher, G. Talbot, J. Bradley, J. Edwards Jr, D. Gil- bert, L. Rice, M. Scheld, B. Spellberg and J. B. B. Bartlett, "An update from the infectious diseases Society of America.," *Clin. Infect. Dis.,* vol. 48, pp. 1-12, 2009.

[38] Nikaidio H., "Prevention of drug access to bacterial targets: permeability barriers and active efflux," *Science,* vol. 264, pp. 382-388, 1994.

[39] Wachino, J. and Y. Arakawa, "Exogenously acquired 16S rRNA methyltransferases found in aminoglycoside-resistant pathogenic Gram-negative bacteria: an update.," *Drug Resist.Updat,* vol. 15, pp. 133-148, 2012.

[42] Beauclerk, A, and E. Cundliffe, "Sites of action of two ribosomal RNA methylases responsible for resistance to aminoglycosides," *J.Mol.Biol,* vol. 193, pp. 661-671, 1987.

[43] Tolmasky, Maria S. Ramirez and Marcelo E., "Aminoglycoside modifying enzymes," *Drug Resist.Updat.,* vol. 13, pp. 151-171, 2011.

[44] Navon-Venezia, S., Kondratyeva, K. and Carattoli, A., "Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance.," *FEMS Microbiol. Rev.,* vol. 41, p. 252–275, 2017.

[45] S. Motta, P. Cluzel and M. Aldana, "Adaptive resistance in bacteria requires epigenetic inheritance, genetic noise, and cost of efflux pumps.," *PLoS. One,* vol. 10, p. e0118464, 2015.

[46] K. Poole, "Stress responses as determinants of antimicrobial resistance in Gram-negative bacteria.," *Trends Microbiol.,* vol. 20, pp. 227-234, 2012.

[47] W. Krzyzanski and G. Rao, "Multi-scale model of drug induced adaptive resistance of Gram-negative bacteria to polymyxin B.," *PLoS. One,* vol. 12, pp. 1-17, 2017.

[48] L. Fernández and R. Hancock, "Adaptive and mutational resistance: role of porins and efflux pumps in drug resistance.," *Clin. Microbiol. Rev.,* vol. 25, pp. 661-681, 2012.

[49] D. Dar and R. Sorek, "Regulation of antibiotic-resistance by non- coding RNAs in bacteria.," *Curr. Opin. Microbiol.,* vol. 36, pp. 111- 117, 2017.

[50] S. Džidić, J. Šušković and B. Kos, "Antibiotic resistance mechanisms in bacteria: biochemical and genetic aspects.," *Food Technol. Biotechnol.,* vol. 46, pp. 11-21, 2008.

[51] J. Galán, F. González-Candelas, J. Rolain and R. Cantón, "Antibiotics as selectors and accelerators of diversity in the mechanismsof resistance: from the resistome to genetic plasticity in the β-lactamases world.," *Front Microbiol.,* vol. 4, no. 9, pp. 1-17, 2013.

[52] H. Flemming and J. Wingender, "The biofilm matrix.," *Nat. Rev. Microbiol.,* vol. 8, pp. 623-633, 2010.

[53] Daniela Santos Pontes, Rodrigo Santos Aquino de Araújo, Natalina Dantas, Luciana Scotti, Marcus Tullius Scotti, Ricardo Olímpio de Moura and Francisco Jaime Bezerra Mendonça-Junior, "Genetic Mechanisms of Antibiotic Resistance and the Role of Antibiotic Adjuvants," *Current Topics in Medicinal Chemistry,* vol. 18, pp. 42-74, 2018.

[54] George Y. Lesher, Ernest J. Froelich, Monte D. Gruett, John Hays. Bailey, and R. Pauline. Brundage, "1,8-Naphthyridine Derivatives. A New Class of Chemotherapeutic Agents," *J. Med. Chem.,* vol. 5, no. 5, p. 1063–1065, 1962.

[55] Lewin, C.S., B.M. Howard, and J.T. Smith., "Protein- and RNA-synthesis independent bactericidal activity of ciprofloxacin that involves the A subunit of DNA gyrase.," *J.Med.Microbiol,* vol. 34, pp. 19-22, 1991.

[56] Aldred KJ, Kerns RJ and Osheroff N., "Mechanism of quinolone action and resistance.," *Biochemistry,* vol. 53, p. 1565–1574., 2014.

[57] Redgrave LS, Sutton SB, Webber MA and Piddock LJ, "Fluoroquinolone resistance: mechanisms, impact on bacteria, and role in evolutionary success," *Trends. Microbiol.,* vol. 22, p. 438–445, 2014.

[58] GA, Hooper DC and Jacoby, "Topoisomerase inhibitors: fluoroquinolone mechanisms of action and resistance," *Cold. Spring. Harb. Perspective. Med.,* vol. 6, p. a025320, 2016.

[59] Jacoby GA, Chow N. and Waites KB, "Prevalence of plasmid24 mediated quinolone resistance.," *Antimicrob. Chemother Agents.,* vol. 47, pp. 559-562, 2003.

[60] Meradi L., Djahoudi A., Abdi A., Bouchakour M., Perrier Gros Claude D. and Timinouni M., "Resistance to quinolones of types qnr, aac (6 ′) - Ib-cr in Enterobacteriaceae isolated in Annaba, Algeria Qnr and aac (6 ′) - Ib-cr types quinolone resistance among Enterobacteriaceae isolated in Annaba, Algeria.," *Pathology Biology,* vol. 59, pp. e73-e78, 2009.

[61] Martínez-Martínez L., Eliecer Cano M., Manuel Rodríguez-Martínez J., Calvo J. and Pascual A., "Plasmid-mediated quinolone resistance.," *Expert. Rev. Anti. Infect. Ther.,* vol. 6, p. 685–711, 2008.

[62] Hansen LH, Johannesen E., Burmolle M., Sorensen AH and Sorensen SJ, "Plasmid encoded multidrug efflux pump conferring resistance to olaquindox in Escherichia coli.," *Antimicrob. Chemother Agents.,* vol. 48, pp. 3332-3337, 2004.

[63] Holt KE, Wertheim H., Zadoks RN, Baker S., Whitehouse CA and Dance D., "Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, anurgent threatto public health.," *Proc. Natl. Acad. Sci. ,* vol. 112, p. E3574 – E3581, 2015.

[64] Beauclerk, A, and E. Cundliffe, "Sites of action of two ribosomal RNA methylases responsible for resistance to aminoglycosides.," *J.Mol.Biol,* vol. 193, pp. 661-671, 1987.

[65] Sally R Partridge, Guy Tsafnat, Enrico Coiera, Jonathan R Iredell, "Gene cassettes and cassette arrays in mobile resistance integrons," *FEMS Microbiol Rev .,* vol. 33, pp. 757-784, 2009.

[66] Tolmasky, M. E., and J.H. Crosa, "Tn1331, a novel multiresistance transposon encoding resistance to amikacin and ampicillin in Klebsiella pneumoniae.," *Antimicrob Agents Chemother,* vol. 31, pp. 1955-1960, 1987.

[67] Castaneda-Garcia A., Blazquez J. and Rodriguez-Rojas A., "Molecular mechanisms and clinical impact of acquired and intrinsic fosfomycin resistance.," *Antibiotics,* vol. 2, pp. 217-236, 2013.

[68] Ma Y., Xu X., Guo Q., Wang P., Wang W. and Wang M., "Characterization of fosA5, a new plasmid-mediated fosfomycin resistance gene in Escherichia coli.," *Lett. Appl. Microbiol.,,* vol. 60, pp. 259-264, 2015.

[69] Guo QL, Tomich AD, McElheny CL, Cooper VS, Stoesser N., Wang MG, SluisCremer N. and Doi YH, "Glutathione-S-transferase FosA6 of Klebsiella pneumoniae origin conferring fosfomycin resistance in ESBL-producing Escherichia coli.," *Journal of Antimicrobial. Chemotherapy.,* vol. 71, pp. 2460-2465, 2016.

[70] Falagas ME, Vouloumanou EK, Samonis G. and Vardakas KZ, "Fosfomycin.," *Clin. Microbiol. Rev.,* vol. 29, p. 321–347, 2016.

[71] Kahan FM, Kahan JS, Cassidy PJ and Kropp H., "The mechanism of action of fosfomycin (phosphonomycin).," *Ann. NY Acad. Sci.,* vol. 235, p. 364–386, 1974.

[72] Avery, O. T., Macleod, C. M., et McCarty, M., "Studies on the chemical of inducing transformation of pneumoccal types. Induction of transformation by a desoxyriboncleic acid fraction isolated from Pneumococcus type III.," *Journal of Experimental Medcine,* vol. 79, pp. 137-158, 1944.

[73] Watson, J. D., et Crick, F. H., "Molecular structure of nucleic acids, a structure for deoxyribose nucleic acid.," *Nature,* vol. 171, pp. 737-738, 1953.

[74] J. Heinrich Matthaei, Oliver W. Jones, Robert G. Martin, Marshall W. Nirenberg, "CHARACTERISTICS AND COMPOSITION OF RNA CODING UNITS," *Proc Natl Acad Sci U S A.,* vol. 48, no. 4, p. 666–677, 1962.

[75] Maxam, A.M. and Gilbert, W., "A new method for sequencing DNA.," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 74, pp. 560-4, 1977.

[76] Sanger, F., Nicklen, S. and Coulson, A.R., "DNA sequencing with chain-terminating inhibitors.," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 74 , pp. 5463-7, 1977.

[77] Cochrane, G., Akhtar, R., Boniefld, J., Bower, L., Demiralp, F., Faruque, N., Gibson, R., Hoad, G., Hubbard, T., Hunter, C., "Petabyte-scale innovations at the European Nucleotide Archive.," *Nucleic Acids Res,* vol. 37, pp. D19-25, 2009.

[77] Dennis A. Benson, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, Eric W. Sayers, "GenBank," *Nucleic Acids Research,* vol. 37, p. D26–D31, 2009.

[78] Lipman, W R Pearson and D J, "Improved tools for biological sequence comparison," *PNAS ,* vol. 85, no. 8, pp. 2444-2448, 1988.

[79] Thompson, J. D., Higgins, D. G., et Gibson, T.J., "CLUSTL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position- specific gap penalities and weight matrix choice.," *Nucleic Acids Res ,* vol. 22, pp. 4673-4680, 1994.

[80] Altschul, S. F., Madden, T.L., Schaffer, A.A., Ahang, Z., Miller, W., et lipman, D.J., "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res,* vol. 25, pp. 3389-3402, 1997.

[81] Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., "Initial sequencing and analysis of the human genome.," *Nature,* vol. 409, pp. 860-921, 2001.

[82] IHGSC, "Finishing the euchromatic sequence of the human genome.," *Nature,* vol. 431, pp. 931- 945, 2004.

[83] R. Wu, "Nucleotide sequence analysis of DNA. I. Partial sequence of the cohesive ends of bacteriophage lambda and 186 DNA.," *Journal of molecular biology,* vol. 51, pp. 501-21., 1970.

[84] C. Hutchison, "DNA sequencing: Bench to bedside and beyond.," *Nucleic acids research,* vol. 35, pp. 6227-37, 2007.

[85] Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, "The diploid genome sequence of an individual human.," *PLoS Biol ,* vol. 5, p. e254, 2007.

[86] Pushkarev, D., Neff, N.F. and Quake, S.R., "Single-molecule sequencing of an individual human genome.," *Nat Biotechnol,* vol. 27, pp. 847- 852, 2009.

[87] 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., "Accurate whole human genome sequencing using reversible terminator chemistry.," *Nature,* vol. 456, pp. 53-59, 2008.

[88] Ahn, S.M., Kim, T.H., Lee, S., Kim, D., Ghang, H., Kim, D.S., Kim, B.C., Kim, S.Y., Kim, W.Y., Kim, C., "The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group.," *Genome Res,* vol. 19, pp. 1622-1629, 2009.

[89] Kim, J.I., Ju, Y.S., Park, H., Kim, S., Lee, S., Yi, J.H., Mudge, J., Miller, N.A., Hong, D., Bell, C.J., "A highly annotated whole-genome sequence of a Korean individual.," *Nature,* vol. 460, pp. 1011-1015, 2009.

[90] Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Guo, Y., "The diploid genome sequence of an Asian individual.," *Nature,* vol. 456, pp. 60- 65, 2008.

[91] M. Metzker, "Sequencing technologies - the next generation.," *Nat Rev Genet,* vol. 11, pp. 31- 46, 2010.

[92] McAdam P, Richardson EJ, Fiztgerald JR., "High throughput sequencing for the study of bacterial pathogen biology.," *Curr Opin Microbiol. ,* vol. 19, pp. 106-13, 2014.

[93] Levinthal C, Davison PF., "Degradation of deoxyribonucleic acid under hydrodynamic shearing forces.," *J. Mol.Bio ,* vol. 3, pp. 674-683, 1961.

[94] P.Hernández, G.Dorado S.Gálvez H.Budak T.Unver, Encyclopedia of Biomedical Engineering, 2019, pp. 443-460 .

[95] Bell, E.F.,Boghossian, N.S., Page, G.P., Stoll, B.J., Murray, J.C., Cotten, C.M., "Late-onset sepsis in very low birth weight infants from singleton and multiple- gestation births.," *J Pediatr. ,* vol. 162, p. 1120–1124., 2013.

[96] Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, "Performance comparison of benchtop high-throughput sequencing platforms.," *Nat. Biotechnol. ,* vol. 30, no. 5, p. 434–9, 2012.

[97] Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J,, "Updating benchtop sequencing performance comparison.," *Nat. Biotechnol. ,* vol. 31, no. 4, p. 294–6, 2013.

[98] Afshin Ahmadian, Maria Ehn, Sophia Hober, "Pyrosequencing: History, biochemistry and future," *Clinica Chimica Acta,* vol. 363, p. 83 – 94 , 2006.

[100] Fedurco M, Romieu A, Williams S, Lawrence I & Turcatti G., "BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies.," *Nucleic Acids Res.,* vol. 34, p. 22, 2006.

[101] Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., "Real-Time DNA Sequencing from Single Polymerase Molecules.," *Science ,* vol. 323 , p. 133–138, 2009.

[102] Shendure J., "The beginning of the end for microarrays?," *Nat Methods. ,* vol. 5, pp. 585-587, 2008.

[103] Mellmann, A., Harmsen, D., Cummings, C.A., Zentz, E.B., Leopold, S.R., Rico, A., Prior, K., Szczepanowski, R., Ji, Y., Zhang, W., "Prospective genomic characterization of the German enterohemorrhagic Escherichia coli 0104:H4 outbreak by rapid next generation sequencing technology.," *PLoS One,* vol. 6, p. e22751, 2011.

[104] Gupta PK., "Single-molecule DNA sequencing technologies for future genomics research.," *Trends Biotechnol. ,* vol. 26, pp. 602-611, 2008.

[105] Orlando L. et al., "True single-molecule DNA sequencing of a pleistocene horse bone.," *Genome Res. ,* vol. 21, p. 1705–1719, 2011.

[106] J. Sengenès, "Développement de méthodes de séquençage de seconde génération pourl'analyse des profils de méthylation de l'ADN.," *Thèse de doctorat, Université Pierre et Marie Curie,* p. 216p., 2012.

[107] Chin, C. S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C.,, "Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data.," *Nat. Methods,* vol. 10, p. 563–569, 2013.

[108] G. I. McGinn S, "DNA sequencing - spanning the generations.," *New Biotechnol.,* vol. 30, no. 4, p. 366–72, 2013.

[109] Gaynor, A.M., Nissen, M.D., Whiley, D.M., Mackay, I.M., Lambert, S.B., Wu, G., Brennan, D.C., Storch, G.A., Sloots, T.P. and Wang, D., "Identification of a novel polyomavirus from patients with acute respiratory tract infections.," *PLoS Pathog,* vol. 3, p. e64, 2007.

[110] Pevzner, P. A., H. Tang, and M. S. Waterman., "An Eulerian path approach to DNA fragment assembly.," *Proc Natl Acad Sci U S A,* p. doi:10.1073/pnas.171285098., 2001.

[111] Li, Heng, and Nils Homer., "A survey of sequence alignment algorithms for next-generation sequencing.," *Briefings in Bioinformatics,* vol. 11, no. 5, pp. 473-83, 2010.

[112] Li, Heng, and Richard Durbin., "Fast and Accurate Short Read Alignment with Burrows- Wheeler Transform," *Bioinformatics (Oxford, England),* vol. 25 , no. 14, pp. 1754-60., 2009.

[113] Miller, Jason R., Sergey Koren, and Granger Sutton., "Assembly Algorithms for Next-Generation Sequencing Data.," *Genomics ,* vol. 95 , no. 6, pp. 315-27., 2010.

[114] R. Staden, "A strategy of DNA sequencing employing computer programs.," *Nucleic Acids Research,* vol. 6, no. 7, pp. 2601-10. , 1979.

[115] M. Pop, "Genome Assembly Reborn: Recent Computational Challenges," *Briefings in Bioinformatics,* vol. 10, no. 4, pp. 354-66., 2009.

[116] Compeau, Phillip E. C., Pavel A. Pevzner, and Glenn Tesler., "How to Apply de Bruijn Graphs to Genome Assembly.," *Nature Biotechnology,* vol. 29, no. 11, pp. 987-91., 2011.

[117] Zerbino, Daniel R., snd Ewan Birney., "Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs," *Genome Research,* vol. 18, no. 5, pp. 821-29., 2008.

[119] Zhou, X. and A. Rokas,, "Prevention, diagnosis and treatment of high-throughput sequencing data pathologies.," *Mol Ecol,* vol. 23, no. 7, pp. 1679-700, 2014.

[120] E. Beyne, " Règles de cohérence pour l'annotation génomique : développement et mise en œuvre in silico et in vivo," 2008.

[121] Goffeau A. et al., "Life with 6 000 genes.," *Science.,* vol. 274, no. 5287, pp. 546,563-567, 1996.

[122] J. Venter et al, "The sequence of the human genome.," *Science,* vol. 291, no. 5507, pp. 1304-1351, 2001.

[123] Medini D, Donati C, Tettelin H, Masignani V, and Rappuoli R., "The microbial pan-genome.," *Current opinion in genetics & development. ,* vol. 15, no. 6, p. 589–94, 2005.

[124] Reinhardt J, Baltrus D, Nishimura M, Jeck W, Jones C, and Dangl J., "De novo assembly using lowcoverage short read sequence data from the rice pathogen Pseudomonas syringae pv. oryzae.," *Genome Research.,* vol. 19, no. 2, p. 294–305, 2009.

[125] Aurélia Caputo, Vicky Merhej , Kalliopi Georgiades , Pierre-Edouard Fournier , Olivier Croce , Catherine Robert and Didier Raoult, "Pan-genomic analysis to redefine species and subspecies based on quantum discontinuous variation: The Klebsiella paradigm," *Biology Direct,* vol. 10, p. 55, 2015.

[126] Bhattacharyya A, Stilwagen S, Ivanova N, D'Souza M, Bernal A, Lykidis A, Kapatral V,Anderson I,Larsen N,Los T,Reznik G,Selkov E Jr,Walunas TL,Feil H,Feil WS, Purcell A, Lassez JL, Hawkins TL, Haselkorn R, Overbeek R, Predki PF, Kyrpides NC, "Whole-genome comparative analysis of three phytopathogenic Xylella fastidiosa strains.," *Proc Natl Acad Sci U S A.,* vol. 99, no. 19, pp. 12403-8, 2002.

[127] Buchrieser C, Rusniok C, Kunst F, Cossart P, Glaser P, and Consortium L., "Comparison of the genome sequences of Listeria monocytogenes and Listeria innocua: clues for evolution and pathogenicity.," *FEMS Immunol Med Microbiol,* vol. 35, no. 3, pp. 207-13, 2003.

[128] Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al., "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing," *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology,* vol. 19, no. 5, pp. 455-77., 2012.

[129] Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler., "QUAST: Quality Assessment Tool for Genome Assemblies," *Bioinformatics (Oxford, England) ,* vol. 29 , no. 8, pp. 1072-75., 2013.

[130] Schatz, Michael C., Arthur L. Delcher, and Steven L. Salzberg., "Assembly of Large Genomes Using Second-Generation Sequencing.," *Genome Research ,* vol. 20 , no. 9, pp. 1165-73., 2010.

[131] Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V,, "The RAST server: Rapid Annotations using Subsystems Technology.," *BMC Genomics ,* vol. 9, p. 75., 2008.

[132] Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S., "Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates.," *J Clin Microbiol,* vol. 43, no. 8, pp. 4178-82, 2005.

[133] Bortolaia V, Kaas RF, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AR, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, "ResFinder 4.0 for predictions of phenotypes from genotypes.," *Journal of Antimicrobial Chemotherapy.,* 2020.

[134] Kortine Annina Kleinheinz, Katrine Grimstrup Joensen, and Mette Voldby Larsen, "Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and E. coli virulence genes in bacteriophage and prophage nucleotide sequences," *Bacteriophage.,* vol. 4, p. e27943, 2014.

[135] Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H., "In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing.," *Antimicrob Agents Chemother.,* vol. 58, no. 7, p. 3895–3903, 2014.

[136] Rice LB., "Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE.," *J Infect Dis.,* vol. 197, p. 1079– 81, 2008.

[137] Lopez, L. L., Rusconi, B., Gildersleeve, H., Qi, C., McLaughlin, M., Scheetz, M. H., Seshu, J., & Eppinger, M., "Genome Sequences of Five Clinical Isolates of Klebsiella pneumoniae.," *Genome announcements,* vol. 4, no. 2, pp. e00040-16, 2016.

[138] Villa, L., Feudi, C., Fortini, D., Iacono, M., Bonura, C., Endimiani, A., Mammina, C., & Carattoli, A., "Complete Genome Sequence of KPC-3- and CTX-M-15-Producing Klebsiella pneumoniae Sequence Type 307.," *Genome announcements,* vol. 4, no. 2, pp. e00213-16, 2016.

[139] Chen L, Mathema B, Chavda KD, Deleo FR, Bonomo RA et al., "Carbapenemase-producing Klebsiella pneumoniae: molecular and genetic decoding.," *Trends Microbiol,* vol. 22, p. 686–696, 2014.

[140] Fursova NK, Astashkin EI, Gabrielyan NI, Novikova TS, Fedyukina GN, Kubanova MK, Esenova NM, Sharapchenko SO, Volozhantsev NV., "Emergence of Five Genetic Lines ST395NDM-1, ST13OXA-48, ST3346OXA-48, ST39CTX-M-14, and Novel ST3551OXA-48 of Multidrug-Resistant Clinical Klebsiella pneumoniae in Russia.," *Microb Drug Resist.,* vol. 26, no. 8, pp. 924-933, 2020.

[141] Mairi, A., A. Pantel, F. Ousalem, A. Sotto, A. Touati, and J.P. Lavigne., "OXA-48-producing Enterobacterales in different ecological niches in Algeria: clonal expansion, plasmid characteristics and virulence traits.," *J. Antimicrob. Chemother.,* vol. 74, p. 1848–1855, 2019.

[142] Marcade, G., S. Brisse, S. Bialek, E. Marcon, V. LeflonGuibout, V. Passet, R. Moreau, and M.-H. NicolasChanoine., "The emergence of multidrug-resistant Klebsiella pneumoniae of international clones ST13, ST16, ST35, ST48 and ST101 in a teaching hospital in the Paris region.," *Epidemiol. Inf. ,* vol. 141, p. 1705–1712, 2012.

[143] Wrenn, C., D. O'Brien, D. Keating, C. Roche, L. Rose, A. Ronayne, L. Fenelon, S. Fitzgerald, B. Crowley, and K. Schaffer., "Investigation of the first outbreak of OXA-48-producing Klebsiella pneumoniae in Ireland.," *J. Hosp. Infect. ,* vol. 87, p. 41–46, 2014.

[144] Navarro, K. Diestra E. Miró C. Martí D. Navarro J. Cuquet P. Coll F., "Multiclonal epidemic of Klebsiella pneumoniae isolates producing DHA-1 in a Spanish hospital.," *Clinical Microbiology and Infection,* vol. 17, p. 1032–1036, 2011.

[145] D. Elhani, L. Bakir, M. Aouni, V. Passet, G. Arlet, S. Brisse and F.-X. Weill, "Molecular epidemiology of extended-spectrum beta-lactamase-producing Klebsiella pneumoniae strains in a university hospital in Tunis, Tunisia, 1999-2005.," *Clinical Microbiology and Infection,* vol. 16, p. 157–164, 2010.

[146] Oteo J, Cuevas O, López-Rodríguez I, Banderas-Florido A, Vindel A, Pérez-Vázquez M, Bautista V, Arroyo M, García-Caballero J, Marín-Casanova P, González-Sanz R, Fuentes-Gómez V, Oña-Compán S, García-Cobos S, Campos J., "Emergence of CTX-M-15-producing Klebsiella pneumoniae of multilocus sequence types 1, 11, 14, 17, 20, 35 and 36 as pathogens and colonizers in newborns and adults.," *Journal of Antimicrobial Chemotherapy,* vol. 64, p. 524–528, 2009.

[147] Nielsen JB, Skov MN, Jørgensen RL, Heltberg O, Hansen DS, Schønning K., "Identification of CTX-M15-, SHV-28-producing Klebsiella pneumoniae ST15 as an epidemic clone in the Copenhagen area using a semi-automated Rep-PCR typing assay.," *European Journal of Clinical Microbiology & Infectious Diseases,* vol. 30, p. 773–778, 2011.

[148] Patrick Musicha, Chisomo L Msefula, Alison E Mather, Chrispin Chaguza, Amy K Cain, Chikondi Peno, Teemu Kallonen, Margaret Khonga, Brigitte Denis, Katherine J Gray, Robert S Heyderman, Nicholas R Thomson, Dean B Everett, Nicholas A

Feasey, "Genomic analysis of Klebsiella pneumoniae isolates from Malawi reveals acquisition of multiple ESBL determinants across diverse lineages.," *Journal of Antimicrobial Chemotherapy.,* vol. 74, no. 5, p. 1223–1232, 2019.

[149] Wejdene Mansour, Raoudha Grami , Anis Ben Haj Khalifa , Safia Dahmen , Pierre Châtre , Marisa Haenni , Mahjoub Aouni , Jean-Yves Madec, "Dissemination of multidrug-resistant blaCTX-M-15/IncFIIk plasmids in Klebsiella pneumoniae isolates from hospital- and community-acquired human infections in Tunisia," *Diagnostic Microbiology and Infectious Disease,* vol. 83, no. 3, p. 298–304, 2015.

[150] Yun-Tae Kim, Ji-Hyun Jang , Hyun-Chul Kim , Hyogyeong Kim , Kyoung-Ryul Lee , Kyung Sun Park , Hee-Joo Lee & Young-Jin Kim, "Identification of strain harboring both aac(6')-Ib and aac(6')-Ib-cr variant simultaneously in Escherichia coli and Klebsiella pneumoniae," *BMB Rep.,* vol. 44, no. 4, pp. 262-6, 2011 .

[151] N. L. Johnson TJ, "Pathogenomics of the virulence plasmids of Escherichia coli.," *Microbiol Mol Biol Rev.,* vol. 73, no. 4, pp. 750-74., 2009 .

[152] Villa L, García-Fernández A, Fortini D, Carattoli A., "Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants.," *J Antimicrob Chemother.,* vol. 65, no. 12, p. 2518–2529, 2010.

[153] Riley MA, Gordon DM., "A survey of Col plasmids in natural isolates of Escherichia coli and an investigation into the stability of Col-plasmid lineages.," *J Gen Microbiol,* vol. 138, p. 1345–1352, 1992.

[154] Zhang J, van Aartsen JJ, Jiang X, Shao Y, Tai C, He X, Tan Z, Deng Z, Jia S, Rajakumar K, Ou HY., "Expansion of the known Klebsiella pneumoniae species gene pool by characterization of novel alien DNA islands integrated into tmRNA gene sites.," *J Microbiol Methods. ,* vol. 84, no. 2, pp. 283-9, 2011.