VNIVERSITAT DE VALÈNCIA

Genomic epidemiology and antimicrobial resistance of *Klebsiella pneumoniae* in the Comunitat Valenciana

PhD thesis

Doctoral Programme in Biomedicine and Biotechnology

Neris García González

Supervisor:
Fernando González Candelas

Valencia, May 2023

Dr. Fernando González Candelas, Catedrático de Genética de la Universitat de València

CERTIFICA

Que Neris García González ha realizado bajo su dirección y supervisión el trabajo que lleva por título "Genomic epidemiology and antimicrobial resistance of *Klebsiella pneumoniae* in the Comunitat Valenciana", para optar al Grado de doctora en Biomedicina y Biotecnología por la Universitat de València. La memoria de este trabajo cumple los requisitos científicos y formales para presentar la misma en depósito y proceder a la defensa de la Tesis Doctoral.

Y para que conste, en cumplimiento de la legislación presente, firma el presente informe favorable en Valencia, nueve de Mayo de 2023.

Dr. Fernando González Candelas

Agradecimientos

Esta tesis ha sido un viaje largo, lleno de desafíos y momentos inolvidables, que, afortunadamente, no he recorrido sola. Me gustaría expresar mi más sincero agradecimiento a todas las personas que han formado parte de este increíble camino.

En primer lugar, quiero agradecer de manera especial a Fernando. Esta tesis no habría sido posible sin su dedicación y compromiso con el proyecto y conmigo. Desde que entré al laboratorio como alumna colaboradora, depositaste toda tu confianza en mí. A lo largo de estos años has dedicado mucho tiempo a resolver dudas, buscar oportunidades para mi futuro y en los últimos meses, a darme ánimos en la recta final. Gracias a tu apoyo, he aprendido mucho más de lo que pueden reflejar estas páginas tanto a nivel científico como personal. Siempre te estaré agradecida.

A todos los microbiólogos clínicos que han participado en el proyecto, por su esfuerzo y dedicación. En particular a Marosi y Bárbara por enseñarme aquellos primeros cultivos de *Klebsiella* resistente. Y a Bego por ser mi teléfono de emergencias de la microbiología, siempre dispuesta a resolver mis dudas.

A toda el área de Genómica y Salud, y en especial a mi grupo Epimol. Muchos de ellos contribuyeron directamente a esta tesis con ayuda y consejos, y muchos otros brindaron risas y buenos momentos que hicieron este viaje mucho más divertido. Agradezco especialmente a Marta y Andrea por recorrer este camino conmigo. A Pascual por su infinita paciencia y su disponibilidad 24/7 para ayudarme con cualquier problema en el servidor. A Jordi porque algunas de estas páginas no se hubieran podido acabar sin su ayuda y trabajo. Y a Irving por todos los cafés de estos últimos meses cargados de últimos empujones. Espero que a partir de ahora los carguemos de "*pole pole*" y *"hakuna matata"*.

To the Evolutionary genomic lab in the Institut Pasteur. My biggest thanks to Eduardo Rocha for giving me the opportunity to do the internship in his lab, and to all the members of the group for the scientific discussions and the very

enjoyable moments, including those in Pornichet. Each moment I spent in Paris was filled with learning and joy.

A Bea por estos 10 años de amistad llenos de risas, apoyo incondicional y esa afinidad tan difícil de encontrar. Sabes que siempre he admirado tu forma de ser y de encarar la vida, y si tenerte como amiga no tiene precio, trabajar contigo ha sido un verdadero regalo. Gracias por toda la ayuda, esta tesis no sería lo mismo sin ti, y gracias por todas esas tardes en el laboratorio compartiendo jornadas interminables. Espero que no tengamos que hacer más conjugaciones en Navidad, pero si así fuera, que al menos sea contigo. Aunque ahora nuestros caminos se separen, estoy convencida de que encontraremos la manera de que se vuelvan a cruzar.

A Sandra por ser mi refugio emocional en el laboratorio, por escucharme y por esos abrazos que me recargaban las pilas.

Y a todos mis amigos que, aunque quizás no sepan hasta qué punto me han ayudado, quiero agradecerles el haber estado ahí en cada paso del camino, ayudarme a desconectar y animarme a seguir adelante.

A Ángel, aunque contigo las palabras se queden cortas, gracias por cuidarme tanto y sostenerme cada día, pero sobre todo por hacerme reír cada uno de los días que escribía algo en estas páginas. Estoy profundamente agradecida de tenerte a mi lado. Nuestro viaje continúa, y aunque todavía no sabemos hacía dónde se dirige, sé que seremos felices.

Y por último a mi familia, a quién les debo todo y dedico esta tesis. A mis padres, por darme una vida muy feliz y extraordinaria. Gracias por enseñarme a pensar a lo grande pero poco a poco y sin agobios y por minimizar los problemas y multiplicar las alegrías. Habéis hecho que este viaje haya sido mucho más fácil. A mi hermano, por celebrar cada uno de mis logros como si fueran suyos propios. Tete, este logro también es tuyo. A mi tía Anabel y Atimi por su inmenso apoyo y cariño incondicional. Y a todo el resto de la familia que siempre me ha deseado lo mejor y me han brindado su apoyo y cariño.

Abstract

Antimicrobial resistance (AMR) is a major public health threat worldwide. The overuse and misuse of antibiotics has led to the rise and spread of antibiotic-resistant infections. As a consequence, clinicians are facing a decreasing number of effective treatments that lead to more severe infections, longer hospital stays, higher healthcare costs, and increased mortality. Globally, AMR is estimated to cause 1.27 million deaths annually. The projections suggest that this number could reach 10 million by 2050 if urgent action is not taken.

A few pathogens are responsible for most of bacterial AMR-associated deaths and among those ***Klebsiella pneumoniae*** represents one of the most concerning pathogens. Currently, *K. pneumoniae* is among the world's most common nosocomial pathogens. Along with its high prevalence, *K. pneumoniae* has gained notoriety due to its potential to rapidly acquire and accumulate AMR. In the last decades, prevalence rates for *K. pneumoniae* non-susceptible against the most commonly used antibiotics have increased alarmingly. In 2017, the WHO recognized *K. pneumoniae* as one of the most critical priority of pathogens for which new antimicrobial development was urgently needed. Indeed, *K. pneumoniae* was ranked as the second cause of AMR-attributable deaths in 2019.

Among the acquired AMR in *K. pneumoniae*, strains that have acquired resistance to **third-generation cephalosporins (3GC)** and **carbapenems (CRKp)** are the major concern. Currently, there is a high prevalence of infections caused by 3GC-resistant *K. pneumoniae* strains worldwide, which has resulted in an increased use of carbapenems, a last resort treatment option. While carbapenem-resistance has been historically low, in recent years, a global increase has been reported, with carbapenem-resistant *K. pneumoniae* identified as the fastest-growing AMR pathogen in clinical settings. Carbapenem-resistant bacteria limit the availability of effective antibiotics to last-resort drugs, which are frequently associated with reduced efficacy and/or increased toxicity, resulting in high mortality and morbidity rates. This highlights the urgent need to address the growing threat of 3GC- and carbapenem-resistance in *K. pneumoniae*.

In this dissertation we have investigated the epidemiology of 3GC- and carbapenem-resistant *K. pneumoniae* using the genomic information gathered under the **Surveillance of *Klebsiella pneumoniae* in the Comunitat Valenciana** (SKPCV) project. This project is the main element of this dissertation. The project was carried out under the Networked Laboratory for Surveillance of Antimicrobial Resistance (NLSAR) framework and enrolled 8 of its hospitals. Under the SKPCV ESBL- and/or carbapenemase-producing *K. pneumoniae* isolates were collected for 3 years (2017 - 2019). Moreover, to provide context and establish a collection that would allow us to elucidate the relationships between SKPCV, Spanish and global *K. pneumoniae* isolates, we included isolates collected previously to the SKPCV in some NLSAR hospitals, as well as external data from three different databases: RefSeq, GenBank, and ENA. Using these data, we collected more than 13,000 genomes.

In chapter 4, we addressed the quality control analysis of the included genomes in addition to the genomic description of lineages and AMR determinants.

Working with large datasets, and ensuring data quality can be challenging, therefore we created a quality control pipeline of tiered filters that assessed taxonomic assignment and interspecies contamination, assembly quality, intraspecies contamination, and finally, genomic similarity across the entire collection. By implementing this tiered filtering approach, we were able to eliminate 8% of the SKPCV and NLSAR-collected genomes and 3% of those retrieved from public databases. Although the number of bad-quality genomes may seem low, they can have a significant negative impact on downstream analysis. That is why we emphasise the importance of conducting a thorough filtering workflow before performing any analysis using your own data or genome sequences from public databases.

Using this quality pipeline, we ended with a large SKPCV collection with 1,604 genomes. Additionally, we included 395 retrospective isolates collected from three NLSAR hospitals and more than 10,000 global genomes available in the public databases. When comparing the collections, we found that the SKPCV and NLSAR lineage compositions were very diverse but also similar to that of the Spanish genomes deposited in the databases. Indeed, most HRCs in the NLSAR isolates were

found to be related to isolates collected in other Spanish regions, which suggests similar evolutionary histories. This global genome collection is a perfect framework to further study and understand the genomic epidemiology of *K. pneumoniae* in the CV.

In chapter 5, we used the genomic data from the global collection, but specifically those collected during the SKPCV, to understand the burden of AMR *K. pneumoniae strains* in the CV. To deeply study the situation in the CV, we first contextualised the SKPCV genomes in the epidemiological context of AMR *K. pneumoniae* in the CV. For this, we interrogated the RedMIVA database for all entries of *K. pneumoniae* antimicrobial susceptibility testing (AST) performed from 2010 to 2022 in the CV. We found that the CV has a high incidence of 3GC-resistant *K. pneumoniae* infections, while remaining a region with low incidence of carbapenem resistance.

We found that the 3GC- and carbapenem-resistant AMR determinants and the corresponding carrier lineages were found to be distinctly distributed in the hospitals. And that, regardless of ST307 carrying $bla_{\text{CTX-M-15}}$, most lineages and AMR determinant combinations were found mostly limited to one hospital. Indeed, hospital populations were different among each other.

Our findings suggest that the burden of AMR and *K. pneumoniae* in this region was the result of a diverse set of factors, including singleton lineages likely originating from the community or from the previous microbiota of patients, as well as a complex interplay between lineage transmission between hospitals and the local proliferation of problematic clones within each hospital.

Our analysis revealed that only one lineage, the ST307, was responsible for most 3GC- and carbapenem-resistant infections and interhospital transmissions. Our investigation of the evolution of this lineage, in which we included global isolates, led us to conclude that the high prevalence of ST307 in the SKPCV collection was primarily due to multiple introductions of the lineage into different settings.

In chapter 6, we have assessed how the emergence and initial colonisation of carbapenem-resistant isolates occurred in a hospital setting. Our findings show that the initial carbapenem resistance emergence and dissemination of the HGUV

occurred during a short period of one year and it was very complex. We found six different lineages comprising the majority of the CRKp population at the HGUV, disseminating different resistance mechanisms (AmpC, OXA-48, NDM-1, and NDM-23) in different plasmid variants. These lineages were in local clonal expansion, with several cases of possible direct transmission within the hospital. Moreover, we demonstrate that most of the lineages found in the HGUV were previously present in other Spanish regions, those being probably the source of the relevant lineages at this hospital.

In chapter 7, we used genomic epidemiology to describe the emergence and dissemination in several hospitals of a new *bla*$_{\text{NDM-23}}$ carbapenemase gene. We were able to elucidate the phenotypic effect and the genetic environment of the gene. The gene was carried in a multidrug-resistant plasmid with 18 additional AMR genes, which produced a multidrug resistance phenotype. The gene and the plasmid were found in a ST437 strain. We found that the plasmid was non-mobilizable, so the *bla*$_{\text{NDM-23}}$ was disseminated through clonal expansion. The dissemination of this ST437 lineage carrying the *bla*$_{\text{NDM-23}}$ affected at least four different hospitals of the Comunitat Valenciana from 2016 until at least 2019, when our sampling concluded. Using bioinformatic and phylogenetic analyses, we have been able to trace the evolutionary and epidemiological route of the new allele, the hosting plasmid, and the strain that carried both from Pakistan to Spain.

This study provides compelling reasons to establish routine genomic surveillance of bacterial pathogens in the CV and throughout the country, particularly in the context of the concerning AMR situation. As shown, genomic surveillance projects enable understanding the prevalence of resistant strains, their genetic determinants, and their transmission patterns. This information is valuable for implementing targeted interventions to prevent the spread of AMR strains. Additionally, our findings underscore the importance of implementing surveillance strategies based on complete genome sequencing in hospitals. As shown during this dissertation, we observed different lineages of the same ST with distinct dynamics and associated resistance determinants. Traditional methods lack the required level of resolution. In addition, WGS enables the performance of evolutionary analyses, which aid in understanding the origins and evolution of bacterial populations. Through this

approach, we were able to gain valuable insights into the evolution of the worldwide emerging MDR lineage, ST307, and to trace the evolutionary and epidemiological route of the new carbapenemase allele, *bla*$_{\text{NDM-23}}$.

The NLSAR and SKPCV projects have enabled all tertiary hospitals in the CV to perform WGS on pathogens, eliminating the need to rely on other reference laboratories. This initiative marks a significant step towards more efficient and effective surveillance and control of AMR regionally in the CV and nationally in Spain. The WGS capabilities will enable hospitals to rapidly identify and track outbreaks of AMR infections, allowing for timely and targeted interventions to prevent further spread. Additionally, the data generated from the nationwide surveillance system conducted by the PRAN and the RedLabRA will contribute to global efforts to tackle the global AMR threat.

Abbreviations

3GC	Third-generation cephalosporins
AMR	Antimicrobial resistance
AST	Antimicrobial susceptibility testing
CG	Clonal group
cgMLST	Core genome multi-locus sequence typing
cKp	"Classical" *K. pneumoniae*
CRKp	Carbapenem-resistant *K. pneumoniae*
CTX-M	Cefotaximase from Munich β-lactamase
CV	Comunitat Valenciana
EARS-Net	European Antimicrobial Resistance Surveillance Network
ESBL	Extended-spectrum beta-lactamase
EUCAST	European Committee on Antimicrobial Susceptibility Testing
FISABIO	Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana
HAV	Hospital Arnau de Vilanova
HCUV	Hospital Clínico Universitario de Valencia
HDM	Hospital Dr. Moliner
HGT	Horizontal gene transfer
HGUA	Hospital General Universitario de Alicante
HGUC	Hospital General Universitario de Castellón
HGUE	Hospital General Universitario de Elche
HGUV	Hospital General Universitario de Valencia
HRC	High-risk clones
HRyC	Hospital Ramón y Cajal
HULR	Hospital Universitario La Ribera
HUPLF	Hospital Universitario y Politécnico La Fe
hvKp	"Hypervirulent" *K. pneumoniae*
ICEs	Integrative conjugative elements
IMP	Active against imipenem; imipenemase
In	Integron
IS	Insertion sequence

K type	Capsular type
KL type	Capsule locus type
KPC	*Klebsiella pneumoniae* carbapenemase
KpSC	*K. pneumoniae* species complex
LMIC	Low and middle-income countries
LPS	Lipopolysaccharide
MBL	Metallo-β-lactamases
MDR	Multidrug resistant
MGE	Mobile Genetic Element
MIC	Minimum inhibitory concentration
MLST	Multi-locus sequence typing
MOB	Plasmid mobility genes
MPF	Mating pair formation
NDM	New Delhi metallo-β-lactamases
NLSAR	Networked Laboratory for Surveillance of Antimicrobial Resistance
O type	O-antigen type
OECD	Organisation for Economic Co-operation and Development
OXA	Oxacillinase β-lactamase
pAmpC	Plasmid-encoded AmpC
PBRT	PCR-based replicon typing
PDR	Pan-drug resistance
PRAN	Spanish National Plan against Antibiotic Resistance
PTG	Possible transmission group
RedLabRA	Spanish Network of Laboratories for the Surveillance of Resistant Microorganisms
RedMIVA	Red de Vigilancia Microbiológica de la Comunitat Valenciana
SHV	sulfhydryl reagent variable β-lactamase
SKPCV	Surveillance of *Klebsiella pneumoniae* in the Comunitat Valenciana
SL	Sublineage
SNP	Single nucleotide polymorphism
SRA	Sequence Read Archive
ST	Sequence Type
TEM	Temonerira β-lactamase

tMCRA	time to the most common recent ancestor
Tn	Transposon
VIM	Verona integron-encoded metallo-β-lactamase
WGS	Whole genome sequencing
WHO	World Health Organization

Table of contents

CHAPTER 1
INTRODUCTION

1.1 Back to the pre-antibiotic era?

Antibiotics are chemical compounds that kill bacteria or slow their growth. They have been produced by microorganisms for billions of years as a survival strategy to feed and/or defend themselves from other organisms [1]. Since their first use in medical practice [2], they have become one of the pillars of modern medicine [3]. Antibiotics have not only been widely used to treat life-threatening infections, but they have also facilitated the development of safe and advanced medical procedures, including most surgical procedures such as cancer treatment, organ transplants, and Caesarean sections. These, coupled with their beneficial use in animal health and husbandry, have made antibiotics one of the most successful medicines ever developed.

However, soon after their discovery (around the 1930s), it became evident that some bacteria could escape their effects and, therefore, become resistant to antimicrobial compounds. There is a long-held belief that **antimicrobial resistance (AMR)** appeared due to the clinical use of antibiotics. Nonetheless, its origin has been found to long predate the human use of antibiotics [4, 5]. AMR is as natural as antibiotics, as it emerged in antibiotic-producing microbes as a self-protective mechanism [4, 6]. AMR can also be acquired by susceptible microbes that become resistant to an antibiotic that would have otherwise killed them. In fact, the term AMR is often used in reference to newly resistant microbes, rather than to those that are naturally resistant to a given antimicrobial drug.

For some time, AMR was not considered a problem, as it was alleviated by the **“golden age”** of antibiotic discovery (1950 - 1970). During these two decades, there was a steady introduction of new and effective antimicrobial compounds that increased the popularity and widespread use of antibiotics. However, since the 1980s, there has been a significant decline in the discovery of new antibiotics, leading to what is known as the **"antibiotic discovery gap**", which continues to this day. In the meantime, resistant strains to every class of antibiotic were described [7], and their incidence increased, primarily due to the selective pressure generated by the widespread use of antibiotics. Nowadays, the burden of AMR is so big that it

is jeopardising the most valuable medical and public health advances achieved in the last two centuries.

1.1.1 Antimicrobial resistance as an urgent threat

In 2019, the World Health Organization (WHO) declared AMR as one of the top ten threats to global public health [8, 9] along with air pollution, climate change and other infectious diseases such as ebola, AIDS, and dengue. That same year, a global survey estimated that AMR caused 1.27 million deaths and that 4.95 million deaths were associated with it [10]. This means that resistant pathogens killed more people than HIV/AIDS (864,000 deaths) or malaria (643,000 deaths), two of the infectious diseases with the highest mortality rates worldwide [11]. In 2020, the European Union (EU) reported 670,000 infections caused by resistant bacteria and approximately 33,000 deaths as a direct consequence of AMR [12].

The global burden of AMR-related deaths is increasing worldwide; moreover, it has been estimated that if no action is taken, deaths associated with AMR could rise to 10 million by 2050, overcoming cancer as a major cause of death [13]. It is worth noting that the burden of AMR goes beyond its negative impact on infected patients, but also from the fact that it can compromise procedures that rely on antibiotics, such as prophylaxis in both human and animal medicine [10].

In addition to the cost in human lives, AMR also represents a very high increase in the economic cost of healthcare. Resistant infections are more expensive to treat. Patients with AMR infections are more likely to require extended hospitalizations and more expensive treatments [14]. Each year, the EU alone spends between 1.1 - 1.5 billion euros on extra healthcare costs due to AMR [12]. The O'Neill report also forecasted that, by 2050, AMR-associated costs will increase as high as US $100 trillion worldwide [13]. However, the Organisation for Economic Co-operation and Development (OECD) calculated that investing as little as 2 US dollars per person per year in implementing measures would reduce 75% of AMR-related deaths by 2050 and save 4.8 billion US dollars per year in the 33 OECD countries [15].

Although AMR is a major health problem globally and must be addressed as such [16], the prevalence and response to AMR varies widely among different

geographical regions and countries. Worrisomely, the prevalence of AMR infections is higher in low and middle-income countries (LMIC) [10, 17] where the resources available to address this issue are often more limited.

1.1.2 AMR priority pathogens

Six pathogens were responsible for about 75% of bacterial AMR-associated deaths in 2019: *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *Acinetobacter baumannii*, and *Pseudomonas aeruginosa* [10].

Regarding AMR, *K. pneumoniae* represents one of the most concerning pathogens. *K. pneumoniae* is among the world's most common nosocomial pathogens [18], accounting for about one-third of all Gram-negative infections [19]. This pathogen contributes significantly to the burden of disease, especially in clinical settings where it is associated with AMR and several life-threatening **hospital-acquired infections** [20]. Along with its high prevalence, *K. pneumoniae* has gained notoriety as a pathogen due to its potential to rapidly acquire and accumulate AMR. In the last decades, non-susceptible prevalence rates for *K. pneumoniae* against the most used antibiotics have increased alarmingly [12]. In 2017, the WHO recognized *K. pneumoniae* as one of the most critical priority of pathogens for which new antimicrobial development was urgently needed [21]. Indeed, *K. pneumoniae* was ranked as the second cause of AMR-attributable deaths in 2019 [10].

1.*2 Klebsiella pneumoniae*

Klebsiella are gram-negative, facultative, anaerobic, non-motile, encapsulated bacilli, of the *Enterobacteriaceae* family [22]. The *Klebsiella* genus comprises a wide diversity of species, many of which have been described recently [23–25] (Figure 1.1). Most *Klebsiella* species are ubiquitous in the environment, being present in soil, water, and plants [26, 27]. However, species within the ***K. pneumoniae* species complex (KpSC)** (Figure 1.1) are also commonly found in humans and other animals, both asymptomatically and causing infections [28–30].

The KpSC is a group with no formal taxonomic designation but is often used in the literature to describe closely related species (95 – 96% average nucleotide identity)

to the species *K. pneumoniae*. Its members share a 90% of average nucleotide identity with the *Klebsiella* species outside the complex. At the time of writing, seven species or phylogroups had been classified within de KpSC [31, 32].

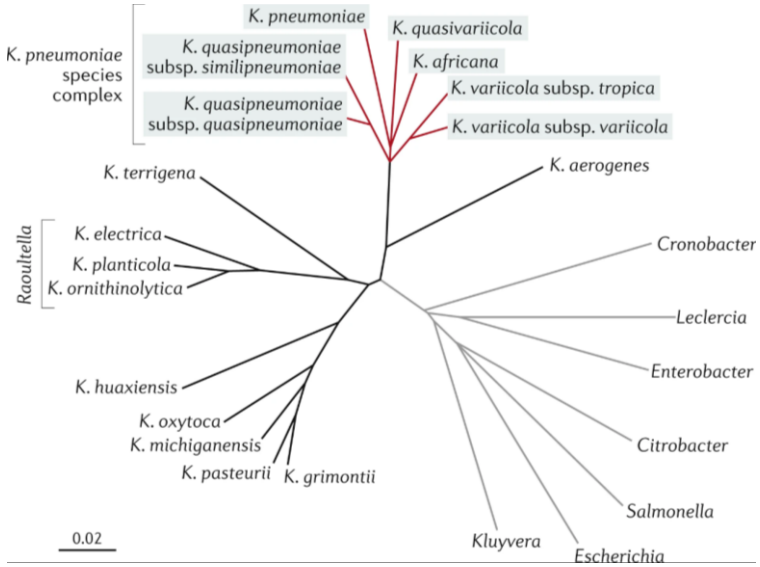


Figure 1.1. Whole-genome tree based on mash distance showing the phylogenetic relationships between *K. pneumoniae*, the *K. pneumoniae* species complex (red branches), other species of the genus (black branches) and family Enterobacteriaceae (grey branches). Scale bar is the estimated average nucleotide divergence. Extracted from [31].

Among members of the KpSC, ***K. pneumoniae* sensu stricto** is the clinically most relevant species, as it accounts for 85% of human infections caused by the complex [28, 33]. Indeed, it is the species with the highest rate of AMR [34].

1.2.1 *K. pneumoniae* population diversity

K. pneumoniae sensu stricto is characterised by a high level of intraspecies diversity [28]. The high diversity of *K. pneumoniae* populations is mainly due to the plasticity of its genome. The average length of a *K. pneumoniae* genome is between 5 to 6 Mbp, encoding from 5,000 to 6,000 genes. Approximately 2,000 of these genes are present in all the strains of the species (**core genome**), whereas more than half (> 3,500) are variably present in different strains and constitute the **accessory genome**. This accessory genome provides metabolic and other capabilities enabling

the adaptability of the lineages to different environments (e.g., metabolism, AMR, virulence). These variable genes are part of a large pool of over 100,000 genes found in the species, most of them rare [28]. This plastic genome is translated into an extensive population diversity, consisting of over 250 phylogenetic lineages (Figure 1.2 A).

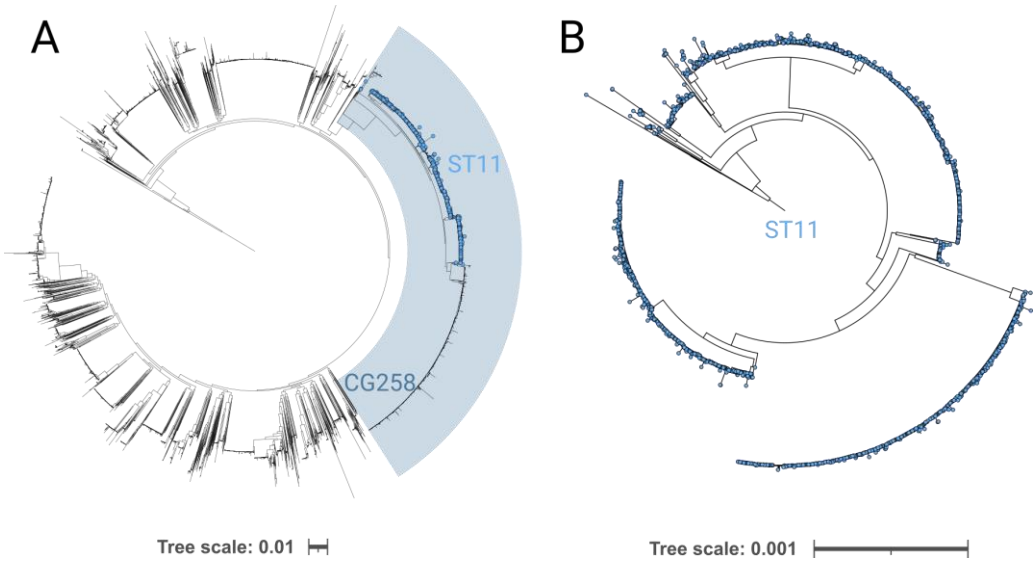


Figure 1.2. A) Phylogeny including 13 thousand genomes of *K. pneumoniae*. CG258 is highlighted in blue, and ST11 with blue dots. B) Zoom of tree A into ST11.

K. pneumoniae lineages are commonly described using Multi Locus Sequence Typing (MLST) that classifies them into sequence types (STs). **MLST** is a nucleotide sequence typing method based on polymorphisms of a few housekeeping/conserved loci that, in the case of *K. pneumoanie*, include *gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB*, and *tonB* [35]. The ST is determined based on the allelic profile of these genes. The MLST system has been adopted by both the research and public health communities to identify and name *K. pneumoniae* lineages [36]. Related STs are grouped into clonal groups (CGs) (Figure 1.2 A-B); however, the relationship level needed to group STs into CG is not well defined. Recently, core genome MLST (cgMLST) has been used to define CGs with more precision. cgMLST schemes are based on the same approach as MLST but extended to a much larger set of genes. In *K. pneumoniae,* different cgMLST schemes have been proposed with

variations in the number of genes included in the scheme (694 - 2,358) [37–39] Currently, the 694-loci described in [37] is the most widely accepted. Of note, the resulting lineages, regardless of the definition system (STs, CGs, etc.), are typically referred to as clones [40].

1.2.2 Epidemiology: from commensal to pathogen

K. pneumoniae is widely known as an opportunistic pathogen, mainly infecting hospitalised patients with underlying medical conditions. Yet, it can also act as a true pathogen, causing severe disease in healthy individuals. These clinical manifestations are caused by two different *K. pneumoniae* pathotypes, termed "**classical**" and "**hypervirulent**" [41, 42]. There is no exact definition for any of the pathotypes but, generally, they can be distinguished by several epidemiological, clinical, and genetic features (Figure 1.3) [41, 43–45].

"**Classical" *K. pneumoniae*** (cKp) strains are primarily opportunistic pathogens, causing **hospital-acquired** infections in immunocompromised patients. cKp are among the most prevalent nosocomial pathogens, accounting for about one-third of all Gram-negative hospital-acquired infections [18], and a leading cause of neonatal sepsis [46]. The most common manifestations are urinary tract, pneumonia, surgical site, and soft tissue infections, any of which can develop into bacteraemia [47]. The severity of these infections has worsened in recent decades and become a health problem due to the acquisition of AMR by cKp strains, resulting in the emergence of **multidrug-resistant (MDR)** strains. Infections caused by drug-resistant strains often lead to therapeutic failure followed by clinical complications that increase mortality rates, which have been reported to reach up to 65% of patients [48–52]. Several predisposing factors have been identified for suffering a drug-resistant *K. pneumoniae* infection: (1) previous colonisation, (2) low profile immune system, (3) prior antibiotic consumption, and (4) presence of indwelling medical devices [53]. These predisposing factors are frequently present in hospitalised patients, posing a major risk among neonates, the elderly, and the immunocompromised.

On the contrary, **"hypervirulent" *K. pneumoniae*** (hvKp) strains generally act like true **pathogens** causing **community-acquired** infections in healthy hosts [54, 55].

Hvkp has emerged as a clinically significant pathogen responsible for highly invasive infections, especially pyogenic liver abscesses that can metastasize and complicate into meningitis, endophthalmitis, and necrotizing fasciitis [55, 56]. These infections have been reported worldwide but most of the cases occur in Asia or in individuals with recent contact with the Asian region, where the prevalence can be as high as 45% of the hospitalised cases [57–61]. Indeed, hvKp is the most common cause of pyogenic liver abscesses in Asia [54, 62]. hvKp strains are generally enriched for certain virulence factors [63], which worsen the severity of the infections leading to high mortality rates, ranging from 3% to 55% [64–66].

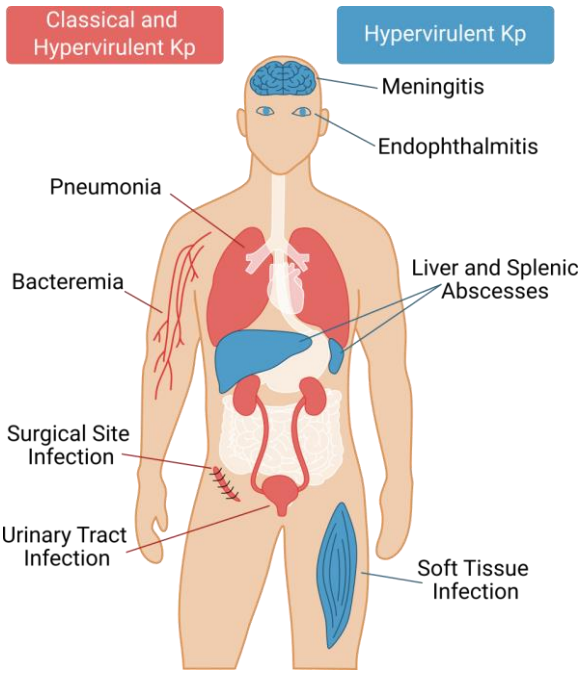


Figure 1.3. Clinical manifestations of both pathotypes. CKp infections are coloured in red whereas hvKp infections are coloured in blue and red. Common colonisation organs are coloured in white. Adapted from [45].

Both pathotypes can be found asymptomatically as common human **commensal** microbes, colonising mucosal surfaces, including the nasopharynx and the intestinal tract [22, 44]. Colonisation prevalence is still not well characterised, and estimated rates vary by pathotype, age group, geographical location, body site, and exposition to healthcare contact [67, 68]. Recent studies have shown that the prevalence of *Klebsiella* colonisation in Western countries can reach up to 25% of hospital

patients, primarily due to cKp with rare cases of hvKp [67, 69]. Moreover, it has been shown that cKp-carriage rates are significantly higher in hospital environments or in individuals with recent healthcare contact than in the community (4–6%) [67, 70]. On the contrary, in Asian countries, hvKp shows very high carriage rates even in the community, reaching up to 87% of the individuals screened [71]. *K. pneumoniae* colonisation is very relevant as it bears a double burden. There is evidence that gastrointestinal colonisation is a relevant source of infection [67, 70, 72], and that colonised patients serve as a major reservoir for further propagation, especially in hospital settings [73, 74].

1.2.3 Pathogenesis and virulence determinants

The pathogenicity of *K. pneumoniae* derives from several virulence factors associated with the ability to colonise the human body, evade the host immune system, and produce invasive infections [68, 75]. Some virulence factors are present in all the strains while others are more frequently found in hvKp. However, there are no exclusive markers for any of the pathotypes [41, 42, 76, 77] (Figure 1.4).

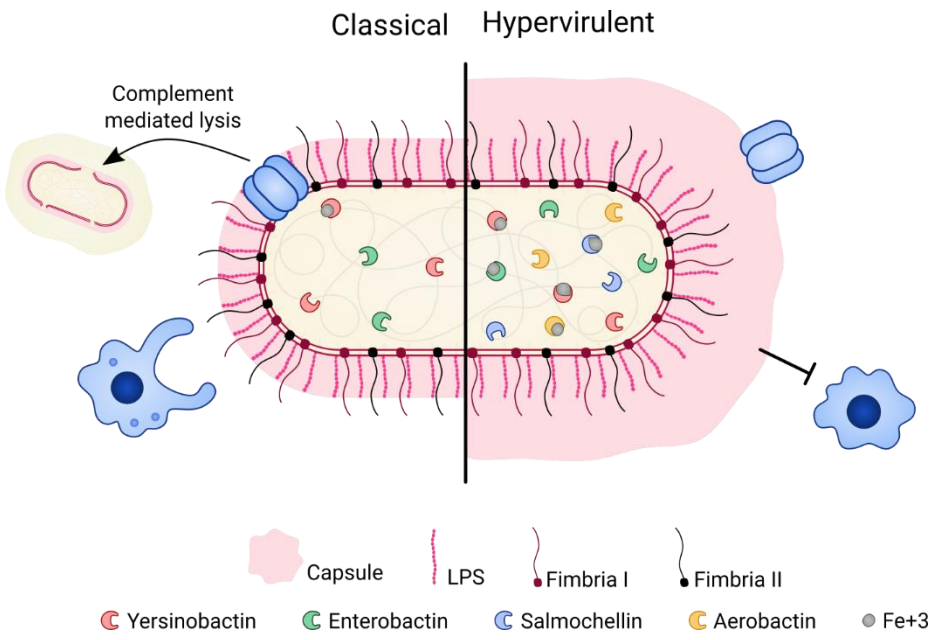


Figure 1.4. Virulence factors present in *K. pneumoniae*. Differences between classical *K. pneumoniae* and hypervirulent strains are shown.

The polysaccharide capsule is the main virulence factor of the species [75, 78]. It interferes with phagocytosis by immune cells, confers resistance to serum bactericidal activity by blocking the action of the complement, and suppresses early inflammatory responses [79]. To date, at least 79 serologically defined capsular types (K types) have been identified [80, 81] with over 146 different capsule loci types (KL types) identified through comparative genomics [82, 83]. Serotypes differ in their degree of virulence. The majority of hvKp strains belong to K1–K5 serotypes [83], which have been shown to be more virulent than other K types that are mainly present in cKp strains [84]. K1-K5 serotypes often present capsule hypermucoviscosity [75, 85]. The mucoid phenotype is related to the expression of two plasmid-born regulators (*rmpA*, *rmpA2*) [80], which are often used as markers of this trait [76]. The hypermucoviscous phenotype can be observed in agar plates and semi-quantified by the “string test” [77]. Because of the hypermucoviscous phenotype that these strains show, hvKp has alternatively been referred to as "hypermucoviscous" or "hypermucoid".

The lipopolysaccharide (LPS) is an essential component of the outer membrane of Gram-negative bacteria that activates the host immune system. To avoid the activation of the immune system, some strains mask their own LPS with the capsule, whereas others modify the LPS to variants that are no longer recognized by the immune system [68]. LPS molecules are composed of lipid A, a core domain, and the O-antigen. Variations in the sugar composition and structure of the O-antigen results in distinct O-serotypes. In *K. pneumoniae* nine O-types have been reported, of which O1, O2, and O3 are responsible for almost 80% of all *Klebsiella* infections [83].

Two major fimbriae have been identified in both *K. pneumoniae* pathotypes, type 1 and type 3 pili, encoded by genes *fim* and *mrk*, respectively. Fimbriae mediate the adhesion to host surfaces, which is a critical step to infection. Both fimbriae are highly associated with *K. pneumoniae* urinary tract infections [75, 86]. Type 1 fimbriae are associated with adhesion to host cells and are essential for the initial establishment of the infection. Type 3 fimbriae have been identified as an important colonisation factor and as strong promoters of biofilm formation, especially on

indwelling urinary catheters, contributing to persistence in the hospital environment [86].

Siderophores are small iron-chelating molecules that microbes use to capture essential metals. They facilitate bacterial growth in conditions of limited availability of free iron, promoting bloodstream or tissue infections [87]. *K. pneumoniae* encodes for different siderophores which are distinctly distributed between cKp and hvKp. The enterobactin siderophore (encoded by *Ent*) is the most common siderophore and is present in both pathotypes as it is encoded in the chromosome of the species. The second most common is yersiniabactin, which is often secreted by both pathotypes [88]. hvKp strains can also synthesise aerobactin and salmochelin [89]. In addition to presenting a larger repertoire of siderophores, hvKp also produces larger amounts of them [90], which has been associated with an increased ability to colonise and spread within the host [91].

Lastly, ~10% of *K. pneumoniae* produces genotoxin colibactin (encoded by *clb* or *pks*) [88,92]. However, the exact mechanisms by which colibactin contributes to pathogenesis of hvKp are largely unknown [93].

1.3 Treatment and AMR

The infections caused by wild type cKp and hvKp strains can be treated efficiently with a variety of antimicrobial agents as they are highly susceptible to antimicrobials [94]. The first line and most used antibiotics to treat these *K. pneumoniae* infections are β-lactams. β-lactams constitute the most widely prescribed group of antimicrobials in human medicine given their therapeutic efficacy, large spectrum of activity, low cost, and low toxicity [95, 96]. Frontline β-lactams to treat *K. pneumoniae* infections include penicillin's combined with β-lactamase inhibitors and cephalosporins [94]. In some cases, depending on the site and severity of the infection, other non-β-lactam antibiotics are also used as therapy, such as aminoglycosides (amikacin or gentamicin), cotrimoxazole (also named trimethoprim/sulfamethoxazole), fosfomycin, tigecycline, and fluoroquinolones (levofloxacin and ciprofloxacin) [94].

The effectiveness of first-line treatments for bacterial infections has been compromised by the emergence of AMR strains, which has made selecting an appropriate antibiotic regimen a major challenge for physicians. In fact, the alarming increase in AMR acquisition in cKp is what has turned into a clinically significant pathogen [97]. As hvKp tends to have none or low levels of AMR and, in consequence, there are currently sufficient treatment options available, this section will focus on cKp.

Among the acquired AMR in cKp, the major concern are strains that have acquired resistance to **third-generation cephalosporins (3GC)** [98] and **carbapenems** [99]. Currently, the worldwide prevalence of *K. pneumoniae* resistant to 3GC is very high and rates are still trending upward (Figure 1.5). In Europe, in 2020, 33.9% of *K. pneumoniae* reported to the European Antimicrobial Resistance Surveillance Network (EARS-Net) were resistant to this group of antibiotics [12]. In Spain, 3GC resistance has reached 40% of the total *K. pneumoniae* hospital isolates [12, 100].

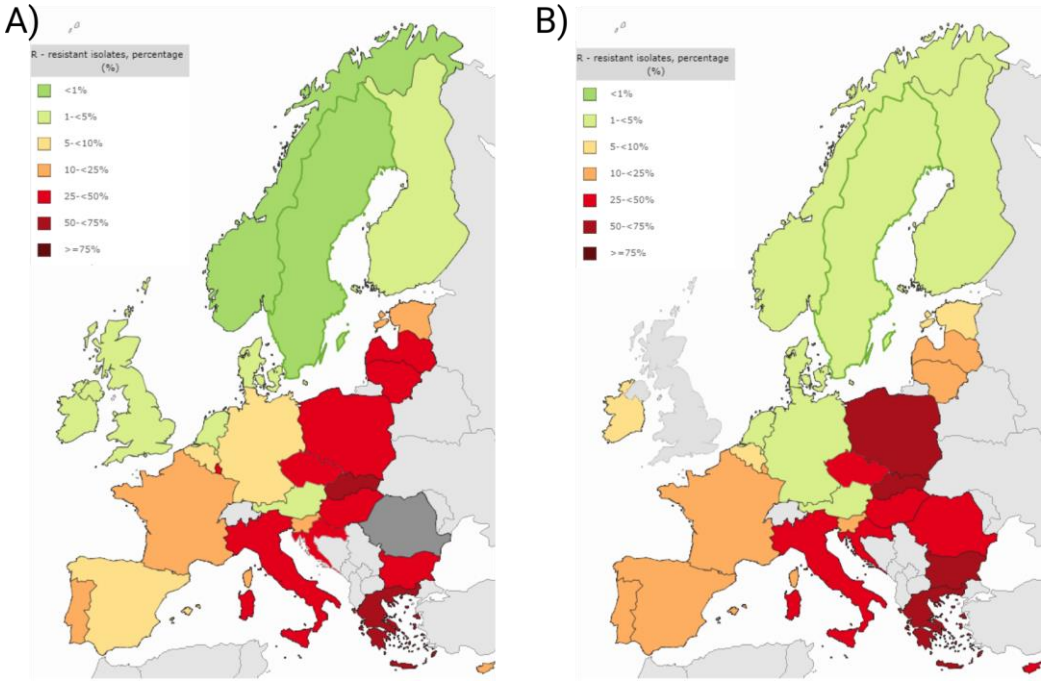


Figure 1.5. Prevalence of combined resistance to third generation cephalosporins, fluoroquinolones and aminoglycosides in Europe in A) 2010 and B) 2020. Extracted from [101].

The infections caused by 3GC-resistant strains can be treated with second-line antibiotics such as carbapenems, aminoglycosides, or fosfomycin. However, 3GC-resistance is more often found in combination with resistance to fluoroquinolones and aminoglycosides than on its own [12]. Hence, the preferred therapeutic option to deal with severe and multidrug-resistant infections caused by 3GC-resistant *K. pneumoniae* are **carbapenems**. As long as carbapenems or other proper therapies are available, 3GC-resistance does not pose a major risk [102]. Nevertheless, an inadequate initial therapy related to the misidentification of 3GC-resistance is associated with high mortality, especially in severe infections such as bacteraemia [103, 104].

The impact of 3GC-resistant *K. pneumoniae* has led to an increased use of carbapenems as therapy, causing a large selective pressure toward carbapenem-resistant strains. Alarming increases in the prevalence of **carbapenem-resistant *K. pneumoniae* (CRKp)** strains have been reported worldwide [10]. In 2005, almost all the European regions were free of CRKp but ten years later, in 2015, it had emerged in many countries [12]. Later, from 2016 to 2020, CRKp showed a significant increase in prevalence [12], reaching endemic rates in some regions (e.g., Italy and Greece) [101] (Figure 1.6). Indeed, in 2019, CRKp was reported as the fastest-growing cause of drug-resistant infections in European hospitals [12, 105]. In 2020, almost a quarter of EU/EEA countries reported more than 10% of CRKp infections [12], and Spanish hospitals reported an average of 5% of CRKp infections [12] (Figure 1.6 B).

Even though the prevalence of CRKp infections remains relatively low, the impact on the health burden is large due to the high attributable mortality and morbidity of these infections. It has been estimated that mortality associated with hospital acquired CRKp infections ranges from 48% to 65% [48–52], significantly higher compared to carbapenem-susceptible infections [106]. This is because resistance to a single class of antibiotics in CRKp is hardly reported and, in addition to resistance to all β-lactams, CRKp strains almost always show resistance to several other key antimicrobial classes [12]. This limits drastically the treatment options and therapy ends up relying on drugs that are less effective, more toxic, and/or not widely available, such as tigecycline, polymyxins, and selected aminoglycosides [96, 107–109].

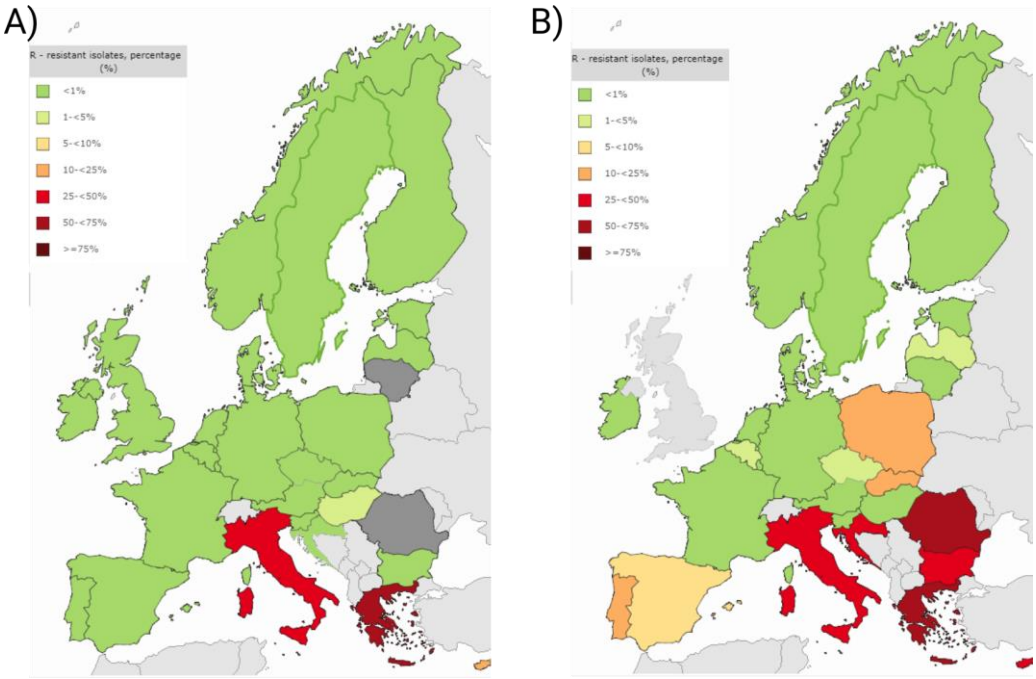


Figure 1.6. Prevalence of carbapenem resistance in Europe in A) 2010 and B) 2020. Extracted from [101].

As a consequence of the rise of CRKp, we face an increasing scarcity of effective treatments. Mild *K. pneumoniae* infections have become very challenging to treat, and more serious infections, such as pneumonia and bacteraemia, have become life-threatening [110]. This situation affects not only immunodeficient individuals but also healthy ones. New antibiotics, such as cefiderocol, or new β-lactam and β-lactamase inhibitors combinations [111], such as ceftazidime-avibactam [112], meropenem-vaborbactam [113], and imipenem-relebactam are starting to be used as a treatment for these infections. Nevertheless, although little is known about possible resistance mechanisms to these new treatments, resistant strains have already been reported [114, 115].

1.4 The mechanisms underlying ARM in *K. pneumoniae*

K. pneumoniae is intrinsically resistant to ampicillin, ticarcillin, and piperacillin due to the chromosomal production of an SHV-1 enzyme. Moreover, it can acquire multiple other mechanisms that confer resistance towards different classes of

antibiotics including other β-lactams, aminoglycosides, quinolones, tigecycline, and polymyxins. Antimicrobial resistance mechanisms include alteration of the drug target, enzymatic modification or degradation of the drug, reduction in cell permeability limiting drug uptake, or active drug efflux reducing antibiotic concentration in the cell (Figure 1.7) [116].

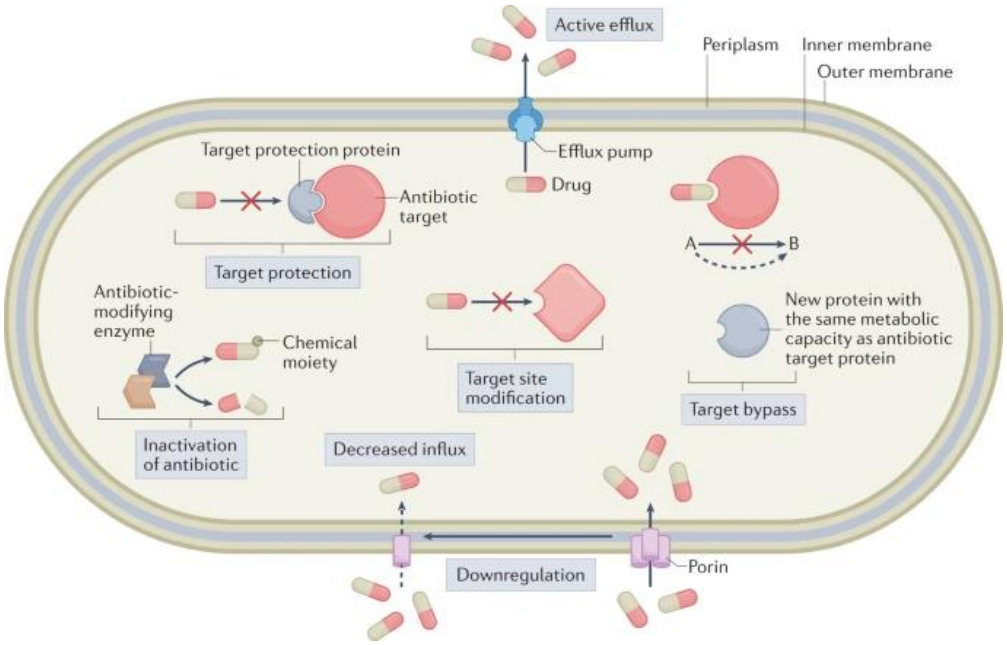


Figure 1.7. Overview of the molecular mechanisms of antibiotic resistance. Extracted from [116].

Antimicrobial resistance in *K. pneumoniae* is mainly caused by the acquisition of new **AMR genes** rather than by spontaneous mutations in chromosomal genes. The number of acquired AMR genes in this species is larger (>400) than in any other AMR pathogen such as other *Enterobacterales* or *P. aeruginosa* [34]. In *K. pneumoniae* these genes are primarily encoded in **mobile genetic elements** (MGEs), such as plasmids and transposons [19].

1.4.1 AMR determinants to β-lactams

β-lactam agents include four main classes of antibiotics: penicillin's, cephalosporins, carbapenems, and monobactams. These agents act by binding the β-lactam ring to

the penicillin-binding proteins of the bacteria responsible for peptidoglycan synthesis. While the penicillin-binding proteins are bonded to the β-lactam rings, bacteria cannot synthesise the cell wall leading to osmotic instability or autolysis [116].

The major AMR mechanism to this antibiotic family is drug modification by the action of β-lactamases [116]. These are enzymes capable of breaking the amide bond of the β-lactam ring inactivating the antibiotic. Since their discovery, hundreds of different β-lactamases have been reported (> 7,800) [116]. Two major classification schemes are currently used for β-lactamases (Figure 1.8). β-lactamases can be classified into four major functional groups (groups 1 to 4), based on their substrate specificity and response to inhibitors [117], or into four molecular classes (A through D), based on the similarity of the amino acid sequence [118].

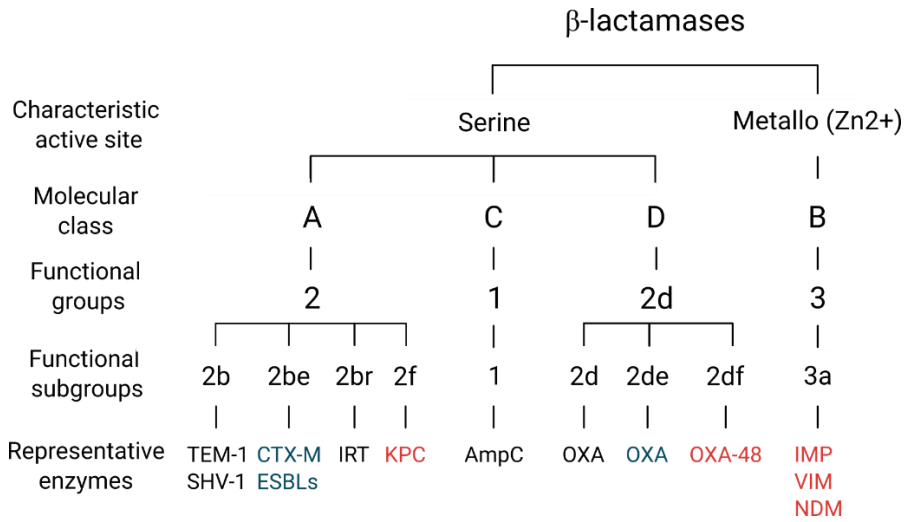


Figure 1.8. Molecular and functional classifications of β-lactamases. Representative enzymes are coloured in blue if they are ESBL or in red if they are carbapenemases. Adapted from [119].

Narrow and extended-spectrum β-lactamases

Molecular class A (functional group 2) was the first group of β-lactamases to be identified. In the early 1960s, the **TEM-1** β-lactamase was isolated and named after Temoneira, the patient from whom it was first reported [120]. Subsequently, in

1974, the **SHV-1** (sulfhydryl reagent variable) β-lactamase was discovered. SHV enzymes are responsible for the intrinsic resistance to ampicillin in *K. pneumoniae* as they are encoded in the chromosome [121]. Both TEM-1 and SHV-1 are classified in subgroup 2b and are only capable of hydrolysing penicillins and narrow-spectrum cephalosporins.

By the mid-1980s, variants of SHV-1 or TEM-1 β-lactamases appeared with a much broader substrate spectrum and were able to hydrolyse extended-spectrum cephalosporins. Hence, the variants were termed **extended-spectrum β-lactamases (ESBLs)** and were classified as 2be where the "e" denotes the extended spectrum (Figure 1.8). There is no standard definition for ESBLs, as they constitute a highly heterogeneous group of enzymes with different substrate spectra [122]. ESBLs refer to the enzymes with the capability of hydrolysing cephalosporins (first, second and third generation), but with no activity against carbapenems or cephamycins, and that are inhibited by β-lactamase classical inhibitors [123].

During the emergence of ESBLs, the major enzymes found in *K. pneumoniae* were SHV and TEM variants [124, 125]. However, in 1989, **CTX-M** (cefotaximase from Munich) enzymes were discovered [126] and by the 2000s they had replaced TEM and SHV enzymes as the most common ESBLs worldwide [127–129]. Various CTX-M variants have been reported, with CTX-M-15 being the most prevalent. The success of CTX-M enzymes is attributed with the efficiency and spread of the plasmids carrying the $bla_{\text{CTX-M}}$ genes [130].

Other class A ESBL families, such as PER, VEB and GES, can also be found in *K. pneumoniae* but with a much lower prevalence in a limited number of countries [131].

In addition to class A enzymes, the ESBL phenotype can be given by class D **OXA (oxacillinase)** enzymes (Figure 1.8)[132]. OXA-type enzymes are a broad group of enzymes with variable amino acid sequences and substrate profiles. Some variants are narrow-spectrum β-lactamases (OXA-1), some are ESBLs, and a few show carbapenemase activity (OXA-48, OXA-162, OXA-163, OXA-181 and OXA-232) [133].

Narrow-spectrum OXA enzymes are classified in functional group 2d, and those with ESBL phenotypes are categorised in the 2de subgroup.

Inhibitor-resistant β-lactamases

Inhibitor-resistant β-lactamases belong to the functional group 2br (Figure 1.8) [117] and are TEM and SHV enzymes that are not effectively inhibited by β-lactamase inhibitors, such as clavulanic acid and sulbactam [134]. Mutations conferring resistance to inhibitors make these enzymes less efficient at hydrolysing cephalosporins; so, only a few variants have been described to maintain the resistance to inhibitors and also have ESBL phenotype (e.g., TEM-152) [135]. These enzymes are not common and have not achieved the relevance of ESBLs in *K. pneumoniae* [135].

Plasmid-mediated AmpC β-lactamases

AmpC enzymes belong to Ambler class C and functional group 1. AmpC enzymes have a similar antibiotic spectrum to that of ESBLs but are not inhibited by β-lactamase inhibitors [136]. They are encoded by the *ampC* gene, which was first found in the chromosome of clinically relevant enterobacterial species including *Citrobacter freundii*, *Serratia marcescens* and *Enterobacter aerogenes*. AmpC chromosomal basal expressions are generally low and therefore do not pose a major risk of AMR. Nevertheless, these enzymes can reach high expression levels when induced or de-repressed and expressed constitutively [137].

In the 1990s, parallel to the ESBL explosion, AmpC enzymes were first reported on plasmids, and plasmid encoded AmpC (**pAmpC**) began to emerge. AmpC is highly expressed when they are encoded in plasmids, mainly due to strong promoters and high gene copy numbers. pAmpC is frequently found in *K. pneumoniae* being the most frequently found enzymes CMY, DHA, FOX, and MOX types [136, 138]. Moreover, when combined with porin loss, pAmpC can increase the MICs of carbapenems, reaching resistance levels [139, 140].

Carbapenemases

Carbapenemases pose the most significant concern among β-lactamases, as they are highly efficient in hydrolysing almost all β-lactams available including carbapenems, the last line of β-lactam treatment [141, 142]. Carbapenemases include enzymes from molecular classes A, B, and D, or functional groups 2f, 3 and 2d, respectively (Figure 1.8). Each class shows different substrate spectra and different carbapenem resistance levels [133, 143].

K. pneumoniae has emerged as a leading cause of carbapenem-resistant *Enterobacteriaceae* infections globally [144, 145] due to the acquisition of KPC and OXA-48 followed by NDM, VIM, and IMP enzymes [146].

Class A **KPC (*Klebsiella pneumoniae* Carbapenemase)** can hydrolyse all β-lactams and clavulanic acid but show low to moderate resistance to carbapenems [147]. KPC-2 emerged in the late 1990s in the United States of America (USA) in a *K. pneumoniae* strain [141]. To date, according to the Reference Gene Catalog of the NCBI [148], 146 different KPC enzyme variants have been identified. Of note, KPC-1 and KPC-2 have been shown to be identical and are referred to as KPC-2. KPC-2 and KPC-3 (differing in just one amino acid) are the most frequent variants and the best studied carbapenemases. Their clonal and global geographical dissemination has been well documented [109, 110, 142, 149, 150]. The endemicity of KPC extended from the USA to South America, Asia, and Europe, especially Greece, Italy, and Poland [149, 150]. Sporadic outbreaks have been reported in every European country, and elsewhere [109, 151].

Class D **OXA-48-like** enzymes can affect carbapenems but with lower hydrolytic efficiency than other carbapenemases [109, 151, 152]. The first OXA enzyme to demonstrate carbapenemase activity was OXA-48, which was initially isolated from a *K. pneumoniae* strain in Turkey in 2001 [153]. Since then, OXA-48 enzymes have become endemic in Turkey, other Middle Eastern countries, and North Africa [154], and have spread worldwide [155]. OXA-48 emerged as the predominant carbapenemase in other Mediterranean countries such as Spain and France, being a common cause of healthcare-associated outbreaks [151, 156]. Although OXA-48 is

found in other *Enterobacterales*, *K. pneumoniae* remains the most common species associated with this class of enzymes [145, 146]. Since its discovery, several variants with similar activity against carbapenems have been reported such as OXA-181, OXA-244, OXA-232, and OXA-162 [133, 143].

Class B or **metallo-β-lactamases (MBL)** are a group of enzymes that hydrolyse all β-lactams, except monobactams (aztreonam), and are not inhibited by clavulanic acid or other similar inhibitors [157]. MBLs in *K. pneumoniae* include IMP (active against imipenem; imipenemase), VIM (Verona integron-encoded metallo-β-lactamase), and NDM (New Delhi metallo-β-lactamases) [158].

IMP-1 was the first MBL to be described. It was identified in Japan in 1988 in a *P. aeruginosa* isolate [159]. Currently, 81 different variants of IMP have been identified [148]. These enzymes have disseminated worldwide [158], yet they are mostly reported in Southern and Eastern Asia [142]. IMP-type carbapenemases are primarily found in *P. aeruginosa*, *Acinetobacter* spp. and, to a lesser extent, in *K. pneumoniae* [160].

VIM-1 was first described in 1997 in Italy in a *P. aeruginosa* isolate [161]. To date, at least 20 VIM variants have been reported [148], with VIM-2 being the most common. Like IMP, *K. pneumoniae* VIM producers have low prevalence; however, they have become especially relevant in the Mediterranean region, especially in Spain [146, 156].

NDM-1 was first identified in a *K. pneumoniae* strain isolated from a Swedish patient transferred from a New Delhi hospital in 2008 [162]. To date, 50 different variants have been reported [148]. Unlike the other MBLs, NDMs have rapidly become prevalent in many regions around the globe. NDM-producing *K. pneumoniae* is considered to be endemic in India, Pakistan, Bangladesh [163, 164], the Balkan states [165], and North African countries [166]. Its global dissemination is strongly associated with travel and receipt of medical care in Asia [167]. NDM-1 is the most prevalent variant and has been reported to cause outbreaks in Europe, America, Africa, China, and Australia [168–172].

Decreased permeability by porin loss

In addition to β-lactamases, a relevant, non-specific AMR mechanism to β-lactam agents is porin loss. Porins are transmembrane channels found in the outer membrane of Gram-negative bacteria that regulate the diffusion into the periplasm of hydrophilic molecules, such as nutrients or antibiotics (Figure 1.7). These channels are the main route of entry of β-lactams. The loss of porins decreases the permeability of the outer membrane, limiting the influx and, therefore, the accumulation of antibiotics in the periplasmic space where they are effective [173].

K. pneumoniae produces two major porins, **OmpK35** and **OmpK36** [174]. Mutational loss of porins has been associated with an increase of the minimum inhibitory concentration (MIC) to multiple β-lactams [175]. Permeability defects coupled with some ESBLs (i.e., CTX-M-15, CTX-M-33) can contribute to reduced carbapenem susceptibility [176].

1.4.2 AMR determinants to other antibiotic classes

In *K. pneumoniae*, resistance to β-lactams is commonly coupled with resistance to other clinically relevant antibiotics such as aminoglycosides, quinolones, polymyxins, tigecycline and fosfomycin. Different AMR determinants are responsible for the resistance to each antimicrobial family. Moreover, overexpression of efflux pumps has been associated with reduced susceptibility to many of these antibiotics [177]. Efflux pumps are active (energy-dependent) systems that, when overexpressed, increase the liberation of antibiotics outside the cell, limiting its cytoplasmic concentration (Figure 1.7) [116]. Detailed mechanisms of resistance for each family are summarised in Table 1.1.

Table 1.1. Resistance mechanisms and determinants reported in *K. pneumoniae* for each antibiotic class.

Antibiotic	Resistance mechanism(s) and determinants (proteins/genes)
Penicillin	Drug modification: Narrow-spectrum β-lactamases (TEM-1, SHV-1*), ESBLs, AmpC and Carbapenemases
3GC	Drug modification: ESBLs, AmpC, Carbapenemases
Carbapenems	Drug modification: Carbapenemases Overproduction of AmpC in combination with porin loss by mutation (OmpK35, OmpK36) ** Specific ESBLs (SHV-12, CTX-M-15) in combination with porin loss by mutation (OmpK35, OmpK36)**
Aminoglycosides	Drug modification by acetylases, aminoacyl transferases, phosphotransferases and streptomycin-inactivating enzymes: aac, aadA, aadB, aph, rmt, strA, and strB Target modification: 16S methylases (armA, rmtB) Efflux pumps: AcrAB-TolC, KPnEF
Fluoroquinolones	Target protection by the DNA-binding protein Qnr: qnrA, qnrB, qnrS, Target modification: Mutations in gene encoding DNA gyrase (gyrA, gyrB) and the topoisomerase IV (parC, parE) Overexpression of efflux pumps: AcrAB-TolC or oqxAB* ** and plasmid-acquired QepA** Drug modification: aac(6')-Ib-cr
Polymyxins	Target modification: chromosomal mutations in the genes involved in the PmrAB-PhoPQ two component systems and their regulators (pmrA, pmrB, phoP, phoQ, mrgB, or crrB) and plasmid-mediated mcr genes
Tigecycline	Upregulation of efflux pumps: acrAB-TolC or oqxAB * ** Mutations in rpsJ Drug modification: Tet(x)
Macrolides	Target modification: erythromycin ribosome methylase (erm) Drug modification: phosphotransferases (mph) Efflux pumps: msr, mef.
Rifampicin	Drug modification: rifampin ADP-ribosyltransferase (Arr)
Sulfonamides	Target modification: dihydropteroate synthase (folP, sul1, sul2, sul3)
Tetracycline	Overexpression of efflux pumps: tetA, tetB
Trimetropin	Target modifications: dihydrofolate reductase (dfr)
Fosfomycin	Drug modification: Glutathione S-transferases (fosA) * ** Decrease of the permeability by mutations affecting transporters (GlpT, UhpT)

* Intrinsic resistance in *K. pneumoniae*

** Low level of resistance

1.5 Understanding the drivers of AMR

Plasmids and other MGEs represent an important fraction of the large accessory genome of *K. pneumoniae* and are responsible for the acquisition of the AMR genes and virulence factors that have shaped its evolution [178, 179]. It has been suggested that *K. pneumoniae* strains may be particularly permissive to plasmid carriage and maintenance. Indeed, *K. pneumoniae* presents larger plasmid loads than other relevant AMR pathogens [34]. Wild-type *K. pneumoniae* strains can harbour several plasmids, between 4 and 6 on average [31]. Plasmids found in *K. pneumoniae* genomes have a broad size range, from a few to several hundred kilobases, including small plasmids in high-copy-number and large plasmids but in low-copy-number [34].

Plasmids are MGEs typically found as extrachromosomal DNA molecules that can replicate semi-independently of their hosts. Plasmids usually present a **backbone** filled with core plasmid functions related to replication, survival, mobility, and stability of the plasmid, recombined with **accessory** functions that provide the tools for the cell to thrive in a particular environment (e.g., metabolic, virulence, AMR). Plasmids are extremely diverse as they are dynamic molecules with the ability to capture, accumulate, and rearrange accessory genes. This plasticity is largely due to the action of other MGEs that move between plasmids and even chromosomes. Common MGEs associated with plasmids are gene cassettes, transposons (Tn), integrons (In), and insertion sequences (IS) [180] (Figure 1.9 A).

The major route of plasmid mobilisation between cells is **conjugation,** a **horizontal gene transfer (HGT)** mechanism that requires cell-to-cell contact through a sex pilus that will transfer the DNA [178]. Plasmids that mobilise through conjugation, or **conjugative plasmids**, are self-transferable as they harbour the repertoire of genes needed to produce the pilus and the transfer machinery. The mobility genes include the origin of transfer (*oriT*), relaxases, type IV coupling proteins (T4CP), and the mating pair formation complex (type 4 secretion system, T4SS) (Figure 1.9 B). Yet not all plasmids are self-transferable or conjugative. **Mobilizable plasmids** lack part of the conjugation machinery but can be transferred by taking advantage of the conjugative apparatus of other genetic elements in the cell whereas **non-**

mobilizable plasmids do not carry enough mobilisation genes and thus cannot be transferred by conjugation [181] (Figure 1.9 B). Nevertheless, there are other HGT mechanisms that can permit DNA mobilisation, such as **transformation** (capture of free extracellular DNA fragments) or **transduction** (transference of genetic material between bacteria mediated by bacteriophages) [182].

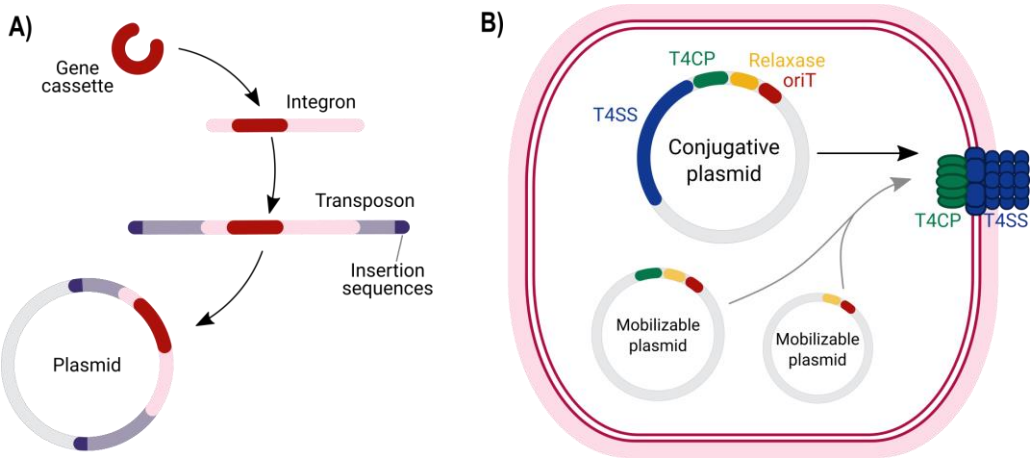


Figure 1.9. A) AMR genes are usually encoded within mobile and transposable elements such as gene cassettes, transposons or integrons, that can jump, integrate and accumulate into the same plasmid. B) Plasmids are **conjugative** if they present all the transfer machinery, **mobilizable** if they can use the transfer machinery of other plasmids or **non-mobilizable** if they do not carry any mobilisation gene. Adapted from [181].

The changing nature of plasmids complicates greatly subtyping and studying the relationships among plasmids even of the same type. Historically, plasmids have been classified based on plasmid incompatibility (Inc) [183, 184]. On this basis, two plasmids that are unable to coexist stably in the same cell line will belong to the same Inc group. Generally, closely related plasmids tend to be incompatible, as they share the same replicon and compete for the same replication mechanism. Therefore, to test the incompatibility of a plasmid, all combinations of plasmids must be tested in competition assays. To replace the competition assays, a more convenient PCR-based replicon typing (PBRT) was proposed in which the classification into an Inc group is based on the sequence of the initiation of replication (Rep) protein [185]. Another typing scheme based on mobility typing (MOB types) has been proposed in which plasmids can be classified as conjugative,

mobilizable or non-mobile, and within types, they can be clustered using the relaxase sequence [186].

Both classification schemes, PBRT and MOB typing, are widely accepted [187]; however, both schemes have limitations. The PBRT methodology must deal with the fact that plasmids frequently carry multiple replicons and, thus, some taxa, especially enterobacteria, carry many non-typable plasmids by this technique. MOb-typing can overcome these problems as plasmids rarely carry more than one relaxase gene and the classification of mobility is independent of the taxa [188]. However, not all plasmids have relaxase or mobility loci, so this classification is also limited to mobile plasmids. Moreover, both are based on a single locus; so, they provide limited resolution and lack phylogenetic depth [189] and do not inform about the genetic content as plasmids of the same type can present different lengths, cargo genes, and even backbones [190 – 192].

The development of long-read high throughput sequencing has provided the possibility of obtaining complete plasmid sequences and, thus, comparing whole sequences rather than loci. Although long-read sequencing has clear advantages, the costs and limitations of the current platforms still prevent its wide application to the study of plasmids. However, the popularity of this type of analysis is growing and as the technology continues to improve and become more cost-effective, it is expected to become more widely used for plasmid analysis [187, 193].

1.5.1 Plasmids in *K. pneumoniae*: AMR and virulence

Plasmid research in *K. pneumoniae* has been primarily focused on AMR. In general, seven major plasmid families have been associated with AMR among *Enterobacteriaceae* and *K. pneumoniae* which are IncF, A/C type (IncA/C), IncL, IncI, IncHI, IncX, and IncN [31, 194]. For instance, carbapenemase genes can be found encoded in different plasmid Inc types or backbones.

The spread of MBL genes is a very complex process, involving multiple mobile genetic elements [172]. $bla_{IMP,}$ bla_{VIM}, and bla_{NDM} genes are found in different genetic environments (Tn, In…) and in different plasmid types [157]. bla_{IMP} and bla_{VIM} are often found as gene cassettes in class 1 integrons whereas bla_{NDM} genes have not

been found associated with any MGE. These genes have been reported in at least 20 different plasmid types [195].

KPC-encoding genes are mainly associated with transposon Tn4401 encoded in many different plasmid backbones [196]. The epidemic ST258 clone carrying *bla*$_{KPC}$, isolated in the early 2000s in the USA, demonstrated that both *bla*$_{KPC-2}$ and *bla*$_{KPC-3}$ genes were encoded on IncFIIk plasmids [197]. In addition to IncFIIk, *bla*$_{KPC-2}$ and *bla*$_{KPC-3}$ were also identified on a wide variety of plasmids, including the IncN, IncX3, IncR, IncHI1 and IncI2 types [198].

Unlike MBLs and KPC, the *bla*$_{OXA-48}$ gene is mainly disseminated by the composite transposon Tn1999 inserted in a conserved plasmid termed pOXA-48a [199]. This is a ~60 kb, highly conserved IncL (IncL/M) conjugative plasmid that does not contain any other AMR genes [199]. The *bla*$_{OXA-48}$ gene has rarely been detected in other Inc plasmids [155]; however, it is one of the most widely distributed carbapenemases in terms of geography and species [155]. Its success is due to the extremely high conjugation rate of the pOXA-48a plasmid, which ensures its transfer even in the absence of antibiotic pressure [109].

In addition to AMR, MGEs such as plasmids and integrative conjugative elements (ICEs) also serve as primary carriers for virulence-related genes in *K. pneumoniae*. For instance, the yersiniabactin siderophore is usually encoded in ICEs known as ICE*Kp* that inserts into the *K. pneumoniae* chromosome [88]. Fourteen distinct ICE*Kp* variants (ICE*Kp1* - ICE*Kp14*) are recognised [88]. ICE*Kp* are present in ~40% of the general *K. pneumoniae* population including MDR strains [32, 88]. ICEs are self-mobilizable structures, which can explain why these virulence factors are commonly extended in both pathotypes [88].

Additionally, hvKp strains usually contain large (~ 200 Kb) hypervirulence plasmids, initially identified as pLVPK and pK2044. These plasmids encode for the siderophores salmochelin (*iro*) and aerobactin (*iuc*) [89] and the mucoid phenotype (*rmpA* and *rmpA2*) [200]. These plasmids are mobilizable but not conjugative [201] which limits their frequency of HGT, thus explaining why they are highly restricted to hvKp lineages.

1.5.2 The role of plasmids in the evolution of *K. pneumoniae*

Plasmids are known to play a critical role in the evolution of AMR [202] and the success of *K. pneumoniae* as an AMR pathogen [203]. While their role in disseminating AMR genes among pathogens in healthcare settings has gained the most attention [204], plasmids have also contributed to the evolution of AMR resistance in several other ways.

MGEs, specially conjugative plasmids, have been suggested to play an important role in moving AMR genes from the chromosomes of environmental bacteria, where they probably originated [205], to clinically relevant bacterial species [206, 207]. *K. pneumoniae* has been suggested as a crucial species in this transfer due to its ubiquitous presence in the environment, frequent occurrence in clinical settings, and high potential for plasmid acquisition [34]. Indeed, the vast majority of AMR genes in this species result from the acquisition of AMR conjugative plasmids [19, 208].

Moreover, plasmids affect the phenotypic resistance levels of the AMR genes encoded therein [209, 210]. This effect is frequently reported in β-lactamases, as their resistance levels often depend on their gene dosage [211]. Gene dosage is higher when these genes are encoded in plasmids than in the chromosome as plasmids are usually in a high copy number [212, 213]. Therefore, AMR genes in plasmids tend to be more highly expressed which has likely triggered the movement of AMR genes from chromosomes to plasmids allowing the successful dissemination of the latter [202].

As plasmids enhance the mobility of AMR genes, they also increase the probability of these genes accumulating in a single plasmid or of multiple AMR plasmids accumulating in a single cell, resulting in multidrug-resistant (MDR) and even pan-drug-resistant (PDR) phenotypes, as it has occurred in *K. pneumoniae* [214]. Furthermore, plasmids have contributed to the recent emergence of convergent strains exhibiting MDR coupled with hypervirulent factors (**MDR-hvKp**) (Figure 1.10) [45, 215 – 217]. Convergence can occur if MDR strains acquire virulence plasmids or hvKp strains acquire resistance plasmids. It can also result from the acquisition of mosaic or hybrid plasmids harbouring both virulence factors and AMR genes [218].

While cases of co-harbouring plasmids have been frequently reported [215, 219], mosaic plasmids have been rarely documented [220 – 222]. However, the emergence of mosaic plasmids poses a significant threat as they can acquire conjugative features, allowing for self-transfer and dissemination [201, 223].

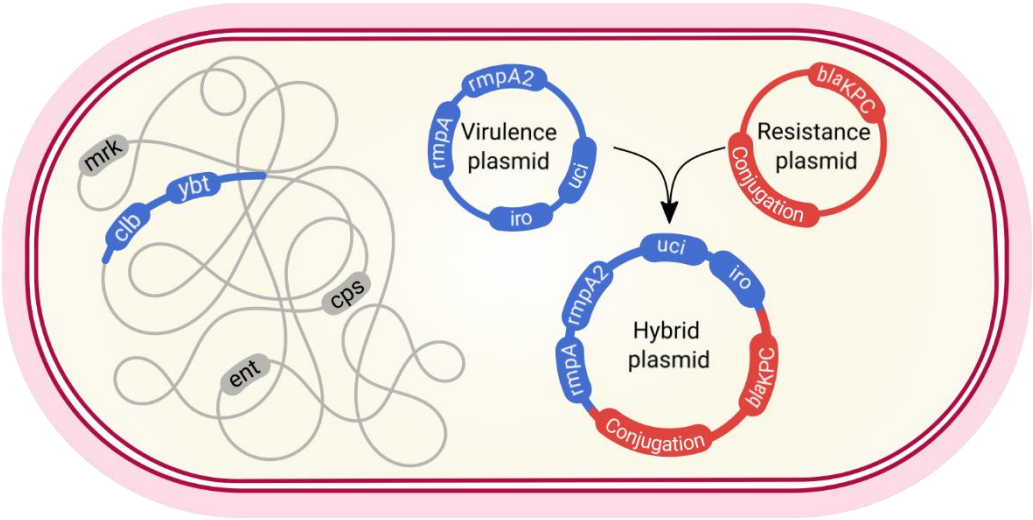


Figure 1.10. The molecular evolution of multidrug and hypervirulent *K. pneumoniae*. MDR-hvKp convergence can occur by the acquisition of both virulence and resistance plasmids or hybrid plasmids harbouring both virulence factors and AMR genes. Adapted from [218] and [217].

1.6 High-risk clones

Although hundreds of *K. pneumoniae* lineages can cause infections in humans [151], only a small fraction of MDR and hypervirulent lineages contribute to most of them [203]. These lineages have been denominated "**high-risk clones**" (HRC) [19] or "global problems" [31, 224], as they are distributed worldwide and are highly prevalent in healthcare settings around the world. There is no precise definition of high-risk clones, but they are generally understood as those lineages that have acquired adaptive traits, usually increasing virulence or AMR, that will enhance their ability to colonise, gaining an advantage to spread [225], and that present a wide or even global distribution [19].

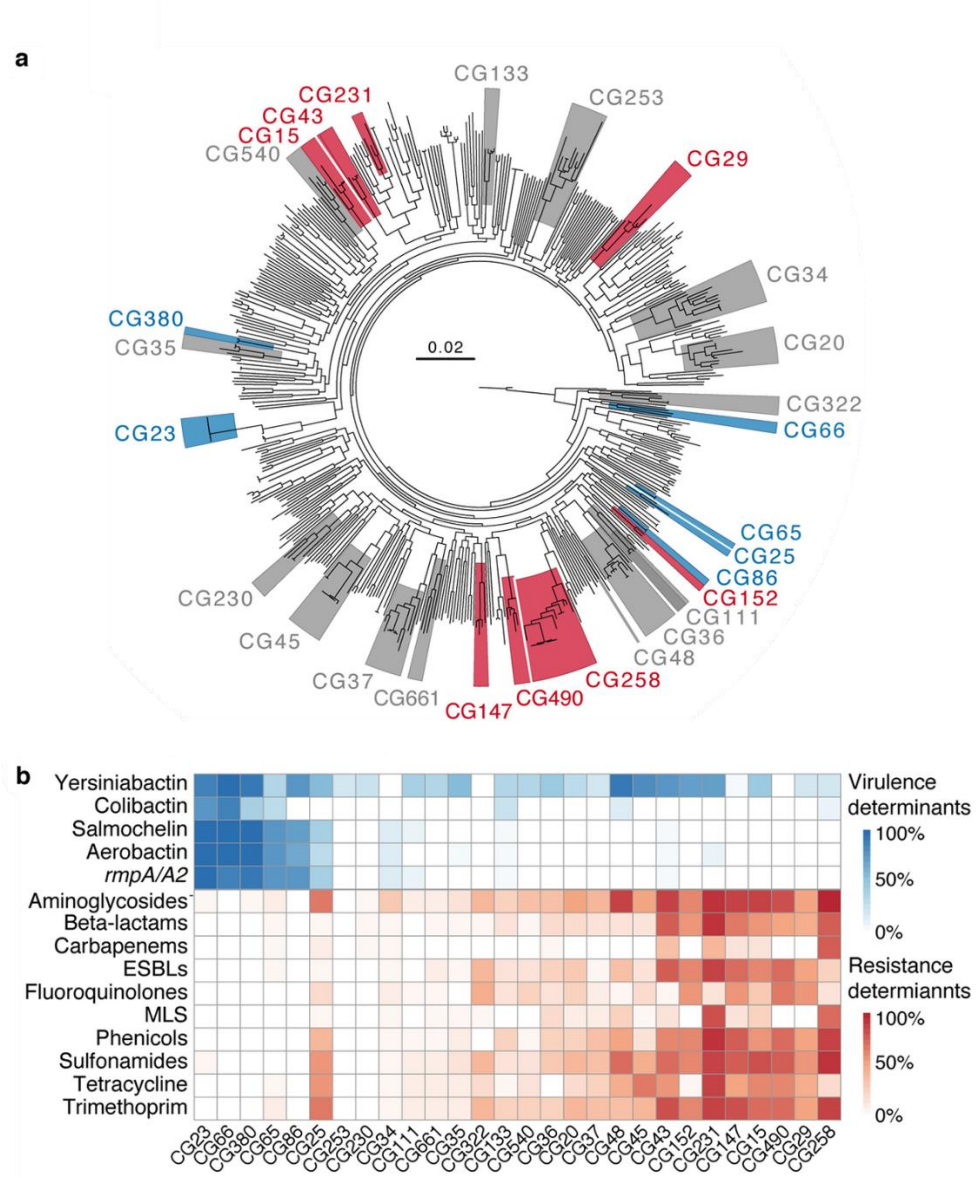


Figure 1.11. A) Phylogenetic tree inferred using maximum likelihood for 2,265 *K. pneumoniae* genomes. Hypervirulent clones are marked in blue, multi-drug resistant in red and unassigned clones in grey. B) Distribution of virulence and resistance determinants by clone. Intensity of box shading indicates the proportion of genomes harbouring the key virulence loci (blue) or acquired genes conferring resistance to different classes of antimicrobials (red). Extracted from [63].

1.6.1 MDR lineages

The emergence of untreatable AMR infections has reached a critical point due to the rise of MDR-HRC in hospital settings. MDR in *K. pneumoniae* is associated with the clonal expansion of a reduced number of lineages whose success has been mainly linked to the acquisition of ESBLs and carbapenemase genes [31,179], notably by the explosive dissemination of CTX-M enzymes after the introduction and massive use of 3GC [129]. Nowadays, most MDR high-risk lineages can be found globally distributed and associated with different MGEs carrying ESBLs, carbapenemases and other AMR genes [31, 32, 179].

MDR high-risk clones include the well-studied CG258, alongside CG307, CG147, CG14/15, and CG101, among others. These lineages account for most 3GC- and carbapenem-resistant infections in many hospital settings around the world, including the Americas, Europe, and Asia [31]. Yet, its prevalence and distribution in many other countries, mostly LMICs, are represented by only a few large cities or tertiary hospitals [226, 227]. Consequently, further studies may change our understanding of these as "global problems".

CG258 (ST258, ST11, ST340, ST437 and ST512) is the most widely recognised high-risk clonal group and has been extensively studied due to its role as the major disseminator of KPC [150, 228–230]. CG258 is a diverse clade formed by several STs [231] being **ST258**, **ST11**, and **ST512** the most relevant ones. The KPC pandemic is largely attributed to ST258 and ST512. Paradoxically, the KPC pandemic began in 1996 in the USA with the first detection of the KPC-2 enzyme in a non-ST258 strain. Afterwards, an ST258 KPC-2 was disseminated throughout the USA, and later spread to South America, Israel and Europe, primarily Italy and Greece [151, 232, 233]. The dissemination ended with ST258 and ST512 carrying KPC-2 and KPC-3 becoming endemic in many countries, mostly in America and Europe [150]. In Asia, especially China, KPC-2 is highly prevalent, but its dissemination is almost strictly due to a different evolutionary history associated with ST11 [234, 235]. Moreover, the genetic environment surrounding the bla_{KPC-2} gene in isolates from China is different from the classic 'Tn4401' and the pKpQIL plasmid found in ST258 isolates [236].

Although the global expansion of KPC has largely been attributed to ST258 and its derivative ST512, these lineages are comparatively rare outside the Americas and southern Europe [151]. On the contrary, ST11 is a much more widely distributed lineage and has contributed to the spread of more diverse carbapenemases and ESBLs, including KPC, VIM, IMP, NDM, OXA-48 and CTX-M-15 [237, 238]. Indeed, ST11 is one of the lineages with higher AMR rates [32]. ST11 is the dominant cause of CRKP in Asia [239, 240] and is highly prevalent in Europe [151, 241], Canada [242], or Brazil [243], and it has been frequently isolated in animal and environmental sources [244–248]. Recently, it has been reported that ST11 is displacing ST258 as the dominant CRKp in some endemic regions [244]. Unlike the other lineages of the CG, ST11 is a highly diverse lineage with more than 15 different KL types [249 – 251]. It has been suggested that ST11 should be segregated into its different sub-lineages rather than be considered as a single clone [251]. Each ST11 sub-lineage shows a different geographic distribution and prevalence. For instance, KL47 and KL64 are the major sublineages in China, whereas in other areas, such as Europe, different and more diverse KLs are found [249, 252].

Apart from CG258, studies are scarce and, although some other lineages are also very prevalent and a serious cause of MDR outbreaks worldwide, they have not received so much attention and little is known about their epidemiology.

CG15 includes ST15 and ST14, with ST15 being more prevalent. This CG is amongst the most common 3GC-resistant lineages in many countries. The success of this CG has been attributed to the CTX-M-15 enzyme. Currently, it has also acquired carbapenemases such as KPC [253], NDM-1 [254], OXA-48 [255], OXA-232 [256], and VIM-1 [257]. As with the other non-CG258 lineages, the evolution, global prevalence, and dissemination of this lineage have not been assessed and all the available information is linked to a few surveillance, outbreaks, or case studies [241, 253, 258–262]. Several studies have shown that ST15 is the dominant 3GC-resistant lineage in the United Kingdom, Ireland [263], Cuba [264], Africa [265], Viet Nam [265, 266], and the second CRKp, after ST11, in China [267–269]. Yet, a recent shift in the dominant ST from ST11 to ST15 has been reported at a Medical Centre in Northeast China [270]. Moreover, ST15 carrying CTX-M-15 was found to be the predominant ESBL lineage (85.5%) causing animal infections in European veterinary clinics [271]. The

population structure of ST15 may be highly diverse, composed of different clades and different capsular serotypes (K24, K60, KL48) [263, 272, 273]. Nevertheless, the ST15 global genomic population structure has not been assessed.

CG307, composed by ST307, is one of the most common 3GC- and carbapenem-resistant clones in hospital settings [151]. It has been reported worldwide but always linked to surveillance studies that were not focused on this lineage [156, 274–278]. This lineage emerged as a major cause of ESBLs outbreaks, associated with the acquisition of the CTX-M-15 enzyme [279] and later of carbapenemases, mostly KPC [280], and OXA-48 [156, 221], OXA-181 [226, 281], although others, such as NDM-1 [282] or VIM-1 [283], can also be found. The abundance of this lineage was largely unnoticed until it was distributed globally [279, 284]. Indeed, recent reports show that the prevalence of ST307 is rising in many regions, and some authors refer to it as an emerging lineage [274,284–287]. Moreover, ST307 has begun to replace ST258 and ST512 as the main KPC producer in the Americas [284], Italy [285, 288], and Argentina [286], and it has emerged as the dominant CRKp lineage in South Africa [226,289]. CRKp ST307 has also been reported in animals, veterinary clinics, and waste waters [245, 290, 291]. However, the epidemiology of ST307 and information about its real burden and global prevalence is lacking. A global genomic analysis of this ST showed that unlike other widespread lineages, ST307 represents a conserved lineage [279]. All its isolates share the same K and O loci (KL102 and O2v2), a π-fimbrial cluster, and a type IV secretion system. The ST307 is almost always associated with the *bla*$_{\text{CTX-M-15}}$ ESBL gene, harboured in an MDR IncFIIK/IncFIB plasmid where several additional AMR determinants, such as *sul2*, *dfrA14,* and *strAB* and *aac(3)-IIa* genes [279]. The levels of sequence conservation and persistence of the *bla*$_{\text{CTX-M-15}}$ plasmid in this lineage are remarkable. It has been suggested that the plasmid is crucial for the evolution of the lineage as, in addition to AMR, it also carries other features such as virulence and metabolic genes, which may lead to increased fitness, persistence, and adaptation as well as limited fitness costs to ST307 hosts [279, 280].

CG147 is composed of ST147, ST392 and ST273. Among them, **ST147** is the most represented and geographically widespread lineage [292]. ST147 emerged around the early 1990s, and as in the other lineages, its dissemination was mainly driven by

the acquisition of the *bla*$_{\text{CTX-M-15}}$ ESBL gene [241, 293]. ST147 has been associated with the spread of different carbapenemase genes such as *bla*$_{\text{OXA-48-like}}$, *bla*$_{\text{KPC-2}}$, *bla*$_{\text{NDM}}$ [294], and *bla*$_{\text{VIM}}$. In fact, surveillance studies have shown that ST147 is the most common ST among VIM-producing *K. pneumoniae* [295]. The global phylogeny of the ST147 lineage has revealed two main clades that differ in the capsular and O antigens (KL64-O2 and KL10-O3a). These clades did not differ in AMR content and were both linked to *bla*$_{\text{CTX-M-15}}$ [292] but they show different geographic distributions: KL10-O3a has been sampled mainly in Asian regions and KL64-O2 genomes have been collected primarily in Europe, but it can also be found in Asia and North America [296]. The other lineages of this CG are not found frequently yet: ST392 has been reported globally in contrast to ST273, which has been predominantly reported in Asia [292].

CG101 includes ST101, ST1685, ST2016, ST2017, and ST2502, but only ST101 is prevalent. This lineage is one of the most prevalent carbapenemase-producing *K. pneumoniae* in Europe [151, 297, 298]. Like the other AMR lineages, ST101 is usually associated with *bla*$_{\text{CTX-M-15}}$ [299]. Regarding carbapenemases, they are mainly associated with *bla*$_{\text{OXA–48}}$ [151], but also KPC-2 and KPC-3 variants [288, 300, 301] and NDM-1 [168] have been reported. ST101 is also associated with the virulence determinant yersiniabactin [88]. The ST101 lineage is emerging and reaching high prevalence in several European countries [263, 302, 303]. In Italy, ST101, along with ST307, is increasing as a KPC producer and it has been suggested that it has started to replace CG258 lineages [288]. Moreover, ST101 is the predominant lineage in Turkey [304], Algeria [305], and Tunisia [306], and many cases have been reported worldwide [301, 307]. A study that assessed the genomic population of the lineage in hospital but also environmental samples found that the lineage is quite diverse, with 5 different KL and O types [308]. However, two serotypes, KL17-O1v1 and KL106-O1v2, were by far the most predominant ones. KL17-O1v1 was the most frequently found among hospital-associated isolates and often found in community-associated strains whereas KL106-O1v2 was predominant in animal-associated and community-associated strains [308].

1.6.2 Hypervirulent lineages

The hypervirulent *K. pneumoniae* population is less diverse than the one causing MDR infections and most of the corresponding infections are due to a small number of lineages and serotypes. Most of these strains belong to serotypes K1, K2 and other less prevalent ones such as K2-K5, which are associated with a variety of lineages [217] including CG23, CG65, ST66, ST86, and CG38 [63, 309, 310]**.**

CG23, which includes ST23, ST26, ST57, and ST163, is the most frequently observed hypervirulent clone [60, 311]. This lineage is closely associated with K1 but is divided into several sub-lineages associated with distinct ICE*Kp* acquisitions. The lineage associated with the ICE*Kp10* is the most prevalent and has become globally distributed, present at least in Asia, Australia, North America, Europe and Africa [312–314]. Similarly, to MDR lineages, the other CGs have not been widely studied so far.

1.6.3 Convergent lineages

MDR *K. pneumoniae* and hvKp pathotypes were considered to be non-overlapping populations since they belong to distinct and unrelated branches in the species phylogeny and exhibit a distinct load of acquired AMR genes and key virulence factors [63]. Nevertheless, MDR-hvKp strains showing both multidrug-resistant and hypervirulent features have been reported [215].

Since the first report in 2015 of a fatal outbreak (100% of mortality, 11 patients) in an intensive care unit of a Chinese hospital due to a hypervirulent KPC-producing ST11 strain [64], MDR-hvKp has gained much interest. The prevalence of MDR and virulence convergence is increasing and poses a significant public health concern as hybrid strains produce severe infections with limited therapeutic options, increasing mortality [64].

MDR-hvKp strains have emerged in unrelated lineages; hence, they present highly diverse genetic backgrounds and antimicrobial resistance profiles [219], covering at least 38 diverse sequence types and 11 serotypes [32, 227, 315]. However, most convergent genomes belong to a few well-known MDR and hypervirulent high-risk lineages, especially ST11 and ST23 [32].

Most convergent strains have been reported in Asia, particularly in China, where both pathotypes are reported to be common and where lineages ST11 and ST23 account for the majority of MDR and hvKp strains, respectively [215, 216]. MDR-hvKp has also been reported in Europe, Africa, and North and South America, but the majority represent sporadic isolations [219, 315]. Indeed, in 2021, a risk assessment by the ECDC raised the issue of emerging ST23 hypervirulent *K. pneumoniae* carrying carbapenemase genes, specially $bla_{\text{OXA-48}}$, in Europe [314].

1.7 Genomics and epidemiology

Whole genome sequencing (WGS) has played a major role in untangling *K. pneumoniae* taxonomy, ecology and evolution, and has become an important tool in clinical microbiology tasks [28, 31, 203].

With the decrease of WGS costs and faster turnarounds, WGS has entered in the routine tasks of Clinical Microbiology laboratories, complementing, or replacing traditional techniques [316, 317]. Common applications of WGS in Clinical Microbiology laboratories include isolate characterization, AMR phenotype prediction, and tracing the spread of infectious agents between patients to the infection sources. Finally, all this information can be combined to perform genome-based pathogen surveillance or **genomic surveillance**, in short.

1.7.1 Isolate characterization

Pathogen typing is essential to understand which species and/or specific problematic strains are causing infections. Strain typing can complement the identification of clinically relevant markers of the pathogen such as antimicrobial resistance and/or pathogenic virulence factors.

Before WGS, the characterization of *K. pneumoniae* isolates required multiple assays to type the strains and identify their AMR gene profile and virulence determinants. However, WGS allows the characterization of all key gene markers for several isolates in a single assay. Hence, WGS has become the most cost-efficient approach for characterising pathogens with as many markers as *Klebsiella* [31, 318, 319].

1.7.2 Genomics for outbreak detection and identification of transmission pathways

One of the first applications of WGS in epidemiology was to identify sources and transmission networks of hospital outbreaks [320–322] and, nowadays, it has become the new gold standard technology for a wide number of pathogens [323, 324].

WGS allows ultrafine genotyping resolution and higher discriminatory power than traditional methods allowing to trace circulating pathogenic strains. However, dealing with bacterial genomic data to reconstruct transmission routes can be complex and poses a series of challenges. One of the biggest limitations to implementing WGS in the Clinical Microbiology setting is that interpreting the large amounts of data produced by WGS requires bioinformaticians, which are usually in research centres rather than hospitals. As a consequence, WGS-informed outbreak tracking is usually performed only retrospectively [316, 325].

The simplest way to determine whether isolates are likely or unlikely parts of the same transmission chain using WGS is to count the number of single-nucleotide polymorphisms (SNPs) that differ between isolates and establish SNP-thresholds that limit likely transmissions. The SNPs threshold is the most common methodology used to test transmission in many pathogens including *K. pneumoniae* [325–328]. Different cut-offs have been proposed to define *K. pneumoniae* transmission events. Most of them vary between 20 to 25 SNPs per approximately 5 Mbp of genome compared [67, 151, 329, 330].

However, there is a concern whether the establishment of a SNP threshold is appropriate to evaluate transmission links. The final number of measured pairwise SNPs distances between isolates can be affected by many different variables. For instance, SNP distances can be highly impacted depending on the dataset, the sequencing technology and, on top of that, the tools employed to estimate the relationships between the isolates [326, 331]. SNP distances are generally estimated using a reference-based approach, where the isolates of interest are compared to a reference genome sequence and SNPs are estimated relative to it. The choice of reference sequence and its relatedness to the isolates being analysed have a large

impact on pairwise SNP distance calculations [332, 333] and, therefore, on which isolate pairs fall below the established threshold [331]. SNP variation can be also given by the biology of the infections. For instance, prolonged asymptomatic colonisation of patients will yield increased variations that may escape low variant cutoffs [263]. Furthermore, limiting the detection of transmission events to the number of SNPs accumulated in the core genome will fail to detect possible plasmid transmission events.

Potential SNP cut-offs tailored for each study may only apply to the specific data set (number of samples, diversity, time) and the sequencing and bioinformatic workflow used to obtain it. Therefore, even when SNPs are very low, SNPs-threshold approaches must be correlated with epidemiological transmission evidence.

There are alternatives to test transmissions avoiding the need to set arbitrary cut-offs. However, they involve complex and sophisticated phylogenetic approaches [334–337] difficult to implement in routine public health workflows [324]. Therefore, implementing WGS for *K. pneumoniae* outbreak detection in routine public health investigations is still a challenge [316]. To implement it and guarantee reproducible results, entire bioinformatics workflows, including tools and analysis parameters, must be standardised and validated [338]

1.7.3 WGS to predict antimicrobial susceptibility

A more recent application of WGS is the *in silico* prediction of antimicrobial susceptibility testing (AST). The detection of antimicrobial resistance is essential to guide treatment decisions and ensure the effectiveness of the drugs of choice. For severe infections, such as bacteraemia, mortality increases every hour that appropriate therapy is delayed [339].

Nowadays, **culture-based AST** is the conventional method employed to test AMR in clinical laboratories, as they predict treatment success with high accuracy [340]. Culture-based AST strategies require isolation, growth, and identification of the pathogen and are only automated for a limited list of pathogens and antibiotic concentrations [341]. Thus, culture-based AST is often combined with other rapid diagnostic tools such as PCR of specific AMR determinants (e.g., carbapenemase

genes). These methods detect only a few resistance markers and are often expensive [342].

To fill this gap, **WGS** has been proposed as a powerful alternative to predict AST profiles (WGS-AST) [343–345]. Several tools using known AMR genetic mechanisms as AST predictive factors have been developed: ResFinder [10], ABRicate (https://github.com/tseemann/abricate), the Comprehensive Antibiotic Resistance Database (CARD) [12] and PATRIC-RAST [14]. WGS can be applied to all types of pathogens and ascertain all the known AMR genetic mechanisms in a single assay [346,347]. However, most predictive models rely on previous knowledge of AMR markers and their phenotypic effects and, therefore, they commonly fail to predict resistance caused by multifactorial mechanisms that are not yet fully understood.

Several studies predicting resistance phenotypes using AMR databases and complete genomes of *K. pneumoniae* have been published. Overall, they reach an accuracy of over 90% for many of the widely used antibiotics [348, 349]. However, accuracy varies greatly among studies and antibiotics, and most studies are completed with only a few antibiotics and isolates.

To improve these results, recent studies have used genome-wide association studies (GWAS) and machine learning to predict AMR phenotypes and even MICs , without relying on a database of pre-existing AMR genes or mutations [350]. Machine learning has been applied to predict *K. pneumoniae* AMR profiles reaching an overall accuracy of 92 % in MIC prediction for 20 antibiotics [351].

Despite the high accuracy predictive values, WGS-AST also has limitations such as the time needed to obtain the results, the need of specialised staff, lack of harmonisation, and standardisation of the methodology or the costs. The major barriers to the application of bioinformatics software for AMR detection in clinical laboratories are the accuracy and discordances between the culture and the WGS based AST, which still limit the implementation of WGS-AST as a standard routine method in Clinical Microbiology laboratories [352].

In addition to AST prediction, genomics and GWAs can be used to pursue the identification of novel genetic mechanisms underlying particular AMR or virulence phenotypic traits of interest or concern.

1.7.4 Surveillance

One of the key applications of WGS is in real-time tracking of infectious disease or surveillance. Surveillance strategies are needed to monitor the local emergence and dissemination of problematic clones and AMR and, therefore, to control their spread in hospital settings, linking the infections to potential reservoirs. In addition, surveillance can help to optimise antimicrobial stewardship and inform treatment decisions [353].

Strain typing and characterization are essential for surveillance in order to understand which strains and associated AMR and/or virulence determinants are entering and circulating. The advantages of the enormous discriminatory power of WGS over traditional methods provide one of the strongest arguments for its implementation in routine surveillance in clinical and public health microbiology laboratories [354–356].

Until now, WGS implementation has been constrained in part by costs, time, lack of necessary equipment, infrastructure, standardisation of methods, and expertise [319, 355]. Nevertheless, the coronavirus disease 2019 (COVID-19) pandemic forced the equipment of clinical laboratories with massive sequencing platforms and specialised staff [357]. Moreover, the COVID-19 pandemic has demonstrated the utility of WGS for surveillance to monitor and control the spread of pathogens locally, regionally, and globally.

Although many hospitals already have the equipment, routine surveillance based on WGS is not still a reality for other pathogens, and WGS is more commonly employed as a second-line tool to investigate specific outbreaks [358]. The biggest limitation for the implementation of WGS for these pathogens is the analytical part, including the need of specialised staff and the standardisation of methods [338].

To enable the automatic and rapid conversion of WGS data into clinically relevant information, several user-friendly web tools have been developed, including Kleborate [32], PathogenWatch [359], and abritAMR [360]. Kleborate is a bioinformatic tool designed specifically for *Klebsiella* typing, capable of detecting species, performing MLST, and genotyping key virulence and AMR determinants

[32]. PathogenWatch is a web application that can perform clustering and phylogenetic analysis, providing context for genome sequences by comparing them to publicly available genomes. By incorporating Kleborate in its workflow, PathogenWatch also offers detailed characterization. abritAMR is an ISO-certified genomic-based pipeline for identifying bacterial AMR genes, tailored to produce customised reports with clinically relevant information on AMR determinants and the antibiotic classes they may impact [361].

However, just as with validating transmission events, it is crucial to validate and standardise bioinformatics workflows for detecting and identifying clinically relevant characteristics such as AMR genes and virulence factors. For instance, the lack of standard pipelines for genomic detection of AMR mechanisms translates in different results between laboratories [362]. Additionally, interpreting and reporting the data to clinicians remains an unresolved issue. Therefore, in order to implement effective surveillance, clinical microbiologists, scientists, and bioinformaticians must collaborate to create standardised pipelines from laboratory procedures to bioinformatics analyses, capable of producing accurate and reproducible results across all laboratories [361].

1.8 Implementation of WGS in Spanish hospitals and the NLSAR project

Currently, the use of WGS is not implemented in the clinical microbiology routine of Spanish hospitals. Nevertheless, the **Spanish National Plan against Antibiotic Resistance** (PRAN) includes provisions to support an important first step towards a wider adoption and integration of WGS in routine clinical practice.

The PRAN aims to reduce the risk of selection and spread of AMR and, consequently, its clinical and public health impact. As part of this plan, the **Spanish Network of Laboratories for the Surveillance of Resistant Microorganisms** (RedLabRA) was created. This is a national network of coordinated microbiology laboratories aimed at carrying out molecular studies of infectious diseases [363]. One of the actions of RedLabRA is to promote the integration of WGS in the routine clinical practice between 2022 to 2024.

The **CARB-ES-19** project was the first nationwide multicentre study conducted under the RedLabRA framework. In this study, they sequenced 377 *K. pneumoniae* genomes from different Spanish hospitals between February to May 2019 [156]. In addition to the studies conducted under the RedLabRA framework, other efforts to apply WGS to the surveillance and characterization of antimicrobial resistance in Spain have been done. While some of these studies are nationwide [152], others have been done in specific regions such as Andalusia [255] or Catalonia [364]. However, most of them have focused on specific lineages such the ST307 [287] or in resistance mechanisms, such as Ampc [365], but specially carbapenemases including OXA-48 [366], KPC [300], NDM [168], and IMP producing strains [367]. It should be noted that the number of isolates sequenced in these studies is relatively low, which leaves a significant knowledge gap in understanding the epidemiology of antimicrobial resistance in Spain.

1.8.1 The Networked Laboratory for Surveillance of Antimicrobial Resistance

The Comunitat Valenciana is an autonomous community located along the Mediterranean coast on the east side of Spain. It is the fourth most populous Spanish autonomous community with more than five million inhabitants where more than 1 million are older than 65 years (consulted in www.ine.es, data from 2021).

As mentioned above, there have been few studies conducted to investigate *Klebsiella* antimicrobial resistance at the national level in Spain, and none has been specifically focused on the Valencian region.

Understanding the AMR patterns in a specific region is critical for developing effective treatment strategies and preventing outbreaks. This is especially important in regions with high population density or areas. Hence, in 2018, the Molecular Epidemiology group of FISABIO and the University of Valencia headed an initiative to establish a framework to perform AMR surveillance in the region, known as the **Networked Laboratory for Surveillance of Antimicrobial Resistance** (NLSAR). The NLSAR is a network of regional microbiology laboratories coordinated by a central node at the Molecular Epidemiology group of FISABIO (Figure 1.12).

The primary objective of the NLSAR was to serve as an initial step towards integrating WGS into the surveillance of AMR pathogens in the Comunitat Valenciana, but also to generate information that can help in understanding the transmission routes, origins, and evolution of the pathogens in the regional clinical settings.

With FEDER (Fondos Europeos de Desarrollo Regional) funds and national projects of the group, the NLSAR provided almost all participating Clinical Microbiology laboratories in CV hospitals with WGS platforms and other necessary resources to perform AMR surveillance. It is within this framework that this dissertation has been developed.

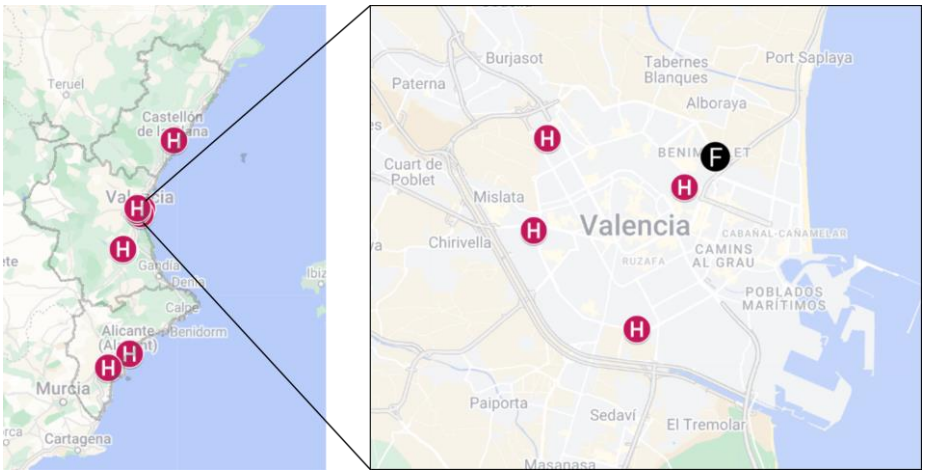


Figure 1.12. Distribution of the hospitals included in the NLSAR in the CV. Participating hospitals are coloured in red whereas the central node in FISABIO in coloured in Black. Figures were made with Google Maps.

CHAPTER 2
OBJECTIVES

2.1 Statement of objectives

The **main objective** of this thesis was to investigate the genomic epidemiology of *Klebsiella pneumoniae* population in the Comunitat Valenciana (CV) through the application of WGS sequencing and genomic analysis, with particular emphasis on the main lineages and determinants of AMR.

The **specific objectives** developed to fulfil the main goal were:

1. **To establish the bases for the analysis of the genomic epidemiology of *Klebsiella pneumoniae* in the Comunitat Valenciana.**

    a. To set up a genomic surveillance project in collaboration with the hospitals of the region to collect representative isolates of the AMR *K. pneumoniae* population.

    b. To create a global genome collection with database genomes to create a context and elucidate possible entering routes.

    c. To develop a pipeline for curating the genome sequences obtained in the project and those downloaded from external databases.

2. **To study the genomic epidemiology of AMR *Klebsiella pneumoniae* in the Comunitat Valenciana.**

    a. To study the prevalence and burden of antimicrobial resistant *K. pneumoniae* infections in the CV.

    b. To describe the distribution of lineages and non-susceptibility genetic determinants of *K. pneumoniae* isolates in the different hospitals of the CV.

    c. To elucidate possible transmission routes at three levels: locally, regionally, and globally.

3. **To understand how the initial emergence and posterior colonisation of carbapenem-resistant determinants occurs in hospitals.**

    a. To analyse the initial population of carbapenem-resistant isolates in a hospital setting elucidating the AMR determinants and lineages.

    b. To study the evolution of the carbapenem-resistant population after the initial emergence, to the establishment or disappearance.

4. **To use WGS to detect the emergence of new threats in hospital settings.**

    a. To describe the emergence and dissemination of a novel carbapenemase gene, the *bla*$_{\text{NDM-23}}$.

    b. To determine the phenotypic effect of the *bla*$_{\text{NDM-23}}$, its genetic environment, and clonal background.

    c. To study the possible origin of the gene, the vector and the lineage it carried.

CHAPTER 3
MATERIAL & METHODS

3.1 The NLSAR and the SKPCV collection

To gain a better understanding of the population of the *K. pneumoniae* ESBL- and carbapenemase-producers in the CV, we launched the **Surveillance of *Klebsiella pneumoniae* in the Comunitat Valenciana** (SKPCV) project.

The SKPCV project was conducted from 2017 to 2019 at a regional scale. It was carried out under the framework of the recently developed NLSAR (§1.9) enrolling 8 of the network hospitals: Consorcio Hospital General Universitario de Valencia (HGUV), Hospital Clínico Universitario de Valencia (HCUV), Hospital Universitario y Politécnico La Fe de Valencia (HUPLF), Hospital Arnau de Vilanova de Valencia (HAV), Hospital Universitario de la Ribera de Alzira (HULR), Hospital General Universitario de Castellón (HGUC), Hospital General Universitario de Alicante (HGUA), and Hospital General Universitario de Elx (HGUE). These are all the tertiary-level hospitals of the CV plus two secondary-level ones (HAV and HULR). The Clinical Microbiology laboratories of these hospitals serve as reference laboratories for primary and other secondary-level hospitals of the CV. In total, they attend 45.5% of the CV population.

3.1.1 The SKPCV study design

The only criterion established for the project was to include, for each hospital and trimester, the 30 first ESBL- and/or carbapenemase-producing *K. pneumoniae* isolates, along with the 10 first 3GC-susceptible clinical isolates. Only one isolate per patient. Both infections and colonisations were considered. Strains collected from sterile sites, prosthetic material, or bronchial aspiration were reported as infections, whereas surveillance cultures were reported as colonisations. Relevant clinical data such as collection date, hospital, ward, patient sex and age, and antimicrobial susceptibility tests were collected.

3.1.2 Additional retrospective strains

For some analyses, we included *K. pneumoniae* isolates collected in the NLSAR hospitals but not in the SKPCV project. These included ESBL producing strains from the HGUC collected during 2014, a non-susceptible to carbapenems collection from

the HGUV from 2015 to 2017, and historical samples from Hospital Dr. Moliner (not in the SKPCV project). These strains were used to provide historical context for the SKPCV strains in chapters 4 and 5, whereas the whole HGUV dataset is analysed in more detail in chapter 6. Unless otherwise indicated, we will denote this dataset as NLSAR isolates/genomes/data.

3.2 Resistance rates of *K. pneumoniae* in the CV

We obtained an overview of the evolution of AMR in *K. pneumoniae* in the CV in the last decade using the information stored in the Microbiological Surveillance Network of the Comunitat Valenciana (**RedMIVA**, from "Red de Vigilancia Microbiológica de la Comunitat Valenciana"). The RedMIVA is an information system of the Conselleria de Sanitat Universal i Salut Pública de la Generalitat Valenciana. Among many other clinical analyses, RedMIVA gathers the results of AST made for clinically relevant microorganisms in the hospitals of the CV [368].

We interrogated the RedMIVA database to collect all the ASTs performed for *K. pneumoniae* isolates from 2010 to 2022. Clinical information on each isolate included: sample type, MICs for several antibiotics, and hospitalisation status. Demographic information about the patient included sex, age, hospital, and ward.

Different episodes of the same patient were categorised according to the criteria established in the protocols for the surveillance of carbapenemase-producing *Enterobacteriaceae* made by the Service of Epidemiological Surveillance and Control of the General Directorate of Public Health of the Comunitat Valenciana. Different episodes were considered when there was a gap of more than 90 days between isolation dates or a 30-day gap with a reported susceptible test in-between. Then, episodes were tagged with the test with highest AMR pattern in the following order of priority: i) combined resistance to carbapenems, 3GC, fluoroquinolones and aminoglycosides; ii) resistance to carbapenems; iii) combined resistance to 3GC, fluoroquinolones and aminoglycosides; and iv) resistance to 3GC.

The interpretation of MIC values from RedMIVA records into qualitative susceptibility categories – S (susceptible), I (susceptible, increased exposure) and R

(resistant) – was performed using the **AMR v2** library for R (https://msberends.github.io/AMR) with the recommended breakpoints of the **EUCAST** 2022 guidelines (https://www.eucast.org/clinical_breakpoints). Resistance to 3GC was considered when isolates were resistant to ceftazidime (MIC > 4 mg/L) or to cefotaxime (MIC > 2 mg/L). Carbapenem resistance was considered when 3GC-resistant isolates were also resistant to ertapenem (MIC > 4 mg/L) or meropenem (MIC > 8 mg/L). Aminoglycoside resistance was determined when the isolates were resistant to gentamicin or tobramycin (MIC > 2 mg/L for both). Fluoroquinolone resistance was reported when the isolates were resistant to ciprofloxacin (MIC > 0.5 mg/L) or levofloxacin (MIC > 1 mg/L). These antibiotics were selected as they were the most frequently tested antibiotics in the CV hospitals and the markers used in the annual AMR report of the ECDC [12]. Combined resistance to several classes was also estimated. For some episodes, resistance to some antibiotics could not be determined, tagging them as "unknown".

3.2.1 Descriptive epidemiology

To study whether 3GC-, carbapenem- or combined-resistance were associated with hospitalisation, we estimated the odds ratio (OR) of each combination with the **fisher.test** function implemented in the R **stats** library v4.1.2. Statistical significance was assessed with the same function, and we considered a p-value of <0.05. The exploitation and statistical analysis of the RedMIVA database detailed previously, was performed by Jordi Sevilla as part of his external practicum, under my direct supervision and co-mentoring with Dr. González-Candelas.

To compare ages between men and women we used Wilcoxon rank-sum test (two-sided) with the R stats function **wilcox.test**. All statistical analyses were performed using R (v4.1.2), and figures were plotted in R using **ggplot2** v3.4.0 [369]

3.3 Isolation, species identification, and AST

All the strains collected by the NLSAR laboratories, regardless of their inclusion in the SKPCV project, were subjected to the same procedures for isolation, AST, and sequencing.

Isolates collection, species identification and AST were performed by the Clinical Microbiology laboratories of each hospital as part of their routine tasks. Clinical isolates were directly smeared on blood agar plates (Beckton Dickinson) and incubated at 37°C for 18–24 h. Species identification was performed by matrix-assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-TOF MS) using the Vitek MS® system (bioMerieux Marcy l'Etoile, France). Antimicrobial susceptibility testing was performed by broth microdilution using the automated system Vitek-2 (BioMérieux, Marcy l'Etoile, France) or MicroScan WalkAway (Beckton Dickinson). MICs were reinterpreted using the 2022 clinical breakpoints recommended by the European Committee on Antimicrobial Susceptibility Testing (EUCAST) and the **AMR v2** library for **R** (https://msberends.github.io/AMR). The antibiotics tested varied depending on the hospital where each sample was collected. The most frequently tested antimicrobials were ampicillin, amoxicillin-clavulanic acid, piperacillin-tazobactam, cefuroxime, cefoxitin, ceftazidime, cefotaxime, cefepime, ertapenem, imipenem, meropenem, ciprofloxacin, levofloxacin, amikacin, gentamicin, tobramycin, fosfomycin, trimethoprim-sulfamethoxazole, and tigecycline. Isolates were classified as MDR based on acquired non-susceptibility to at least one agent in three or more antimicrobial categories, as previously defined [370]. Carbapenemase production was confirmed upon admission to the hospital using Etest and disk diffusion (Oxoid, Hampshire, UK).

3.4 High throughput sequencing

3.4.1 DNA extraction

Single colonies from overnight-grown plates were subjected to DNA extraction mainly at the Clinical Microbiology laboratories. Although each laboratory had a different DNA extraction platform, all their methods were automated systems recommended for Illumina sequencing. At FISABIO, DNA was extracted using an NucliSENS® easyMag® (bioMérieux). DNA quantification was performed using Qubit 3.0 and the High Sensitivity double-stranded DNA kit (Invitrogen™).

3.4.2 Library preparation and sequencing

All the samples were sequenced using Illumina short-read technology. Sequencing was performed in an Illumina NextSeq 500 platform and the barcoded Illumina DNA libraries were prepared using the Nextera XT library preparation kit, generating 150 bp paired-end reads. In each sequencing run, 380 *Klebsiella* isolates were pooled together, expecting an average sequence depth of 50x (~900,000 reads) for each sample. Fragment sizes were evaluated using the Fragment Analyzer (Agilent) with the HS NGS Fragment Kit (Agilent).
In addition, some isolates were also sequenced by long-read sequencing technologies by the Sequencing and Bioinformatics Service in FISABIO. These isolates were representative of different lineages (STs) and/or carbapenem-resistant determinants. Two isolates were sequenced using MinION (Oxford Nanopore Technology, UK), whereas 48 isolates were sequenced using a PacBio Sequel II platform.

To perform long-read sequencing, DNA integrity and quantity were checked using the NanoDrop® ND-1000 UV-Vis (Fisher Scientific) and the Invitrogen™ Qubit™ 3 Fluorometer (Fisher Scientific) spectrophotometers. For the isolates with low DNA quality, bacteria were regrown in Luria-Bertani (LB) culture plates, and a single colony was cultured in LB liquid media overnight. DNA was extracted from 500 µL of liquid LB using the NORGEN Genomic DNA Isolation Kit (Norgen Biotek) following the manufacturer's instructions.

MinION genomic libraries were prepared using the Rapid Barcoding Kit (SQK-RBK004, Oxford Nanopore Technology, UK). Libraries were loaded into an R9.4.1 flow cell (FLO-MIN106D). Sequencing was computed over 36 hours and base-calling was performed using **Guppy** v4.5.4 (Oxford Nanopore Technology, UK).
The sequencing reads generated that have been published were deposited at the European Nucleotide Database (ENA) under project number PRJEB37504.

3.5 Database genomes

To provide a comparative background, we downloaded *K. pneumoniae* genome sequences from three different sources: the GenBank sequence database, the Reference Sequence (RefSeq) collection, and the Sequence Read Archive (SRA).

All *K. pneumoniae* genome assemblies publicly available in the RefSeq collection (accessed on 6/13/2020) were included. The **ncbi-genome-download** (https://github.com/kblin/ncbi-genome-download) tools were used to download the assemblies and the **E-utilities** commands to extract the metadata. To ensure the presence of isolates collected in Spain, all the read sets from the SRA and the assemblies from the GenBank database labelled with "*Klebsiella pneumoniae"* AND "Spain" (accessed on 4 April 2022) were also retrieved. When downloading SRA experiments, only Illumina paired-end and complete genome entries were selected. SRA runs were downloaded using the information of the NCBI Entrez and the NCBI run selector and the **prefetch** tools from the **SRA Toolkit 2.10.7**. SRA files were converted to fastq format using **fasterq-dump**. SRA runs of biosamples included in the downloaded assembled RefSeq genomes were removed.

3.6 Reconstructing the genomes

3.6.1 Read quality, cleaning and trimming

For short-reads, quality was assessed using **FastQC v0.11.9** (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and **MultiQC v1.11** (https://multiqc.info/). **Prinseq-lite v0.20.4** [371] was used to filter and trim the raw reads. Reads with a mean quality below 25, a read length shorter than 60 bp, longer than 500 bp, or optical duplicates were removed. Also, the last three positions in the 3' end and positions with a quality lower than 20 were trimmed. The quality assessment of long Nanopore reads was performed using **Nanostat** (https://github.com/wdecoster/nanostat). Long reads with a length below 3,000 bp were removed.

3.6.2 *De novo* genome assembly and gene annotation

Samples sequenced by short reads and Nanopore were *de novo* assembled using **Unicycler v0.4.8** [372]. For Nanopore reads, we performed hybrid assemblies with their corresponding short-reads. Samples sequenced by PacBIO were assembled using **HGAP** or **Unicycler**, depending on the sequencing yield for each genome. **HGAP** is the assembler implemented in the PacBio SMRT Analysis software suite, SMRT® Link. As **HGAP** tends to remove small plasmids, each **HGAP** assembly was compared with its short-read **Unicycler** assembly to find possible small plasmids. If small, closed plasmids were found in the **Unicycler** assembly but not in the **HGAP** assembly, they were added to the **HGAP** assembly as recommended [373]. When the number of bases in PacBio reads was below 60K, and thus no complete chromosome or plasmid was obtained, a hybrid assembly using the short and long reads was made using **Unicycler**. For both assemblers, HGAP and Unicycler, we used the default parameters. **HGAP** and **Unicycler** assemblies were polished by **pilon** using the short reads and 10 pilon rounds. This software identifies inconsistencies between an assembly and the evidence in the reads. It can be used to fill gaps and fix small indexes and possible local misassemblies. The resulting assemblies were annotated using **Prokka** v1.14.6 [374]. To uniformise the gene names of all the assemblies, the option --proteins was used with a concatenated GenBank file of the *K. pneumoniae* traditional reference genome (Accession number: GCF_000240185.1) and several sequences of curated ICE*Kp* (Accessions numbers: KY454627.1 - KY454638.1) [88]. To annotate ribosomal proteins, we used the **Barrnap** v0.9 (https://github.com/tseemann/barrnap) software implemented **Prokka** v1.14.6.

3.6.3 Sequence type, antimicrobial resistance, virulence determinants and plasmids

Kleborate [32] is a tool to screen KpSC genomes assemblies for most of the main relevant markers for the species. Kleborate v1.0.0 and v2.0.1 were used to determine the ST and identify virulence and antimicrobial resistance genes. **Kaptive** [375], integrated into Kleborate, was used to type the KL type and the LPS O antigen. Additionally, **AMRFinderPlus** [376], **ResFinder** [377], **staramr**

(https://github.com/phac-nml/staramr) and **BLAST+** [378] were used to identify and verify the antimicrobial resistance genes detected. To characterise the mechanisms contributing to cefiderocol resistance, we compared isolates to the *K. pneumoniae* ATCC 13883 reference genome and identified gene mutations of previously proposed cefiderocol resistance targets [379].

For plasmid replicon typing we interrogated each assembly using **PlasmidFinder** [380] whereas for phages we used **PHASTER** [381].

3.7 Isolate filtering

Assessing the quality of WGS data is critical to ensure the reliability of downstream analyses. To evaluate the NLSAR and the downloaded genomes, we developed and applied a five-step filtering approach. These filters consider the quality of the WGS experiments, the taxonomy and possible contamination, the assembly quality and the genome content.

3.7.1 Basic statistics

The short-read sequencing experiments with less than 200,000 reads in total (~10x of the expected average depth) were removed from further analyses. Depth calculations were made using Illumina's recommended formula (www.illumina.com/documents/products/technotes/technote_coverage_calculation.pdf) and 5.4 Mb as genome size.

3.7.2 Taxonomic assignment and interspecies contamination

Species misidentifications and cross-species contamination in short-read data can be identified by taxonomic classification of sequence reads. We used **Kraken2** [382] and its **Standard** database (downloaded on 23/6/2020) to assign taxonomic labels to each short read. In this step, short-read experiments were filtered if: (i) had less than 40% of identified reads, (ii) less than 75% of the identified reads at the genus level were assigned as *Klebsiella,* or (iii) had more than 3% of non-enterobacterial species or more than 15% of any enterobacterial species, except *K. pneumoniae*.

Second, the general taxonomic assignment of the whole genome assembly made by **Kleborate** v1.0.0 to remove isolates that were not identified as *K. pneumoniae sensu stricto*. Kleborate uses mash distances [32] and a custom database for species identification.

3.7.3 Genome assembly and annotation

To assess the quality of the reconstructed genomes, we used **QUAST v5.0.2** [383], which generates assembly metrics. Assembly metrics, especially the GC content and assembly length, vary among species and even strains. Assembly metrics can be used to evaluate *de novo* assembly quality, but also to infer contamination, as this may result in highly fragmented assemblies, unexpected GG content, or deviant genome sizes [384]. Genomes were removed when they fulfilled any of the following conditions: (i) samples with a CG percentage outside the range [55-58.5], (ii) samples outside the 4 - 6.5 Mb range of genome length, (iii) samples with unique predicted genes above 6,000, and (iv) samples with more than 500 contigs.

3.7.4 Intraspecies contamination

Detecting intraspecific contamination is more challenging than interspecific contamination because the Kraken cannot distinguish between strains of the same species. Furthermore, de novo assembly metrics are generally not influenced by mixtures of these strains, making it even more difficult to detect intraspecific contamination. To avoid including in the analysis samples with mixtures of *Klebsiella* strains, the number of genes affected by heterozygous reads was counted. To do so, we mapped the short reads of each isolate using **BWA** v0.7.17 [385] with the mem option against the reference genome, NTUH K2204 (AP006725). This genome has several plasmids and an ICE*Kp1* inserted in the chromosome. As these regions can induce mapping errors unrelated to contamination, only the chromosome was used and the ICE*K*p1 positions (3,394,131 - 3,473,411) were removed. **Samtools** v1.11 and **BCFtools** v1.11 [386,387] were used to call heterozygous positions, defined as those positions where none of the nucleotides was represented in at least 85% of the reads and had a minimum depth of 3 reads. As the mapping approach tends to accumulate heterozygous positions in some regions of the *K.*

pneumoniae genome, we considered the number of affected genes instead of the total of heterozygous positions. Genes were considered when at least one heterozygous position was found. Genomes with more than 1,000 affected genes were removed from further analysis.

3.7.5 Pairwise genomic similarity

The last filtering step was performed to ensure robust results when comparing genomes to each other, either in a pan-genome or in a mapping to a reference analysis.

To determine the similarity between all the isolates of the collection, we estimated pairwise mash distances using **Mash** v2.0 and the **dist** module [388]. Mash reduces assemblies to smaller fragments called k-mers and Mash distances are estimated by comparing the k-mer sets of two assemblies. Large mash distances usually reflect different genomic content. Genomes with median mash distances to all the others above 0.014 (rounded) were removed. This threshold was established by exploring the data and testing how core genomes would result by including or not the genomes in the core genome analysis. Moreover, this value has been shown to differentiate between *K. pneumoniae* subspecies [32].

Phylogenetic trees based on mash distances were made using the neighbour-joining function of the **ape** v5.7.1 R library [389].

3.8 Diversity and rarefaction curves

Population diversity is typically evaluated as the probability that two units (alleles, species, lineages, etc.) randomly chosen from a population will be different [390]. We used the Simpson Diversity Index [391] to measure the STs diversity using the Simpson's diversity index implemented in the R package **vegan** v2.6-4 (https://CRAN.R-project.org/package=vegan). This index results in a probability value between 0 and 1 and considers the number of STs present as well as their relative abundances. As lineage richness and evenness increase, diversity increases and, therefore, so does the probability of randomly choosing different STs in the population.

3.9 Genomic analyses

The main methodologies applied to obtain the relationships between strains were mapping to a reference genome or finding orthologous genes in *de novo* assemblies. Applying one or the other methodology depends on different aspects. Mapping is the gold standard to test for transmission, epidemiological analysis, or recombination [151, 392, 393]. This approach yields better results when a high-quality reference genome is available and if the overall genetic distance of the genomes to be compared is low. However, mapping analysis has the limitation of the reference choice, as the reference chosen can affect which regions are included in the analyses [332] and, therefore, lead to different conclusions [333].

Pan-genome analysis is more adequate when the sample is highly diverse because the bias of the choice of reference is removed. Due to the diversity of *K. pneumoniae* species, core genome analysis is the preferred methodology to compare diverse or large datasets [394]. Nevertheless, it also presents limitations, the most relevant one being that bad-quality assemblies will have a high impact on the results.

Because of the diversity, dataset size, and objectives of each chapter of this thesis, both techniques were applied.

3.9.1 Core genome using *de novo* assemblies

In Chapters 4, 5 and 6, the relationships between isolates were analysed by computing core genomes from *de novo* assemblies. In each chapter, a different pan-genome analysis tool was used.

For Chapters 4 and 5 we used **PanACota** 1.3.1-dev3 [395] whereas in Chapter 6 we used **panaroo** v1.2.10 [396]. This election was based on the size of the datasets analysed in each chapter as **PanACota** runs faster with large datasets. For both tools, the nucleotide identity threshold to infer orthology was set at 80%. This value has been seen as a good point to recognize genes of the same family without increasing noise [395, 396]. Within the pan-genome, strict or relaxed core genomes can be defined. The strict core genome includes the gene families present in all the strains whereas the relaxed core includes gene families present in a specific

proportion of the isolates in the collection, usually 95%, 90%, or 50%. Generally, relaxed core genomes at 90% were used unless indicated otherwise.

After obtaining the gene families present in the core genome, both tools have the option to extract the nucleotide alignment of each gene family and concatenate them to obtain a core-genome alignment. Nevertheless, instead of using the built-in commands of both programs, the alignment and concatenation of the gene families were performed using **mafft** v7.490 [397] with the --adjust-direction option and the **AMAS.py** software [398]. This is because both **panaroo** and **PanACota** share the same problem: they introduce gene sequences in the core-genome alignment as they appear in the contigs, where genes are often found in reverse-complementary orientation, thus inserting a large number of unreal variants in the corresponding alignments.

Furthermore, we observed that the gene annotation software used during our work introduced frameshift errors due to incorrect identification of the start or end of a CDS in fragmented contigs. Frameshift errors in the middle of a gene were reflected in a large accumulation of false SNPs, having a profound impact on the results. This affects epidemiological and phylogenetic estimates, such as transmission clusters, dating analysis or identification of recombination events, because these types of studies rely on mapping analysis [151, 392]. Although it represents a serious problem, these errors have received limited attention and, therefore, none of the usual pan-genome analysis tools tackles them.

This problem was highly significant when **PanACota** was used, as the datasets were larger and, therefore, so was the number of genes affected. To overcome this problem, several popular sequence alignment trimming tools such as **Trimmal** [399] and **CLipKit** [400] were used but no improvement in the final results was detected. To solve this problem, we developed and implemented an algorithm that detects and removes affected sequences from each gene alignment producing error-free core genome sequence alignments. The algorithm is based on constructing each gene phylogeny, searching for unexpectedly long branches, and checking whether those are due to unique SNPs runs in the alignment (Figure 3.11). When a sample was to be removed from a gene family, the corresponding

nucleotides were replaced by gaps, as it has been shown that adding up to 60% of missing data in the alignments can still result in informative alignments [401].

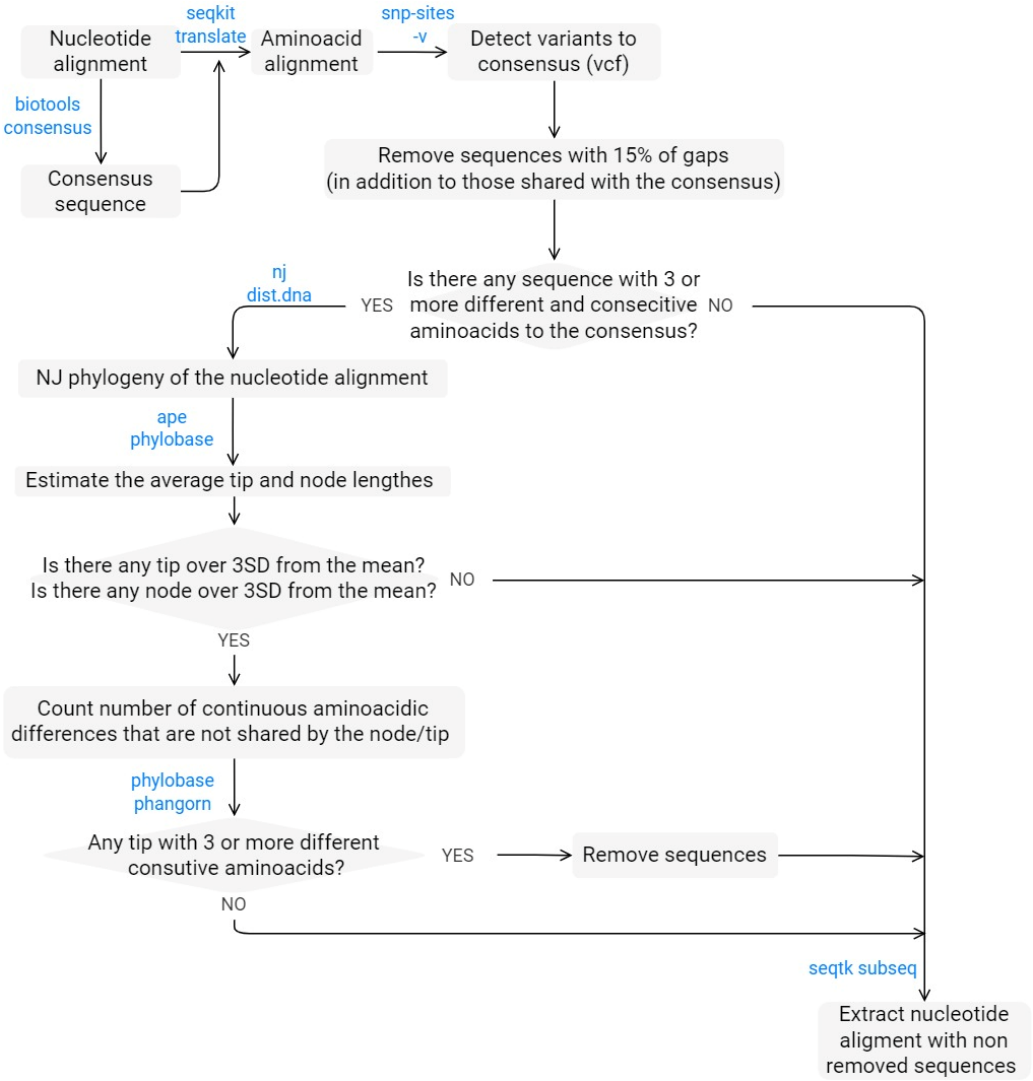


Figure 3.11. Schematic diagram of the algorithm developed to remove sequences accumulating mapping errors in each core gene family alignment.

For **panaroo** frameshift errors were not as significant as the datasets were smaller, thus spurious errors were trimmed with **ClipKit** v1.3.0 [27] with the kpic-gappy mode.

For both programs, the concatenation of the resulting alignment genes was made with **AMAS.py** concat [398] and the quality of each gene and the whole genome sequence alignment were assessed using **AMAS.py** summary [398]. SNP positions were extracted using **snp-sites** v2.5.1 [402].

3.9.2 Mapping to a reference genome

When we used reference-mapping approaches, reads were aligned to the reference genome using **Snippy** v4.6.0 [403]. Minimum quality, minimum site depth, and minimum proportion for variant evidence for calling alleles were set to 60, 5, and 0.75, respectively. Reference genome mapping coverage was calculated using the **genomecov** function of **BEDTools** [404].

3.9.3 Phylogenetic trees

Maximum-likelihood phylogenetic trees were inferred using **IQTREE** v2.0.7 [405]. SNP alignments were used as input. Constant sites were obtained using **snp-sites** and passed to **IQTREE** with the -fconst option. Generally, the substitution model was set to GTR+G except when indicated that the best-fitting was tested using the -TEST option. Ultrafast bootstrap branch supports were assessed after 1000 pseudorandom replicates [406]. When phylogenetic trees were inferred from plasmid sequences, we added the -cbz option in **IQTREE** v2.0.7 [405], which has a different behaviour with gaps.

Phylogenetic trees and annotations were plotted using the **Interactive Tree OnLine** v6 (https://itol.embl.de/) and **R** using **ggtree** [407]**.**

3.9.4 Definition of lineages

Species can be subdivided into any number of lineages, yet there is no threshold to test if two groups of individuals belong to separate lineages or not. cgMLST schemes are popularly used to define lineages within the *K. pneumoniae* populations, especially the 694-cgMLST scheme [37]. This cgMLST scheme was applied to obtain the CG, and the predefined lineages for each sample. This method was preferred over other methods (patristic distances, ST…) because of its reproducibility and the

integration of the resulting lineages with clonal groups or STs already described [408].

Chewbbaca [409] was used to infer the 694-cgMLST profile of the isolates. Profiles and alleles were downloaded from the *Klebsiella* MLST site [410]. After noticing that **Chewbbaca** overestimated new inferred alleles, all new alleles were verified by making a **BLAST+** [378] **search** of each novel allele to all previously identified alleles of the corresponding gene. Only those alleles showing less than 100% identity and 100% coverage with any known allele were kept as new.

Profile distances between samples were obtained by **cgml-dists** (https://github.com/tseemann/cgmlst-dists). cgMLST distances were used to cluster genomes into CG groups, lineages and sublineages (SLs) using the previously proposed thresholds [408]. The clusters were obtained with R functions **hclust** and **cuttree** functions implemented in the **R stats** package v4.1.2. The thresholds used were: 610 allelic mismatches separated species, 585 separated subspecies (or phylogroups), 190 allelic mismatches defined 'sublineages', and 43 allelic mismatches separated clonal groups.

Phylogenetic clades within SLs were defined using **fastBAPS** R package v1.0.6 [411]. We used the optimise_prior function using the "optimise.symmetric" parameter. For the fast_baps function we used default parameters. One important parameter in this function is *k*, the total number of clusters that are expected. When there is no clue about it, the default is set to the number of sequences / 4, which was used. The stability of the inferred clusters was made using the bootstrap method as the developers of the package recommend in the instructions (https://github.com/gtonkinhill/fastbaps).

3.9.5 Within and between nucleotide diversity

To estimate the within population diversity (D) we calculated the nucleotide diversity for each population as the average number of pairwise nucleotide differences (*k*) between all the genomes in the same population, and *k* was calculated using the SNP pairwise distances obtained using **snp-dists** software. The nucleotide diversity per site (π) was calculated dividing *k* by the total length of the alignment used to

obtain the SNP pairwise distances. Nucleotide divergence between populations x and y was obtained using the same procedure for all the pairs of sequences taken from each population (Dxy). To measure the absolute differentiation between populations we estimated the "net" nucleotide differences between two populations, or Da, as

$$Da = Dxy - (Dx+Dy)/2,$$

with D taking values of k or π for absolute or per site differentiation, respectively.

3.9.6 Definition and identification of transmissions

To avoid the limitations of mapping approaches, and to compare as large a core-genome alignment as possible, we performed a core genome analysis of the closely related clusters identified by **fastbaps** and studied the SNP distances between them. We defined as **possible transmission groups** (PTG) those isolates with pairwise SNP distances equal or fewer than 5 SNPs per Mbp compared [67, 413]. Of note, we did not remove recombinant regions, as it has been shown that masking them can inappropriately reduce the number of SNPs between truly distant isolates, incorrectly overestimating possible transmission events [331, 412].

To determine PTGs, we used the SNPs distances obtained by **snp-dist** followed by clustering using the R **stats** v4.1.2 functions **hclust** and **cuttree**. To build the transmission networks we used the R package **network** v1.18.1 and to plot them we used the **ggplot2** v3.4.2 and **ggnetwork** v0.5.10 libraries.

3.9.7 Temporal signal

To examine the relationship between genetic divergence and time (temporal signal) we used **TempEST** [414]. The temporal signal refers to the degree of correlation between the genetic divergence inferred from a phylogenetic tree and the temporal information associated with the samples, such as collection dates or sampling times. A strong temporal signal indicates that the genetic divergence is well explained by the temporal information. TempEst provides a means to investigate the potential temporal correlation between a tree, which describes the genetic divergence, and the dates associated with the tips. As the temporal signals obtained were high

enough, we used the estimation of the time to the most common recent ancestor (tMRCA) made by TempEst.

3.10 Plasmid analysis

3.10.1 Relationships among plasmids using wGRR

Comparing plasmids is difficult due to two main problems. Firstly, plasmids are often difficult to reconstruct and almost impossible when assemblies are based only on short reads. Obtaining complete plasmid sequences is mostly restricted to the use of long-read sequencing [415]. Secondly, the plasticity of plasmids' gene content complicates sequence comparisons between plasmids because usually there is only a small percentage of shared homologous genes between them, even when they belong to the same Inc type.

Hence, for plasmid comparison, a custom plasmid database was created. For this, all the assemblies from the NLSAR collection and those downloaded from databases with less than 20 contigs were kept, corresponding mostly to experiments including long-reads sequencing (Supplementary Figure 3.1). Although at this stage chromosomes may not be closed, plasmids are expected to be complete. The contigs of each assembly were interrogated using **mlplasmids v2.1.0**, a machine-learning tool that predicts if a contig is plasmid-borne or chromosomal. Mlplasmids was run using *K. pneumoniae* as the species model and with default values, which were a minimum contig length of 1,000 bp and a posterior probability threshold of 0.7 [416]. To extract the predicted plasmid contigs from the genome assemblies, we used **seqtk** subseq (https://github.com/lh3/seqtk). To enrich the database with possible closed plasmids of the NLSAR collection, we interrogated all the short-read assemblies to add to the plasmid custom database all the contigs that were circularised and predicted to be plasmids. Most of them were short plasmids (3-6 Kb); however, we rescued some larger ones.

Plasmid relationships were calculated with the weighted Gene Repertoire Relatedness score (**wGRR**) [417] as described [418, 419]. The wGRR is often used to find relationships between pairs of MGEs, such as ICEs or plasmids. This score varies

between 0 and 1, and it represents the number of homologous proteins between two elements, weighted by their sequence similarity. It is 0 if there are no orthologs between the elements and 1 if all genes of the smaller element have an ortholog 100% identical to the other element. To perform wGRR between plasmids we made a **BLAST+** [378] search between all the proteins detected by **Prokka** v1.14.6 [374] in those plasmids.

To reduce computation time, wGRR scores were estimated only between plasmids longer than 10 Kb. Short plasmids constituted half of the custom database and did not carry AMR genes.

3.10.2 Plasmid identification and classification

Complete plasmid sequences were characterised using different tools. Incompatibility types were identified using **PlasmidFinder** whereas *in silico* predictions of the relaxase, *oriT*, and mate-pair formation type and predicted transferability of the plasmid **Mob-typer** v.3.1.4 [420]. Mob-typer classifies plasmids as 'conjugative' if they contain at least a relaxase and a mate-pair formation marker. Plasmids that lack mate-pair formation markers but contain either a relaxase or an *oriT* are classified as 'mobilizable'. Plasmids that lack both, a relaxase and an *oriT*, are classified as 'non-mobilizable'. Plasmids were also interrogated for AMR, virulence factors and stressors content using **AMRFinderPlus** [376]. Plasmid sequences were annotated using **Prokka** [374].

To identify similar plasmids publicly available we used **BLAST+** [378] with the NCBI standard nucleotide collection and the **PLSDB** online tool [421].

Plasmid figures were drawn using the CGview server [422] and the **genoplotR** [423] library.

3.11 Computer resources

Bioinformatic analyses were performed in the computing cluster of the NLSAR. It consists of two computation nodes, Bull X800 with 6 TB and 3 TB of RAM memory and 384 and 192 cores, respectively. The former system includes two GPU cards

Volta V100S of 32 GB and they are connected to a Lustre storage system with 360 TB. The cluster runs with a CentOS version 8 operating system and a SLURM system for the management of jobs.

3.12 *bla*$_{\text{NDM-23}}$ transformation and conjugation assays

To evaluate the antimicrobial susceptibility of the relevant *bla*$_{\text{NDM}}$-carrying isolates, we performed cloning and transformation assays. The complete sequences of *bla*$_{\text{NDM}}$ genes were amplified as previously described [424]. PCR fragments were cloned using the TOPO TA Cloning kit (Life Technologies) and transformed into *E. coli* TOP10 background (Life Technologies). Transformants carrying the *bla*$_{\text{NDM}}$ genes were selected on plates containing kanamycin (50 µg/mL) and confirmed by PCR as above. Antimicrobial susceptibility testing of the transformants was performed using the MicroScan WalkAway (Beckman Coulter) automated system. For carbapenems, the minimum inhibitory concentration (MIC) was checked by the broth microdilution method. Cefiderocol was tested using MIC Test Strips (Liofilchem™ MTS™). We made three independent replicates for each isolate. Growth curves were represented using **ggplot2** v3.4.0 [369].

To study the mobility of the plasmid carrying the *bla*$_{\text{NDM}}$ gene, we performed conjugation assays using the clinical strains encoding carbapenemases as donors and the plasmid-free azide-resistant *E. coli* J53 strain as the recipient. Three *K. pneumoniae* donors were studied: isolate 179KP-HG, carrying *bla*$_{\text{NDM-1}}$, isolate 146KP-HG, carrying *bla*$_{\text{NDM-23}}$, and isolate 262KP-HG, carrying *bla*$_{\text{NDM-23}}$ and *bla*$_{\text{OXA-48}}$. Donor and recipient cells were incubated separately overnight at 37°C with moderate shaking (150 rpm) in 1 mL of LB broth. Overnight cultures were ten-fold diluted and combined in a 1:1 donor-to-recipient ratio. Precipitated conjugation mixtures were cultured overnight at 37°C in a nitrocellulose membrane (0.22 µm pore diameter). Finally, recipients, donors, and transconjugants were selected on LB agar containing 0.5 µg/mL of ertapenem, 100 µg/mL of sodium azide, or both. Technical and biological replicates for each isolate were performed. Mobilisation of the carbapenemase genes into *E. coli* J53 was confirmed by CHROMagar and PCR. PCR for the *bla*$_{\text{OXA-48}}$ gene was performed as previously described [425].

CHAPTER 4
From regional to global genomics of *K. pneumoniae*

To study the genomic epidemiology of *K. pneumoniae* in the Comunitat Valenciana, we conducted the SKPCV project in the NLSAR hospitals (see details in §3.1).

Genomic surveillance projects, such as the SKPCV project, require the integration of "contextual" genomes at local and global scales. External data provide a more complete snapshot of the dissemination of lineages that cross geographic boundaries frequently [426]. Moreover, database genomes can help to decipher the temporal and geographic origin of emerging lineages and AMR genes [263]. To gather all possible contextual information, we downloaded more than 11,000 *K. pneumoniae* genomes from different public databases.

The availability of large genome databases is an advantage and a valuable source of information. Nevertheless, working with large datasets, especially when dealing with draft genomes, can be a handicap for the reliability of the results if they are not high quality data [427,428]. For this, we designed a workflow of tiered filtering steps that allowed us to remove genomes that corresponded to other taxa, were contaminated, had bad-quality statistics or had a negative effect on downstream analyses. We applied it to the strains sequenced in our laboratory and to those downloaded from different databases.

In addition, in this chapter we provide an overview of the resulting genome collections that will be exploited in the following chapters. We detail the metadata associated with the genomes, such as collection date and origin, and the genetic features, with special emphasis in population diversity, AMR, and virulence factors. Moreover, we address the contextualization of the NLSAR collection within the downloaded genomes.

4.1 Initial genome collection

4.1.1 The NLSAR and the SKPCV

A total of 1,774 isolates were collected under the SKPCV project from 2017 until 2020. In addition, we included 414 isolates from three different hospitals collected under the NLSAR from 2004 to 2017. In this chapter, as both collections were treated in the same way, we will refer to both as NLSAR isolates/collection to simplify the

nomenclature. The strains were subjected to whole genome sequencing and *de novo* assembly (see §3.4 and §3.6).

4.1.2 Database genomes

A total of 11,553 *K. pneumoniae* genomes downloaded from RefSeq (n=10,283), GenBank (n=589) and ENA (n=681) were included in the dataset. ENA SRA experiments were subjected to the same assembly procedure (see §3.6) as the NLSAR collected strains.

4.2 WGS results and filtering

To keep only high-quality samples and assemblies for downstream analyses, various filtering steps were used. For the isolates sequenced in this project (NLSAR), the filtering steps included the analysis of the number of reads, taxonomy assignment by Kraken and Kleborate, assembly quality, heterozygous positions, and mash distances. For isolates downloaded from the ENA (SRA), the filtering process included the analysis of the number of reads, taxonomy assignment by Kleborate, assembly quality, heterozygous positions, and mash distances. For assemblies downloaded from the GenBank database (CGF) and ENA (GCA), the filtering steps included taxonomy assignment by Kleborate, assembly quality, and mash distance. Filters are detailed in §3.7 and the results for each genome, analysis and filter is gathered in Supplementary Table 4.1.

4.2.1 Basic sequencing statistics

The overall average number of reads per sample in the NLSAR collection was 1,309,321 (~73x), yet 293 isolates had an average depth of over 100x. The mean depth of SRA samples was higher, with an average of 2,899,387 reads (161x). We discarded 13 NLSAR and 3 SRA samples with a mean depth lower than 10x (fewer than 200,000 reads) (Figure 4.1).

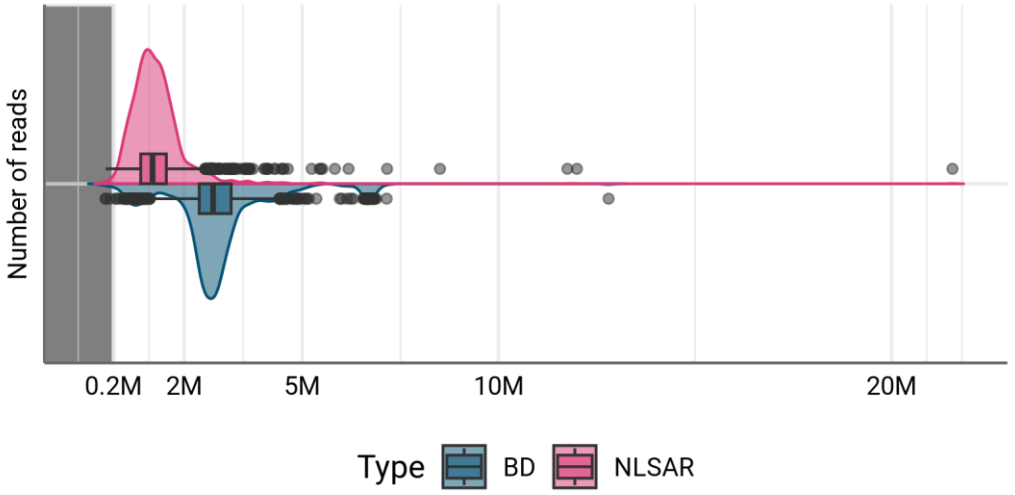


Figure 4.1. Distribution of the number of reads per sample for the NLSAR collection and the SRA downloaded experiments displayed as box plots and density curves. Lower and upper hinges of the box plot correspond to the first and third quartiles (the 25th and 75th percentiles) and the vertical bar shows the median. Data beyond 1.5 * IQR are plotted as dots. Filter threshold and removed isolates fall into the grey area.

4.2.2 Taxonomic assignment and interspecies contamination

In order to remove species other than *K. pneumoniae sensu stricto* or contaminated samples, the short-reads from the 2,175 sequencing experiments were identified taxonomically with Kraken2 and the corresponding assemblies were assigned to species with Kleborate. Firstly, we removed 15 samples that had less than 40% of the reads identified (Figure 4.2 A). The threshold was adapted to the data as, in general, the percentage of reads identified per sample was low (average 61.1%, range 3.3 - 98.4%).

To assign a species to each sample, we estimated a ratio by dividing the number of reads assigned to *Klebsiella* by the number of reads assigned to other genera. 76 samples had a larger portion of reads assigned to other bacterial species. The alternative genera most frequently found were *Escherichia* (n=19), *Proteus* (n=12), *Enterobacter* (n=8), *Pseudomonas* (n=7), and *Acinetobacter* (n=7) (Figure 4.2 A) and the corresponding samples were removed from further analyses. In addition, we examined in detail those samples with *Klebsiella* to other genera ratios above 75%

to evaluate whether the remaining 25% were due to interspecies contamination (Figure 4.2 B). 22 samples showed a notable percentage of reads assigned to other genera, primarily *Proteus*, *Salmonella*, *Escherichia*, and *Serratia,* among others. As *K. pneumoniae*, these bacterial species are also clinically relevant, highly prevalent in hospitals and frequently isolated in their Clinical Microbiology laboratories [18, 429]. Therefore, it is not surprising that they were misidentified as *K. pneumoniae* or, alternatively, were contaminants in supposedly "pure" cultures. It should be noted that almost all the samples always had a low proportion of reads assigned to *Escherichia* or, alternatively, to *Salmonella*. Although these are also highly prevalent pathogens in clinical settings, this effect is probably occasioned by the high over-representation of these two genera in the databases, including the database used by Kraken2 to assign taxa to reads.

We also removed 48 samples that were identified as other *Klebsiella* species by Kleborate, mostly *K. quasipneumoniae* (n=22, 45%). Of note, 2 samples (HRA2_01 and HGUA6_54) were identified as unknown species. HRA2_01 might be a new *Klebsiella* species as the Kraken results showed that 91% of the total identified reads were *Klebsiella*. HGUA6_54 might belong to a new genus, as only 56% of its reads were identified and, of those, only 52% belonged to *Klebsiella* spp.

Kleborate was also applied to the database genomes, removing 68 genomes, of which 45 were samples from *Escherichia* and *Enterobacter* and 23 were from other *Klebsiella* species, with a large representation of *K. quasipneumoniae*.

Curiously, the percentage of total identified reads differed between *Klebsiella* species. *K. pneumoniae sensu stricto* and *K. quasipneumoniae* subsp. *similipneumoniae* genomes had an average of total identified reads of 60.2% (range 42.7 - 70.4%), whereas other *Klebsiella*, including *K. aerogenes*, *K. michiganensis*, *K. oxytoca, K. variicola,* and *K. quasipneumoniae* subsp*. quasipneumoniae,* showed very high values (average of 88.8%, range 80.3 - 97.0%). The latter can be observed in Figure 4.2 A as a separate cloud in the upper right corner of the chart, which is also probably due to a database content effect.

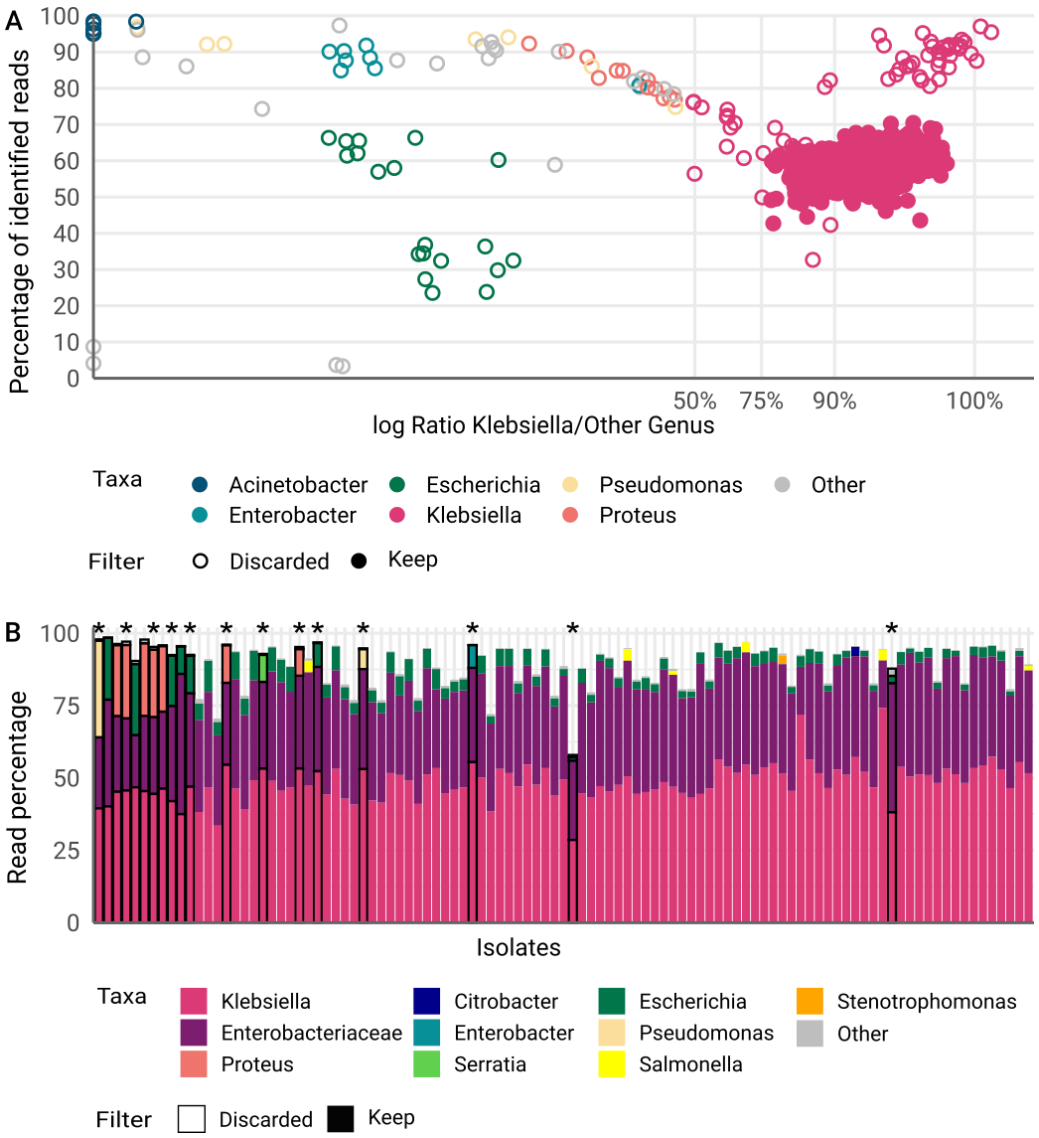


Figure 4.2. A) Each dot represents each strain collected under the NLSAR program. The x-axis represents the percentage of reads assigned to *Klebsiella* relative to the number of reads assigned to other genera. Dots are coloured with the most prevalent genera identified and are filled if the strain was kept for further analysis. The x-axis was transformed to Log10. B) Zoom in on the samples with Klebsiella to other genus ratios between 75% and 90% in (A). Each bar represents an isolate and colours represent different taxa. Isolates that have been removed are indicated by contorted bars and by stars at the top of the bar.

4.2.3 Assembly quality

After applying the taxonomy filter, the genome of the remaining samples was reconstructed by *de novo* assembly. Highly fragmented assemblies or assemblies with different values from those expected for the species under study are possibly due to sequencing errors or contamination and, therefore, are not suitable for genome comparison analysis [384]. By this means, we discarded 17 NLSAR and 313 database isolates (Table 4.1, Figure 4.3 A-D). NLSAR isolates were removed exclusively by the number of contigs, which was expected because they had already been filtered by taxonomy and all the remaining isolates were supposed to be *K. pneumoniae*. On the contrary, 7 of the 313 database assemblies were removed due to unexpectedly different genome length or GC content values for *K. pneumoniae* genomes. As misidentification was already evaluated using Kleborate, but taxonomic contamination was not, these genomes were likely contaminated.

The NLSAR genomes retained were assembled in 125 contigs on average (range 1-364) whereas genomes from the database were assembled in a slightly lower number of contigs (average 103, range 1 - 499). The rest of the parameters were highly similar between databases, length (5.5 Mb ± 172 Kb), number of genes (4,874 ± 150), and GC content (57.1% ± 0.223) (Figure 4.3 A-D).

4.2.4 Intraspecies contamination

Previous metrics would not be affected if contamination was due to closely related strains. Hence, to remove mixtures of different *K. pneumoniae* lineages, we evaluated the number of genes affected by heterozygous positions. By this method, we removed 8 isolates with more than 1,000 genes affected, which are likely to be mixtures of *K. pneumoniae* strains (Figure 4.3 E). The remaining isolates had an average of 260 (± 85.8) genes affected by heterozygosity, which was higher for NLSAR samples than for those downloaded from the ENA. Although this may appear to be a shortcoming, we have to take into account that *K. pneumoniae* genomes usually carry a large load of MGEs and inserted prophages [19, 430]. Plasmids, phages and other MGEs have duplicated regions that they also share with each other and that are prone to accumulate sequencing and reference-based mapping errors, which can be detected as heterozygous positions.

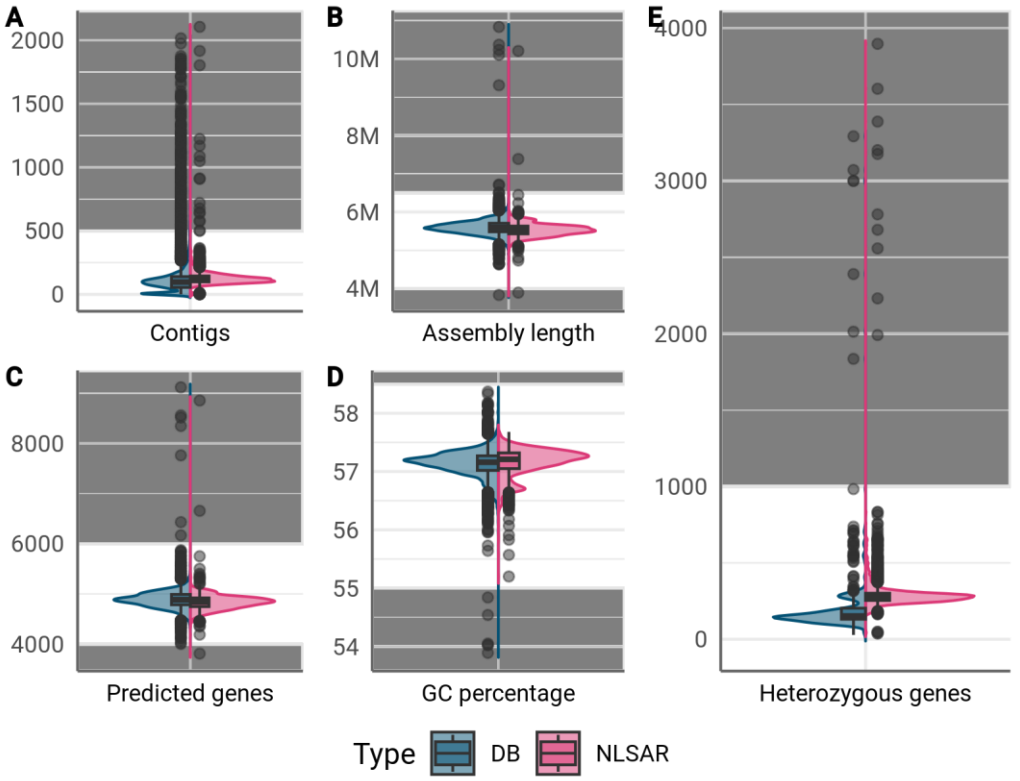


Figure 4.3. Distribution of different assembly statistics in the NLSAR and downloaded genomes: A) contigs, B) assembly length, C) predicted genes, D) GC percentage, E) Genes affected by heterozygosity (in this case database genomes only include SRA experiments). The lower and upper hinges of the box plot correspond to the first and third quartiles (the 25th and 75th percentiles) and the horizontal bar shows the median. Data beyond 1.5 * IQR are plotted as dots. Filter thresholds and removed isolates fall into the grey area.

4.2.5 Pairwise genomic similarity

The last filtering step was to check whether the isolates that had passed all the quality and contamination filters had "*K. pneumoniae*"- like genome content. For this, we used Mash distances, which estimate the genetic relationships between isolates by comparing their k-mer content. We removed 4 NLSAR and 31 RefSeq assemblies that showed a large mean distance from the other samples.

We also used Mash distances coupled with Kleborate results to check whether corresponding short and long-read sequencing had been performed with the same isolate. We removed isolate HCV2_51, which had a notable mash distance between the long and short assemblies and, although they belonged to the same lineage (ST), had different KL and O types.

4 2.6 Filtering summary

Finally, of the 13,741 genomes initially considered, 13,133 (95.6%) genomes passed the tiered quality-control filter: 1,999 (91.4%) NLSAR and 11,134 (96.3%) database assemblies (Table 4.1, Supplementary Table 4.1). The entire collection, regardless of the source, was highly affected by contamination, misclassification, and bad-quality assemblies.

Table 4.1. Summary of isolates included and removed after applying each of the filters mentioned.

QC Filter	RefSeq Assemblies	GenBank Assemblies	SRA Raw Data	NLSAR (and SKPCV) Genomes
Total samples	**10.283**	**589**	**681**	**2.188**
Low number of reads	-	-	3	13
Kraken taxonomy	-	-	-	98
Kleborate	13	1	54	48
Contigs	301	4	1	17
Assembly length or genes	2	0	0	0
Assembly GC	0	0	5	0
Heterozygous positions	-	-	4	8
Mash distance	31	0	0	4
Long vs short sequencing	-	-	-	1
Samples that met all criteria	**9.936**	**584**	**614**	**1.999**

- : Filter not applied.

The NLSAR collection showed a high level of contaminant or misidentified isolates with other clinically relevant pathogens widely found in Clinical Microbiology laboratories (Figure 4.2. A-B). Hence, it is not surprising that the most effective filters

for NLSAR genomes were the taxonomic evaluation of sequencing reads using Kraken2 and the species identification by Kleborate. Unexpectedly, taxonomic filtering was also important in database genomes, where we found frequent contamination and misidentification. In our dataset, 14 NCBI genomes were misidentified with other species, whereas 31 were likely contaminated. Moreover, 54 SRA experiments were also misidentified or contaminated.

Our results show that whereas the quality of assemblies can be easily assessed, contamination, especially with highly similar strains, requires much more effort. We had to apply three different filters to get rid of contaminant isolates: read taxonomic classification (§4.2.2), intraspecies contamination (§4.2.4), and genomic similarity between isolates (§4.2.5).

To test the essentiality of the finer taxonomic filter we checked whether those isolates would have been removed by species identification by Kleborate or assembly metrics. The metrics of the contaminated samples were generally within the ranges observed in samples considered to be “pure”, except for the samples that clearly corresponded to other species or were highly contaminated. Moreover, in our analysis, none of the filtering steps fully replaced any of the others. We performed subsequent analyses avoiding some filtering steps, but the results were not as reliable as with all the filters. Each step eliminates a different problem that may affect distinctly different metrics. Avoiding a step translates into errors that affect posterior genomic comparisons [395, 396].

4.3 Final working datasets

The final NLSAR collection comprises 1,604 isolates collected under the SKPCV project in 8 different CV hospitals (Figure 4.4 A-B) and 395 isolates from other retrospective projects. The isolation dates start as early as 2004, spanning 16 years, whereas the SKPCV project was conducted between 2017 and 2019 (Figure 4.4 C).

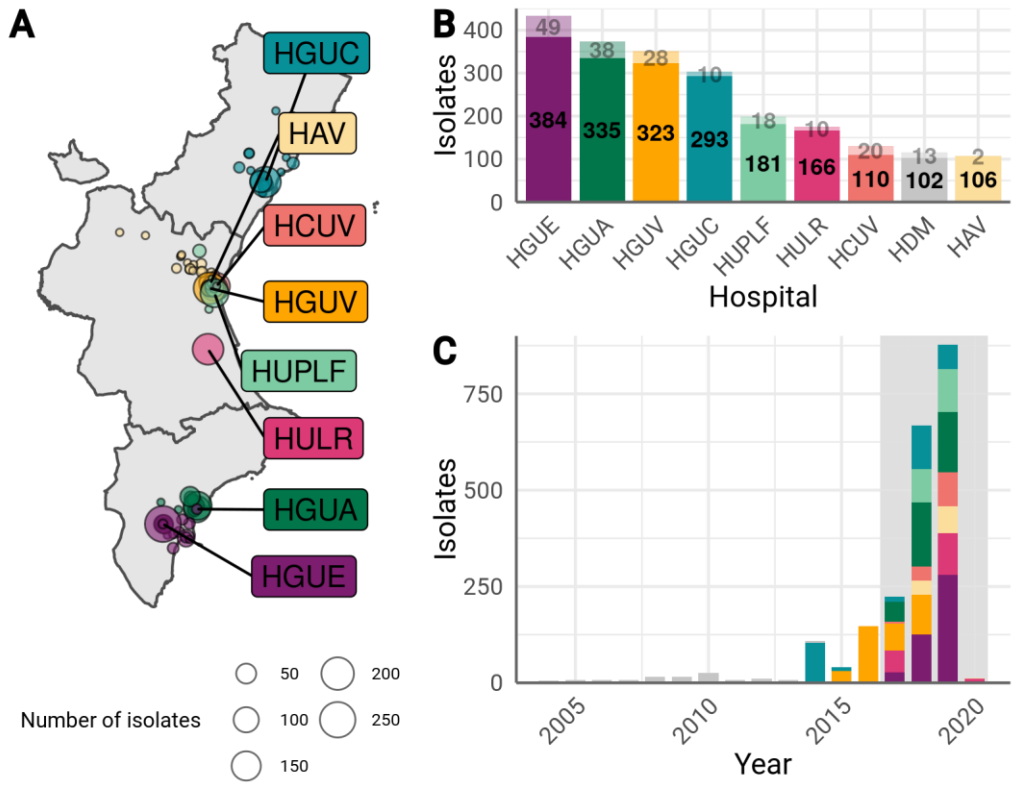


Figure 4.4. A) Geographic distribution of the sampling sources of the SKPCV project. Each dot represents a sampling point. The size of the dot indicates the number of isolates collected in each source. Reference hospitals are labelled. B) Isolate contribution per reference hospital. In each bar it is shown the isolates that passed the genome quality filters and those that did not (lighter colours). C) Distribution of the final NLSAR collection through the years and hospitals. The SKPCV sampling period is highlighted in grey.

The database genome data, comprising a total of 11,134 genomes, incorporates isolates from at least 94 countries spanning at least 109 years (Figure 4.5). For most of the samples, we obtained the origin (97%), collection year (85%), and isolation source (90%). Most of the genomes collected were recovered from humans (n=9,345, 83%) and/or were clinically related. Only a few isolates were collected from other animals (n=315) and the environment (n=315). Animal samples were mostly collected in farms, whereas environmental samples were mostly collected from sewage or hospital surfaces. The majority (10737/11260, 95%) were collected between 2011 and 2019 (Figure 4.5 B) and the oldest genome was collected in 1911

in the USA (NCTC418, GCF_900635995.1). To better contextualise the NLSAR genomes in the global collection, we grouped the genomes by continent except for the NLSAR and Spanish isolates, which were incorporated as different groups (Figure 4.5). Each genome was included in only one group (NLSAR > Spain > Europe).

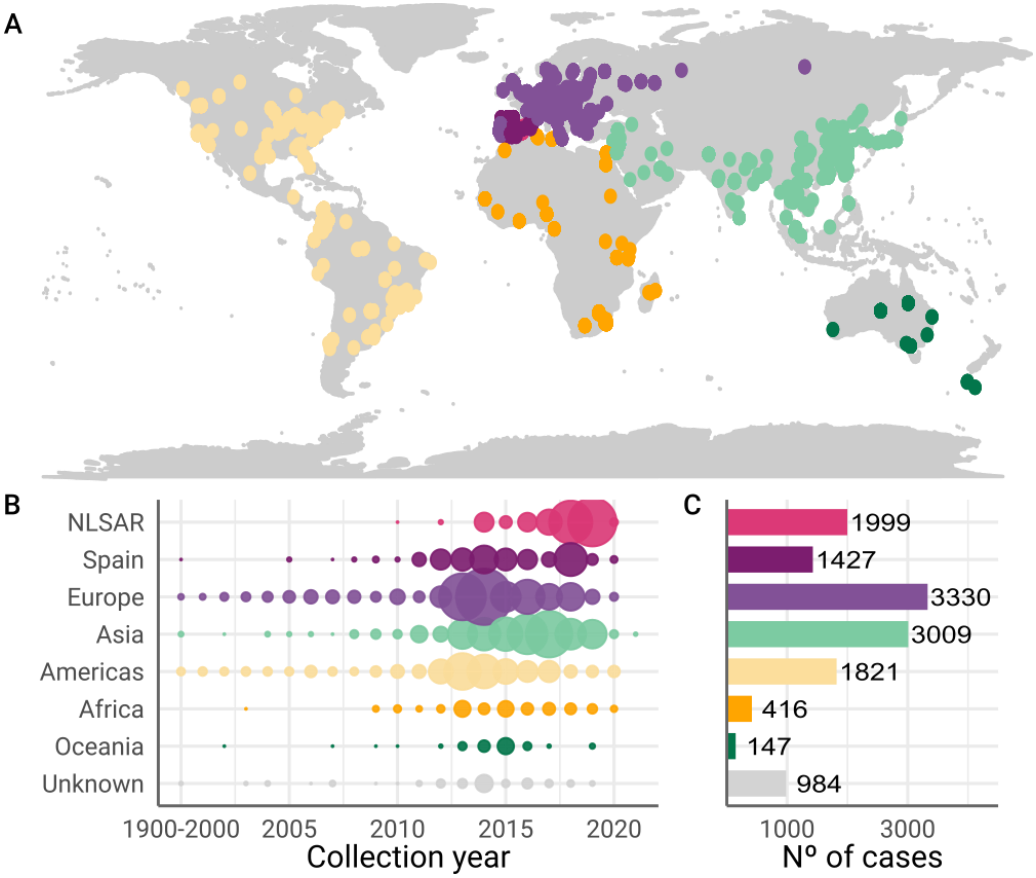


Figure 4.5. A) Global distribution of the final working dataset. Dots represent each genome. B) Distribution of genomes by date and continent. D) Number of genomes per region.

The distribution in and within continents was far from uniform, with some specific countries being overrepresented. Most isolates were collected in Europe (3,330/11,134, 29%, or 43%, 4,757/11,134 if those from Spain are also included) (Figure 4.5 D), especially in Italy (n= 669) and the United Kingdom (n= 655). Notably, half of the European genomes (51.1%) were collected in a European-wide surveillance study of carbapenem-non-susceptible *K. pneumoniae* (PRJEB10018)

[151] remarking on the relevance of major surveillance projects. Spain was poorly represented in RefSeq, with only 274 assemblies. By including the Spanish genomes available in the GenBank and SRA databases, we improved the Spanish representation with 1,153 assemblies in total. The Asian genomes were mainly from China (1449/3009, 48%) whereas those of America were mostly collected in the USA (1333/1821, 73%). Despite the large number of genomes collected, many regions and countries remain underrepresented or not represented at all.

4.4 Genomic description

To contextualise the NLSAR samples in the final working dataset, we also assessed their genomic diversity, including phylogenetic lineages (STs), virulence, and antimicrobial resistance markers, especially those related to 3GC- and carbapenem-resistance.

4.4.1 Population diversity and high-risk clones

To study the population diversity of each region we performed lineage accumulation curves and Simpson's diversity index for each region (Figure 4.6 A).

Simpson's diversity index (Di) measures the probability that two strains randomly selected from a sample belong to the same lineage. So, the higher the index, the greater the diversity. Curiously, Africa was the most diverse collection (Di = 0.96), despite being the second-least sampled region. The Asian, Spanish, and European regions also showed high diversity values (Di > 0.9). On the contrary, America and the NLSAR collections showed the lowest values (Di = 0.82 and 0.86, respectively) (Figure 4.6 A). Diversity values are a result of lineage richness and evenness. Lower values can be explained by actual low diversity or highly biased collections.

The accumulation curves showed that, despite differences in sample sizes, no region reached the saturation point (a plateau) that represents its total diversity. Despite the high diversity already reported in the species [31], further sequencing projects focused on screening diversity would be needed to capture the whole *K. pneumoniae* population diversity of each region.

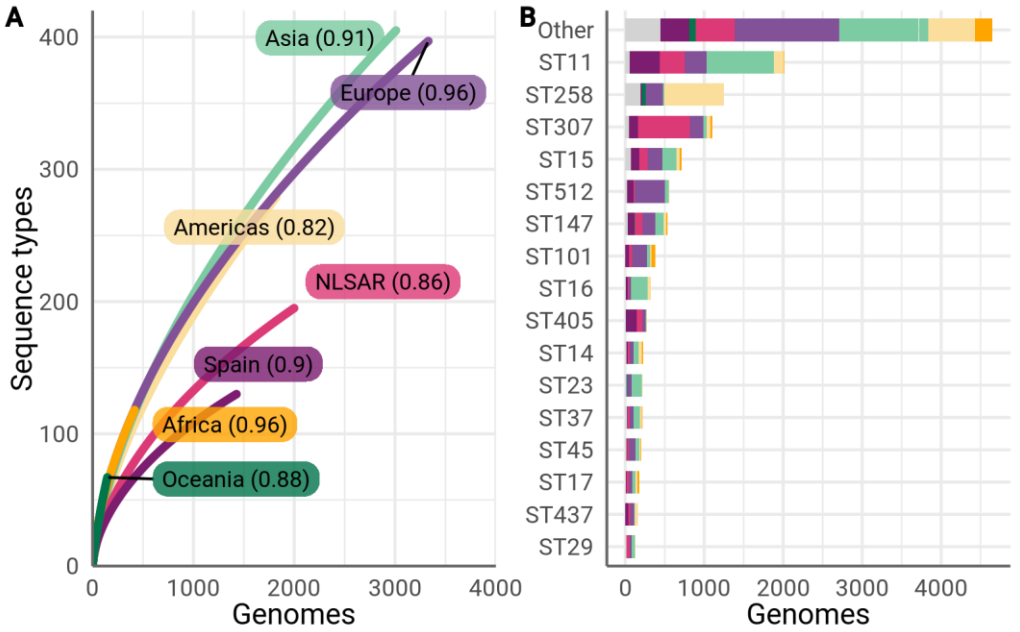


Figure 4.6. A) Rarefaction curves show the accumulation of new STs in each region as more genomes were sampled. The Simpson's diversity index (Di) of each region is shown in parentheses. b) Distribution of the sequence types in all the databases coloured by region.

Regarding the lineages, we identified 930 known STs in the database collection, of which 50.1% were represented by just a single genome. In the NLSAR collection, we found 168 different STs and 24 locus variants of known STs. Notably, 48.2% of the STs were singletons. Overall, the same major lineages were observed in the NLSAR and the global collections (Figure 4.6 B), which are recognised as MDR- HRC lineages [31].

To better represent the lineage distribution, we examined the global distribution of the top five most prevalent STs in each region, which resulted in a total of 14 STs (Table 4.2). In most regions, a single lineage comprised a large portion (>25%) of the collection, except for Europe and Africa where no single ST dominated and the most frequent STs had a prevalence below 15%. Africa and Oceania showed ~ 55% of other STs that were not frequently found elsewhere.

The NLSAR showed a similar composition to the Spanish collection, which was composed mostly of lineages **ST307** and **ST11** (>70%). Lineage **ST307** was the most

prevalent lineage in the NLSAR collection (n=659 genomes, 32.97%) and the third most frequent in the whole dataset. However, this can be attributed mostly to the NLSAR collection as it comprises 59.8% of all ST307 isolates (Figure 4.6 B). Without these genomes, this ST would have been sampled at the same level as ST101 or ST16. Despite being widely distributed, it was found at much lower frequencies in other regions, including Spain (Table 4.2). In contrast, lineage **ST11,** which is the second most prevalent in both the NLSAR (n=317, 15.86%) and the whole dataset, was frequent in many regions. Indeed, it was the most prevalent lineage in Asia (n=844, 28.22%) and Spain (26.70%). Other lineages found in the NLSAR collection, such as ST15, ST147, ST405 and ST101, were also found in other countries. ST15 and ST147 were found at a low prevalence but widely distributed with a similar frequency in the different collections. ST405 was found primarily in the NLSAR and the Spanish collections, where it was the second most prevalent lineage. ST101 was found in Europe, Spain and was remarkably prevalent in Africa (12.5%), where it was the most frequently found lineage.

Table 4.2. Global distribution based on the 5 most frequent STs in each region. STs proportion values are relative to each region. STs are ordered by prevalence in the whole collection.

ST	NLSAR (1999)	Spain (1427)	Europe (3330)	Asia (3009)	Americas (1821)	Africa (416)	Oceania (147)	Unknown (984)
ST11	15.86	26.70	8.35	28.22	6.75	2.64	0.68	5.39
ST258	0.10	0.91	6.55	0.50	41.46	0.00	33.33	19.92
ST307	32.97	7.92	5.20	1.20	2.75	6.01	0.00	4.67
ST512	0.75	5.75	11.38	1.66	0.66	0.00	0.00	2.44
ST16	0.80	1.61	0.84	6.91	2.09	0.96	0.68	0.71
ST101	2.15	2.94	5.62	1.46	0.55	12.50	0.00	0.30
ST15	5.35	7.57	5.44	5.95	2.25	6.73	0.68	7.11
ST405	3.75	9.46	1.32	0.13	0.38	0.24	0.00	0.81
ST23	0.20	0.35	1.65	4.22	0.55	0.24	2.04	1.42
ST147	4.85	6.03	5.05	3.36	1.70	5.29	1.36	3.15
ST437	1.35	2.73	1.47	0.30	2.25	0.00	0.00	0.10
ST152	0.35	0.21	0.39	0.10	0.38	5.05	0.68	0.61
ST17	1.70	0.21	1.20	1.36	1.43	5.05	0.00	1.12
ST86	0.00	0.00	0.57	0.76	0.55	0.48	6.12	1.63
Other	29.82	27.61	44.97	43.87	36.25	54.81	54.43	50.62

Despite being the second and fifth most frequent lineages in the whole collection, **ST258** (n=1,248) **and ST512** (n=562) were very rare in most regions. They account for only 5.75% of the genomes in Spain, while they represent only 0.45% of the NLSAR collection. In Oceania, ST258 accounts for 33% of the isolates, but this is due to a specific sequencing project (PRJNA415530), and very likely not to a real epidemiological situation. These lineages only had a big impact in the USA (45%), Italy, and Greece, but not elsewhere [150]. The prevalence of these lineages can be explained because of the intense sequencing efforts to screen the KPC pandemic [32, 110]. The large sampling bias introduced by this CG explains the low diversity index found in the Americas (Figure 4.6 A).

Other rare clones were found to be specific to some regions, including hypervirulent lineages ST23 and ST16 in Asia, ST86 in Oceania, and MDR lineages ST17 and ST152 in Africa.

To further study the relationship between the global and the NLSAR collections, we obtained a global phylogeny using the core genomes (§3.9.1). As previously reported, the global population of *K. pneumoniae* is highly diverse, comprising hundreds of deep-branching lineages [28]. The NLSAR genomes fell into multiple branches of the global tree, showing a population as diverse as any of the other populations (Figure 4.7 A). Moreover, the NLSAR isolates were found to be closely related to many isolates from other regions. We also found some clades composed mainly of NLSAR genomes. This may suggest multiple sources and introductions to the CV but also indicate further spreading within the region (Figure 4.7).

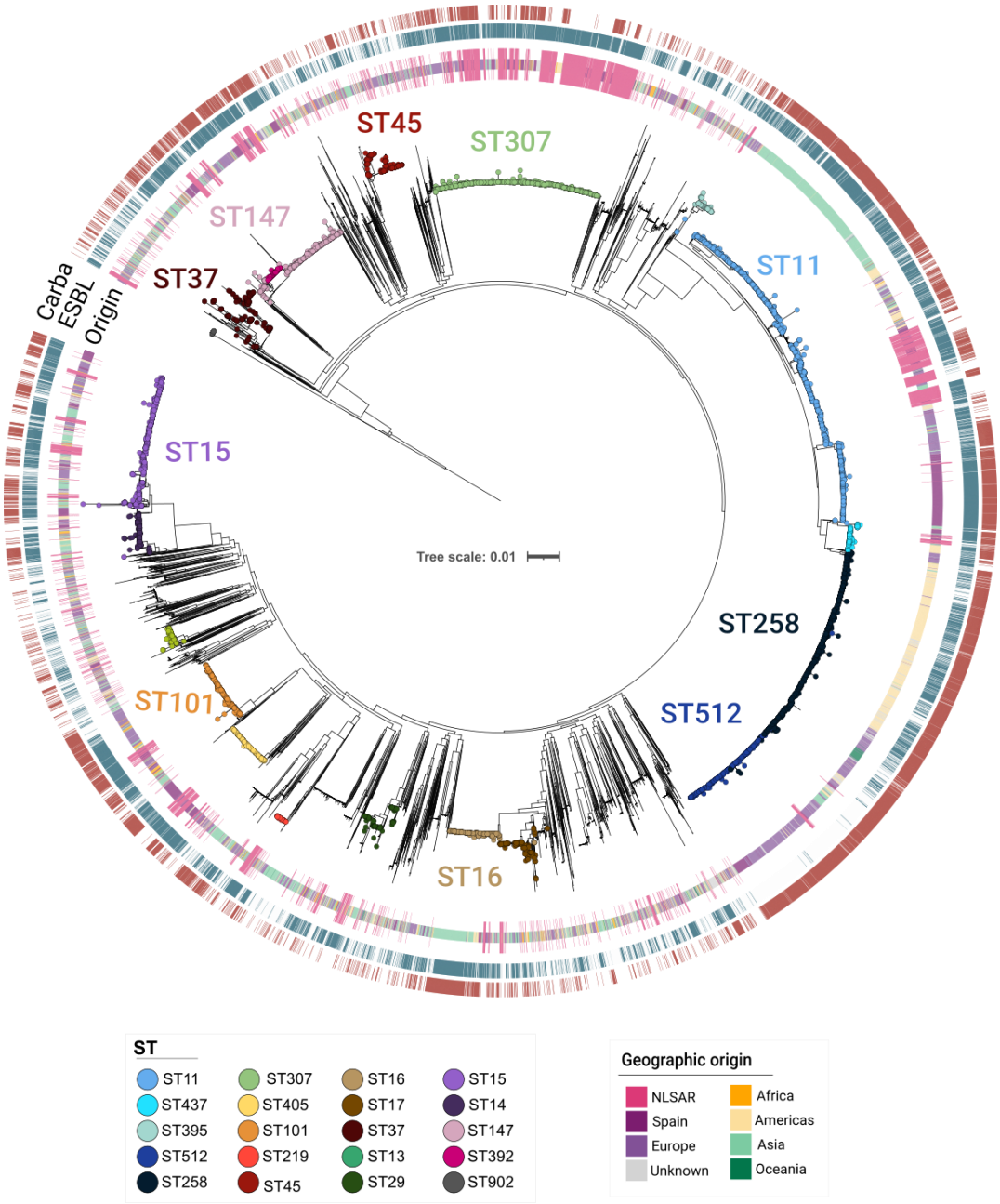


Figure 4.7. Maximum likelihood tree obtained with the core genome alignment of the 13,133 genomes (730 genes in at least 50% of the isolates). The tree was midpoint rooted and nodes with 200 to 500 leaves were collapsed. The inner ring shows the geographic origin and the second and third rings show the presence of ESBL and carbapenemase genes, respectively.

4.4.2 Antimicrobial resistance and virulence determinants

We inspected the *Klebsiella* genomes for antimicrobial resistance determinants to different antibiotic classes and virulence genes and we divided the dataset into different categories regarding their resistant/virulent nature: ESBL or carbapenemase producers, virulence factor carriers, convergent strains (carrying both resistant and virulence determinants), and wild type (WT), which encompasses those not included in the previous categories [32] (Figure 4.8).

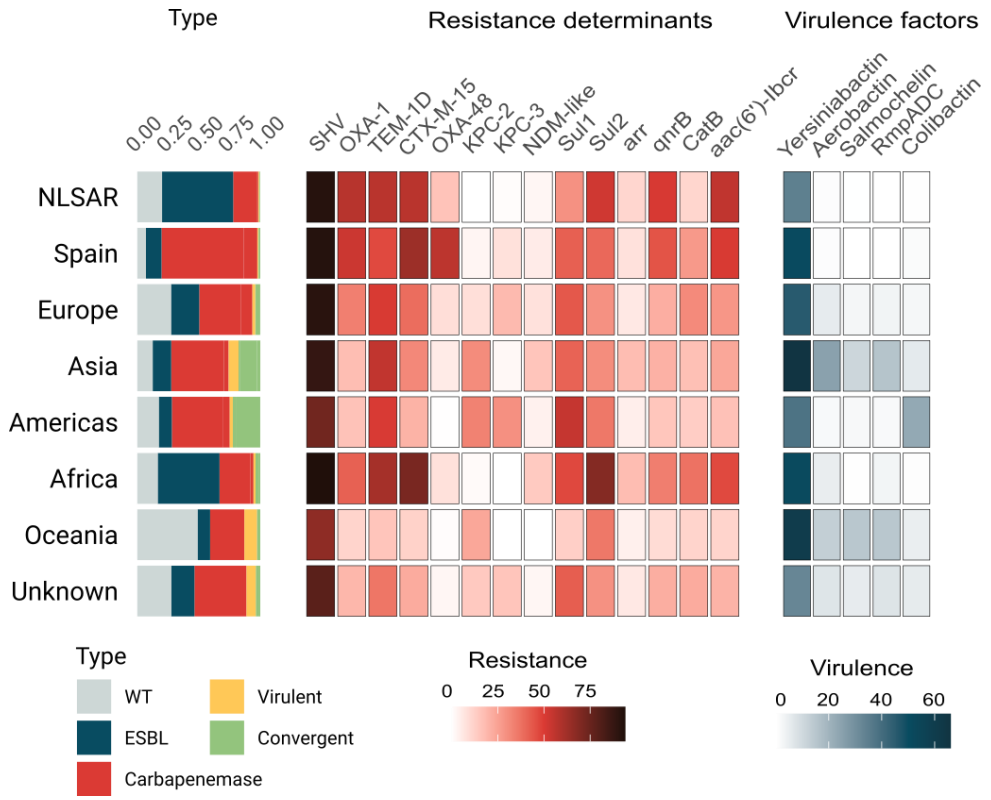


Figure 4.8. Distribution by regions of the different types of genomes regarding the presence of genetic resistance and virulence factors and the prevalence of the most relevant factors.

In the NLSAR collection, most genomes carried ESBL genes (n=1,156), carbapenemase genes (n=408), or were wild type (WT) strains (n=402). In almost every country, with the exception of Africa, carbapenemase producers were the most frequently sampled type, even though it has been shown that those are still

less prevalent than ESBL producers in all of them [12, 431]. Virulent strains or convergent genomes were only frequently sampled in Asia and America. In the NLSAR collection, only 13 isolates were virulent, and 20 isolates were convergent.

Antimicrobial resistance

The number of resistance genes per isolate in the global collection was bimodal, distinguishing between low- and high-AMR gene carriers. Low AMR gene-carriers were represented by a reduced number of isolates (25.7%) that carried an average of 4 AMR genes corresponding mainly to the intrinsic AMR genes: *bla*$_{SHV-1}$, *fosA*, *oqxA*, and *oqxB*. The high carriers had an average of 14 unique AMR genes known to confer resistance to β-lactams, fluoroquinolones, and aminoglycosides, among others (Figure 4.8). Regarding β-lactam resistance genes, the most prevalent acquired penicillinases were *bla*$_{TEM-1}$ (n=7,163). AmpC were generally rare but the most frequently found were *bla*$_{DHA}$ (n=467) and *bla*$_{CMY}$ (n=201). ESBLs were largely encoded by *bla*$_{CTX-M}$ (n=7,747), especially *bla*$_{CTX-M-15}$ (n=5,730), followed by *bla*$_{OXA-1}$ (n=4,185). Except for *bla*$_{CMY}$, these were also the most relevant β-lactam resistance genes in the NLSAR collection. Carbapenemase determinants were found in 6,390 isolates of the global collection and in 419 of the 1,999 isolates in the NLSAR collection. As carbapenemases are frequently found coupled with an ESBL (65% of isolates), often *bla*$_{CTX-M-15}$ (58%), we studied their joint distribution (Figure 4.9 A-B).

***bla*$_{OXA-48-like}$** were the carbapenemase genes most frequently found in the NLSAR collection (n=325) and were also highly prevalent in the global collection (n=2,210). This enzyme was highly associated with *bla*$_{CTX-M-15}$. This combination was widely distributed among regions and among lineages, but it was more relevant in Spain (n=846) and among ST11 strains. This carbapenemase was also found in combination with other ESBLs (e.g., CTX-M-1, CTX-M-14). Of note, the combination of *bla*$_{OXA-48-like}$ with *bla*$_{CTX-M-1}$, was almost exclusive of the CV and the ST11 lineage.

In addition to *bla*$_{OXA-48}$, we found ***bla*$_{VIM-1}$** (n=37), different ***bla*$_{NDM}$** variants (n=49), and ***bla*$_{KPC}$** (n=18) in the NLSAR collection. These enzymes were also frequent in the rest of the genomes. In the global collection, ***bla*$_{VIM}$** genes were less frequent (n=229) than any of the others. They were mostly *bla*$_{VIM-1}$ (80%) without an ESBL companion, and distributed primarily through Europe (90.1%, including Spain and the NLSAR

collection). On the contrary, ***bla*$_{NDM}$** genes were highly prevalent (n=1,010). Different ***bla*$_{NDM}$** variants (primarily NDM-1, -5, -7), associated or not with ESBL, were found distributed among all the regions, especially Asia (46.5 %) and Europe (including Spain and the NLSAR collection, 38.71%). In the NLSAR collection, we only found ***bla*$_{NDM-1}$** (n=41) and ***bla*$_{NDM-23}$** (n=8) variants. Remarkably, the ***bla*$_{NDM-23}$** variant was only found in the NLSAR collection.

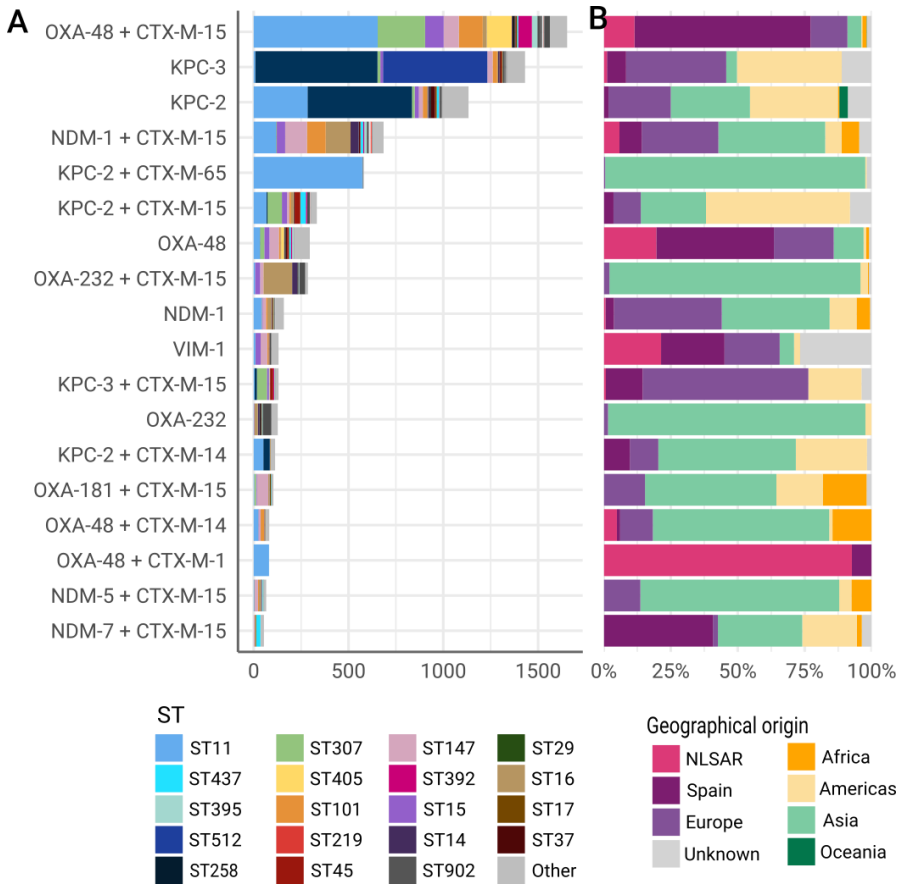


Figure 4.9. A) Prevalence of the most frequent carbapenemase and ESBL combinations and their distribution by sequence types, and B) regions. Only those combinations present in more than 50 genomes are shown.

***bla*$_{KPC}$** enzymes were the most prevalent carbapenemases in the whole collection (n=3,575) but they were the least present among NLSAR isolates. Overall, the *bla*$_{KPC}$ variants most frequently found were *bla*$_{KPC-2}$ and *bla*$_{KPC-3}$, and both were highly

associated with CG258. KPC-3 was mainly found in ST258 isolates in Europe and America, and in ST512, mostly in Europe. *bla*$_{KPC-2}$ was mostly found among ST258 strains in Europe and America, and among ST11 strains mainly in China. In the NLSAR collection, we only found 18 isolates carrying *bla*$_{KPC-3}$, mainly in ST512 (83.3%). It is surprising that the *bla*$_{KPC}$-CG258 association only exists when KPC genes are not associated with ESBLs. The same genes associated with *bla*$_{CTX-M-15}$ were much more diverse in terms of lineage spread (Figure 4.9 A-B). *bla*$_{CTX-M-15}$ may have been a mediator in the spread of *bla*$_{KPC}$ into other genetic backgrounds, but it has not been studied.

Most carbapenem and ESBL gene combinations were widely detected across lineages and regions, reflecting their mobile nature and their dissemination potential. Overall, lineage ST11 presents the largest number and diversity of carbapenem and ESBL gene combinations followed by ST147, ST16, ST15, ST37, ST14, ST17, ST307, and ST101. This result largely contradicts the role assigned to lineages ST258 and ST512 as carbapenemase disseminators since they have only played a role in the dissemination of KPCs (Figure 4.9 A-B). Notable exceptions to wide lineage and geographic distributions were CTX-M-65, which was largely associated to KPC-2 (98%), lineage ST11 (98.2%), Asia (96.9%), and *rmtB* (86.4%) [249,432–434].

Virulence loci and convergent genomes

The siderophore yersiniabactin was the virulence factor most frequently found in the genomes of the collection (n=6,199, 47.2%), including the NLSAR database (n=692, 34%) whereas most other virulence factors were rare. The other virulence factors were distributed mainly in Asia and America. Virulence siderophores, aerobactin (n=1,133), and salmochelin (n=563) and the hypermucoidy locus *rmpADC* (n=753) were mostly collected in genomes from Asia. Otherwise, the colibactin toxin was found in 782 genomes, largely associated with America and the carbapenemase KPC (49%) (Figure 4.10).

Convergent genomes were rare outside Asia and America. In Asia, especially in China, both MDR and hypervirulent *K. pneumoniae* infections are reported to be common, and cases of convergence have been reported to be on the rise [215, 216].

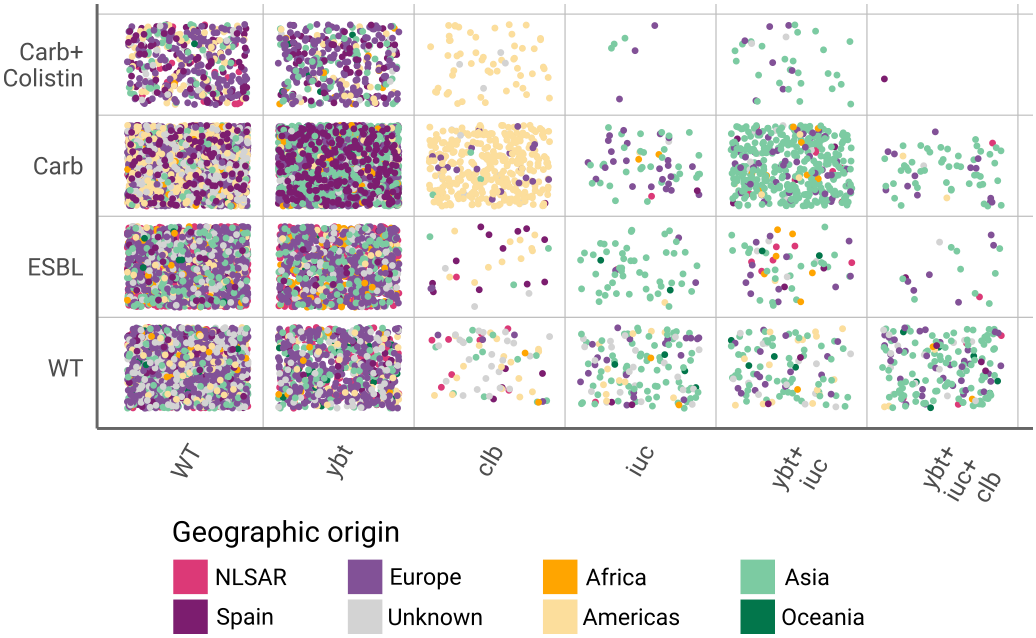


Figure 4.10. Distribution of genomes, for each combination of resistance and virulence factors. Each genome dot is coloured by region.

On the opposite, in America, hypervirulent or convergent isolates are reported as sporadic cases [435]. The same occurs in Spain [436] and Europe [314]. In the NLSAR collection, we found five different convergent lineages. We found 15 ST11 genomes carrying the $bla_{\text{OXA-48}}$ carbapenemase gene, $bla_{\text{CTX-M-1}}$ ESBL gene, and *ybt* 16 (ICE*Kp*12) and *iuc*-5 virulence factors. Convergent ST11 lineages have been reported to be prevalent in China [216], thus supporting different origins of these lineages. We also found 5 convergent strains belonging to ST268 (n=2), ST23 (n=1), ST199 (n=1), and ST48 (n=1) among the NLSAR isolates. Except for ST23, these lineages showed a high load of AMR genes (range 6 to 9). Except for ST199, they showed most hypervirulence markers, including siderophores *ybt*, *uic*, and *iro*, the toxin colibactin determinant (*clb*), or the hypermucoviscous regulator *rmpA*. Unlike ST11 convergent isolates, these were collected mostly in Asia and America.

4.4.3 Plasmids

As it is almost impossible to reliably reconstruct plasmid sequences from only short-read sequencing experiments [415], we used plasmid replicon typing as an indicator

of plasmid load and diversity. We found plasmid replicon types in 9,876 of the 13,133 genomes. The average number of replicons per entry was 4.6 (range 1-37) which was highly similar for the NLSAR genomes (4.5) to that already published [63]. We found a high variety of replicon types (n=68), with most of them present in less than 10% of the plasmid carriers (82.35%). Two replicons were found in a large portion of the genomes, IncFIB in 81.7% and IncFII in 74.1%, and many genomes co-carried them (68.5%). These two replicons have been largely found associated with AMR genes such as *bla*$_{KPC}$ and *bla*$_{CTX-M}$ [19, 437]. We also detected other replicons associated with AMR genes, specially carbapenemases and ESBLs, such as IncR (n=3,190), IncL (n=1,378), IncHI1B (n=1,190), IncFIA (n=1,095), and IncX3 (n=1,049) [19].

Although replicon typing gives insights into the diversity patterns and distribution of plasmid replicons, it does not inform about the possible structure of the plasmids or their cargo genes and, therefore, the similarity between plasmids. Hence, to investigate the relationships between the AMR plasmids collected in the NLSAR collection with those found in the global collection we created a custom plasmid database. This database contained predicted plasmids of a selection of highly assembled samples. To ensure the comparison with the plasmids of interest in the NLSAR collection, we performed long read sequencing with representatives of each AMR gene and lineage. Moreover, the rest of the short-read sequence NLSAR samples were interrogated to extract complete and closed plasmids. The methodology is detailed in §3.10. Finally, we obtained a database with 4,420 plasmids larger than 10 Kb, of which 1,046 plasmids were extracted from NLSAR genomes and 3,374 were extracted from the global database.

Among the 4,420 plasmids, 1,026 carried ESBL genes and 992 carried carbapenemase genes. We focused on those carrying the main carbapenemase genes in the NLSAR collection: *bla*$_{OXA-48}$, *bla*$_{NDM}$, *bla*$_{VIM}$, and *bla*$_{KPC}$. For each gene, we studied the relationships of the plasmids found to be carrying it in the NLSAR collection with the entire collection. Plasmid relationships were calculated with the wGRR (as detailed in §3.10.1).

***bla*$_{OXA-48}$** was found in 309 plasmids (11 from the NLSAR collection). As expected, the plasmids carrying this gene were highly similar to each other, independently of the source or bacterial background (average wGRR=0.85, range 0.037 -1). We found 363 plasmids similar to those collected in the NLSAR carrying *bla*$_{OXA-48}$ (pairwise wGRR above 0.85) (Figure 4.11 A-B), and most of them were conjugative (n=343), IncM/L (n=347), and carried the *bla*$_{OXA-48}$ (n=287) (Figure 4.11 E-G). These features are in line with what has already been reported for the traditional pOXA-48a plasmid, which is highly conserved but can lose the *bla*$_{OXA-48}$ gene. These plasmids were encoded in 62 different bacterial STs and mainly found in Spanish genomes (n=219) (Figure 4.11 C-D). In fact, the OXA-48 carbapenemase has been reported as the principal carbapenemase in Spain [146, 366, 438]. Although this is a very conserved plasmid, there were 8 different groups delineated by a wGRR > 0.9. We found NLSAR isolates with plasmids in 4 of the 8 groups, which indicates that slightly different variants of this plasmid were present in the NLSAR collection. These variations were mainly due to structural changes, especially gains/losses of genes.

***bla*$_{VIM}$** genes were detected in 36 plasmids of the custom database. These plasmids showed a low average wGRR (0.21, range 0 - 1) and differed in major plasmid features such as length, GC content, incompatibility types (Incl/M, IncR, IncC, IncA, ...), mobility, and resistance carriage. This suggests a highly diverse plasmid background of this gene. When comparing the two *bla*$_{VIM}$ plasmids found in the NLSAR collection with all the databases (pairwise wGRR above 0.8) (Figure 4.12 A-G), we observed that both plasmids from the NLSAR collection were highly different to each other (wGRR=0.17).

The plasmid extracted from isolate HLF1_65 (Group a) belonged to a mobilizable IncP type carrying AMR genes *bla*$_{VIM-1}$, *aadA1*, *aac(6')-Ib4*, *dfrB1*, *sul1*, and *catB2*, and was very different to the rest. Although we did not find any similar plasmid, IncP plasmids carrying *bla*$_{VIM}$ genes have been reported previously in *P. aeruginosa* strains [439, 440]. In fact, mob typing of this plasmid showed a multi phyla predicted host range, primarily on *Pseudomonadota*, which may suggest an interspecies transfer of the plasmid.

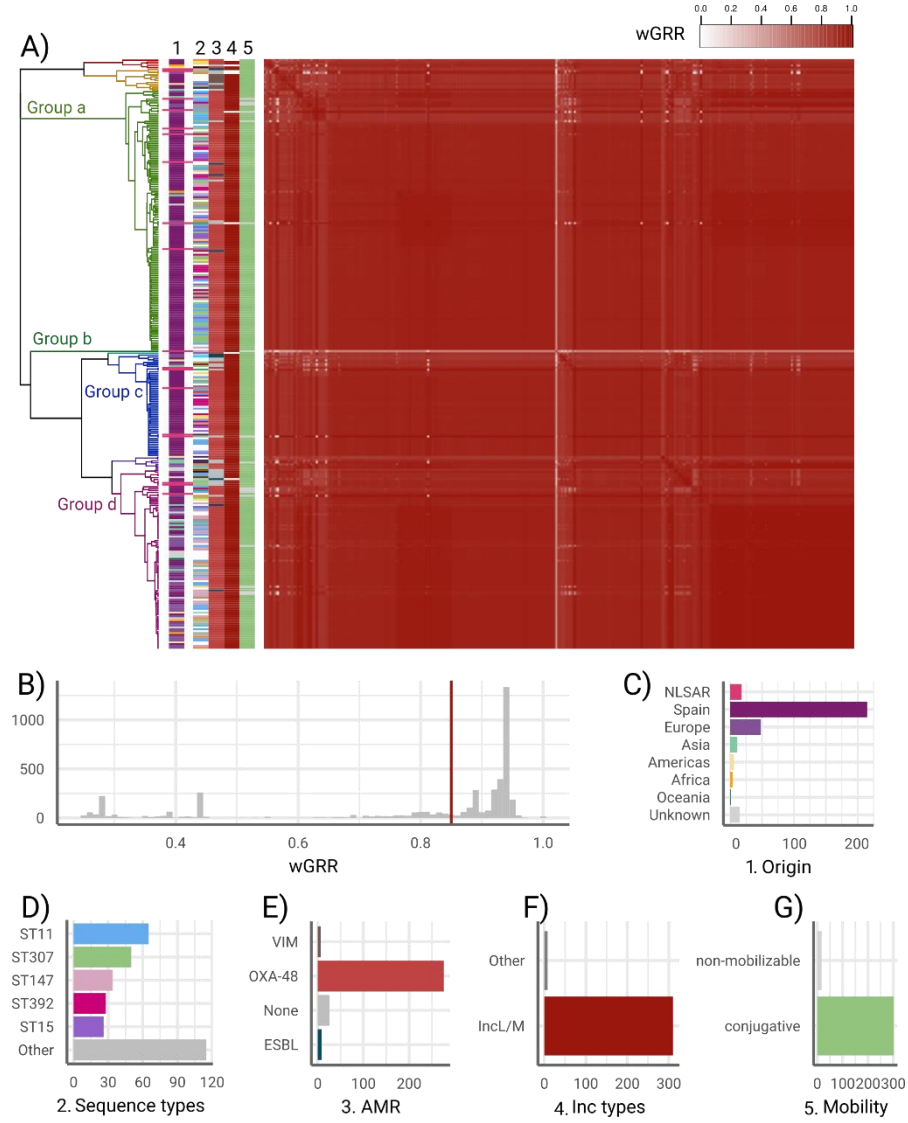


Figure 4.11. A) Heatmap and dendrogram representing the wGRR values between the plasmids similar to those in the NLSAR collection carrying the *bla*$_{\text{OXA-48}}$ gene (wGRR similarity above 0.85, marked as a red line in B). The groups (coloured branches) in the dendrogram are delineated by a wGRR value of 0.9. The columns indicate, from 1 to 5, the origin of the genomes, the ST of the bacteria carrying the plasmid, the AMR genes present in the plasmid, the replicon type, and the predicted mobility for each plasmid. B) Distribution of wGRR pairwise distances between NLSAR isolates carrying the *bla*$_{\text{OXA-48}}$ gene and the rest of the custom database. Only plasmids with similarity values above 0.2 are shown. The data summarised in plots C-G is detailed for each plasmid in A (columns 1-5).

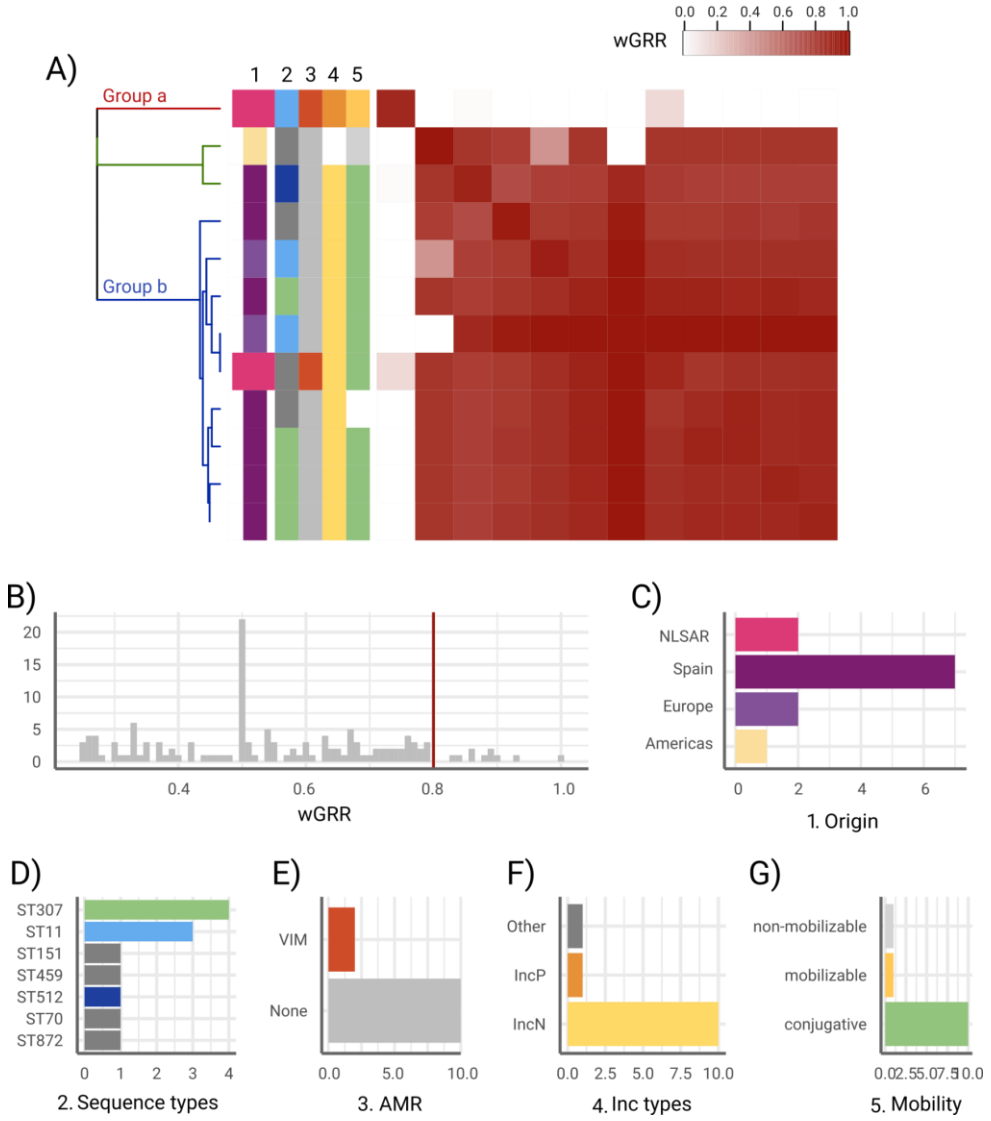


Figure 4.12. A) Heatmap and dendrogram representing the wGRR values between the plasmids similar to those in the NLSAR collection carrying the *bla*$_{VIM}$ gene (wGRR similarity above 0.8, marked as a red line in B). The groups (coloured branches) in the dendrogram are delineated by a wGRR value of 0. The columns indicate, from 1 to 5, the origin of the genomes, the ST of the bacteria carrying the plasmid, the AMR genes present in the plasmid, the replicon type, and the predicted mobility for each plasmid. B) Distribution of wGRR pairwise distances between NLSAR isolates carrying the *bla*$_{VIM}$ gene and the rest of the custom database. Only plasmids with similarity values above 0.2 are shown. The data summarised in plots C-G is detailed for each plasmid in A (columns 1-5).

The other plasmid, extracted from genome KP_HGUA03_124, was found in a highly related cluster formed by 9 IncN and conjugative plasmids (Group b). Contrarily to IncP, IncN plasmids have been widely reported associated with *bla*$_{VIM}$ genes in *K. pneumoniae* [441, 442]; however, no member of the cluster, except for the NLSAR plasmids, has been found to carry the *bla*$_{VIM}$ gene.

None of the 34 VIM plasmids of the custom database was similar to those found in the NLSAR collection, which suggests that the plasmid backbone of the *bla*$_{VIM}$ genes in the NLSAR collection is different to those found in the closed genomes of the global collection.

A total of 8 different variants of ***bla*$_{NDM}$** genes were identified in 125 plasmids. These plasmids showed a low overall wGRR similarity (average 0.15, range 0-1). They belonged to more than 30 different Inc types or combinations. The most frequent ones were the combination of IncFIB + IncHI1B (n=12), IncX3 (n=11), IncFIB+ IncFII (n=9), IncC (n=8), and IncFII (n=8). When comparing the 5 NDM plasmids collected in the NLSAR collection with the whole plasmid database (wGRR> 0.7) we found them in 5 of the 11 groups delineated by a wGRR value of 0.85 (Figure 4.13 A-G).

We found one NLSAR plasmid in Group a (HLF2_47). This group included mostly conjugative IncFII plasmids of around 160 Kb with a high number of AMR genes in addition to *bla*$_{NDM-1}$, (*bla*$_{TEM}$, *aph(6)-Id*, *aac(6')-Ib*, *catB3*, *bla*$_{OXA-1}$, *bla*$_{CTX-M-15}$, *aac(6')-Ib-cr5*, *rmtF*) and metal-resistance genes (*sil*, *pco*, *ars*). All of them were collected in Spain and belonged to ST101 (Figure 4.13 A).

In Group b we found two NLSAR plasmids (from genomes 146KP-HG and 179KP-HG) and it included non-mobilizable non-typable plasmids with an average length of 90 Kb (79,857 - 97,784 bp) carrying *bla*$_{TEM-1}$, *bla*$_{OXA-1}$, *bla*$_{CTX-M-15}$, *dfrA27*, *aadA16*, *aadA1*, *aph(6)-Id*, *aph(3'')-Ib*, *aac(6')-Ib-cr5*, *sul1*, *sul2*, *catB3*, *tet(A)*, *qnrB1*, and *arr3*. In this group, we found plasmids carrying *bla*$_{NDM-1}$ and *bla*$_{NDM-23}$. These plasmids were collected mainly in Europe and Spain and most of them were extracted from bacteria belonging to ST437 (Figure 4.13 A). A detailed study of these plasmids is undertaken in Chapter 7.

NDM plasmids extracted from the NLSAR HLF1_29 genome were similar to only one plasmid from the database (Group c). These plasmids were extracted from ST11 bacteria in Europe; however, they showed different genetic content (Figure 4.13 A).

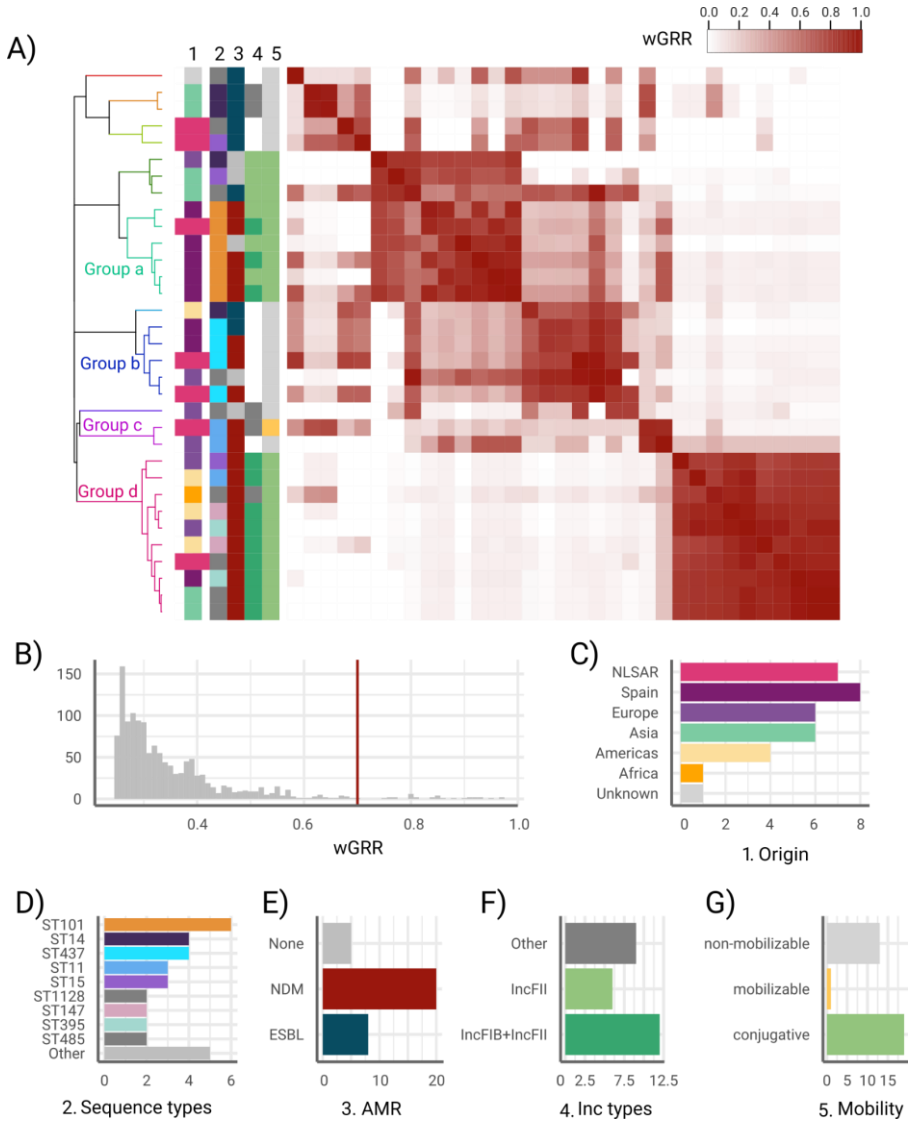


Figure 4.13. A) Heatmap and dendrogram representing the wGRR values between the plasmids similar to those in the NLSAR collection carrying the *bla*$_{NDM}$ gene (wGRR similarity above 0.7, marked as a red line in B). The groups (coloured branches) in the dendrogram are delineated by a wGRR value of 0.85. The columns indicate, from 1 to 5, the origin of the genomes, the ST of the bacteria carrying the plasmid, the AMR genes present in the plasmid, the replicon type, and the predicted mobility for each plasmid. B) Distribution of wGRR pairwise distances between NLSAR isolates carrying the *bla*$_{NDM}$ gene and the rest of the custom database. Only plasmids with similarity values above 0.2 are shown. The data summarised in plots C-G is detailed for each plasmid in A (columns 1-5).

The NLSAR plasmid was an IncFIA and IncR mobilizable plasmid carrying AMR genes (*bla*$_{NDM-1}$, *bla*$_{OXA-1}$, *bla*$_{CTX-M-15}$, *aac(6')-Ib-cr*, *aac(3)-IIe*, and *dfrA14*) and mercury-resistance genes (*mer*).

The last NLSAR NDM plasmid (extracted from HLF2_47) was found in Group c, a cluster with high heterogeneity regarding origin and ST, but not in their gene content (Figure 4.13 A). All the plasmids had IncFIB and IncFII replicons, were conjugative, and all of them carried the carbapenemase gene *bla*$_{NDM-1}$ and the 16S rRNA methyltransferase gene *rmtC*, which confers resistance to all aminoglycosides.

We found many different NDM plasmids in the NLSAR collection, in line with the large variability of plasmids carrying this gene, as previously reported [172].

***bla*$_{KPC}$** genes were highly frequent in the global collection but rare in the NLSAR collection. Hence, from the 505 plasmids carrying *bla*$_{KPC}$ genes only two were from the NLSAR collection. These plasmids were found in many different backgrounds (IncFIB, IncFII, IncFIA, IncR, IncP, IncU, IncX3) but not in different bacterial backgrounds as most of them were found in CG258 (n=367). When we compared the two NLSAR KPC plasmids with the whole database, we observed different groups (clustered at wGRR >0.85) (Figure 4.14 A-G). Although in different groups, the plasmids carrying *bla*$_{KPC}$ genes showed highly similar features such as replicon types, GC content, or AMR genes. The plasmid extracted from NLSAR genome HLF2_108 (ST512) fell in Group a, a large cluster with highly similar characteristics to those reported for the IncFIB pKpQIL plasmids associated to the KPC epidemic in Europe and America [197, 443]: 53.93% CG content, 113 Kb size, the replicon types, and the conjugative system and the gene content (*bla*$_{KPC}$, *bla*$_{TEM-1}$, truncated *bla*$_{OXA}$, merC).

The other plasmid, extracted from an ST258 NLSAR genome (HLF1_17), fell into a different cluster, Group b (Figure 4.14 A). This plasmid was larger (201 Kb) and showed a different gene content: in addition to the *bla*$_{KPC}$, *bla*$_{TEM-1}$, and the truncated *bla*$_{OXA}$, this plasmids also carried *aadA2*, *sul1*, *catA1*, *dfrA12*, *mph(A)* and *aph(3')-Ia* AMR genes.

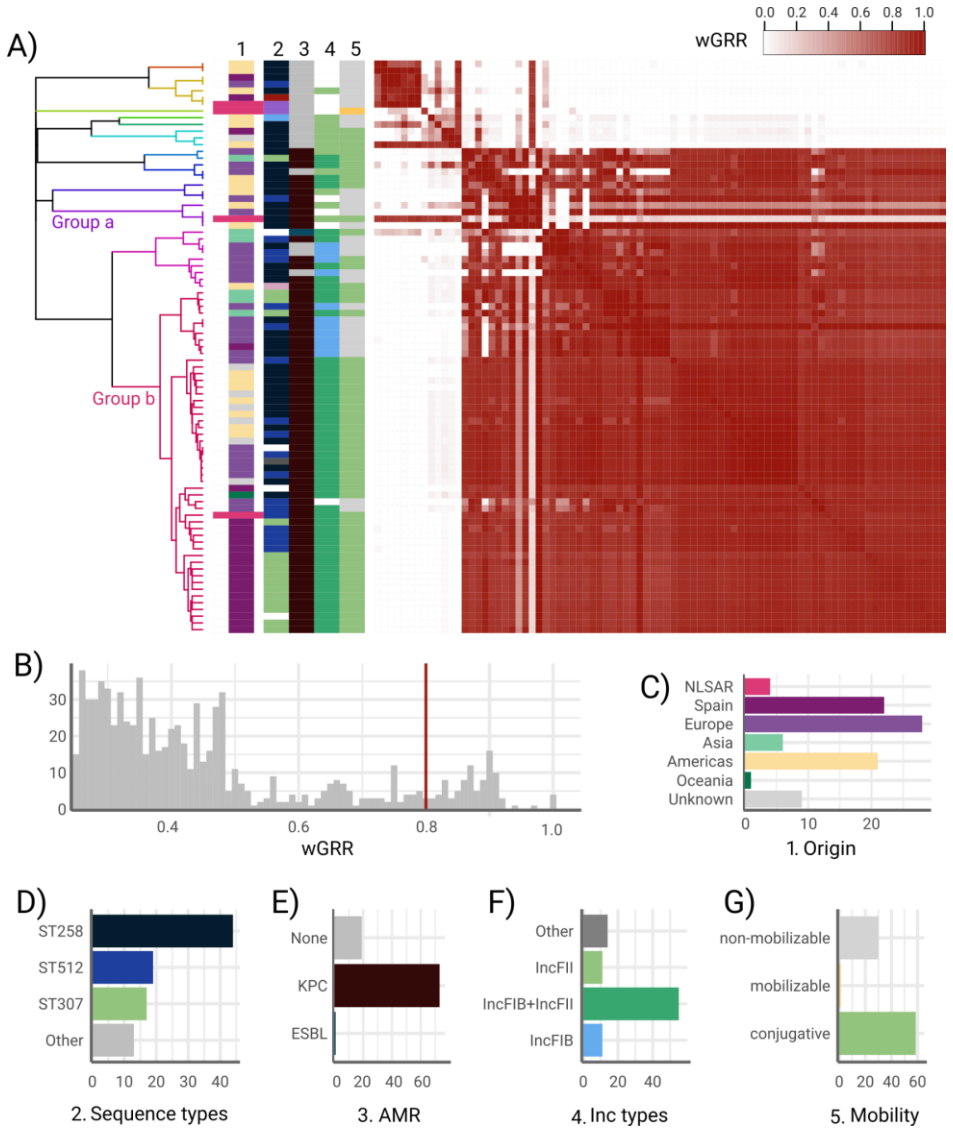


Figure 4.14.. A) Heatmap and dendrogram representing the wGRR values between the plasmids similar to those in the NLSAR collection carrying the *bla*$_{KPC}$ gene (wGRR similarity above 0.8, marked as a red line in B). The groups (coloured branches) in the dendrogram are delineated by a wGRR value of 0.85. The columns indicate, from 1 to 5, the origin of the genomes, the ST of the bacteria carrying the plasmid, the AMR genes present in the plasmid, the replicon type, and the predicted mobility for each plasmid. B) Distribution of wGRR pairwise distances between NLSAR isolates carrying the *bla*$_{KPC}$ gene and the rest of the custom database. Only plasmids with similarity values above 0.2 are shown. The data summarised in plots C-G are detailed for each plasmid in A (columns 1-5).

4.5 Concluding remarks

In this chapter, we have demonstrated the importance of performing a careful examination regardless of the origin of the data. These filters allowed us to obtain a local collection of almost 2,000 *K. pneumoniae* genomes free of contaminants and bad-quality assemblies. We also applied some of these filters to obtain a global collection of more than 10,000 database genomes. When comparing the two collections, we found that the NLSAR lineage composition was very diverse and similar to that of Spanish genomes and that most of the HRC in the NLSAR isolates were found to be related to isolates collected in other regions, which suggests similar evolutionary histories. This is highly valuable as it will allow the investigation of possible links and transmission events between the Comunitat Valenciana and other regions of the globe. Moreover, when looking for resistance determinants and plasmids, most of the carbapenemases and ESBLs found in the NLSAR collection have already been reported elsewhere as well as the plasmids that carried them. However, we found some exceptions. The NDM-23 enzyme and the plasmid that carried it were only found in the NLSAR collection or in Spanish genomes, and the ST11 lineage co-carrying OXA-48 and CTX-M-1 enzymes has been hardly reported outside the NLSAR collection or Spain. In addition, we did not find similar plasmids in the custom plasmid database to those carrying *bla*$_{VIM}$ genes and a *bla*$_{KPC}$ gene.

CHAPTER 5
Genomic analysis of AMR *K. pneumoniae* in the CV

Spain has been identified as one of the European regions with the highest rates of AMR in *K. pneumoniae* [444], particularly to 3GC and carbapenems due to the production of ESBLs and carbapenemases, respectively. The first ESBL-producing *K. pneumoniae* in Spain was isolated in 1988 in a hospital in Madrid [445], whereas the first detection of a carbapenemase-producer was reported in 2003 [446]. Since then, resistance rates have increased dramatically [444]. In 2022, the reported resistance rates to 3GC reached 40% of the total *K. pneumoniae* hospital isolates, whereas carbapenem resistance rates reached 5% [100].

Several studies have investigated the nature of *K. pneumoniae* ESBL- and carbapenem-resistance in Spain, revealing a high rate of SHV-12, CTX-M-15, CTX-M-1 [447, 448], and OXA-48 [151, 152, 366], as well as a wide variety of other resistance determinants, including other carbapenemases such as VIM, KPC, or NDM [151, 152]. Furthermore, a nationwide study found that the ST307, ST11, and ST15 lineages were responsible for dispersing these AMR determinants in Spain [156].

To gain insights into the dynamics and evolution of the AMR *K. pneumoniae* population in the Comunitat Valenciana, the SKPCV was established, and 1,604 *K. pneumoniae* isolates were collected from eight different hospitals. This collection provided a detailed overview of the AMR population in each hospital and allowed us to study the dynamics within the region and transmission between hospitals. Additionally, information obtained from over 11,000 genomes in public databases and 400 genomes collected earlier in the same hospitals (NLSAR) helped to elucidate possible origins and to track global outbreaks.

5.1 AMR rates of *K. pneumoniae* in the Comunitat Valenciana.

To contextualise the SKPCV in the epidemiological context of AMR *K. pneumoniae* in the CV, we studied its prevalence and evolution for the last decade. For this, we interrogated the RedMIVA database for all entries of *K. pneumoniae* AST performed from 2010 to 2022 in the CV (methods detailed in §3.2). In total, 131,120 infection episodes corresponding to 93,785 patients, were reported. Patients were distributed in 34 different hospitals: 5 from Castellón, 16 from Valencia, and 12 from

Alicante (Figure 5.1 A-B, Supplementary Table 5.1). Only 6.7% of the patients suffered several non-susceptible episodes, suggesting few cases of reinfection, or relapses.

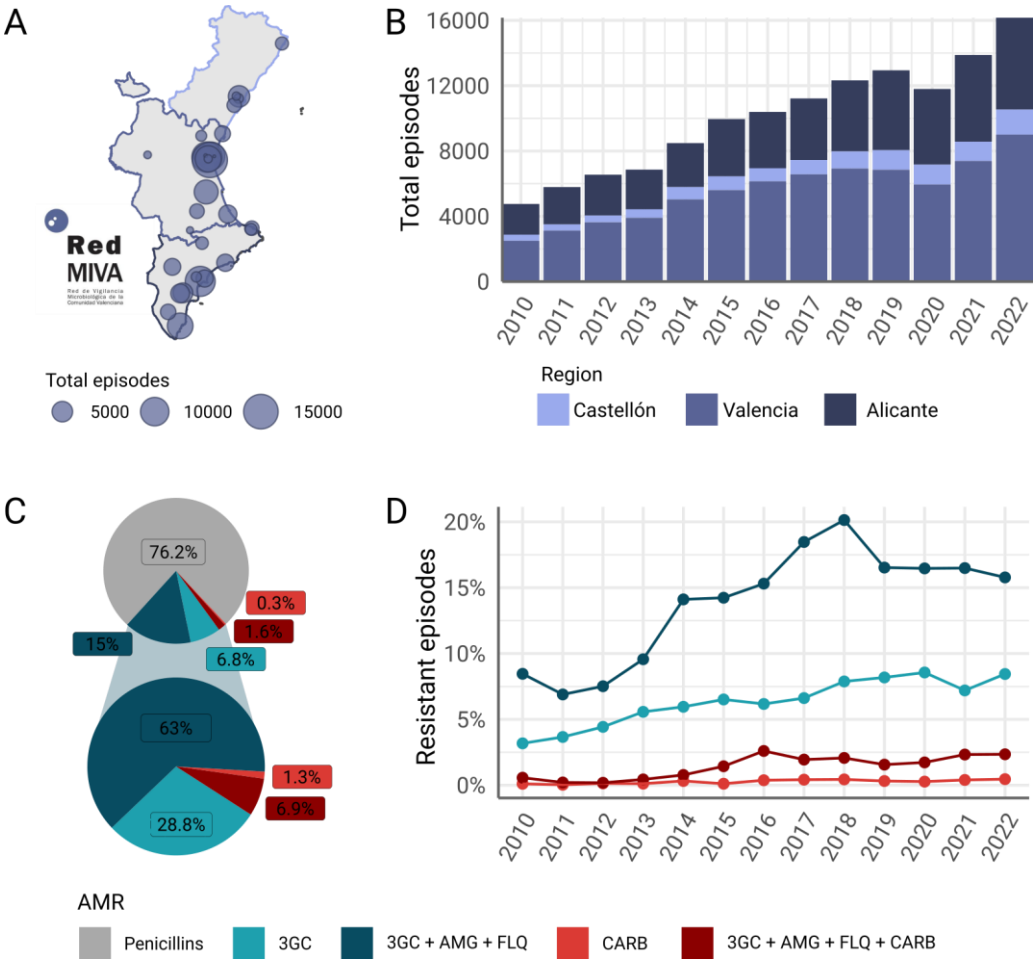


Figure 5.1 A) Map of the Comunitat Valenciana with the number of reported episodes per hospital included in the RedMIVA database. B) Number of *K. pneumoniae* episodes for each year and in each province. C) Total cases admitted or not to the hospital per resistance type. D) *K. pneumoniae* resistant episodes for each of the three provinces among the 10-year period.

During the study period, most infection episodes were caused by antimicrobial-susceptible strains (Figure 5.1 C). The most prevalent resistance pattern involved 3GC-resistance combined with fluoroquinolone- and aminoglycoside-resistance, or

3GC-resistance alone, reaching a prevalence of over 27% (Figure 5.1 D). Fortunately, during the last four years (2019 to 2022), there was a reduction in the 3GC non-susceptibility rates, which have remained stable at around 21%. Nevertheless, the high proportion of 3GC-resistant isolates likely led to an increased use of carbapenems, with carbapenem-resistance episodes increasing shortly after the increase in 3GC-resistance. However, the prevalence of carbapenem-resistance remained low (~2.5%) throughout the period. Before 2013, reported carbapenem non-susceptibility episodes were limited to a few sporadic cases. It is important to note that in the early 2010s, carbapenem-resistance was not considered a threat in the region and, therefore, was not analysed routinely. With the increase in the number of episodes, the screening effort intensified, and it is possible that the number of episodes during those years was underestimated.

During the SKPCV study period (2017 to 2019), the prevalence of non-susceptibility rates to 3GC reached a peak of 27.5% of the episodes, while carbapenem-non-susceptibility remained stable at around 5%. This prevalence is similar to that reported in Spain during those years [10, 12].

Using the RedMIVA database, we were also able to assess whether having an episode of resistant *K. pneumoniae* was a risk factor for hospitalisation in the region. For this, we estimated the odds ratio (OR) for each resistance group using a Fisher test (§3.2.1). All episodes with AMR infections had a significant, positive OR (Supplementary Table 5.2). The odds of being hospitalised for patients with carbapenem-resistant episodes were 5.09 times higher than that of those with susceptible episodes. Remarkably, these were higher than for patients with MDR carbapenem episodes (4.73, range 4.13 - 5.45). Patients with susceptible episodes were negatively correlated to hospitalisation (0.47, range 0.46 - 0.48).

5.1.1 AMR rates in the SKPCV hospitals

The prevalence of *K. pneumoniae* AMR episodes differed across the 8 hospitals that participated in the SKPCV project (Figure 5.2). While resistance to 3GC, fluoroquinolones, and aminoglycosides was most prevalent in all hospitals, the rates varied significantly both over time and between hospitals. Some hospitals

reported resistance peaks reaching up to 40% of AMR episodes (such as the HGUV and the HULR), whereas in the remaining hospitals, the maximum prevalence reached 20% (in the HAV and the HUPLF). However, prevalence was not constant through time, as all the hospitals showed peaks, indicating possible ongoing outbreaks. Fortunately, in 2022, resistance rates had reduced to under 20% in all the hospitals, except for the HGUV.

Carbapenem-resistance rates were generally low in all the hospitals (Figure 5.2). However, most hospitals experienced peaks of prevalence at some point, such as the HGUA, HGUV, and HUPLF in 2016. Of concern, some hospitals, including the HGUV and the HCUV, showed an increase in carbapenem resistance rates in 2022.

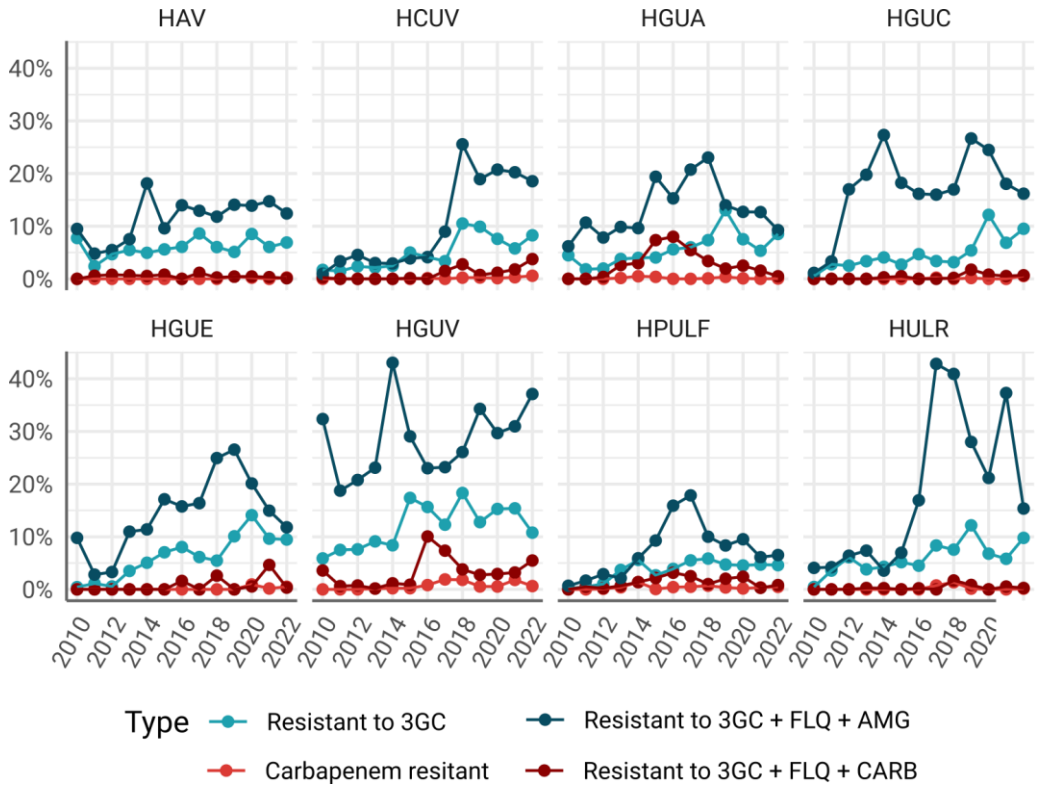


Figure 5.2. Results from the RedMIVA database for the hospitals included in the SKPCV surveillance program.

5.2 The SKPCV project

Under the SKPCV project, a total of 1,604 *K. pneumoniae* isolates were collected from 8 different hospitals between January 2017 and February 2020 (Figure 5.3 A-B). Clinical data was available for approximately 1,481 isolates (depending on the variable). Missing information was mainly due to HGUA and spurious samples from each hospital (Supplementary Table 5.3).

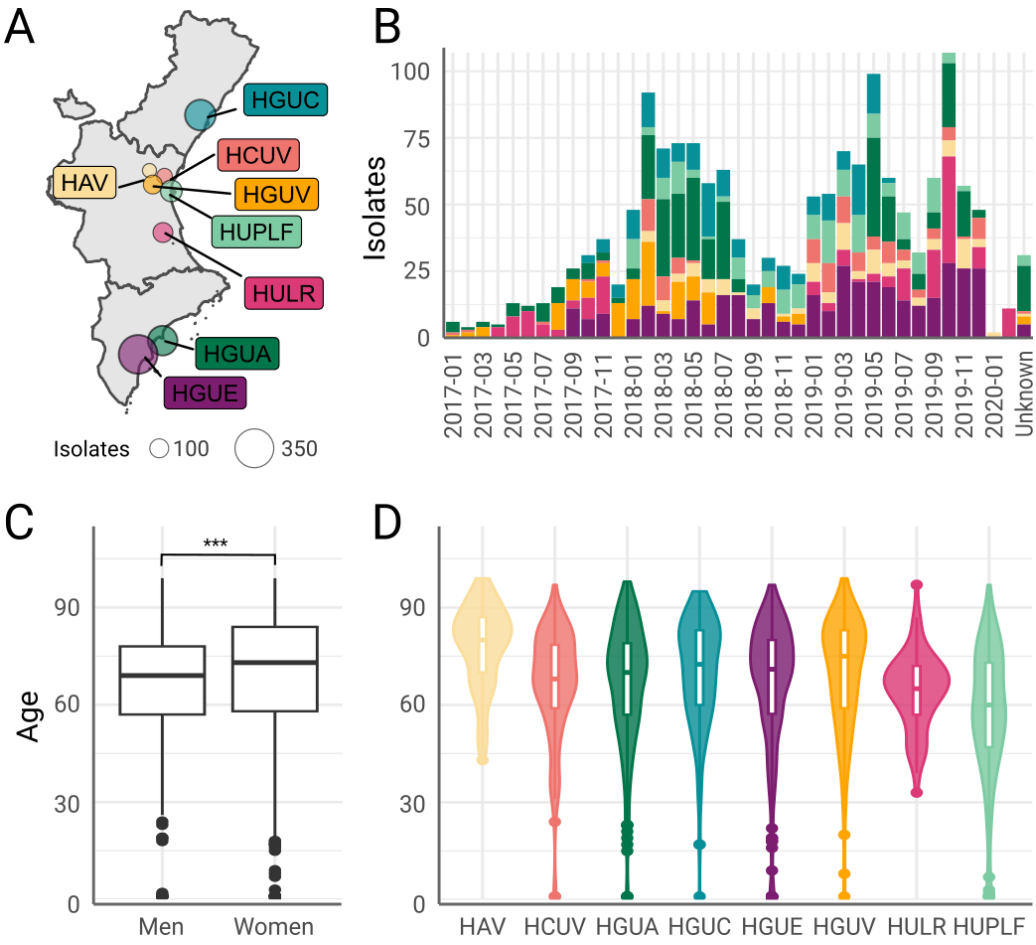


Figure 5.3. A) Map representing the hospitals included in the SKPCV. B) Distribution of samples during the collection period coloured by hospitals (coloured as A). C) Differences in the age of females and males, wilcox.test. D) Distribution of patient ages by hospital.

Patients were mainly elderly, with an average age of 71.3 years (range 0 to 94; IQR 57-81), and women were significantly older than men (69.8 vs 66.2, p-value = 1.233e-06) (Figure 5.3 B). The age distribution of patients varied among the hospitals. Some hospitals, such as HGUA, HPLF, and HGUE, had a significant number of neonatal patients. On the contrary, HAV and HULR did not collect any isolates from patients younger than 30. The observed differences in age distribution can be attributed to the type of hospital and, therefore, the wards they have, rather than actual epidemiological differences.

Isolates were obtained from infections (n=693) and surveillance cultures (n=870). The most frequent *K. pneumoniae* infection type was urine tract infections (UTI) (n=192) followed by bloodstream infections (n=149).

5.2.1 Antimicrobial susceptibility

As expected, due to the selection criteria, most of the strains were resistant to 3GC (80%). Yet, we found high susceptibility to other cephalosporins, such as cefoxitin (89.1%), to the three carbapenems, ertapenem (73.7%), imipenem (90.5%) and meropenem (87.1%), and, also, to other front-line antibiotic classes such as fosfomycin, (60.8%), amikacin (90.6%), and tigecycline (88.2%). Although tigecycline was only tested in 41% of the strains, we only found a few resistant isolates (Figure 5.4 A). Despite the relatively low levels of resistance to individual antibiotics observed, most of the strains were MDR (n=1,331). The non-MDR corresponded mostly to the control strains.

The distributions of carbapenem- and 3GC-resistant isolates were different among hospitals (Figure 5.4 B). In general, the combined resistance to 3GC, fluoroquinolones, and aminoglycosides was the most prevalent in all the hospitals (Figure 5.4 B), as seen with the RedMIVA data (Figure 5.2). And, although the collection was enriched in carbapenem-resistant isolates, these represented a small proportion in most hospitals with the exceptions of the HGUV, HULR, and HUPLF.

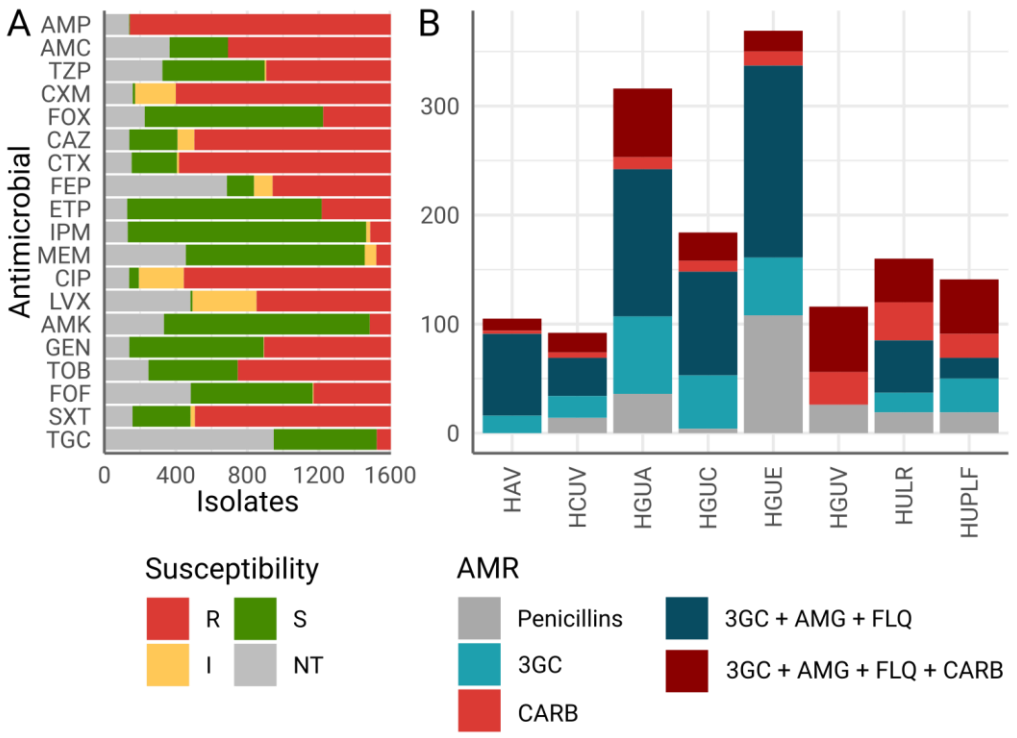


Figure 5.4. A) Distribution of SKPCV isolates categorised as resistant, intermediate and susceptible by the EUCAST recommendations, for each antibiotic tested: ampicillin (AMP), amoxicillin-clavulanic acid (AMC), piperacillin-tazobactam (TZP), cefuroxime (CXM), cefoxitin (FOX), ceftazidime (CAZ), cefotaxime (CTX), cefepime (FEP), ertapenem (ETP), imipenem (IPM), meropenem (MEM), ciprofloxacin (CIP), levofloxacin (LVX), amikacin (AMK), gentamicin (GEN), tobramycin (TOB), fosfomycin (FOF), trimethoprim-sulfamethoxazole (SXT), and tigecycline (TGC). B) Number of Isolates collected in each hospital coloured by the AMR pattern: 3GC-resistance, carbapenem-resistance (CARB) or both combined with aminoglycosides and fluoroquinolones resistance.

5.3 3GC- and carbapenem-resistant population structure and determinants

We used genomic tools to characterise the 1,604 strains of the SKPCV collection. Isolates were characterised using MLST, cgMLST, surface antigens (KL and o types), AMR genes, virulence determinants, and plasmid content.

By applying the criteria established by [408] (detailed in §3.9.4) to use the cgMLST scheme to delineate populations we grouped the 1,604 genomes into 178 different sublineages (SLs) (Supplementary Figure 5. 1), which corresponded to 132 known CGs and 188 unique STs. Of the 178 SLs, more than half (95 isolates, 53.3%) were singletons (Supplementary Figure 5. 1) and 28.5% of them were non-susceptible to 3GC or carbapenems. Only 24 SLs were found with at least 5 samples (Figure 5.5). The most frequently found STs were those represented by sequence types ST307, ST11, ST15, ST147, ST405, and ST219.

Resistant isolates were not equally distributed among STs (Figure 5.5). **3GC-resistant strains** (n=1,316) were found in most of the STs (54%) whereas **CRKp** (n=371) were found in 38 STs (20%) (Figure 5.5). However, only 8 STs included more than 5 carbapenem non-susceptible isolates. These STs accounted for 87% of the CRKp and 91% of the total carbapenemase-producing strains (n=325). Of note, to reduce the effect of missing data, we included in the 3GC- and carbapenem-resistant groups those isolates without AST information but detected as ESBL or carbapenemase gene carriers.

The 3GC- and carbapenem-resistant determinants and the corresponding carrier lineages were found to be distinctly distributed among hospitals (Figure 5.6).

bla$_{\text{CTX-M-15}}$ was the ESBL gene most frequently detected (n=981) (Figure 5.6). It was found in every hospital, in different bacterial clonal backgrounds (n=53 STs), and in combination with many other AMR determinants such as *bla*$_{\text{OXA-48}}$ (n=130), *bla*$_{\text{SHV-76}}$ (n=14), *bla*$_{\text{NDM}}$ variants *bla*$_{\text{NDM-1}}$ and *bla*$_{\text{NDM-23}}$ (n=15), *bla*$_{\text{DHA-1}}$ (n=2) and *bla*$_{\text{KPC-3}}$ (n=1). Despite its diversity, this gene was strongly associated with lineage ST307. This lineage carrying *bla*$_{\text{CTX-M-15}}$ was found to be highly frequent in all the SKPCV hospitals (above 21% of all the isolates collected in each hospital). The other STs found to carry *bla*$_{\text{CTX-M-15}}$ were less widely distributed, with high incidence of specific gene-ST combinations in different hospitals. For instance, when related to *bla*$_{\text{CTX-M-15}}$, ST15, which was the third most frequently found lineage, was mainly found in the HGUE; with ST29, it was found in the HULR; with ST902, it was restricted to the HGUC; and with ST33, to the HULR.

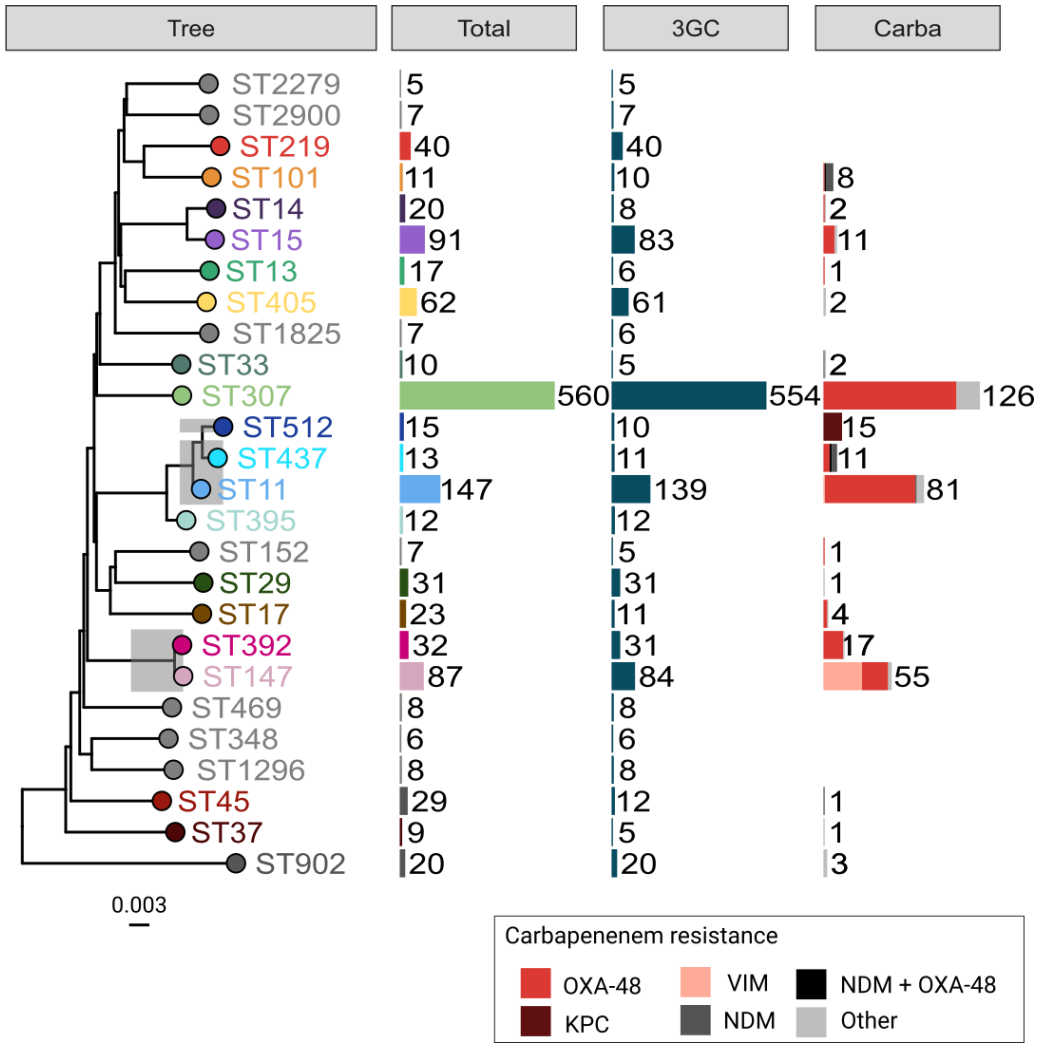


Figure 5.5. A simplified version of the maximum likelihood tree of the 1604 genomes of the SKPCV using the core-genome (90%) alignment (Supplementary Figure 5. 1). The tree has been simplified by showing only those lineages with more than 5 isolates and collapsing sublineages (SLs) and STs. The grey background represents the same SL. Each branch is labelled with the most prevalent ST present in the SL. The bar plots indicate i) the total number of isolates of each SL; ii) the number of them resistant to 3GC and iii) resistant to carbapenems. The latter is coloured by the mechanisms of resistance (carbapenemase or others).

In addition to *bla*$_{CTX-M,}$ we found other ESBL genes, such as *bla*$_{SHV-12}$ (n=54), and *bla*$_{SHV-2}$ (n=8). We also found the AmpC encoding gene *bla*$_{DHA-1}$ in 37 isolates, all belonging to ST11, but in different hospitals. However, when found together with the *bla*$_{CTX-M-3}$ gene, it was mostly restricted to the HGUC.

Regarding the mechanism of carbapenem resistance, acquired carbapenemases were the primary mechanism with 88% of the total (Figure 5.6). The *bla*$_{OXA-48}$ was the most prevalent carbapenemase gene (n=257) and it was found in combination with five different ESBLs (CTX-M-15, -1, -14, -9, and SHV-12) and in different genetic backgrounds (21 STs) being the most prevalent ST307 (n=107), ST11 (n=73), ST147 (n=21), ST392 (n=16), ST15 (n=9), and ST437 (n=7). Most of these were widely distributed in the hospitals; yet, when the *bla*$_{OXA-48}$ gene was found in combination with *bla*$_{CTX-M-9}$ and *bla*$_{SHV-12}$, it was always found in lineage ST147 and at the HUPLF hospital.

The next carbapenemase most frequently found was *bla*$_{VIM}$ (n=37), mainly associated with ST147 (n=31, 83.7%) and the HGUA (n=31, 83.7%), individually or in combination with *bla*$_{CTX-M-19}$ (n=18) and *bla*$_{SHV-12}$ (n=22). We found this gene in 6 other STs but only in three other hospitals, the 4 isolates in the HUPLF, one isolate in the HCUV and another one in the HGUC.

The *bla*$_{KPC-3}$ gene was detected in 18 samples, always without ESBL, and primarily in the HUPLF (n=15) and restricted to CG258 (94.4%): ST512 (n=15) and ST258 (n=2).

Lastly, we found two variants of *bla*$_{NDM}$, *bla*$_{NDM-1}$ (n=10) and *bla*$_{NDM-23}$ (n=8) both in all forms of combination with *bla*$_{CTX-M-15}$ and *bla*$_{OXA-48}$. They were found in 5 STs, mostly in ST101 (n=7) and ST437 (n=6), and generally in the HGUV and the HUPLF.

As shown previously, regardless of ST307 carrying *bla*$_{CTX-M-15}$, the distribution of the other 3GC- and carbapenem-resistance determinants and the lineages carrying them was different among the hospitals (Figure 5.6). Moreover, most STs and AMR determinant combinations were found mostly limited to one hospital. Therefore, we can conclude that we have a mixed situation in the regional hospitals, with lineage ST307 and genes *bla*$_{CTX-M-15}$ being widely distributed, whereas the rest of the

3GC- and carbapenem-resistance burden being caused by different local gene-lineage combinations in each hospital.

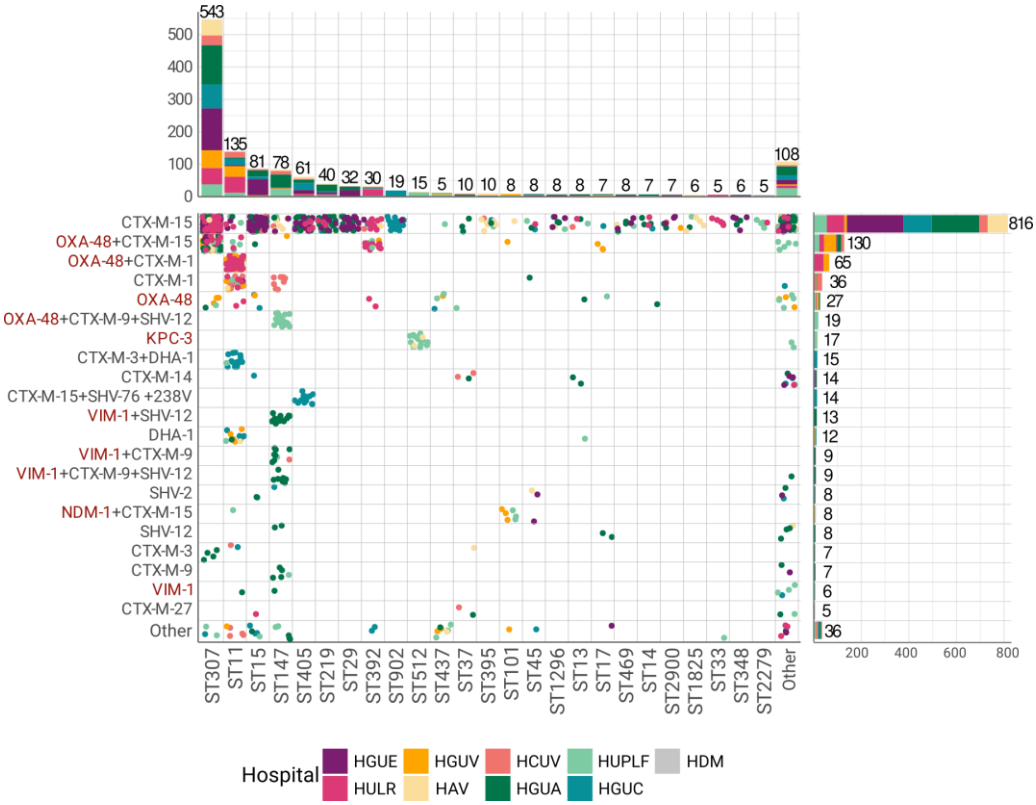


Figure 5.6. Distribution of genomes depending on the ST and the AMR determinants to 3GC or carbapenems. Each dot represents a genome and is coloured by the source hospital. Labels corresponding to carbapenemase genes are highlighted in red.

5.4 AMR determinants and lineage diversity within and among hospitals

To test how similar the populations were among hospitals, we estimated the genetic diversity within and between hospitals by calculating the nucleotide diversity (k), and the nucleotide diversity per site within (π) and between hospitals (Dxy). We also calculated the net between-hospital divergence (Da). The higher the Da values, the larger the differences between the populations compared. Diversity formulae are detailed in §3.9.5. For these analyses, we used the core genome alignment (see

§3.9.1) of the 1,604 SKPCV isolates that resulted in a total of 3.9 Mb of alignment length with 319 Kb variant positions.

Our results showed that the total within-hospital diversity was very high in all the hospitals (Dx average 19,930 SNPs, range 18,124 - 23,458, $\pi \simeq 0.005$) (Supplementary Table 5.4), which contrasts with low values of the net interhospital divergence (Da average 653 SNPs, range 86 - 1,472) (Supplementary Table 5.4).

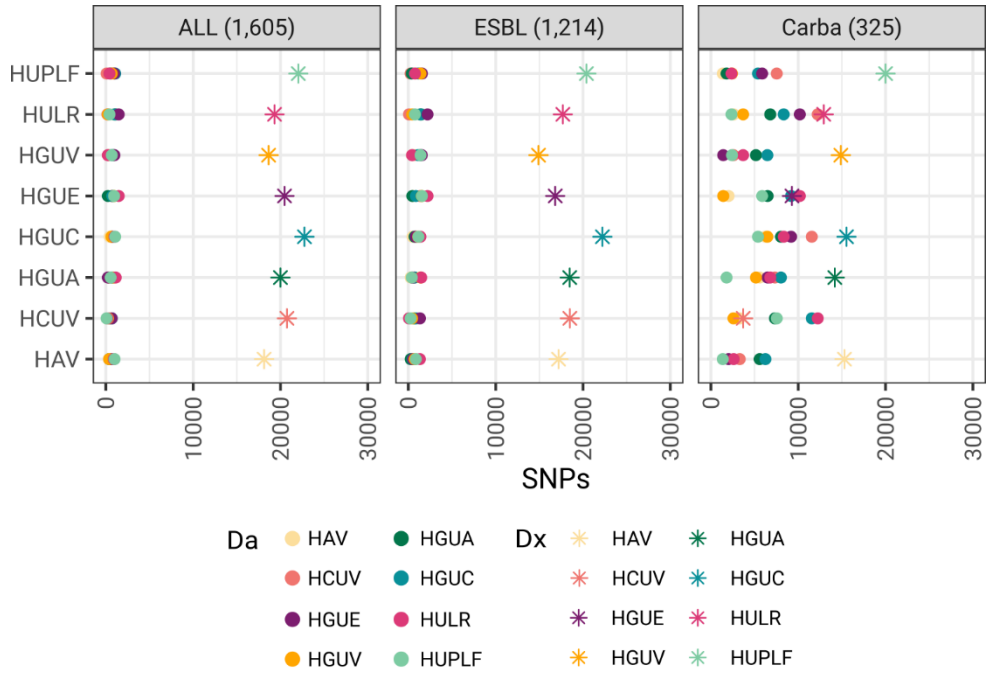


Figure 5.7. Genetic diversity, shown as SNPs, of SKPCV isolates within and between hospitals. For each hospital, the within diversity is represented as asterisks, and the net between hospital divergences (Da) is represented in circles. From left to right, the different blocks show the estimates for the whole population, the 3GC non-susceptible population, and the carbapenem-resistant population.

As most isolates were ESBL producers, the diversities for this group of isolates showed a very similar distribution to that of the total, with high values of intra-hospital diversity and low values of the net divergence between hospitals. The carbapenemase-producing population also showed high levels of within-hospital diversity (Dx average 13,224, range 3,698 - 19,997) but, contrarily to the overall and the ESBL-producing population, it also showed high levels of net divergence

between hospitals (Da average 5,192 SNPs, range -302 to 12,235) (Figure 5.7, Table 5.1). Moreover, in this case, we observed a large difference in the within diversity among hospitals: whereas the HUPLF had the largest within diversity (19,997 SNPs) others, such as the HCUV, showed much lower values (3,698 SNPs). The Da estimates also uncovered possible similarities between hospitals. For instance, the HAV showed very low values of divergence with the HGUV, and the HGUE with the HCUV. However, these similarities were not explained by geographic location or interhospital transmission, but because the burden of the carbapenem-resistant isolates is very low in these hospitals, and they only reported 6 (HAV) and 8 (HGUE) isolates with carbapenemase genes.

Table 5.1. Genetic diversity estimated for the carbapenem non-susceptible population. The lower hemi matrix shows the nucleotide diversity per site between hospitals (Dxy), and the diagonal shows the nucleotide diversity (π) per site within each hospital. The upper matrix shows the net divergence (Da) estimated between hospitals.

	HAV	HCUV	HGUA	HGUC	HGUE	HGUV	HULR	HUPLF
HAV	**0.003894**	0.000840	0.001422	0.001588	0.000515	-0.000077	0.000664	0.000347
HCUV	0.003258	**0.000941**	0.001867	0.002941	-0.000074	0.000659	0.003114	0.001919
HGUA	0.005175	0.004144	**0.003612**	0.002042	0.001653	0.001313	0.001731	0.000458
HGUC	0.005511	0.005387	0.005824	**0.003952**	0.002337	0.001646	0.002122	0.001373
HGUE	0.003643	0.001577	0.004640	0.005494	**0.002362**	0.000361	0.002592	0.001495
HGUV	0.003764	0.003023	0.005013	0.005516	0.003436	**0.003788**	0.000940	0.000614
HULR	0.004256	0.005230	0.005182	0.005743	0.005418	0.004479	**0.003290**	0.000601
HUPLF	0.004839	0.004935	0.004809	0.005894	0.005221	0.005053	0.004791	**0.005090**

However, as we have shown, most hospitals showed a different distribution for most AMR-resistance determinants and lineage combinations (Figure 5.6). Therefore, we would have expected larger divergence values between hospitals. As we had found that the ST307 lineage carrying the *bla*$_{\text{CTX-M-15}}$ genes was widely distributed and highly prevalent in all the hospitals, we thought that this lineage might be masking the differences between hospitals. After removing isolates of lineage ST307 carrying *bla*$_{\text{CTX-M-15}}$ from the comparisons, the divergence between hospital populations shifted to higher values for all the populations studied, the total population, the 3GC and carbapenem non-susceptible populations (Figure 5.8). It is important to highlight that, as the carbapenemase burden in the HGUE and the

HCUV were mostly due to lineage ST307, carrying *bla*$_{\text{CTX-M-15}}$ and *bla*$_{\text{OXA-48}}$, no comparisons were obtained for these hospitals.

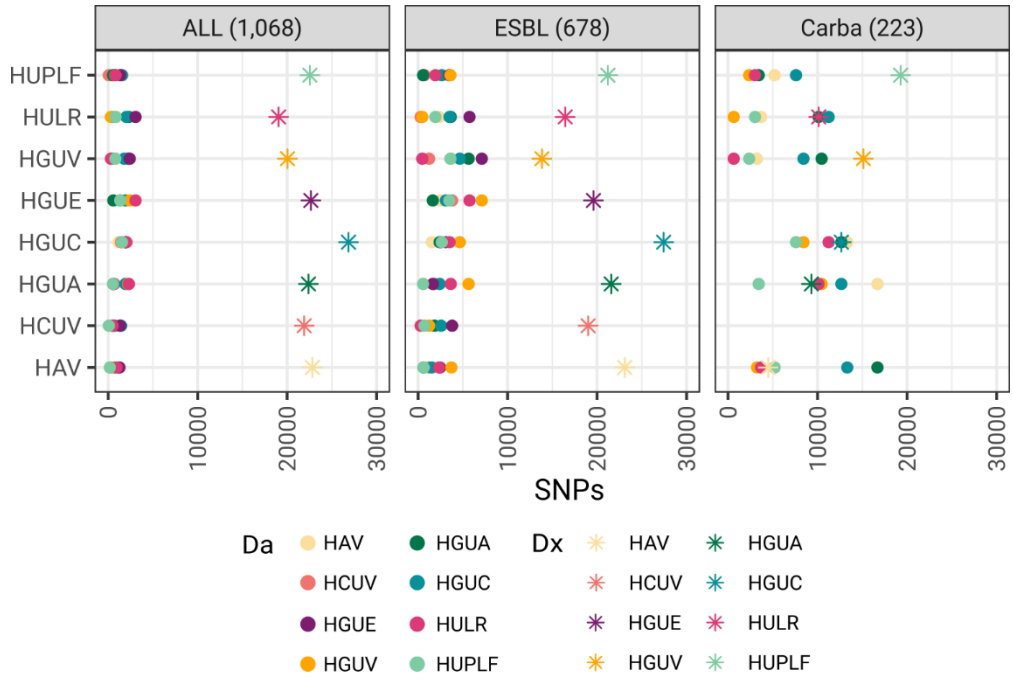


Figure 5.8. Genetic diversity in SNPs of SKPCV isolates within and between hospitals after removing ST307 isolates carrying *bla*$_{\text{CTX-M-15}}$ genes. For each hospital, the within diversity is represented as asterisks, and the net between hospital divergences (Da) is represented in circles. From left to right, the different blocks show the estimates for the whole population, the 3GC non-susceptible population, and the carbapenem-resistant population.

These findings suggest that there is a complex interplay between lineage transmission between hospitals and the local proliferation of problematic clones within each hospital. This is supported by the differences in the distribution of resistant lineages observed between hospitals (Figure 5.6), and further supported by the results of the divergence analyses, which showed that including the ST307 lineage in the analysis led to a low level of divergence, whereas its removal resulted in a higher degree of divergence. This points to a major role in the transmission of ST307 strains among hospitals in the SKPCV project.

5.5 Possible transmission groups

Due to our interest in elucidating the dispersion of ESBL- and carbapenem-resistant strains in the hospitals of the CV, we further studied those SLs with more than 5 3GC or carbapenem non-susceptible isolates. In total, 24 SLs (encompassing 35 STs) were selected and subjected to further core-genome analysis together with the database genomes of the global collection detailed in §4.3 that share the same SL (Supplementary Table 5.6). Using each core-genome tree, we delineated **clusters** within SLs (§3.9.4). We kept for further analysis those clusters containing at least 5 SKPCV isolates, for a total of 31 clusters (31 STs) comprising 1,235 SKPCV genomes, 359 that were obtained in NLSAR hospitals (denoted as NLSAR) and 3,712 genomes obtained from the global database.

Finally, to investigate transmissions, we repeated the process of constructing the core genomes at the 90% for each resulting cluster. This process generated alignments larger than 4 Mb, with only a few exceptions where the size was at least 3.5 Mb (Supplementary Table 5.7). Within each cluster, we defined **possible transmission groups (PTGs)** by grouping isolates with pairwise SNPs below 25 SNPs / 5 Mbp, as described in [67] and [413] (details in §3.9.5).

Our analysis identified 205 PTGs comprising 847 (52.80%) SKPCV isolates in 30 of the 31 phylogenetic clusters included in the analysis (Supplementary Table 5.8). These numbers already highlight the relevance of possible transmission events in the spread of AMR isolates in the CV. Within those 205 PTGs, we also found 135 NLSAR and 67 global database genomes involved. Most (n=129) of the PTGs consisted of more than two isolates (Figure 5.9 A). The majority of SKPCV isolates involved in PTGs belonged to phylogenetic clusters of ST307 (n=410), followed by ST11 (n=113), ST15 (n=62), ST147 (n=45), and ST405 (n=38), among others (Figure 5.9 A-B). Most of these PTGs were found in phylogenetic clusters associated with ESBL producers (n= 129), whereas a lower proportion was associated with carbapenemases (n=64) (Figure 5.9 B).

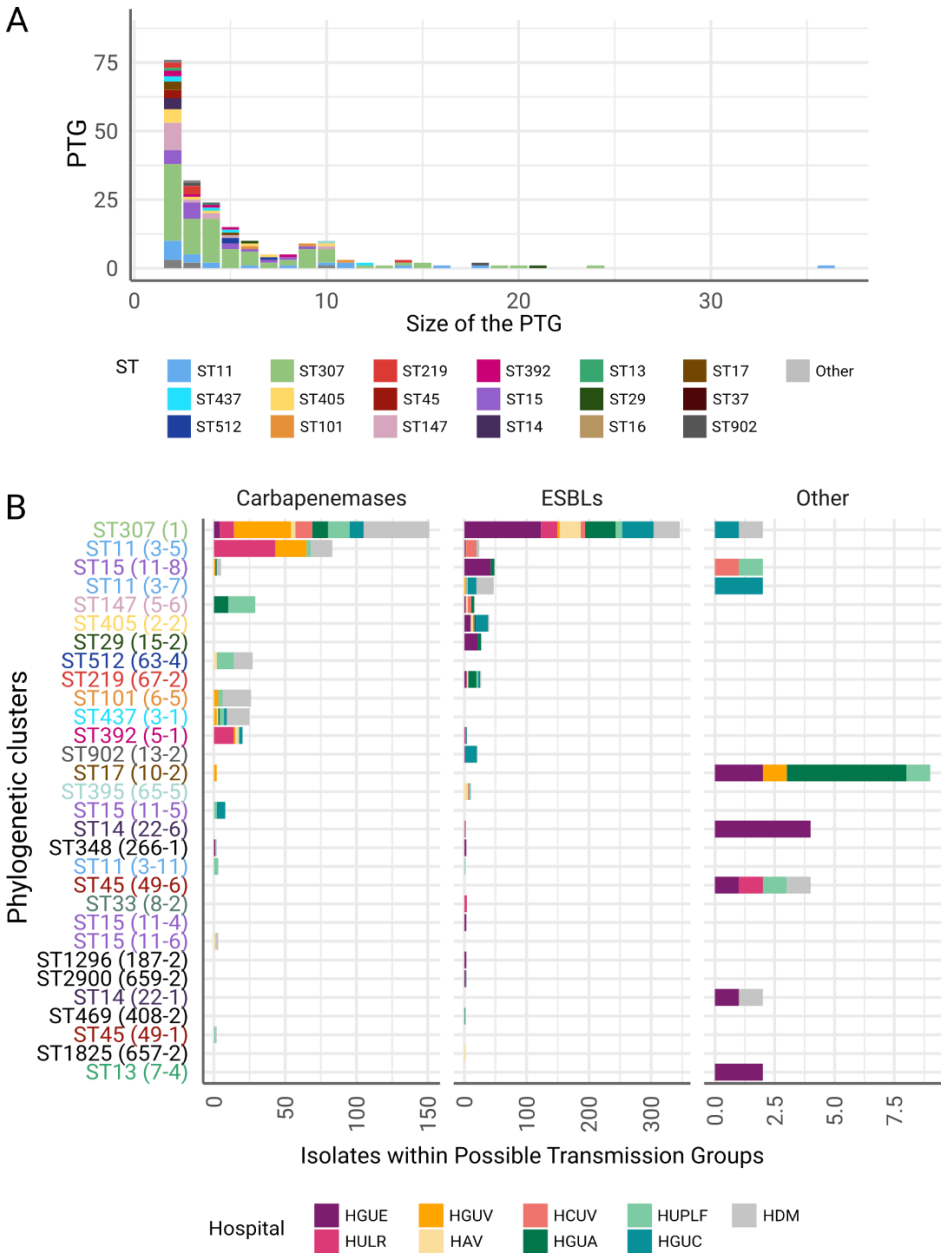


Figure 5.9. A) Distribution of the size (isolates involved) of each possible transmission group identified. Groups are coloured by the ST. B) Distribution of isolates in possible transmission groups (PTGs) including at least one SKPCV isolate according to phylogenetic cluster and presence of carbapenemase or ESBL determinants. For each PTG, the most frequent ST in the cluster is used to denote it. Labels are coloured by ST. Phylogenetic sublineages and clusters are indicated in parenthesis.

As an example, we will provide details on some of the largest and most significant PTGs. The largest PTG (consisting of 36 isolates) belonged to SL cluster 3-5, which corresponds to ST11 and KL13 type. This PTG was highly associated with the presence of $bla_{\text{CTX-M-1}}$ (94.4%) and $bla_{\text{OXA-48}}$ (91.6%) genes. The PTG was found in two hospitals, 34 isolates from the HULR and 2 isolates from the HUPLF, collected from 2017 to 2019. Another large PTG (n= 16) was identified within the same SL cluster (3-5), also belonging to ST11 and the KL13 type. This PTG, in addition to carrying $bla_{\text{CTX-M-1}}$ (100%) and $bla_{\text{OXA-48}}$ (81.5%), also carried virulence factors yersiniabactin (*ybt16* in an ICE*Kp12*, 100% identity) and aerobactin (*iuc5*, 100%), making it a convergent MDR-hypervirulent PTG. This PTG was mainly limited to the HGUV from 2017 to 2018. Five isolates of the databases fell within this PTG, but all were collected in the same hospital. This is remarkable since, to our knowledge, no convergent PTG has been detected in Spain [45]. Indeed, although both PTGs within the SL 3-5 did not conform a single PTG, they were found in the same clade of the SL phylogeny, indicating that they may be closely related (Supplementary Figure 5. 2). We found that these convergent strains originated from previous isolates in the hospital by acquiring a virulence plasmid (data not shown).

The second largest PTG (24 isolates) belonged to SL 1, corresponding to ST307 and KL102 type. This PTG was associated with $bla_{\text{CTX-M-15}}$ (91.6%) and $bla_{\text{OXA-48}}$ (100%) genes. Although $bla_{\text{CTX-M-15}}$ was highly associated with ST307 in most of the SKPCV isolates, this was the largest ST307 cluster carrying $bla_{\text{OXA-48}}$. The remaining ST307 $bla_{\text{OXA-48}}$ carriers were limited to a few isolates, as detailed in the next section. This PTG was mostly collected in 2016 (under the NLSAR and not the SKPCV collection) and restricted to the HGUV. Only one isolate of the SKPCV collection fell in this PTG.

We were particularly interested in identifying possible transmission chains within and between hospitals. We found 141 PTGs associated with within-hospital transmissions, comprising 591 isolates, 475 of which were from the SKPCV project. Additionally, we found 64 PTGs that involved multiple hospitals comprising 81.2% (n= 372) of the SKPCV isolates, indicating inter-hospital transmissions (Supplementary Figure 5. 3).

We observed that most hospitals shared at least one PTG, providing evidence of inter-hospital circulation of strains (Table 5.2). The only exception was the HGUC, which did not share any PTG with the HGUE or the HULR. The HUPLF exhibited the highest level of shared PTGs with other hospitals, with a maximum value shared with the HGUE. The two hospitals shared 11 PTGs with an average pairwise SNP difference of 7.99 (Table 5.2).

Table 5.2. Possible transmission groups (PTGs) shared by hospitals. The diagonal shows the number of PTGs identified in each hospital and the average pairwise SNPs distances in parenthesis. The upper hemi matrix shows the number of PTGs found shared by hospitals whereas the lower hemi matrix shows the average pairwise SNPs distances between those groups.

	HAV	HCUV	HGUA	HGUC	HGUE	HGUV	HULR	HUPLF
HAV	**26 (7.88)**	3	5	5	3	5	4	9
HCUV	9.60	**19 (9.35)**	6	4	0	2	0	6
HGUA	9.53	10.4	**46 (11.9)**	6	11	2	3	8
HGUC	9.80	8.55	11.3	**35 (7.69)**	2	9	4	10
HGUE	9.13	-	7.99	13.4	**55 (7.16)**	4	1	2
HGUV	10.4	10.6	10.3	12.1	14.0	**29 (9.32)**	1	8
HULR	12.8	-	12.9	10.8	7.07	10.7	**23 (10.2)**	7
HUPLF	9.43	10.3	11.9	9.15	10.0	11.8	12.5	**44 (8.66)**

We also assessed the genetic similarity between the SKPCV isolates in PTGs that were restricted to a single hospital (Within-Within, WW) and those in PTGs involving multiple hospitals. The later were subdivided in those collected within the same hospital (Between-Within, BW), and those collected from different hospitals (Between-Between, BB) (Figure 5.10 A). Our results showed that there was no significant difference in the average pairwise SNP distances between the WW and BW groups (8.97 vs 8.28, t-test p-value = 0.369). However, we observed significant differences between each of these two groups and the BB group (average 11.7 SNPs, t-test p-values of 0.000212 and 0.000225, respectively). These results indicate that there is a higher degree of genetic relatedness among the isolates collected within hospitals compared to those between hospitals, as indicated by the lower pairwise SNP distances observed within hospitals. This suggests that the possible transmission events between isolates from the same hospital may involve a higher proportion of direct, or nearby, transmissions, while transmissions between

different hospitals may involve more complex routes, possibly involving intermediate, undetected spreaders.

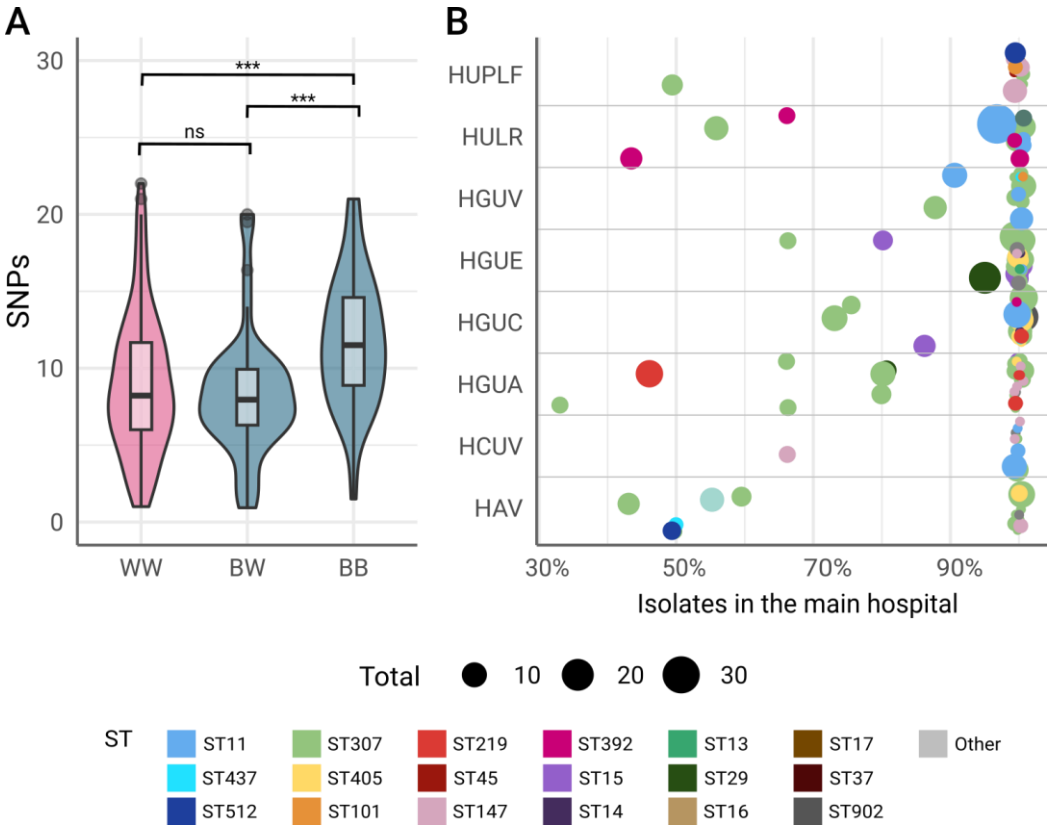


Figure 5.10. A) Distribution of pairwise SNPs between isolates of possible transmission groups (PTGs) restricted to a single hospital (WW), isolates of multihospital PTUs that belong to the same hospital (BW) or to different hospitals (BB). B) Each dot represents a PTG. The y-axis refers to the main hospital is and in the x-axis the percentage of samples in the PTG belonging to that hospital. The ratio of the point reflects the size of the PTG. Dots are coloured by ST.

However, it is important to determine the frequency of isolates moving between hospitals to truly understand the role of inter-hospital transmission. Despite finding that inter-hospital transmission is common and associated with a low average distance, we observed that inter-hospital cases were limited within PTGs. Most of the PTGs found involving between hospitals spread were primarily restricted to a single hospital, with an average of 1.96 cases (range 1 - 7 isolates) from different hospitals (Figure 5.10 B). PTGs involving multiple hospitals comprised mainly ST307

and a few of ST45, ST395, ST437 and ST512 (Figure 5.10 B). Additionally, we observed that the size of the PTG did not necessarily determine inter-hospital transmission, as we identified small clusters that involved more than one hospital (Supplementary Figure 5. 3). Moreover, we did not find any clear association between geographic proximity of the hospitals and the number of shared PTGs, except for the HGUE, which shared a significant number of PTGs with the nearby HGUA.

As previously mentioned, only 67 global isolates were grouped in the same PTGs as the SKPCV isolates. We observed that 94% of external isolates were collected in Spain, mostly from Valencia 67,4%. This implies that the contribution of the global population to the ongoing transmission in the analysed period in the CV is low. This suggests that transmissions occurred predominantly at the regional level before expanding nationally. As we increased the distance threshold to consider PTGs, global isolates gradually emerged, particularly from Europe and countries such as Greece, United Kingdom, France, Italy, and Portugal. Moreover, we also found American isolates.

5.6 The successful ST307 lineage carrying CTX-M-15 in the CV

The ST307 was the most frequently found in the SKPCV collection with 559 isolates (34.8%) (Supplementary Figure 5. 5). Moreover, as shown in the diversity and transmission analyses, this ST was driving most of the within and between hospital transmission. In fact, most isolates collected from this ST (80.3%) were involved in PTGs (Supplementary Figure 5. 4). Out of all PTGs, 93 (45.3%) comprising 499 isolates were attributed to the ST307 cluster. Moreover, 30 of the 64 PTGs involved in inter-hospital transmission were also attributed to this clone.

To conduct a more in-depth analysis, we investigated the relationships among the ST307 isolates collected in the global database. Although the ST307 clone was one of the most frequently sampled STs in the global collection (n= 1,110 genomes), this was primarily due to the high number of isolates collected under the SKPCV project (n= 559) and the NLSAR collection (n=100). Moreover, most of the remaining global

isolates were collected in Europe (n=173) or Spain (n=113), with only a small number of genomes (n=111) sampled from other locations.

We made a phylogenetic analysis using the core genome of the 1,110 global genomes. The core genome at the 90% had 4,393 genes and resulted in an alignment of 4,268,326 bp with 16,936 variant positions (~ 3,967 SNPs/Mbp). We found that the genome of the global ST307 population was highly conserved and showed less variation than other SLs. Our analysis of other SLs with similar sampling numbers, such as the ST258 or ST11, showed a larger number of SNPs positions per Mbp (12,051 and 18,825 SNPs/Mbp, respectively).

The AMR gene load was also highly conserved within the ST307 lineage. The *bla*$_{\text{CTX-M-15}}$ ESBL gene was found in 978 (88.1%) genomes, usually in combination with *bla*$_{\text{TEM-1D}}$ (89.9%), *bla*$_{\text{OXA-1}}$ (87.6%), *sul2* (93.4%), *dfrA14* (93.4%), *strAB* (93.4%), *aac(6')-Ib-cr* (87,3%) and *aac(3)-IIa* (70.6%). This is due to this lineage being highly associated with a MDR plasmid that carries all these genes (detailed below). Contrarily to the other AMR genes, carbapenemase genes were found to be occasionally present and diverse. We found 262 *bla*$_{\text{OXA-48}}$ and 159 *bla*$_{\text{KPCs}}$, but MLBs were rarely found (n= 10). The *bla*$_{\text{OXA-48}}$ genes were not distributed equally throughout the phylogeny, as they mostly accumulated in a few clades, one of them an SKPCV clade with most isolates collected in the HGUV.

To better study the AMR context and the MGEs responsible for the dissemination of these genes we used the sequence of the isolate HGV2_418 obtained with PacBio SMRT technology. This isolate belongs to the largest ST307 PTG identified in our study. Our analysis revealed the presence of two plasmids, the pCTX-M-15 plasmid carrying the *bla*$_{\text{CTX-M-15}}$ gene, and the pOXA-48 plasmid carrying the *bla*$_{\text{OXA-48}}$.

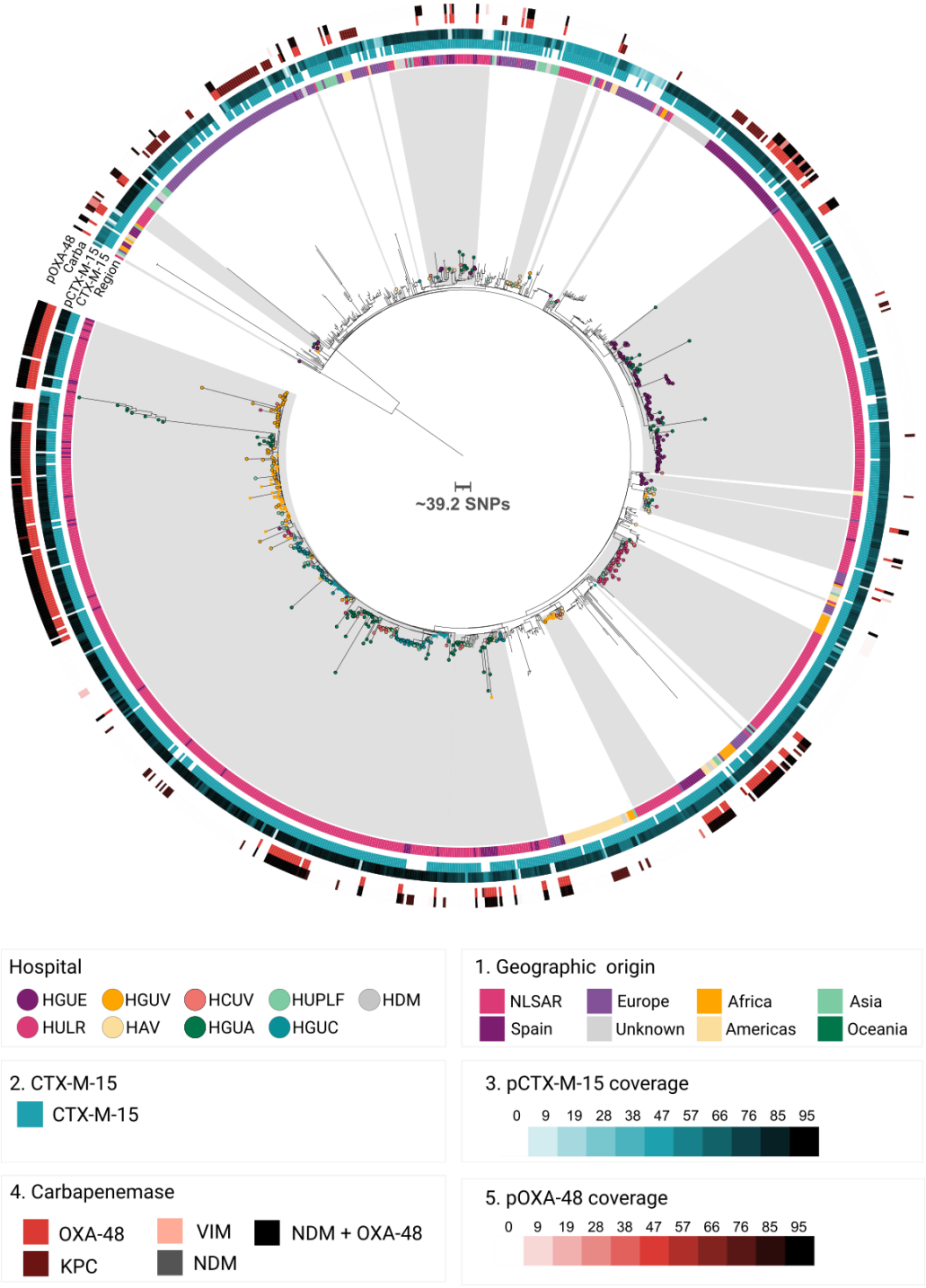
The plasmid pCTX-M15 has two replicons, being a IncFIBk/IncFIIk plasmid with 194,649 bp of length. This is a MDR plasmid as it carries *strA*, *strB, aac(3)-IIa, aac(6')-Ib-cr, qnrB1 sul2, dfrA14, tet(A), bla*$_{\text{TEM-1B}}$, *bla*$_{\text{CTX-M-15}}$, and *bla*$_{\text{OXA-1}}$ genes in a multidrug resistance region. We also found copper- and arsenic-resistance genes and the transfer region (Supplementary Figure 5. 6). Highly similar plasmids to this one have been reported previously [279, 280]. We found this plasmid with an average coverage of 73.1% (range 1-100) in all the genomes carrying *bla*$_{\text{CTX-M-15}}$ genes,

but for a few exceptions. Regardless of these exceptions, the level of plasmid conservation within the whole ST307 lineage is remarkable.

The pOXA-48 plasmid found had one replicon type IncL and had 63,544 bp of length. The plasmid was highly similar to the conserved pOXA-48a previously reported, which only carries the *bla*$_{\text{OXA-48}}$ gene. The mean coverage of the isolates carrying these genes was 91.8% (range 30 - 95%), with the SKPCV isolates having an average coverage of 94.61% (range 91 - 95%).

The global phylogeny of the lineage indicated that the SKPCV isolates were not grouped together in a single clade, but rather dispersed across different groups, (Figure 5.11) which corresponded to the groupings of different PTGs identified (Supplementary Figure 5. 5). This suggests that the current distribution of the ST307 in the CV is due to multiple introductions of the lineage into the region, followed by subsequent dissemination and diversification. However, to fully understand the evolution of this clade in the CV, further analyses are needed.

Figure 5.11. (Next page). Tree of the ST307 genomes of the total collection. 16,936 SNP position, 4,268,326 pb WG. Clades containing SKPCV isolates are highlighted in grey. The dots at the end of the tips are coloured by the hospitals where they were collected. Isolates from the SKPCV are rounded whereas isolates from the NLSAR are not. The first ring it's of Geographic origin. The second is the presence or absence of the *bla*$_{\text{CTX-M-15}}$ and the second the coverage of the pCTX-M-15 assembled in this work. The fourth is the presence of *bla*$_{\text{OXA-48}}$ whereas the fifth ring is the coverage of the pOXA-48 assembled in this work.


pOXA-48
Carba
pCTX-M-15
CTX-M-15
Region
~39.2 SNPs
Hospital
HGUE
HGUV
HCUV
HUPLF
HDM
HULR
HAV
HGUA
HGUC
1. Geographic origin
NLSAR
Europe
Africa
Asia
Spain
Unknown
Americas
Oceania
2. CTX-M-15
CTX-M-15
3. pCTX-M-15 coverage
0 9 19 28 38 47 57 66 76 85 95
4. Carbapenemase
OXA-48
VIM
NDM + OXA-48
KPC
NDM
5. pOXA-48 coverage
0 9 19 28 38 47 57 66 76 85 95

5.7 Concluding remarks

The results of this chapter highlight that the Comunitat Valenciana has a high incidence of 3GC-resistant *K. pneumoniae* infections, while remaining a region with low incidence of carbapenem resistance. In this context, we conducted the SKPCV project, in which we collected and sequenced 1,604 isolates from eight hospitals in the region between 2017 and 2019. Our genomic analyses revealed a high-level population diversity both overall and within each hospital, and a wide variety of ESBLs and carbapenemases. Yet, we found that the AMR population differed in each hospital, with distinct lineages and AMR genes distributed among them.

The burden of AMR and *K. pneumoniae* in this region was the result of a diverse set of factors, including singleton lineages likely originating from the community or from the previous microbiota of patients, as well as transmission and clonal dissemination both within and between settings.

Finally, our analysis revealed that only one lineage, ST307, was responsible for most 3GC- and carbapenem-resistant infections and the interhospital transmissions. Our investigation of the evolution of this lineage, in which we included global isolates, led us to conclude that the high prevalence of ST307 in the SKPCV collection was primarily due to multiple introductions of the lineage into different settings.

CHAPTER 6
Initial dissemination of carbapenem-resistant *K. pneumoniae* clones in a tertiary hospital

This chapter has been published in *Microbial Genomics*:

García-González N, Fuster B, Tormo N, Salvador C, Gimeno C, González-Candelas F;. Genomic analysis of the initial dissemination of carbapenem-resistant *Klebsiella pneumoniae* clones in a tertiary hospital. *Microbial Genomics.* 2023 Jun;9(6). DOI 10.1099/mgen.0.001032

Reducing AMR and preventing the spread of resistant pathogens requires an understanding how AMR pathogens emerge and spread in hospital environments. Most studies focus on understanding the first detected cases of specific clones or AMR determinants, especially carbapenemases [449–451], or to understand their epidemiology once they are found spreading in hospital settings [151, 322]. Yet, little is known about the emergence, initial colonisation, and subsequent dispersion of these AMR pathogens in clinical settings.

Understanding this process, coupled with active genomic surveillance of resistance, can provide clues on how to prevent and control new entries and the spread of AMR in the future. In addition, it is important to understand what the relative contribution of new introductions with respect to the local spread of endemic strains or genes is.

In this chapter, we assess the initial colonisation and dissemination evolution of carbapenem-resistant *K. pneumoniae* in a hospital in Valencia (Spain). We approach this issue in the tertiary HGUV with a reference population of almost 360,000 inhabitants. In this hospital, CRKp infections were first detected in 2015 and became endemic since then [452]. Here, we have analysed the first *K. pneumoniae* isolates non-susceptible to carbapenems detected (January 2015) and extended these analyses to 4 years later (until December 2018).

6.1. Isolates, clinical data, and WGS

A total of 183 carbapenem non-susceptible *K. pneumoniae* isolates and 41 carbapenem-susceptible control strains were included in this study. These isolates included the initial carbapenem-resistant isolates collected between January 2015 and 2017 at the same hospital, and all the strains isolated under the SKPCV project in this hospital during 2017 and 2018.

It is important to note that the increase in CRKp occurred shortly after 2014, when a large peak of cephalosporin-resistant strains was reported in the HGUV (up to 45% of the collected isolates) (Supplementary Figure 6.1). Increased selective pressure, driven largely by the use of carbapenems to treat infections by cephalosporin-resistant isolates, could explain the rapid increase in CRKp strains. A highly selective

environment provides the perfect setting for CRKp lineages to succeed and disseminate.

Patients were mostly elderly, with a median age of 74 years (range 23 to 94). Regarding isolation sources, the most frequent ones were those from urine (98 samples, 44%), followed by axillary/faecal carriage (68 samples, 30%) (Supplementary Table 6.1).

6.2. Carbapenem non-susceptibility is related to 4 major STs, and OXA-48 and NDM carbapenemases

In silico typing identified 21 different known STs, among which ST307, ST11, ST101, and ST437 were the most prevalent ones and accounted for 88.5% (162/183) of all the CRKp (Figure 6.1 A). These four STs had already been identified in Spain as carbapenemase carriers, and they are among the STs widely distributed across Europe [151]. They were related to the acquisition of several antimicrobial-resistant plasmids in different European countries in the late 2000s [289].

The 183 CRKp isolates were classified into four different Carba-R groups by the nature of their non-susceptibility to carbapenems (Figure 6.1 A).

Group 1 encompassed the largest group of carbapenem-resistance and included 131 isolates with acquired carbapenemases. It was mainly associated with the most prevalent STs: ST307, ST11, ST101, and ST437. Two different carbapenemase genes were detected in group 1: $bla_{\text{OXA-48}}$, in 101 isolates (mainly ST307 and ST11), and two variants of bla_{NDM} genes, $bla_{\text{NDM-1}}$ and $bla_{\text{NDM-23}}$, in 34 and 3 isolates, respectively. While $bla_{\text{NDM-1}}$ was found in ST101 and ST437, $bla_{\text{NDM-23}}$ was only found in ST437. Remarkably, we found 7 isolates from the two STs, ST101 and ST437, harbouring both $bla_{\text{OXA-48}}$ and bla_{NDM} genes (Figure 6.1 B).

Group 2 included carbapenem-resistant strains due to the acquisition of an *ampC* gene. This group included 30 isolates of ST11. AmpC was found to be mediated only by DHA-1 types.

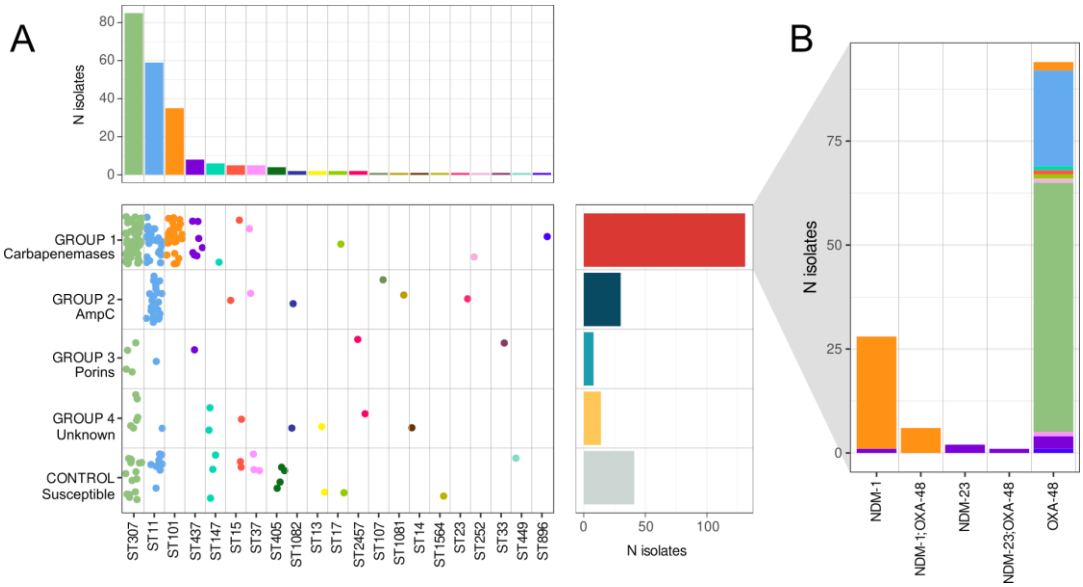


Figure 6.1. A) Isolate distribution in Carba-R groups and STs. Dots in the central plot and bars in the upper bar plot are coloured by ST, while bars in the right bar plot are coloured by Carba-R group. b) Distribution of the different Carba-R genes found in Group 1. Bars are coloured by ST.

Group 3 included missing or truncated porins (8 isolates) and was related to 5 different STs. Porin truncations were detected only in the *Ompk36* gene. We found one isolate with a Gly134Asp135 duplication in loop 3 (OmpK36GD). The antibiotic resistance levels produced by this duplication are similar to those produced by the loss or truncation of the porin [453].

Group 4 included 14 isolates for which none of the previously described mechanisms could explain low carbapenem susceptibility. In 11 of these isolates, belonging to 8 different STs, resistance to carbapenems might be attributed to the presence of ESBL genes (11/14), especially to $bla_{\text{CTX-M-15}}$ (8/11).

Carba-R groups differed in their population diversities. Although some isolates from major lineages, such as ST101 and ST437, were restricted to a single group, those in other lineages, such as ST307 and ST11, were found in different subgroups, indicating possible different populations. Conversely, some Carba-R groups, such as 3 and 4, were highly diverse, presenting several different STs, with no single lineage associated with them.

6.3. Carbapenem-resistant strains in the HGUV were disseminated by six lineages.

In mid-2015, a rapid increase of carbapenem non-susceptible *K. pneumoniae* isolates was noticed at the hospital (Supplementary Figure 6.1). The "outbreak" was mostly associated with the expansion of certain STs. To understand the relationships between the STs found in the HGUV, we obtained the core genome and derived the corresponding ML phylogenetic tree (Figure 6.2). The phylogeny, capsular types, and AMR genes revealed that ST307, ST11 and ST437 could be divided into different lineages. All these lineages are globally distributed clones of *K. pneumoniae* and are considered MDR [31].

To facilitate the ensuing description of the sublineages, we named them following their capsular types and AMR genes. The ST307 was divided into two different clades, ST307-OXA48 and ST307-CONTROLS. The ST11 was divided into three lineages: ST11-KL13 (31 isolates), ST11-KL105 (27 isolates), and ST11-KL24 (2 isolates). The ST437 clade was split into two clades, ST437-OXA48 and ST437-NDM. Finally, ST101 was retained as a single lineage.

The genetic diversity within lineages was much lower than that between lineages (Supplementary Tables 6.5 - 6.8), likely supporting different clades. This indicates that each sublineage corresponds to a single dissemination following a clonal expansion and not multiple, independent introductions and subsequent spreads. In fact, some isolates were highly similar (pairwise SNP distances < 5), possibly belonging to the same transmission chains.

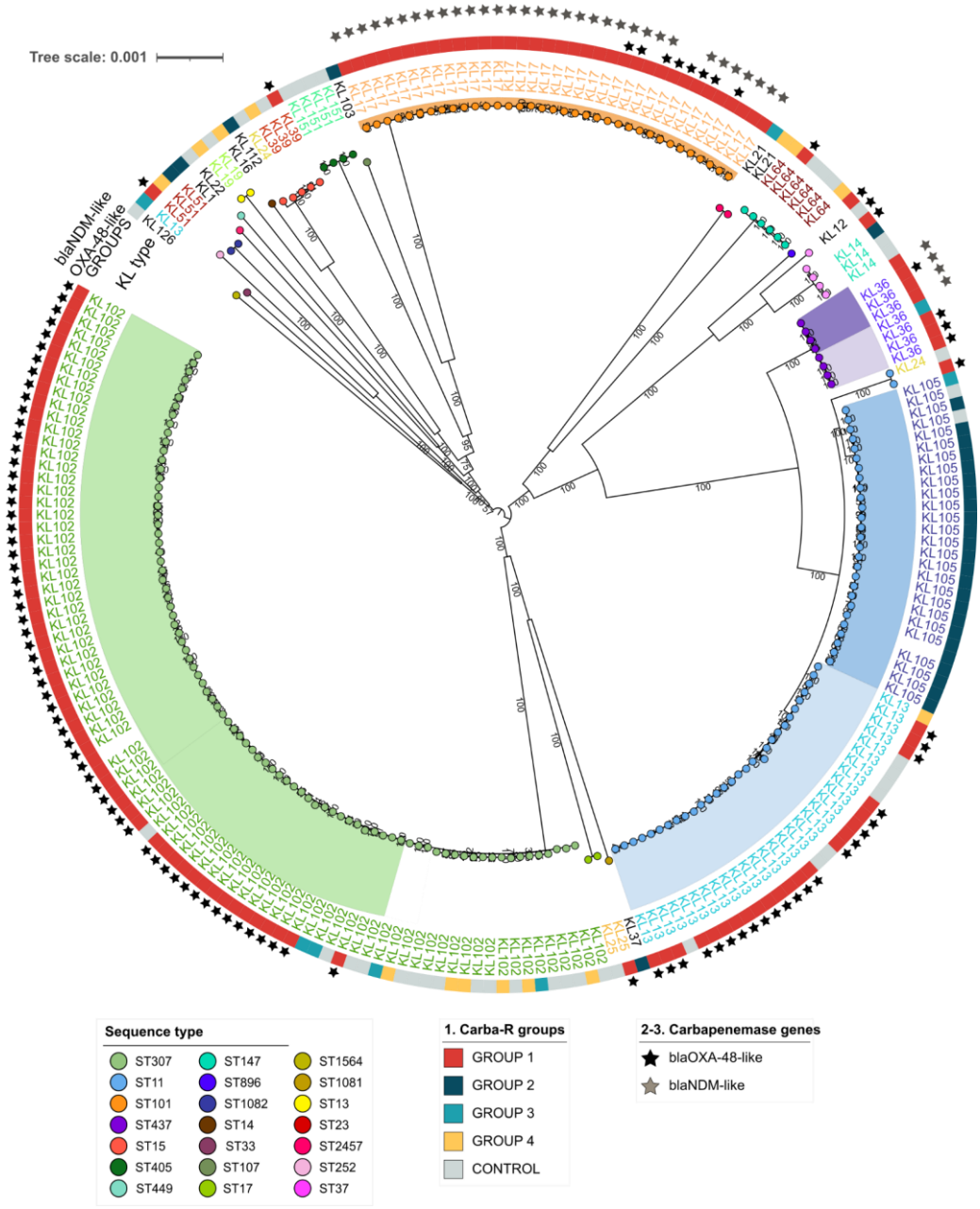


Figure 6.2. Maximum-likelihood phylogenetic tree of the *K. pneumoniae* isolates collected in the HGUV from 2015 to 2018. Tree was constructed from the core genome (90%) alignment (3,570,188 bp) of the 224 isolates. STs, capsular locus, Carba-R groups, and the presence of $bla_{\text{OXA-48}}$ and bla_{NDM} genes are shown.

6.4. Antimicrobial susceptibility of Carba-R groups and sublineages and association with other genetic features (AMR genes and plasmids)

All the isolates, including the controls, were MDR strains as they were non-susceptible to at least one drug in three or more antimicrobial categories (Supplementary Table 6.1). In general, all the isolates were resistant to penicillins, cephalosporins, and quinolones, and most of them were also resistant to cotrimoxazole and ertapenem. Colistin was the antibiotic with the highest susceptibility rates (69.5%, MICs lower than 2 mg/L), followed by tigecycline (69.5%) (Figure 6.3 A).

We observed an association between Carba-R groups and sublineages, susceptibility to antimicrobials, and AMR gene repertoire (Figure 6.3 A-C, Supplementary Table 6.2).

The most prevalent carbapenemase gene was *bla*$_{\text{OXA-48}}$, which was linked to lineages ST307-OXA48, ST11-KL13, and ST437-OXA-48.

Isolates in lineage ST307-OXA48 were resistant to most of the antibiotics tested and had a high proportion of tigecycline resistance (52.5%). Almost all the isolates in this lineage carried *bla*$_{\text{SHV-28}}$, *bla*$_{\text{TEM-1D}}$, *bla*$_{\text{OXA-1}}$, *bla*$_{\text{CTX-M-15}}$, *sul2*, *qnrB1*, *catB4*, *aac(3)-IIa*, *acc(6)-Ib-cr*, *strA*, and *strB* genes.

Isolates in lineage ST437-OXA-48 had a similar antimicrobial susceptibility profile to those in the ST307-OXA48 lineage, yet the AMR genes present were different, as they carried *bla*$_{\text{OXA-1}}$, *bla*$_{\text{SHV-12,}}$ *bla*$_{\text{OXA-48,}}$ *sul1*, arr3, CatB3, *aac6'-Ib,* and *aph(3)-Ia.*

Lineage ST11-KL13 showed a highly different susceptibility pattern, with a lower load of AMR genes than ST307-OXA48, reflected in lower resistance levels to aminoglycosides and tigecycline. The genes found present in all the isolates of this clade were *bla*$_{\text{SHV-11}}$, *bla*$_{\text{CTX-M-1,}}$ *bla*$_{\text{OXA-48}}$ and, in some isolates, *bla*$_{\text{TEM-1D}}$ (18.1%), *Sul1* (63.3%), *Sul2* (22.7%) and *aadA*2 (54.5 %). These three lineages had different plasmids but share the presence of IncL/M(pOXA-48) plasmid, highly associated with *bla*$_{\text{OXA-48}}$ genes.

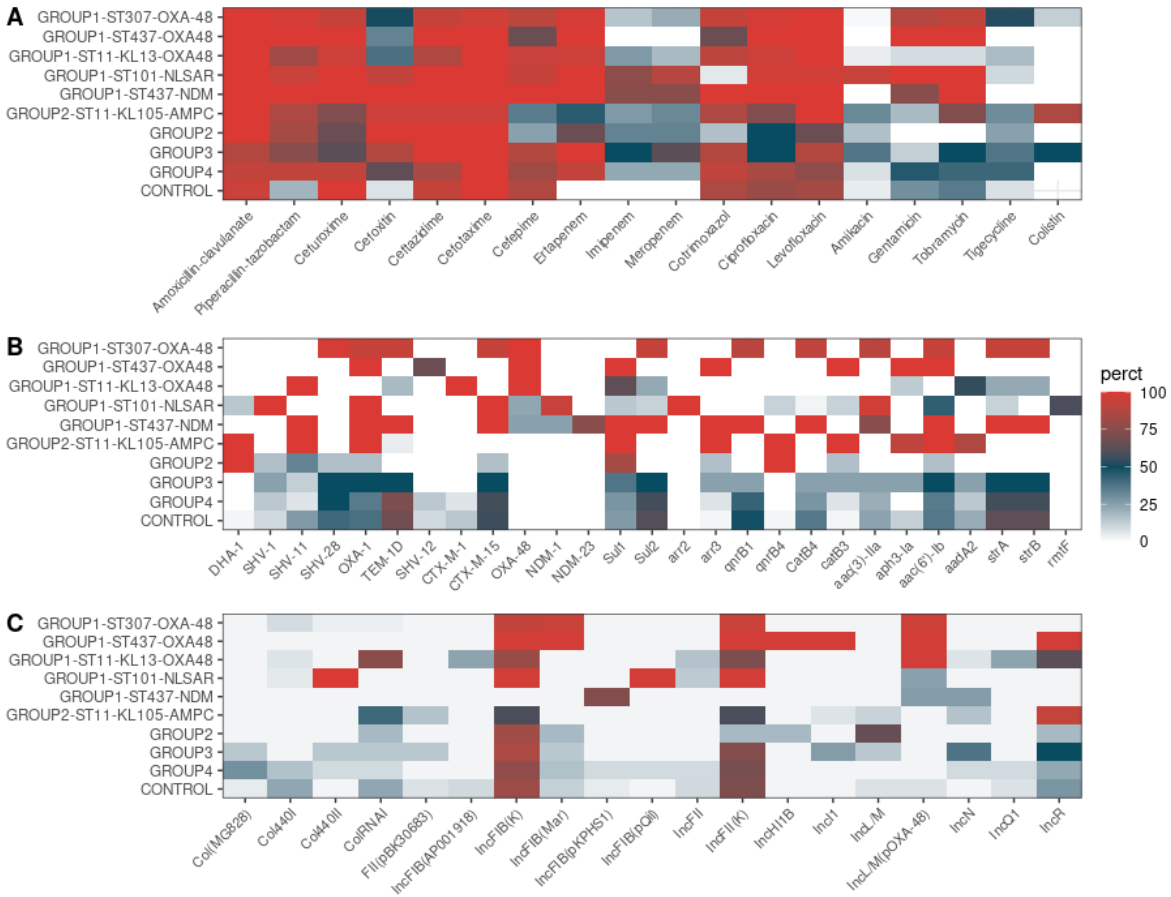


Figure 6.3. The proportion of isolates in each Carba-R group and sublineage that A) showed phenotypic resistance to each antimicrobial tested, B) carried the antimicrobial resistance genes identified, and C) harboured any of the plasmid incompatibility groups detected.

The other carbapenemases found, NDM-1 and NDM-23, were associated with group 1 lineages ST101 and ST437-NDM. Isolates belonging to the ST101 lineage were resistant to almost all the antibiotics tested except colistin, cotrimoxazole, and tigecycline, due to the combined effect of the genes *bla*$_{\text{NDM-1}}$, *bla*$_{\text{SHV-1}}$, *bla*$_{\text{OXA-1}}$, *bla*$_{\text{CTX-M-15}}$, *arr-2*, *aac(3)-IIa*, and, in some cases, also genes *bla*$_{\text{OXA-48}}$ (23.5%) and *acc(6)-Ib* (44.1%). Moreover, this lineage was the only one carrying *rmtF* (58.8%), which explains the amikacin resistance hardly seen in the other lineages. Although this lineage was largely associated with *bla*$_{\text{NDM}}$ genes (94.1%), we found some isolates that had lost the *bla*$_{\text{NDM}}$ gene but had kept *bla*$_{\text{OXA-48}}$. The coexistence of *bla*$_{\text{OXA-48}}$ and *bla*$_{\text{NDM}}$ genes in the same isolate is seldom reported worldwide [454, 455].

Lineage ST437-NDM isolates had a similar antimicrobial susceptibility profile to those of ST101, with the difference that they were resistant to cotrimoxazole and susceptible to amikacin. Isolates in ST437-NDM had a large number of AMR genes consisting of *bla*$_{\text{SHV-11}}$, *bla*$_{\text{OXA-1}}$, *bla*$_{\text{TEM1-D}}$, *bla*$_{\text{CTX-M-15}}$, *sul1*, *sul2*, *arr-3*, *qnrB1*, *CaB4*, *acc(6)-I, strA, strB.* All the isolates in this lineage carried *bla*$_{\text{NDM}}$ variants, 25% carried *bla*$_{\text{NDM-1}}$, and 75% *bla*$_{\text{NDM-23}}$. Moreover, 25% of those isolates also carried *bla*$_{\text{OXA-48}}$. As expected, only those lineages that carried *bla*$_{\text{NDM}}$ genes were resistant to the three carbapenems tested (ertapenem, meropenem, and imipenem).

Although lineages ST101 and ST437-NDM were associated with NDM carbapenemases, they carried different plasmids and none of them was found in both lineages, thus suggesting that those carbapenemases are encoded in different plasmids in these lineages.

Group 2 was mainly formed by isolates of lineage ST11-KL105, which included almost all AmpC-producers. This lineage showed lower levels of resistance than the groups mentioned above, with more susceptibility to cephalosporins and cotrimoxazole. Almost all the isolates of this group carried *bla*$_{\text{DHA-1}}$, *bla*$_{\text{SHV-11}}$, *bla*$_{\text{OXA-1}}$, *sul1*, *arr3*, *qnrB4*, *catB3*, *aac6'-Ib, aph(3)-Ia*, and *aadA2* genes. The remaining isolates of Group 2 belonged to several different lineages, and this is reflected in the variation in the presence of AMR genes. The only gene present in all the isolates was *bla*$_{\text{DHA-1}}$ and *qnrB4.*

Isolates in group 3 (defective porins), group 4 (unknown mechanism), and controls showed a highly diverse composition of STs (Figure 6.1 A) and were very variable in their antimicrobial susceptibility profiles and AMR gene repertoires. No common AMR gene was found to be associated with any of these groups. Nevertheless, we found some enrichment in *bla*$_{\text{TEM-1D}}$, *bla*$_{\text{OXA-1}}$, *bla*$_{\text{CTX-M-15}}$, *sul2*, *qnrB1*, *aac(3)-IIa*, *acc(6)-Ib*, *strA*, and *strB*. Isolates in group 3 (defective porins) and controls showed more susceptibility to β -lactams than any other group.

6.5. Most lineages were also found in other Spanish regions

To better understand and contextualise how these lineages emerged in the HGUV, we added Spanish genomes (n=360) to the global analysis (Supplementary Figure 6.2, Supplementary Table 6.3, Supplementary Table 6.4) and to those of each major ST (Figure 6.4). These genomes were collected between 2008 and 2018. However, most of those isolates were collected in a multi-regional study in 2018 [366].

The two major STs in the HGUV, ST11 and ST307, were also frequently reported in Spain, whereas the other major lineages found in the HGUV were seldom found in other locations. Moreover, lineages ST147, ST405, ST15, ST392, and ST512, which were found to be frequently reported in Spain, were not detected in the HGUV.

For lineages ST11-KL105, ST11-KL13, ST101, and ST437-NDM, there were related isolates from other Spanish regions (Figure 6.4). Lineage ST11-KL105 had related isolates from 2018 collected in North-western Spain (Ferrol) and the Balearic Islands (Palma de Mallorca). These isolates did not carry *bla*$_{\text{DHA1}}$, but they carried *bla*$_{\text{OXA-48}}$ and *bla*$_{\text{VIM-1}}$. However, the clade was very homogeneous, with an average of 39.6 pairwise SNPs (range 1–119). Related isolates of clade ST11-KL13 were collected in 2013–2014 from the centre of Spain (Guadalajara, Toledo, Madrid, and Ciudad Real) and in 2018 from other Valencian hospitals. However, the samples from the centre of Spain carried the *bla*$_{\text{KPC-2}}$ gene but not the *bla*$_{\text{OXA-48}}$ gene, as did the ST11-KL13 isolates collected in the HGUV and other Valencian hospitals. Nevertheless, the clade had a low diversity (71.8 pairwise SNPs on average, range 0–192). The other ST11 Spanish samples fell into a new clade, ST11-SPAIN, with most isolates carrying a *bla*$_{\text{OXA-48}}$ and collected in 2018 throughout Spain (Madrid, Guadalajara, Cadiz, and Santander). Only two isolates found in the HGUV fell into this clade.

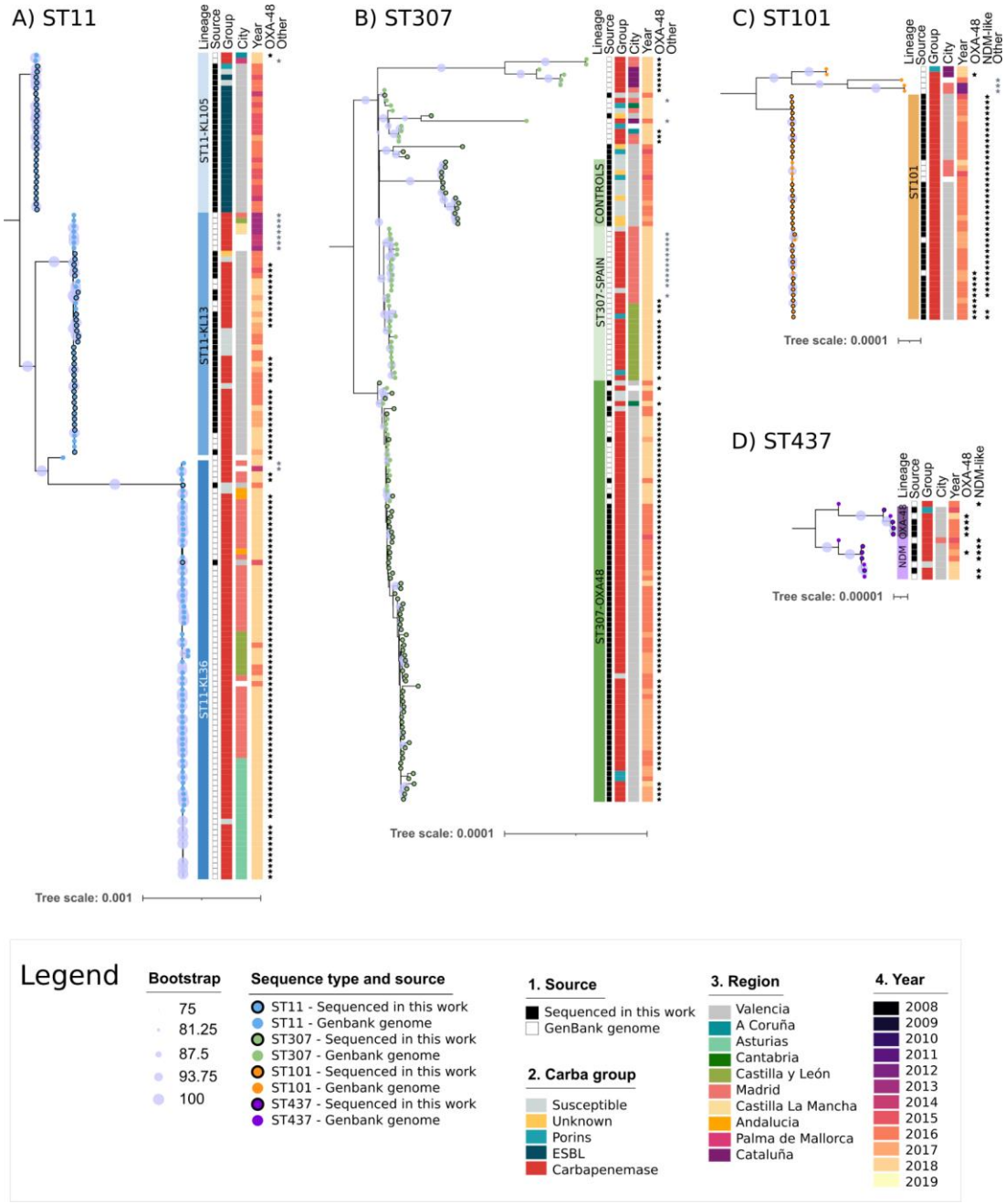


Figure 6.4. Maximum-likelihood phylogenetic trees (and length of the alignments used) for the A) ST11 (4,081,060 bp), B) ST307 (4,461,655 bp), C) ST101 (4,278,965 bp), and D) ST437 (4,346,386 bp) isolates from the HGUV and downloaded from GenBank with Spanish origin.

The three ST11 clades were very distant from each other (>2000 SNPs), suggesting three different introductions in Spain (Supplementary Table 6.5, Supplementary Figure 6.3).

Lineage ST101 included three genomes collected in Madrid between 2015 and 2018 and in Valencia in 2017. The average pairwise SNPs within the clade was 14.9 (range 0-65) in 4 years, suggesting a rapid dispersion of the founder clone (Supplementary Table 6.6). ST437-OXA48 included only isolates sequenced from the HGUV and other Valencian hospitals and was a remarkably homogeneous clade (average 10.3 pairwise SNPs, range 0–21) (Supplementary Table 6.7, Supplementary Figure 6.3).

ST437-NDM had a basal genome collected in 2015 in Madrid but it was phylogenetically distant to the rest of the clade (64.1 pairwise SNPs on average) (Supplementary Table 6.7, Supplementary Figure 6.3).

Some database genomes fell into the ST307-OXA48 clade. However, all of them (except one from Oviedo) were from Valencia. Spanish genomes with *bla*$_{\text{OXA-48}}$ and *bla*$_{\text{KPC-3}}$ from Madrid and Guadalajara formed a new clade (ST307-SPAIN) while other isolates collected throughout the Spanish territory (including Valencia) did not fall into any lineage. The ST307-OXA48 clade (including the database genomes), maintained a low diversity (average of 25.2 pairwise SNPs, range 0- 102), meaning that all those genomes likely belonged to the same clonal expansion group. ST307-SPAIN isolates were more similar to the ST307-OXA48 clade than to the controls of the same hospital (201 vs 349 SNPs on average, respectively). This suggests that the origin of the ST307-OXA48 was not in the controls collected in the hospital (Supplementary Table 6.8, Supplementary Figure 6.3).

Remarkably, we detected different carbapenemase-coding genes (*bla*$_{\text{OXA-48}}$, *bla*$_{\text{NDM-1}}$, *bla*$_{\text{NDM-23}}$, *bla*$_{\text{VIM-1}}$, and *bla*$_{\text{KPC-3}}$) within highly related clades which involved both Spanish and HGUV strains (Figure 6.4, Supplementary Figure 6.2). This supports the idea that plasmid dissemination is a major factor responsible for the rapid spread of these carbapenemases in the HGUV and in Spain, as also shown in other studies [193].

6.6. *bla*$_{\text{OXA-48}}$ and *bla*$_{\text{NDM}}$ genes were disseminating in 3 and 2 different plasmids, respectively

To understand the variability and dissemination of the different carbapenemase genes in the different lineages, we studied the plasmids carrying *bla*$_{\text{OXA-48}}$ and *bla*$_{\text{NDM}}$ genes. For that, short reads were mapped to different references depending on the carbapenemase gene they carried. For the OXA48-producing isolates, we used the reference plasmid NC_019154, while for the NDM-producing strains, we used plasmid p146KP_NDM23 (NZ_CAKLAW010000003) extracted from the GCF_920939505 (146KP-HG) genome, a closed genome obtained in this work (§ 7) [456].

To study the spread of *bla*$_{\text{OXA-48}}$ in the hospital, we obtained a maximum likelihood tree from the core alignment of reads mapped to the reference plasmid NC_019154. This resulted in a 61 kb alignment with a median mapping coverage of 96% (range 30 - 96%) of the total plasmid length. Two isolates (33KP-HG and 66KP-HG) mapped only 30% of the plasmid length. These samples had the *bla*$_{\text{OXA-48}}$ as an integron or inserted in a different plasmid backbone and thus were removed from subsequent analyses. The phylogeny of the *bla*$_{\text{OXA-48}}$ plasmid (Figure 6.5 A) showed that at least 3 different plasmids or variants of the plasmid appeared almost simultaneously and coexisted at the hospital in several distinct lineages. We only considered as plasmid variants those that differed in at least 2 shared SNPs from any other variant (Supplementary Figure 6.5). We observed several cases of interlineage transfer of different plasmid variants. Nevertheless, one variant was mainly associated with lineages ST307 and ST101, another variant was mainly associated with the ST11-KL13 lineage, and the third variant was associated with lineages ST11-KL13 and ST437-OXA48.

To understand the dissemination of the NDM-plasmid, we mapped the sequencing reads of the samples to an NDM-plasmid (NZ_CAKLAW010000003) collected at the HGUV and obtained the maximum-likelihood tree shown in Figure 6.5 B. *bla*$_{\text{NDM}}$ genes were detected in two lineages, ST101 and ST437-NDM. All the isolates in the ST437-NDM lineage mapped with more than 80% coverage to the reference plasmid, whereas isolates in the ST101 lineage barely reached a coverage of 60%.

Based on the differences in the coverage and the number of SNPs between both (Supplementary Figure 6.6) lineages, ST101 and ST437-NDM, we inferred that they hosted different NDM plasmids, although these two lineages appeared in the hospital simultaneously.

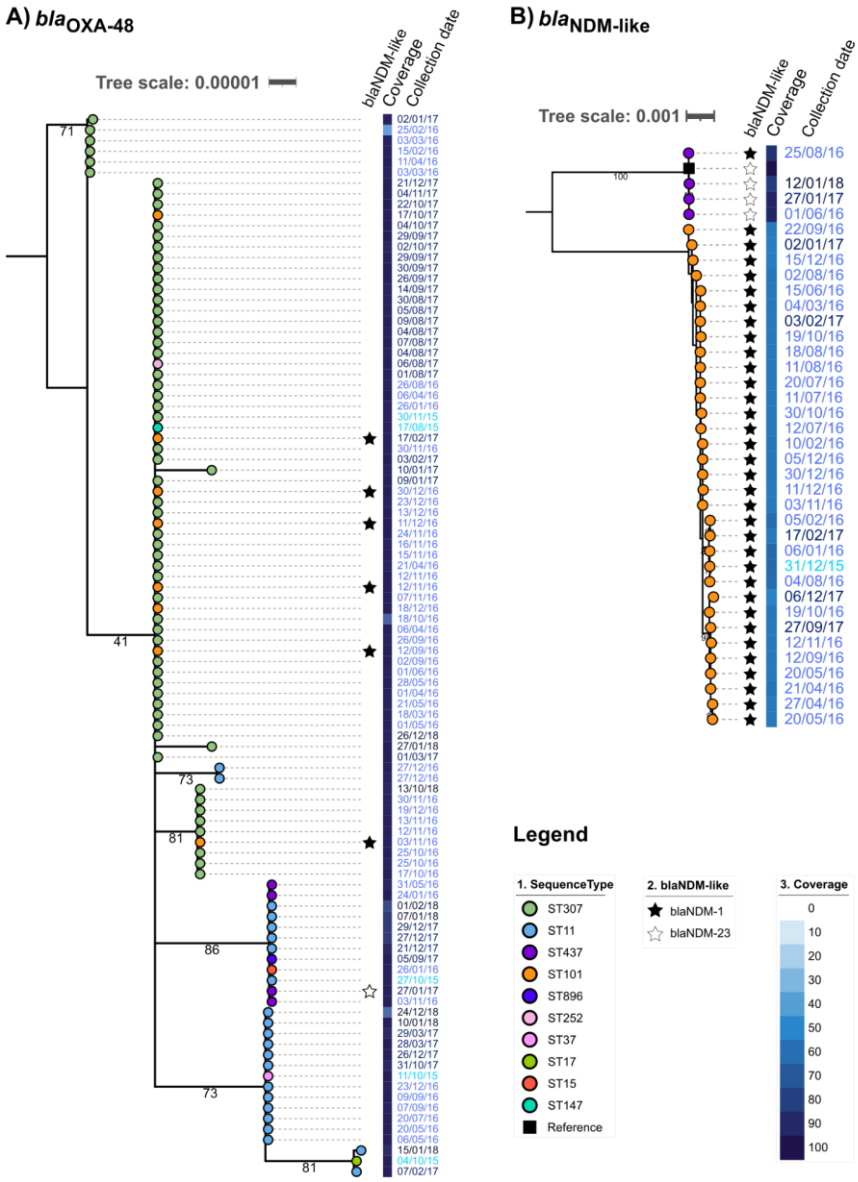


Figure 6.5. Maximum likelihood phylogenetic trees for plasmids carrying carbapenemase genes A) *bla*$_{\text{OXA-48}}$, and B) *bla*$_{\text{NDM}}$-like.

6.7. The emergence of carbapenem resistance in the hospital was associated with travel and dissemination from other Spanish regions

The lineages found in the HGUV were detected almost simultaneously, but they dispersed differently through time.

In the initial phase, before 2016, when the rise of isolates non-susceptible to carbapenem was noticed, and the major source of non-susceptibility to carbapenem was found to be the AmpC-producing ST11-KL105 lineage, along with occasional cases of isolates with truncated porins (Figure 6.6).

At some point in 2016, the AmpC-producing ST11-KL105 lineage was rapidly replaced by carbapenemase-producing isolates with two different carbapenemase types, OXA-48 and NDM. It was replaced by five different lineages that included two NDM-like-producing lineages, ST101 and ST437-NDM, and three OXA-48-producing lineages, ST437-OXA-48, ST307, and ST11-KL13.

For both the ST437-NDM and ST101 lineages, we found previous genomes in Madrid in 2015 that linked these cases to an already reported transmission chain [457]. Moreover, from epidemiological information, we could identify that the index patient in Madrid was the same as in the HGUV. In both hospitals, both lineages entered simultaneously in the same patient which was also related to travel. This patient was admitted to the Madrid hospital on his way home to Valencia from Pakistan, and a few months later he was transferred to the HGUV with the infection [457]. Lineage ST11-KL13 has basal strains collected years before, from 2013 to 2014, in the centre of Spain, which indicates that probably this lineage was disseminated from other Spanish regions to the HGUV. The ST307 lineage and ST11-KL105 show patterns of possible dissemination throughout the country but their emergence in the HGUV cannot be ascertained with the available data.

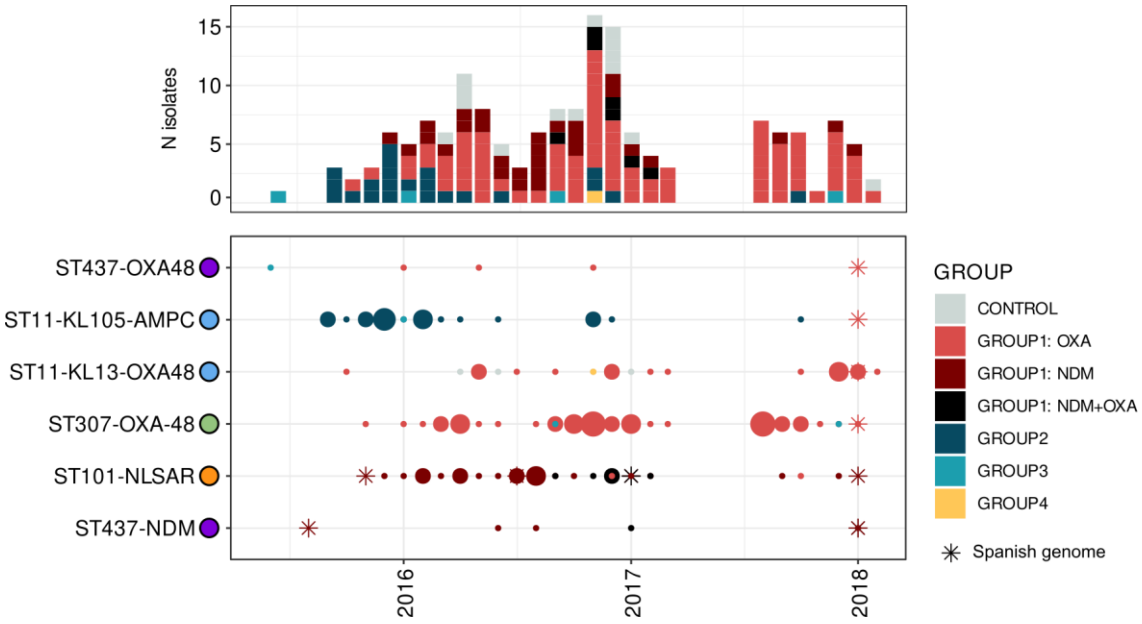


Figure 6.6. Distribution of Carba-R groups in each lineage throughout the sample collection period.

6.8 Concluding remarks

Our findings show that the initial carbapenem resistance emergence and dissemination of the HGUV occurred during a short period of one year and it was very complex. We found six different lineages comprising the majority of the CRKp population at the HGUV: ST307, ST11-KL105, ST11-KL13, ST101, ST437-OXA48, and ST437-NDM, disseminating different resistance mechanisms: AmpC, OXA-48, NDM-1, and NDM-23 in different plasmid variants. These lineages were in local clonal expansion with several cases of possible direct transmission within the hospital. Moreover, we demonstrate that most of the lineages found in the HGUV were previously present in other regions of Spain, those being probably the source of the relevant lineages at this hospital.

CHAPTER 7
Tracking the emergence and dissemination of the *bla*$_{\text{NDM-23}}$ gene

This chapter has been published in *Microbial Spectrum*:

García-González N, Beamud B, Fuster B, Giner S, Domínguez MV, Sánchez A, Sevilla J, Coque TM, Gimeno C, González-Candelas F; Networked Laboratory for Antimicrobial Resistance Surveillance of Comunitat Valenciana (Spain). Tracking the Emergence and Dissemination of a *bla*$_{\text{NDM-23}}$ Gene in a Multidrug Resistance Plasmid of *Klebsiella pneumoniae*. Microbiol Spectr. 2023 Apr 13;11(2):e0258522. doi: 10.1128/spectrum.02585-22.

One of the essential objectives of genomic surveillance is to inform about new threats and facilitate the improvement of intervention strategies to control their spread.

While searching for AMR genes in the NLSAR collection, we detected several isolates with a previously unreported variant of the NDM enzyme, denominated NDM-23. At the time of the finding, this variant was not available in NDM variants sequences have been submitted to the databases, but only 21 of those have been studied in detail [172]. Remarkably, some of them show an enhanced carbapenemase activity [458]. At that time, the *bla*$_{\text{NDM-23}}$ allele had been deposited as a complete coding sequence (CDS) in GenBank (accession MH450214.1). Yet, no information about its genetic context, bacterial lineage(s) host, or phenotypes had been published. Consequently, one goal of this thesis was to characterise this new NDM allele at different levels, from the nucleotide sequence to its epidemiology and evolutionary trajectory.

This is the first report of *bla*$_{\text{NDM-23}}$ in which its carbapenemase activity, genetic context, emergence, and dissemination are elucidated. First, we give a detailed characterization of the novel *bla*$_{\text{NDM-23}}$ gene and the antimicrobial resistance phenotype it confers. Secondly, we describe its immediate genetic context and the plasmid found to be carrying it. In addition, we postulate a possible origin of this plasmid. Lastly, we describe how this gene disseminated through the clonal spread of a *K. pneumoniae* ST437 strain, and how this strain emerged in the region.

7.1 Epidemiology of *bla*$_{\text{NDM-23}}$

To perform this analysis, we used the strains collected under the SKPCV project until March 2019. Note that, at this point, the SKPCV project had not concluded; therefore, not all SKPCV isolates were included in this analysis. Moreover, we also included 226 isolates from the NLSAR collection. These isolates were collected between 2004 and 2015 from two of the SKPCV hospitals.

A total of 8 patients were found to carry a *K. pneumoniae* isolate encoding a *bla*$_{\text{NDM-23}}$ (Table 7.1). We identified an isolate (179KP-HG) harbouring a *bla*$_{\text{NDM-1}}$ gene that was highly similar in both the chromosomal and the NDM plasmids (detailed

later), so we included it in the analysis. Remarkably, when all these patients were screened for carbapenem-resistant pathogens, three of them yielded negative results (Supplementary Table 7.1). This was likely because some of the methods were PCR-based with targets to the most common, and previously known, NDM variants, and NDM-23 was not in the database. We confirmed that after an update of the kits, NDM-23 was detected.

Table 7.1. Clinical data of the 8 patients with a *K. pneumoniae* NDM-23-producing infection and the one patient with NDM-1 included.

Sample	Carbapenemase	Isolation date	Hospital	Age	Sex	Diagnostic	Outcome
179KP-HG	NDM-1	25/08/2016	HGUV	85	Male	Sepsis of cutaneous origin	Exitus
146KP-HG	NDM-23	01/06/2016	HGUV	42	Male	Colonization of ventricular catheter	Unknown
262KP-HG	NDM-23; OXA-48	27/01/2017	HGUV	88	Male	UTI	Recovery
KP_HGUA01_012	NDM-23; OXA-48	20/04/2017	HGUA	65	Male	Cardiac transplant	Unknown
HGV2_364	NDM-23	12/01/2018	HGUV	77	Male	UTI	Unknown
HGV2_363	NDM-23	12/01/2018	HGUV	58	Male	Recurrent UTI	Relapse
HLF1_07	NDM-23	23/01/2018	HUPLF	83	Male	Diabetic foot	Recovery
HAV1_09	NDM-23; OXA-48	09/02/2018	HULR	99	Male	UTI by indwelling urinary catheter	Recovery
HLF2_50	NDM-23	30/01/2019	HUPLF	45	Male	Urinary tract disease in spinal cord injury	Recovery

Urinary Tract Infections (UTI)

All patients were male and with ages ranging from 42 to 99 years. Most were hospitalised (Table 7.1). Four of these patients were non-hospitalized but maintained a close relationship with the hospital as they had indwelling catheters or other health problems to check frequently. Thus, the infection was more likely to be hospital-acquired than community-acquired.

Worrisomely, all 8 *bla*$_{\text{NDM-23}}$-carrying strains were non-susceptible to almost all the antimicrobials tested, except to a few second-line and last-resort antimicrobials such as colistin, amikacin and fosfomycin and, occasionally, tigecycline (Table 7.2). Most of the patients recovered from the infection after receiving treatment. Although most clinicians choose aminoglycosides as treatment, different therapies were reported.

Table 7.2. Antimicrobial Sensitivity Testing results for the NDM-23 carrying strains. Cells are coloured regarding if they are interpreted as Resistant in red, Susceptible in green, Intermediate in yellow or unknown in white. MIC values are given in mg/L.

Sample	179KP-HG	262KP-HG	HLF1_07	KP_HGUA01_01	HAV1_09	146KP-HG	HLF2_50	HGV2_364	HGV2_363
Ampicillin	>16	>16	>8	>16	>16	>16	>=32	>16	>16
Amoxicillin-Clavulanic acid	>16/8	>16/8	>32	>16/8	>16/8	>16/8	>=32	>16/8	>16/8
Piperacillin-Tazobactam	>64	>64	>1	>64	>64	>64		>64	>64
Ceftazidime	>8	>8	>32	>16		>8	>=64	>16	>16
Cefotaxime	>16	>16	>32	>32	>16	>16	>=64	>32	>32
Cefuroxime	>16	>16		>16	>16	>16	>=64	>16	>16
Cefepime	>16	>16	>8	>16		>16	>=32	>16	>16
Cefoxitina	>16	>16	>16	>16		>16	>=64	>16	>16
Ertapenem	>1	>1	>1	32	>64	>1	>=8	>1	>1
Imipenem	8	>8	>8	32	>16	8	0.75	>8	>8
Meropenem	>8	>8	>32	6	>8	>8	0.5	>8	>8
Ciprofloxacin	>2	>2	>1	>2	>8	>2	>=4	>2	>2
Levofloxacin	>4	>4	>1	>4	16	>4		>4	>4
Gentamicin	>8	>8	>4	<=2	>8	>8	>=16	<=2	<=2
Tobramycin	>8	>8	>4	16	>1	>8	>=16	>8	>8
Amikacin	<=8	<=8	<=8	8	>8	<=8		<=8	<=8
Cotrimoxazole	>4/76	>4/76	<=32	>4/76	>2	>4/76	>=320	>4/76	>4/76
Fosfomycin	>256		<=32	<=32	>4			<=32	<=32
Colistin	1		<=2	<=2			1 mg/L	<=2	2
Tigecycline	<=1	2	<=1	2		2	0.5 mg/L	2	2

7.2 Properties of *bla*$_{NDM-23}$

The *bla*$_{NDM-23}$ sequence differs from that of *bla*$_{NDM-1}$ in one non-synonymous substitution at codon 101 (I101L). To test if this amino acid substitution changed the antimicrobial susceptibility profile of this gene in comparison to that of *bla*$_{NDM-1,}$ we selected two isolates of the collection with highly similar clinical and genomic backgrounds but with different *bla*$_{NDM}$ alleles: isolate 179KP-HG carried *bla*$_{NDM-1}$ whereas isolate 146KP-HG carried *bla*$_{NDM-23}$.

Table 7.3. Results of antimicrobial susceptibility tests for the clinical isolates and the TOP10 *E. coli* transformants carrying *bla*$_{NDM-1}$ and *bla*$_{NDM-23}$ genes. MIC values are given in mg/L.

	Clinical isolates				Transformants			
	179KP-HG (NDM-1)		146KP-HG (NDM-23)		TOP10-NDM-1		TOP10-NDM-23	
Ampicillin	>8	R	>8	R	>8	R	>8	R
Amoxicillin-clavulanate	>32	R	>32	R	>32	R	>32	R
Piperacillin-tazobactam	>8	R	>16	R	>16	R	>16	R
Cefuroxime	>16	R	>8	R	>8	R	>8	R
Cefotaxime	>8	R	>32	R	>32	R	>32	R
Cefepime	>16	R	>8	R	>8	R	>8	R
Cefiderocol [a]	0.75	S	>256	R	0.5	S	0.75	S
Aztreonam	>4	R	>4	R	<=1	S	<=1	S
Ertapenem [b]	>64	R	>64	R	>64	R	>64	R
Imipenem [b]	>64	R	>64	R	>64	R	>64	R
Meropenem [b]	>64	R	>64	R	>64	R	>64	R
Amikacin	<=8	S	<=8	S	<=8	S	<=8	S
Ciprofloxacin	>1	R	>1	R	<=0.06	S	<=0.06	S
Fosfomycin	<=16	S	<=16	S	<=16	S	<=16	S
Tigecycline	<=8	S	>2	S	<=1	S	<=1	S
Trimethoprim	>4		>4		<=2		<=2	
Cotrimoxazol	>4/76	R	>4/76	R	<=2/38	S	<=2/38	S
Colistin	<=2	S	<=2	S	<=2	S	<=2	S

[a] The minimum inhibitory concentration (MIC) was checked by test strips.
[b] The MIC was checked by the broth microdilution method.

We transformed these genes into TOP10 *E. coli* bacteria and found that the single amino acid variant does not produce changes in the susceptibility to carbapenems. The MIC values and growth curves for TOP10 *E. coli* transformants carrying *bla*$_{\text{NDM-1}}$ and *bla*$_{\text{NDM-23}}$ were coincident for all the antibiotics tested (Table 7.3, Supplementary Figure 7.1). *bla*$_{\text{NDM}}$ genes were responsible for a large part of the β-lactam-resistant profile of the clinical strains; however, these genes did not confer resistance to other antimicrobial classes. Remarkably, when we tested the activity of the new antibiotic cefiderocol, the antibiotic showed a high *in vitro* activity in most of the cases except for the clinical isolate 146KP-HG (*bla*$_{\text{NDM-23}}$), which had a CMI >256 (Table 7.3). The NDM-TOP10 transformants did not show cefiderocol-hydrolysing activity, meaning that neither NDM-1 or NDM-23 hydrolyse cefiderocol by themselves (Table 7.3).

7.3 *bla*$_{\text{NDM-23}}$ is located in a non-mobilizable, non-typeable, multidrug resistance plasmid

To decipher the genetic context of the *bla*$_{\text{NDM-23}}$ gene, we performed long-read sequencing of both isolates, 179KP-HG and 146KP-HG. The long hybrid assembly of isolate 146KP-HG revealed that it carried 4 plasmids and that the *bla*$_{\text{NDM-23}}$ gene resides in a 97 kb plasmid hereafter referred to as plasmid p146KP-NDM23 (Figure 7.1). The other three plasmids corresponded to a 118 Kb phage-like plasmid and 5 Kb and 3 Kb plasmids, none of which had AMR genes.

p146KP-NDM23 is a multidrug-resistant plasmid that, in addition to *bla*$_{\text{NDM-23}}$, has 18 additional genes associated with reduced susceptibility to nine different antimicrobial families, including β-lactams, sulphonamides, aminoglycosides, trimethoprim, fluoroquinolones, tetracycline, chloramphenicol, fosfomycin, tunicamycin, and rifampicin. All these genes are inserted in a large multiresistant region (MRR) that contains different putative transposable modules (Figure 7.1 A), which have been previously detected in IncFII plasmids [459, 460]. The MRR region includes the gene array comprising *bla*$_{\text{OXA-1}}$, a truncated *catB* gene, and *aac(6')-Ib-cr* flanked by two copies of IS*26*; the tetracycline resistance gene, *tetA,* associated with the Tn*As1* transposase; genes *aac(3)-IIa* and *tmrB* linked to a copy of IS*26*; and a

ΔTn*3* (*bla*$_{TEM-1B}$) with the transposase (*tnpA*) disrupted by an IS*Ecp1*-*bla*$_{CTX-M-15}$, with *sul2*, *strA,* and *strB* genes located downstream of *bla*$_{TEM-1B}$.

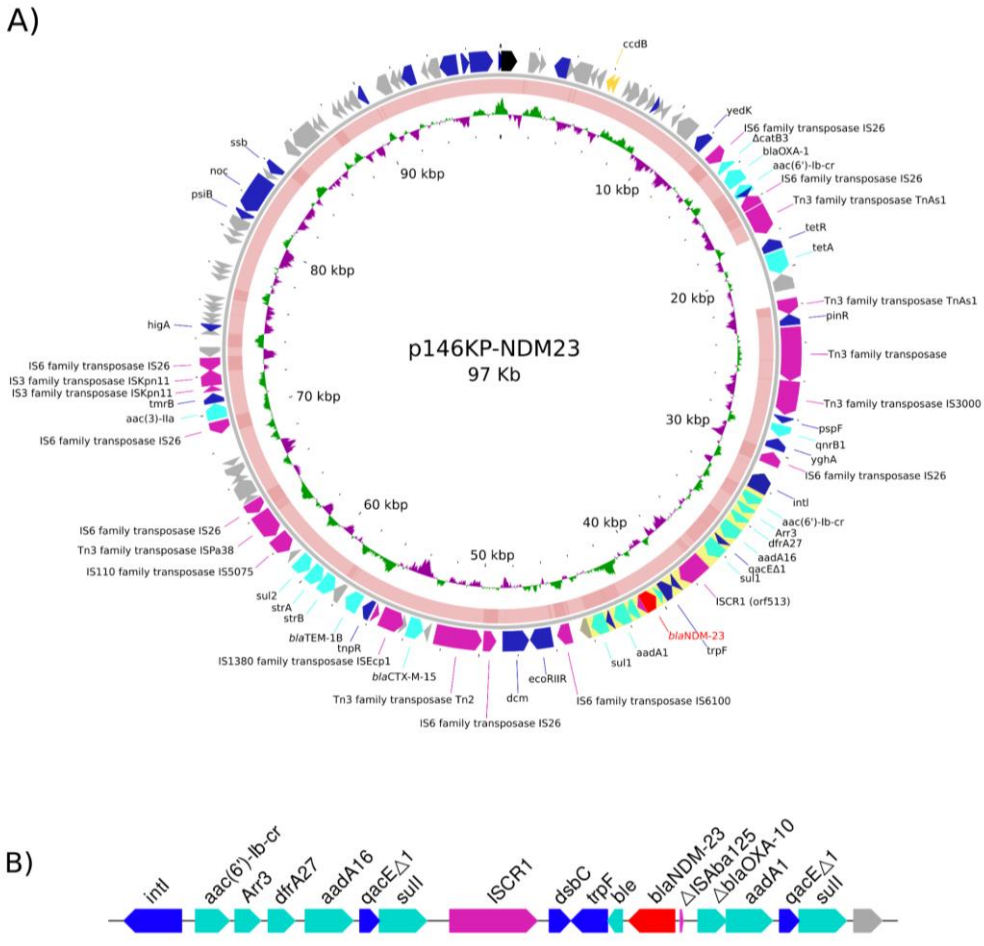


Figure 7.1. A) Gene content and structure of plasmid p146KP-NDM23. The *bla*$_{NDM-23}$ embedded in the second variable part of an ISCR1 complex class 1 integron is highlighted in yellow. The middle ring shows the BLAST comparison between p146KP-NDM23 and p179KP-HG. The inner ring shows the GC skew of the plasmid. B) ISCR1 Complex class 1 integron where the *bla*$_{NDM-23}$ gene is embedded.

The *bla*$_{NDM-23}$ gene is also in the MRR region but in a complex class 1 integron associated with the ISCR1 (Figure 7.1B). Similar genetic contexts have been described previously with other *bla*$_{NDM}$ variants [172]. The complex class 1 we found was structured as follows. The first variable region of the integron consisted of

acc(6')-Ib-cr, arr3, drfA27, and *aadA16* gene cassettes. Then, the ISCR1 region comprised an *aadA1* gene, a truncated bla_{OXA-10} gene, and a truncated IS*Aba125* downstream of the bla_{NDM-23} gene, while upstream we found a ble_{MBL} gene, a phosphoribosyl anthranilate isomerase gene (*trpF)*, and an oxidoreductase *dsbC*.

We could not identify the type of plasmid p146KP-NDM23 by replicon (Inc) nor relaxase (MOB) typing. Although a **BLAST+** [378] search with RepA returned 24 plasmids with 100% identity to the query sequence (Supplementary Table 7.2), no incompatibility type could be assigned to any of them. Apart from the *repA* gene, no similarity was found for the rest of the sequence of p146KP-NDM23 and any of the 24 plasmids.

We tested the plasmid mobility by *in vitro* and *in silico* analysis. We could not detect any relaxases or transfer genes in p146KP-NDM23. Thus, mobilisation of p146KP-NDM23 could only occur in trans with the help of co-resident plasmids; nevertheless, conjugation assays showed no mobilisation to the receptor cells. Although conjugation assays for p146KP-NDM23 were negative, our bioinformatic analysis revealed the presence of the complex class 1 integron in different plasmids in the *Enterobacteriaceae* family such as *K. pneumoniae*, *P. faecalis*, *E. coli* and *C. freundii* [461, 462], thus suggesting that this integron has mobilised to diverse genetic backgrounds, breaking the species barrier in the course of its evolutionary history. This might be explained either because the integron is flanked by IS26 transposases, which could mediate its transposition, or, although it has not been proven experimentally, by the independent mobilisation of ISCR1 elements [463]. Nevertheless, further work will be needed to understand the mobilisation of these genetic elements in different strains and species.

7.4 Origin of the p143KP-NDM23 plasmid

To understand how this plasmid appeared in the hospitals of this study, we wanted to decipher its origin. When comparing the plasmid p146KP-NDM23 with the plasmid carrying bla_{NDM-1} in the 179KP-HG clinical strain, we found that the plasmid was almost identical to p146KP-NDM23 but with the bla_{NDM-1} allele and lacking a TnAs1 transposable element that carried a *tetA* resistance gene (Figure 7.1A). We named this plasmid p179KP-NDM1.

Regarding the origin of the plasmid, we found that plasmid pKDO1 (accession JX424423) was the most similar plasmid to p146KP-NDM23 in terms of identity and shared genes (100% coverage and identity, Supplementary Table 7.3) [464]. pKDO1 contains almost all the genes found in plasmid p146KP-NDM23 rearranged but lacks the ISCR1 complex class 1 integron where the *bla*$_{NDM-23}$ gene is embedded (Figure 7.2). Plasmid pKDO1 was described in an ST416 *K. pneumoniae* strain in the Czech Republic in 2009 [464] whereas p146KP-NDM23 was isolated in our study from an ST437 *K. pneumoniae*. Although p146KP-NDM23 and pKDO1 share the *repA* gene, pKDO1 could be typed as an FII(K) plasmid whereas p146KP-NDM23 could not.

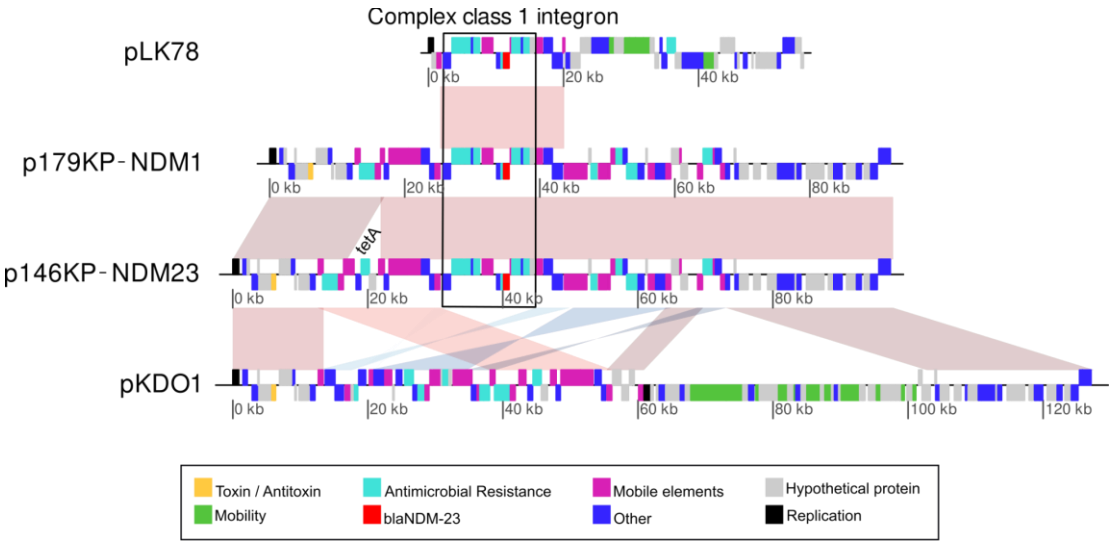


Figure 7.2. Structure comparison of plasmids pLK78, p146KP-NDM23, pKP179-NDM1, and pKDO1. pLK78 shares similarity with the integron sequence found in pKP146-NDM23, pKP179-NDM1, while pKDO1 shares similarity with the backbone of those plasmids.

Although the complex class 1 integron was not found in pKDO1, the integron associated with the *bla*$_{NDM-23}$ gene has been described previously [172]. Nevertheless, the structure around the *bla*$_{NDM-23}$ gene in p146KP-NDM23 is not the usually conserved structure found around *bla*$_{NDM}$ genes. In this case, we found the presence of a truncated *bla*$_{OXA-1}$ downstream of the *bla*$_{NDM}$-ΔIsaba125. We found, by BLAST analysis, that an exact complex class 1 integron, but carrying a *bla*$_{NDM-1}$, was already identified in an IncN3 plasmid, denoted as pLK78 (accession KJ440075)

(Supplementary Table 7.4) [461]. Plasmid pLK78 was isolated in 2012 in Taiwan and collected from a *K. pneumoniae* strain [461]. Plasmids pLK78 and p146KP-NDM23 only have the ISCR1 complex class 1 integron with the bla_{NDM} gene in common (99.9% of identity and 100% of coverage), while the rest of the plasmid has no similarity (Figure 7.2). The identity between these two integrons is not perfect because of the single substitution that differentiates the bla_{NDM-1} allele from bla_{NDM-23}.

Although it is not possible to learn the details about how the p146KP-NDM23 backbone and integron converged, we guess that, at some point, they were found together in the same cell. We detected the presence of the ISCR1 complex class 1 integron, and pKDO1 and pLK78 plasmid sequences with coverages of 100%, 99.83%, and 92.39%, respectively, in the genome of an ST101 *K. pneumoniae* isolate named PN4 [462]. This strain has coverage of 95.1% for the p146KP-NDM23 plasmid. However, we could not decipher the structure of each mobile genetic element in PN4 because it was available only as a draft genome assembly. PN4 is a strain involved in a multispecies outbreak in 2010 in Pakistan. Thus, we hypothesise that the PN4 lineage could be a possible origin where pKP146-NDM23 could have formed as we have identified all the necessary structures in a single isolate. Moreover, we found a conjugation module in pKDO1 but not in pKP179-NDM1 nor pKP146-NDM23, suggesting that the latter plasmids lost their mobilisation capabilities.

7.5 The recent evolutionary history of bla_{NDM-23}-producing isolates

To understand the epidemiological and evolutionary dynamics of the new bla_{NDM-23} gene and its carrier plasmid, we obtained a phylogeny of all the ST437 isolates collected in the NLSAR surveillance program, carrying bla_{NDM} genes (8 isolates) or not (14 isolates) (see Material and methods §3.9.2). We also included all the ST437 genomes from the RefSeq (121 isolates, accessed on 27/04/2021) which include 33 isolates from five different Spanish regions, and isolates from areas considered as the main reservoirs of bla_{NDM}-producers, such as India, Pakistan, and the Balkans (Serbia and Croatia) (Supplementary Table 7.5).

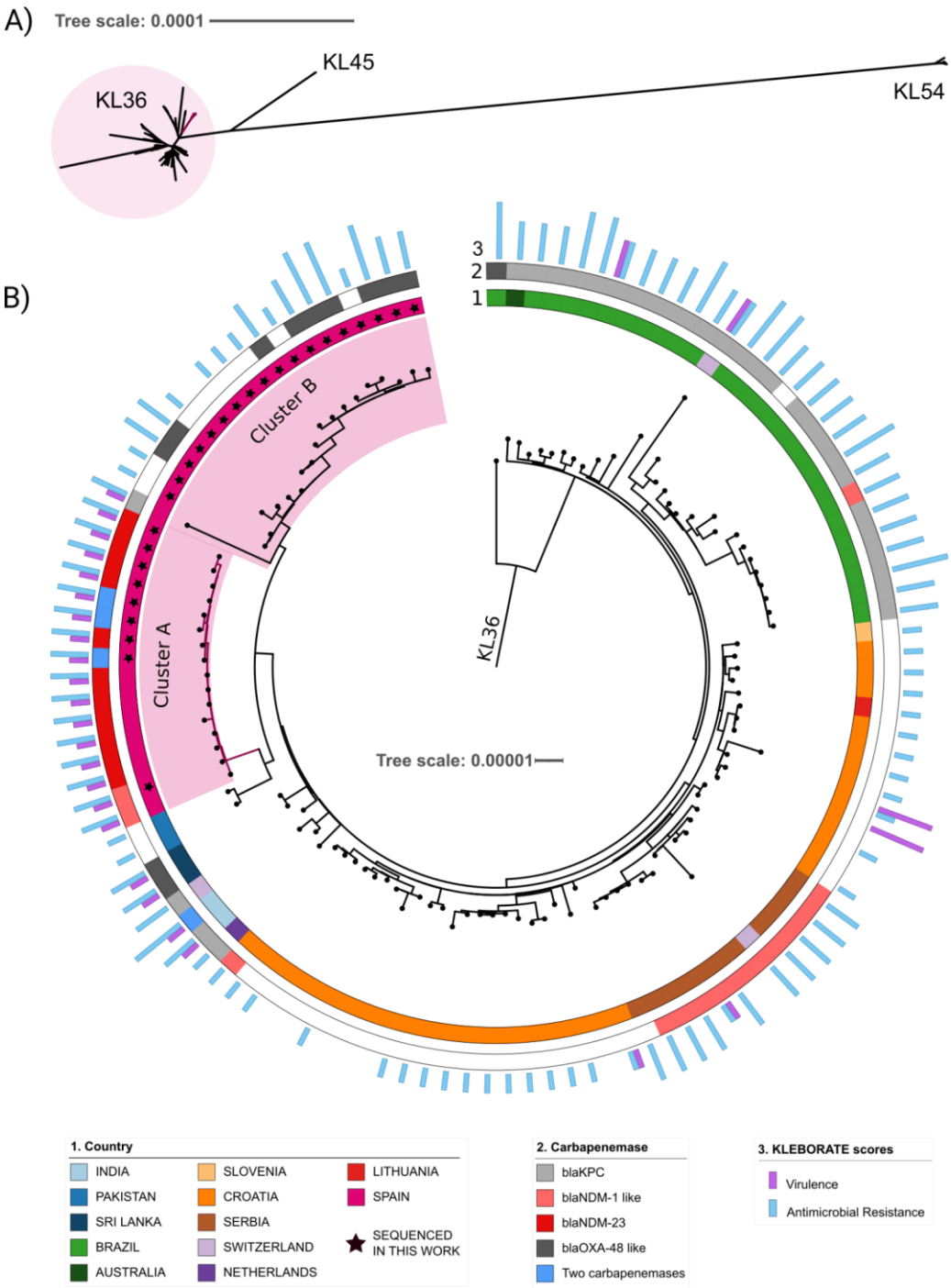
To decipher the relationships among these isolates, we used a mapping approach using the chromosome of the 146KP-HG genome (Accession: GCF_920939505.1) as

a reference. This genome was sequenced by long and short reads and assembled in this work (§3.6). To avoid including highly divergent regions in the analyses, five complete prophages identified by PHASTER and an ICEK*p* yersiniabactin inserted into the reference chromosome were removed from the genome alignment (Supplementary Table 7.6). Mapping statistics for each genome and pairwise distance matrix are documented in Supplementary Tables 7.7 and 7.8, respectively.

The phylogenetic analysis revealed that the ST437 global population collected in this work is divided into three main groups corresponding to different capsular types KL36, KL54 and KL45 (Figure 7.3 A, Supplementary Figure 7.2). Only the isolates belonging to capsular type KL36 (Figure 7.3 B) were kept for further analysis as all the *bla*$_{\text{NDM-23}}$-producers were found in KL36 isolates.

All *bla*$_{\text{NDM-23}}$ isolates were clustered in Clade A (Figure 7.3 B). This cluster contains 16 isolates, all of them with a *bla*$_{\text{NDM}}$ gene in the accessory genome and collected in Spain. This clade comprises two basal isolates (GCA_011684095 and 179KP-HG) carrying the carbapenemase gene *bla*$_{\text{NDM-1}}$ and 14 isolates that carry the novel *bla*$_{\text{NDM-23}}$ allele. The low number of SNPs among cluster A genomes (0 to 30, with an average of 11) and the highly conserved accessory genome of this clade suggest a recent common ancestry and a case of recent clonal dissemination.

Figure 7.3. (Next page). A) Whole genome maximum likelihood tree of ST437, with the KL36 clade highlighted in pink. B) Whole genome maximum likelihood tree of ST437 – KL36 isolates. Spanish isolates are highlighted in light pink. Colours in the inner ring (1) indicate the country of origin of each sample. A star inside this ring indicates that the isolate has been sequenced in this work. Colours in the middle ring (2) represent carbapenem-resistant genes. In the outer ring (3), blue bars show the resistance score while purple bars show the virulence score obtained from Kleborate (17). The scale represents nucleotide divergence as substitutions per site.


A)
Tree scale: 0.0001
KL36
KL45
KL54
B)
3
2
1
Cluster B
Cluster A
KL36
Tree scale: 0.00001
1. Country
INDIA
PAKISTAN
SRI LANKA
BRAZIL
AUSTRALIA
SLOVENIA
CROATIA
SERBIA
SWITZERLAND
NETHERLANDS
LITHUANIA
SPAIN
SEQUENCED IN THIS WORK
2. Carbapenemase
blaKPC
blaNDM-1 like
blaNDM-23
blaOXA-48 like
Two carbapenemases
3. KLEBORATE scores
Virulence
Antimicrobial Resistance

Clade A showed a highly conserved accessory genome. All these isolates presented a yersiniabactin virulence factor (*ybt9*) in an ICE*Kp3* island inserted in the chromosome and largely the same plasmids. Nevertheless, we found changes during the dissemination process, with gains and losses of complete mobile genetic elements. For instance, samples 262KP-HG, HAV1_09, and KP-HGUA01_12, have an additional IncL plasmid harbouring a $bla_{\text{OXA-48}}$ gene. Contrarily, some samples have lost a few complete plasmids: sample KP-HGUA01_12 lacks 118 kb, 5 kb, and 3 kb plasmids, and samples HLF2_50, HGV2_363, and HGV2_364 lack the 118 Kb plasmid. Remarkably, we found 100% identity for the p146KP-NDM23 plasmid among all the isolates showing complete conservation at the nucleotide level during its recent and short (5-year) dissemination. Yet, we found several differences at the structural level with an average coverage of 97.5%, ranging from 84% to 100% (Supplementary Table 7.9, Figure 7.4), meaning that the plasmid has accrued several structural changes which were mainly deletions (Figure 7.4).

Notably, we found differences in chromosomal content that affected cefiderocol resistance determinants. Although little is known about the resistance mechanisms to this new antimicrobial compound, resistance has been reported for *K. pneumoniae* strains that carry a combination of mutations in *cirA* genes and NDM production [115], or the combination of β-lactamases with mutations in the *baeS* gene (baeSR regulon) [379,465]. The difference in the MICs can be explained because 179KP-HG lacks the baeSR regulon while 146KP-HG and the remaining isolates of the clade carry the regulon with a mutation in the *baeS* gene (F3L) recently related to reduced susceptibility to cefiderocol, but with a lower MIC than previously reported (4 mg/L vs 256 mg/L) [379].

7.6 Origin of the $bla_{\text{NDM-23}}$ clade

Spanish isolates clustered in two well-defined clusters A and B (Figure 7.3 B), where the remaining isolates collected in Spain and in the NLSAR surveillance program were found. Although cluster B contains NLSAR isolates collected from the same hospitals and dates as those in cluster A, the phylogenetic tree shows that these clusters likely represent two different introductions of ST437-KL36 in the CV. This is also supported by the on average 141 SNPs (range 84 - 275) and the very different

accessory genomes differing between cluster A and cluster B (range 224 - 1,281 genes). Hence, the Spanish non-NDM-producing ST437 isolates sequenced in this work were not the founder population of the *bla*$_{NDM-23}$-producing isolates.

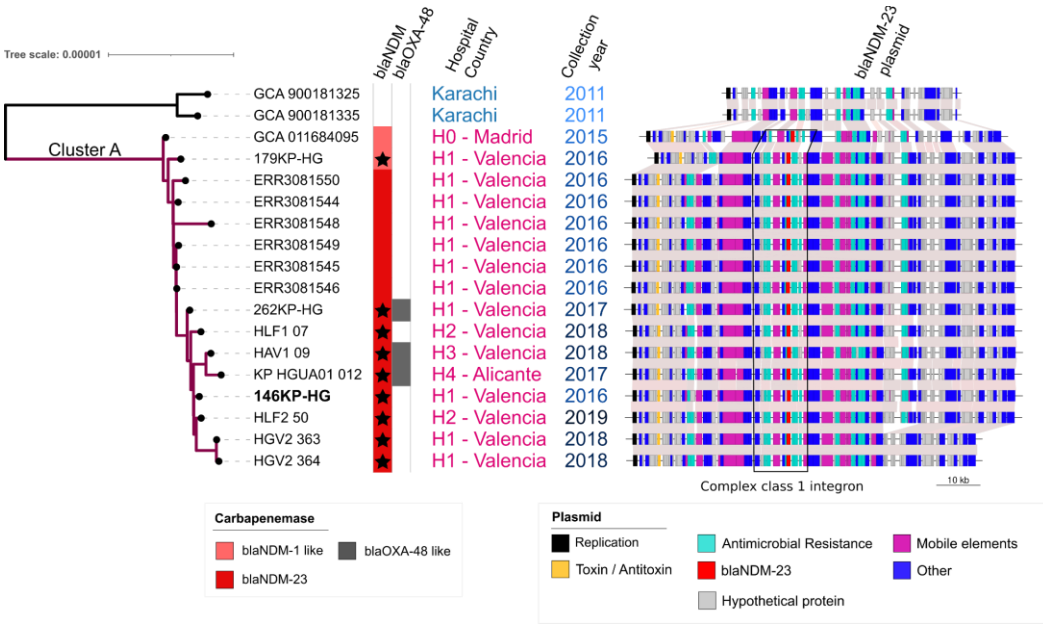


Figure 7.4. Dissemination of the ST437 clone carrying the *bla*$_{NDM-23}$ gene involves 5 different hospitals over 5 years. Isolates sequenced in this work are marked with a star inside the *bla*$_{NDM}$ strip. For each isolate, the structural variants of plasmid p146KP-NDM23 are shown.

The phylogeny (Figure 7.3 B) shows that the closest isolates to those in cluster A are two samples from Pakistan collected in 2011 (GCA_900181335 and GCA_900181325). These isolates show an average difference of 146 SNPs (range 131 - 157) with cluster A isolates. These SNPs are dispersed over the whole genome, and they do not cluster in one or a few loci. Hence, we can eliminate the possibility that they have arisen from recombination or horizontal gene transfer. In addition, the Pakistani and cluster A samples share an *E. coli* phage, phiV10 (accession NC_007804), inserted in the chromosome that is not present in any of the other samples, further supporting their common ancestry.

The index case of the dissemination of this clone, GCA_011684095, is the most basal genome of cluster A and was also the index case of a multi-clonal and multispecies

outbreak detected in hospital Ramón y Cajal (HRyC) in Madrid [457]. This patient was a 36-year-old man, native of Pakistan with residence in Valencia. The patient had received healthcare assistance in Pakistan following a traffic accident in November 2014. On 26 August 2015, upon return from Pakistan, he was admitted to the neurosurgery ward in the HRyC in Madrid. In December 2015, the patient was transferred to Valencia and admitted to the HGUV. A few months later, this patient went through the emergency ward in the HUPLF and was admitted to the general ward. This explains the fast spread of the lineage through the region but also points to Pakistan as the possible origin of the strain causing this dissemination event.

The close similarity of the chromosome sequences of isolates in clade A (Figure 7.4) and two isolates from Pakistan, the origin of the p146KP-NDM23, and the epidemiological link between the index case in Spain and that country strongly suggest the origin of cluster A sequences in Pakistan and not in other ST437 isolates previously found in Spain (cluster B). TempEst analysis estimated the tMCRA between the Pakistani and cluster A samples in 1999 (R^2=0.78).

7.7 Transmission of plasmid p146KP-NDM23 carrying *bla*$_{\text{NDM-23}}$ by recent clonal dissemination of ST437

According to the phylogenetic analysis and the dates of isolation of the relevant samples, the *bla*$_{\text{NDM-23}}$ mutation had to occur between the arrival of the clones at the hospital and the first detection of *bla*$_{\text{NDM-23}}$ genes in that same hospital. Isolate 179KP-HG, being basal and carrying variant *bla*$_{\text{NDM-1}}$, suggests that the emergence of the *bla*$_{\text{NDM-23}}$ allele was likely due to a point mutation in the plasmid during the expansion of the lineage. Temporal analysis showed a very recent tMCRA of cluster A in mid-year 2013 (R^2 = 0.47), thus implying that the mutation had to occur between 2013 and 2016. After that, the p146KP-NDM23 plasmid remained genetically stable over time.

The dissemination of this ST437 isolate carrying the p146KP-NDM23 and the *bla*$_{\text{NDM-23}}$ affected HRyC in Madrid (GCA_011684095) and at least four different hospitals (HGUV- HGUA) in the Comunitat Valenciana, spanning from 2016 until at least 2019, when our sampling concluded. Eight *bla*$_{\text{NDM-23}}$ isolates were collected in

the NLSAR surveillance program, and six isolates were downloaded from the database. The database genomes were isolated in Valencia in 2016 [168, 457] and all of them carried the *bla*$_{\text{NDM-23}}$ allele. This implies that the *bla*$_{\text{NDM-23}}$ allele was already in Spain and Valencia before we noticed it in the surveillance of NLSAR.

7.8 Concluding remarks

The dissemination of this ST437 carrying the *bla*$_{\text{NDM-23}}$ affected at least four different hospitals of the Comunitat Valenciana from 2016 until at least 2019, when our sampling concluded.

Here, we have used genomic epidemiology to describe the emergence and dissemination into several hospitals of a multi-drug-resistant plasmid carried in a *K. pneumoniae* ST437 strain carrying a new *bla*$_{\text{NDM-23}}$ carbapenemase gene and 18 more antimicrobial resistance genes. By means of bioinformatic and phylogenetic analyses, we have been able to trace the evolutionary and epidemiological route of the new allele, the hosting plasmid, and the strain that carried both from Pakistan to Spain.

CHAPTER 8
DISCUSSION

The rise and spread of antibiotic-resistant infections pose a global health threat [8, 9], as the decreasing number of effective treatments leads to more severe infections, longer hospital stays, higher healthcare costs, and increased mortality [10]. Currently, *K. pneumoniae* is a major contributor to the burden of antimicrobial resistance in hospital settings, with 3GC and carbapenem-resistant strains considered priority pathogens that require urgent research and development for new antibiotics [10, 18, 21].

The high prevalence of infections caused by 3GC-resistant *K. pneumoniae* strains, particularly in Europe [12], has resulted in an increased use of carbapenems, as they are the last resort treatment option for infections caused by multidrug-resistant bacteria, including 3GC-resistant *K. pneumoniae* strains. However, the widespread use of carbapenems has led to the emergence of carbapenem-resistant strains, which represent a significant obstacle in the management of *K. pneumoniae* infections and many other clinically relevant species, including *Enterobacteriaceae*. Carbapenem-resistant bacteria limit the availability of effective antibiotics to last-resort drugs, which are frequently associated with reduced efficacy and/or increased toxicity [107–109], resulting in high mortality and morbidity rates [48, 49]. While carbapenem-resistance has historically been low, in recent years, a global increase has been reported [10, 431], with carbapenem-resistant *K. pneumoniae* identified as the fastest-growing AMR pathogen in clinical settings [12, 105]. This highlights the urgent need to address the growing threat of 3GC- and carbapenem-resistance in *K. pneumoniae*.

For this, it is crucial to comprehensively understand the emergence and spread of carbapenem-resistant *K. pneumoniae* in hospitals and communities [10]. Previous studies have highlighted the role of hospital transmission in the spread of this pathogen [151, 263, 466], while others have investigated the importation of cases from endemic regions [167], the community spread [254, 467, 468], and the interspecies transfer of AMR plasmids [469]. Recently, the One Health approach has been employed to address these questions. However, several studies have shown limited transmission between clinical and non-clinical settings [330, 470], suggesting

that there could be other factors responsible for the emergence and spread of this pathogen in clinical settings.

Despite the extensive research on *K. pneumoniae*, many questions remain unanswered about how resistant strains of *K. pneumoniae* emerge and colonise hospital settings, the relative contribution of each different colonisation/entering route, the role that hospitals play as disseminators at the regional level, and the global contribution to local epidemics.

In this thesis, we have investigated the epidemiology of 3GC- and carbapenem-resistant *K. pneumoniae* isolates in several hospitals in the Comunitat Valenciana, utilising genomic data to shed light on several unanswered questions. This study has made significant contributions to: i) the creation of a comprehensive collection for the genomic surveillance of *K. pneumoniae* in the CV; ii) unravel the population structure and genetic diversity of *K. pneumoniae* in clinical settings of the CV, as well as the evolution of the global HRC ST307; iii) a deeper understanding of the complex colonisation of carbapenem-resistant *K. pneumoniae* strain in hospital settings; iv) elucidate the emergence of a novel $bla_{\text{NDM-23}}$ allele, and demonstrate how carbapenemase-imported cases can rapidly develop into regional transmission through clonal expansion.

8.1 Comprehensive collection for genomic surveillance in the CV

From 2017 to 2019, we implemented the SKPCV program to achieve a comprehensive representation of the hospital population of *K. pneumoniae* and, especially, of that associated with ESBL- and carbapenem-resistance. The program was carried out under the NLSAR framework (§1.9, Figure 1.12) and enrolled 8 of its hospitals. The only inclusion criterion for the project was the selection of the 30 first ESBL- and/or carbapenemase-producing *K. pneumoniae* isolates and the 10 first 3GC-susceptible clinical isolates for each hospital and trimester. This resulted in a large sample collection of 1,774 isolates. Additionally, we included 414 retrospective isolates collected from three NLSAR hospitals from 2004 to 2017, resulting in a total of 2,188 isolates.

To provide context and establish a collection that would allow us to elucidate the relationships between NLSAR and global *K. pneumoniae* isolates, we included external data. As the cost of genome sequencing has decreased, more and more genome data are being generated and the number of available genomes has grown rapidly [471]. By the time we interrogated the databases, we were able to download a total of 11,553 *K. pneumoniae* genomes from three different sources: RefSeq (n=10,283), GenBank (n=589), and ENA (n=681).

Large genomic datasets are a valuable resource, but they can also be an obstacle if not handled properly, as shown in this thesis. Ensuring data quality is essential, as errors in sequencing, assembly, and contamination can affect the accuracy of downstream analyses [428,472]. Therefore, it is widely accepted that appropriate quality control measures must be applied in genomic surveillance pipelines. However, there is no consensus on which filters should be applied or in what order, and which programs and parameters yield better results.

Indeed, quality control measures are crucial when working with clinical samples and species like *Klebsiella*. In our study, the clinical isolates were collected, cultured, and processed by different clinical microbiologists at each hospital. Moreover, the species diversity of *K. pneumoniae*, its complex population structure, and the high horizontal gene transfer rates [31] make contaminant detection challenging. Furthermore, the evolving taxonomy of *Klebsiella* [23–25] results in many genomes in public databases being incorrectly assigned, as demonstrated in our study. The number of genomes available for this pathogen [471] also pose challenges to select appropriate context samples and ensure good quality *K. pneumoniae* genome data.

In chapter 4, we have emphasised the relevance of conducting an exhaustive filtering workflow before performing any analysis using your own data or genome sequences from public databases. To build a well-curated genome database, we had to apply a combination of five filters using six different programs as no single program can tackle all the quality control issues entirely on its own. Initially, we assessed taxonomic assignment and interspecies contamination, followed by assembly quality, interspecies contamination, intraspecies contamination, and finally, genomic similarity across the entire collection. By implementing this tiered

filtering approach, we were able to eliminate 8% of the NLSAR-collected genomes and 3% of those retrieved from public databases. It is worth mentioning that we found misidentification or potential contamination with non-target species in 109 genomes obtained from databases (Table 4.1). These findings highlight the importance of analysing genomes from any source to ensure their quality before using them for downstream analyses.

Generally, in large *Klebsiella* genomic studies, these steps are not implemented and filters are based on assembly metrics and the species identification by Kleborate that, although highly valuable, do not have the power to detect smaller/closer contamination cases [151, 156, 359]. Indeed, despite the well-known issue of misidentification and cross-species contamination in publicly available genomic data, as shown in this work and elsewhere [427, 428, 473], many studies also skip filtering database-downloaded genomes or do not detail them [330, 392].

The use of WGS in Clinical Microbiology laboratories has become increasingly common in recent years, with many laboratories adopting this technology as a routine practice for pathogen surveillance [323, 324]. Yet, as shown here, contamination and misidentification are frequent when working with highly diverse collections in terms of sample preparation or sources. Therefore, we sustain that both quality and contamination aspects should be considered when designing routine protocols in Clinical Microbiology laboratories and that routine genomic epidemiology workflows should implement these or similar tiered control filters.

We are aware that the complexity and number of tiered steps in our workflow may limit its implementation in more general and rapid genomic surveillance pipelines but, when we performed these analyses, there was no tool available that could complete all these analyses together. However, a recent upgrade in Kleborate has added an assembly quality control which assesses most of the features mentioned [32]. Nonetheless, this tool only works for *Klebsiella*, which impedes its application in more general pipelines needed for pathogen surveillance in hospitals. Thus, to implement WGS as a routine surveillance methodology, efforts to standardise key steps such as filtering and contamination removal should be established to ensure accurate and reproducible results across laboratories [361].

The application of these quality control filters has allowed us to obtain a large dataset of curated genomes with the aim of serving as a valuable resource to study the genomic epidemiology of *K. pneumoniae* AMR in the Comunitat Valenciana (Figure 4.5).

8.2 The NLSAR collection in a global context

In chapter 4, we also characterised the genome collection and assessed the similarity of database genomes to those collected in the NLSAR hospitals. The final NLSAR collection comprises 1,604 sequenced isolates, collected under the SKPCV project in 8 different CV hospitals (Figure 4.4), and 395 isolates from other retrospective projects. The NLSAR isolation dates started as early as 2004, spanning 16 years, whereas the SKPCV project was conducted between 2017 and 2019 (Figure 4.5), with some isolates from early 2020 before the Covid-19 pandemic interrupted normal activities at the hospitals. The inclusion of isolates from multiple hospitals over an extended period provides a comprehensive view to better understand the epidemiology of *K. pneumoniae* in the CV region and enables tracking the evolution and spread of antimicrobial resistance in this pathogen through time.

This *K. pneumoniae* genome collection is one of the largest datasets available for understanding the genomic epidemiology of this pathogen in a specific region. Although large genomic studies have been conducted to investigate the distribution of human carbapenem-resistant *K. pneumoniae* infections in Europe (1,717 genomes) [151], in a hospital in Houston, Texas (1,777 genomes) [284] or to explore the overlap between isolates in clinical and non-clinical settings, which included 3,482 genomes [330], genomic epidemiology datasets usually include only a few hundred isolates, either at the national level [156, 227, 263, 474, 475] or focused on a specific hospital or city [262, 392, 466, 476].

In addition, to the NLSAR we included 11,134 database genomes in the global collection. It is important to note that having more genomes does not necessarily equate to having more information or diversity. Despite the vast number of genomes collected in the database, there are still many regions and countries that

remain underrepresented or not represented at all. This is more acute in LMICs even though the burden of AMR and *K. pneumoniae* is higher and more significant there than in most of the highly represented countries [10]. Indeed, we have found that Africa, which was the continent with the least amount of sampling, is the most diverse in terms of lineages (Figure 4.6). Furthermore, we have found that none of the other highly sampled regions reached a plateau in terms of diversity, indicating that there is still much *K. pneumoniae* diversity that has yet to be studied (Figure 4.6). It is worth noting that some of the countries that remain underrepresented have already been identified as reservoirs of carbapenem-resistant strains [166]. Thus, despite intense sequencing efforts, it is possible that many lineages, genetic characteristics, and sources and reservoirs of relevant lineages remain unknown.

Furthermore, it is important to consider that the low diversity observed in regions such as America and Europe may not be fully representative of the actual situation therein, as their collections are highly biased towards specific AMR determinants and high-risk clones (Table 4.2). For instance, the available genomes in public databases for *K. pneumoniae* are heavily biased towards MDR and HRCs, particularly the KPC-producing ST258 lineage in America and Europe, as well as towards convergent genomes in Asia and América (Table 4.2, Figure 4.8, Figure 4.10). This bias is strongly influenced by local problematic strains and the economic resources of each region. Therefore, it is crucial to note that the prevalence of AMR genes, virulence factors, and lineage distribution in the global database do not accurately reflect the real epidemiological situation in any given region. Despite this limitation, the database can be used to track pathogens and identify possible links beyond our region [426]. Our findings about the global populations are highly similar to those found in Lam *et al*. [32] where they present the Kleborate software by applying it to 13,156 publicly available *Klebsiella* genomes. This result is expected at a large scale because of the coincidence among the datasets, despite the inclusion of the NLSAR isolates in our analyses.

Interestingly, we have observed that the NLSAR and the Spanish collections exhibited a similar lineage composition, both being mostly composed of ST307 and ST11 lineages (>70%), along with several cases of ST15, ST405, ST147, and ST101

(Table 4.2). These collections are also similar to the European one, except that the latter is enriched in ST258 and ST512 lineages. We have also performed an analysis of the resistance determinants to 3GC and carbapenems found in the NLSAR genomes, as well as their plasmid vectors. Moreover, we have compared both to those present in other regions. Our findings have revealed that the most prevalent ESBLs in the NLSAR collection are *bla*$_{\text{CTX-M-15}}$, *bla*$_{\text{SHV-12}}$, and *bla*$_{\text{CTX-M-1}}$, while the carbapenem resistance genes are mainly *bla*$_{\text{OXA-48}}$, *bla*$_{\text{VIM-1}}$ (n=37), *bla*$_{\text{NDM}}$ variants including *bla*$_{\text{NDM-1}}$ and *bla*$_{\text{NDM-23}}$, and *bla*$_{\text{KPC}}$ (Figure 4.9). Most of these resistance determinants and plasmids detected in the NLSAR collection were also present in other regions, particularly in Spain and Europe (Figure 4.9). Indeed, the OXA-48 and VIM carbapenemases have been reported as endemic in Spain [151, 156]. Additionally, VIM has been found to be largely restricted to the Mediterranean region, especially to Spain [146, 156]. The discovery of similar features outside the NLSAR collection is highly valuable, as it can aid in investigating possible links and transmission events between the Comunitat Valenciana and other regions across the globe.

We have identified a few exceptions of AMR determinants and lineages that were predominantly limited to our NLSAR collection or Spain. We have observed that the NDM-23 enzyme and the plasmid carrying it were only present in the NLSAR collection or in Spanish genomes (Figure 4.13). Furthermore, we found that the combination of *bla*$_{\text{OXA-48-like}}$ with *bla*$_{\text{CTX-M-1}}$, which was restricted to the ST11 lineage, was almost exclusively observed in the NLSAR collection and database genomes collected in the CV (Figure 4.9). Of note, some of those genomes were described in *K. pneumoniae* and *E. coli* strains collected from wild avian species in Spain [248], which might indicate certain transfer between clinical and non-clinical environments. Additionally, we have not identified plasmids in the custom plasmid database similar to one of the *bla*$_{\text{VIM}}$ NLSAR plasmids (Figure 4.13) and one of the *bla*$_{\text{KPC}}$ NLSAR plasmids (Figure 4.14). Although *bla*$_{\text{KPC}}$ and *bla*$_{\text{VIM}}$ genes are highly associated with *Klebsiella* species, they have also been reported in other enterobacteria [142, 144]. Therefore, we may have missed similar plasmids as the custom plasmid database did not include other species apart from *K. pneumoniae*.

In conclusion, this carefully curated genome collection will not only enhance our understanding of the evolution and transmission of *K. pneumoniae* AMR in the Comunitat Valenciana, but it will also be a valuable resource for investigating potential links and transmission events in and between other regions worldwide.

8.3 Genomic analysis of AMR *K. pneumoniae* in the CV

To gain insights into the genomic epidemiology of 3GC- and carbapenem-resistant *K. pneumoniae* in the Comunitat Valenciana, we used the genomic data from the SKPCV collection (detailed in §3.1 and §5.2) (Figure 5.3). Contextualising the genomic findings within the current epidemiological situation of a region is essential to understand the implications of such findings.

To understand the epidemiologic situation, we examined all the ASTs performed for *K. pneumoniae* in the CV over the span of a decade. Our analysis revealed that the burden of 3GC-resistance was already significant in the CV prior to 2010, with an incidence of approximately 10% of all episodes (Figure 5.1 D). However, it was not until 2013 that the resistance rates began to increase to concerning levels. By 2018, the prevalence of 3GC-resistant episodes had risen to around 20% of all cases. These results are in line with the reported emergence of ESBLs in Spain [100] and are consistent with the national prevalence reported (18.2% in 2021) [477]. However, our data showed a downward trend in recent years, while both global and national data reported steadily increasing episodes [101, 477]. The emergence and spread of CRKp is more recent and, as of today, its prevalence remains low [12, 144]. The rates found in the CV were generally low during the period, with rates slightly lower than the national average (3% vs 5.9%) [101]. On the contrary, unlike the decreasing trend in 3GC resistance, the prevalence of CRKp is not decreasing and may be in line with the national situation where cases are increasing [477].

We conducted a more detailed analysis of the AMR situation in the hospitals included in the SKPCV collection (Figure 5.2). Our findings indicate that most of these hospitals had a high prevalence of 3GC-resistant episodes, and relatively low rates of carbapenem resistance (Figure 5.2), which is in line with the CV, national and

global data [10, 477]. However, we also noted that the rates of resistance were not consistent over time, often appearing as peaks, which may indicate outbreaks.

To investigate the genetic mechanisms underlying the 3GC- and carbapenem-resistance phenotypes, as well as their modes of acquisition and dissemination we analysed the genome data from the SKPCV collection. During the years when the SKPCV isolates were collected, 3GC- and carbapenem-resistance episodes accounted for 27.5% and 5%, respectively, of the total episodes in the CV.

Our study revealed a remarkable population genomic diversity in terms of phylogenetic lineages both within each hospital and across the region. Specifically, we identified 178 distinct phylogenetic sublineages corresponding to 132 known CGs and 188 unique STs (Figure 5.5), representing a substantial portion of those previously described in the literature [31]. This diversity is similar to that found in other surveillance programs at the hospital or national levels [156, 263, 466]. Of note, more than half (53.3%) of the total isolates were found to be singleton lineages. It has been proposed that these lineages most likely originate from the patient's pre-existing microbiome [466], indicating that a significant proportion of the hospital-associated disease burden may not be attributed to healthcare-related colonisation and transmission, but rather to community acquisition or patients pre-existing colonising strains [70, 466]. Indeed, 28.5% of the singleton strains were non-susceptible to 3GC or carbapenems, suggesting a high level of resistance colonisation in patients.

3GC-resistant strains were found in most of lineages identified in the SKPCV (54% of the STs) whereas CRKp were found in 38 STs (20%) (Figure 5.6). The most frequently found lineages were those represented by sequence types ST307, ST11, ST15, ST147, ST405, and ST219. All these STs are known HRCs associated with multidrug resistance worldwide [32] and, in particular, in Europe [151] and Spain [156], and now we know that they also are the major cause of 3GC- and carbapenem-resistant strains in the CV. The majority of the STs, ESBLs and carbapenemase combinations were found to be limited to a single hospital, rather than being widely disseminated across multiple hospitals (Figure 5.7).

In our study, we conducted a diversity analysis to determine whether hospitals shared the same population of AMR *K. pneumoniae* or if they were divergent from one another. Interestingly, we found that the widespread dissemination of the ST307 carrying $bla_{\text{CTX-M-15}}$ along the SKPCV hospitals had a homogenising effect of the hospital diversity, suggesting highly similar populations between hospitals. However, after removing this lineage from the diversity analyses, we observed greater divergence between hospitals, revealing movement of different ESBLs and carbapenemase-producer populations between hospitals (Figure 5.6). These findings suggest a complex scenario in the regional hospitals, where ST307 carrying $bla_{\text{CTX-M-15}}$ genes is widely disseminated, and the remaining burden of 3GC- and carbapenem-resistance is attributed to distinct gene-lineage combinations in each hospital.

It is noteworthy that other lineages, such as ST11 and ST15, were observed to be widely distributed across hospitals, similarly to ST307. However, our analysis demonstrated that these lineages exhibit high levels of genetic structure in deep phylogenetic clusters, with different clusters being present in each hospital (Supplementary Figure 5. 2). These findings highlight the limitations of relying only on MLST, as belonging to the same ST does not always indicate a recent common ancestry of those isolates, even within the same hospital. This can result in erroneous epidemiological surveillance conclusions. This issue has been previously reported [251, 272], and Henart *et al*. have proposed an alternative nomenclature system based on multilevel single linkage (MLSL) clustering and life identification numbers (LINs), the scheme we adopted and reason why we used the SL instead of the ST level to perform the analysis [408].

To understand the contribution of within-hospital transmission, between-hospital transmission, and other sources to the burden of *K. pneumoniae* AMR in the CV, we identified possible transmission groups using the analysis of genomic data. Our analysis identified 205 PTGs comprising 52.80% of the SKPCV isolates; of those, 64 PTGs were involved in inter-hospital transmission. This suggests that clonal dissemination within hospitals is the primary mechanism driving the spread of AMR in the region, in line with findings from other countries [151, 329, 466]. However, our results also suggest that between-hospital transmission is significant and

involves more complex routes of transmission that may include intermediate, undetected spreaders, highlighting the importance community acquired infections. This is a limitation of the study as we did not collect samples from the community, or healthy patients that could serve as intermediaries. In addition, we used the global curated database to assess the contribution of global epidemiology to the region's epidemic situation. We found that transmission was predominantly occurring at a regional level with few national cases, and that the impact of international transmission events was scarce.

An innovative aspect of our study is the use of core genome alignment analysis, rather than genome mapping, to detect potential transmission groups. Core genome alignments are not usually used for this purpose and, as we have shown in our study, we believe it is due to errors produced by annotations in draft genomes, which introduce many spurious SNPs, making it impossible to use such low thresholds to identify transmissions. As a result, many researchers perform a general pan-genome analysis and switch to a mapping approach to study transmissions of a few STs for which closely related and good quality references are available [151]. However, in our study, we have developed an algorithm that removes these errors, enabling us to conduct transmission analysis using the same approach for 31 different STs, including rare lineages for which references are often unavailable. This methodology provides a significant advantage for genomic surveillance of pathogens as it removes the bias of selecting a closely related reference and enables a standardised approach for every pathogen and sublineage. However, we did not corroborate the transmission groups with epidemiological information, such as patient movement or contact tracing data. Nevertheless, as reported in other studies, the genomic analysis largely corresponds with epidemiological information, as the genomic data often provides a more accurate and complete picture of transmission events [329, 331, 413].

8.3.1 The successful spread of ST307 and *bla*$_{\text{CTX-M-15}}$

In our work, we found that most of the burden of 3GC-resistance in the CV was due to the widespread dissemination of this ST carrying *bla*$_{\text{CTX-M-15}}$. Indeed, our phylogenetic study suggests multiple clonal expansions in our region (Figure 5.11).

As detailed in this dissertation and primarily in §4.4.1, the ST307 is a globally disseminated MDR lineage [151, 279], usually associated with the production of ESBLs, especially CTX-M-15 [279, 280], and that been reported to cause outbreaks and nosocomial infections in many countries [156, 274–278]. Indeed, recent reports show that the prevalence of ST307 is rising in many regions as an emerging lineage [274, 284–287].

However, despite the widespread reports of ST307, we observed that it occurred at much lower frequencies outside the CV. Indeed, the ST307 isolates collected in the CV hospitals, including the SKPCV and the NLSAR collections, represented 59.3% of the global ST307 isolates collected.

In 2019, a comprehensive analysis of ST307 was conducted by Wyres *et al*. [279], who studied the 95 genomes of this lineage publicly available at that time. Their evolutionary analysis revealed that ST307 emerged in 1994, in close association with $bla_{\text{CTX-M-15}}$, and has since rapidly disseminated across different countries, becoming a global threat. Interestingly, despite its wide dissemination, this lineage remained largely unnoticed for almost 20 years. Moreover, they pointed to a possible origin in the USA, as the largest genetic diversity was found in this country [284].

However, our results may shift the analysis towards a different direction. We used 1,110 ST307 isolates to understand the success of this lineage in the CV. Only 396 belonged to international isolates and most of them were collected in Europe (n=173), mainly in Spain (n=113). We have been unable to conduct an in-depth evolutionary analysis due to time constraints; however, we believe that the information obtained from the SKPCV genomes will contribute to a better understanding of the evolution of this multidrug-resistant lineage.

Our future perspectives include complementing our current analysis of the ST307 lineage using a phylogenomic approach to gain a better understanding of the role of the CV in the evolution of this lineage. This approach will allow us to determine the dates of introduction and dissemination of resistant strains, which will provide more comprehensive insights into the global evolution of this lineage. Furthermore, this approach may help us understand the factors contributing to its success in terms of resistance profile and transmission dynamics in the CV and worldwide.

In summary, our SKPCV study has provided important insights into the emergence and spread of AMR *K. pneumoniae* in the CV region. We have identified key factors that contribute to the spread of AMR in the CV such as the local population, mobile genetic elements, and transmission dynamics within and between healthcare settings and abroad. Additionally, our findings have contributed to a more detailed understanding of the global threat posed by the ST307 lineage.

8.4 The complexity behind initial CRKp colonisation

Carbapenem-resistant *K. pneumoniae* is a major contributor to antibiotic resistance [431] and is recognized as the fastest-growing pathogen in clinical settings [105]. Despite the many studies that have investigated the epidemiology of the first carbapenem-resistant lineages and determinants found [449], few have examined the emergence and evolution of these strains from initial colonisation to full establishment within a hospital. A better understanding of this process is critical for developing effective strategies to control and prevent the emergence and spread of carbapenem-resistant strains in healthcare settings. A better picture of the epidemiological situation can be obtained by applying complete genome analysis following the first cases until they establish and colonise a hospital setting. Additionally, WGS can help to determine the relative contributions of new introductions and local spread of previously established strains and/or plasmids and/or plasmids.

In Chapter 6, we have shown the complexity of the emergence of non-susceptibility to carbapenems in a hospital by analysing the first *K. pneumoniae* isolates that were non-susceptible to carbapenems detected in the HGUV. We extended this analysis to cover a period of 4 years. The study period began in mid-2015, when the first strains carrying $bla_{\text{OXA-48}}$ were identified at the HGUV. Shortly after, a rapid increase in carbapenem non-susceptible *K. pneumoniae* isolates was observed. The sudden increase of cases led to suspect the existence of a large outbreak; however, our findings have shown that the rise was due to the simultaneous emergence of six different circulating clones involving four common HRC types, including ST307, ST11, ST101, and ST437, different resistance-determinant factors, such as OXA-48, NDM-1, NDM-23, and DHA-1, and different plasmids.

The SNP diversity within lineages showed that they likely corresponded to local clonal expansions with several cases of possible direct transmission within the hospital (pairwise SNPs < 5) (Supplementary Figure 6.3). All these lineages are globally distributed clones of *K. pneumoniae* and are considered MDR [31]. They were related to the acquisition of several antimicrobial-resistance plasmids in different European countries in the late 2000s [289]. These four STs have already been identified in Spain as carbapenemase carriers, and they are among the widely distributed STs across Europe [151]. For three of the six lineages identified in the HGUV, we found related samples from other Spanish regions, reflecting the successful dispersion of these lineages and supporting their interregional spread (Figure 6.4, Supplementary Figure 6.2).

Surprisingly, the initial cases of CRKp in the HGUV were mainly due to non-carbapenemase-producing *K. pneumoniae* strains, rather than carbapenemase producers (Figure 6.6), which are the most commonly found mechanism [151]. The very early cases were due to AmpC and truncated porins, while only a few were related to carbapenemase enzymes. The AmpC mechanism was mediated by $bla_{\text{DHA-1}}$ genes and disseminated by the ST11-KL105 lineage (Figure 6.1). The mutated porins were mainly due to the truncation of the Ompk36 porin. Nevertheless, other mutations, such as OmpK36GD, were also found. Mutated porins were found in different lineages. Although most studies are usually focused on carbapenem genes and enzymes [156, 426], in this work we have shown that other mechanisms are also associated with the resistance phenotype and also need to be controlled.

Although most studies are usually focused on carbapenemase genes and enzymes [156, 426], in this work we have shown that other mechanisms are also associated with the resistance phenotype and also need to be controlled.

The AmpC-carrying lineage was rapidly replaced by carbapenemase-producing isolates (Figure 6.6). Two different types of carbapenemases were found in this hospital, OXA-48 and NDM, with some cases of coexistence of $bla_{\text{OXA-48}}$ and bla_{NDM} genes in the same isolates, which is seldom reported worldwide [454, 455]. Several studies have shown that $bla_{\text{OXA-48}}$, together with $bla_{\text{VIM-1}}$ and bla_{KPC}, are the most frequent carbapenemase genes in Spain [438]. Neither $bla_{\text{VIM-1}}$ or bla_{KPC} were

detected in the HGUV during our sampling period (2015-2018), although they had been reported to be widely spread in Europe at that time [151]. However, we detected two variants of *bla*$_{NDM}$ genes: NDM-1 carbapenemases, which are infrequently reported in Spanish hospitals [168] whereas NDM-23 has never been reported outside the CV [456].

After the first detection of the *bla*$_{OXA-48}$ gene in mid-2015, this gene disseminated through lineages ST307, ST11-KL13, and ST437-OXA48. As previously stated, OXA-48 represents the main carbapenem-resistance determinant in Spain, especially in *Klebsiella* strains [155, 438], and has been reported to be widely spread throughout Spain linked to a few successful lineages, including ST11, ST307 and ST437 [152, 366]. Thus, we suspect that the interregional dispersion of these clones could be the source of the strains detected in the HGUV. Yet, when we analysed the plasmid and not the lineages, we identified several cases of interlineage transfer of plasmids carrying *bla*$_{OXA-48}$ (Figure 6.5). Indeed, these plasmids were found to be circulating simultaneously at the HGUV and among the reference Spanish isolates (Supplementary Figure 6.4).

This plasmid was found to be highly prevalent in co-colonization with different carbapenemase-producing *Enterobacterales* in Spain [478]. There is also evidence of frequent bacterial transfer between patients and pervasive plasmid transfer within patients for this plasmid [479]. In light of our findings and previous works, the most likely scenario is that *K. pneumoniae* ST11 and ST307 clones carrying pOXA-48 have spread among hospitalised patients, colonising their gut microbiota, and spreading the plasmid between different *K. pneumoniae* lineages [478, 479]. However, as this plasmid is highly conserved [199], it is difficult to distinguish whether the different pOXA-48a variants found in the HGUV were due to a single introduction and later evolution in the hospital, to several introductions, or to a combination of both scenarios. Moreover, we could not elucidate the role of other species in their plasmid transfer contribution as we only included *K. pneumoniae* isolates.

bla$_{NDM}$ genes were detected a few months later than the *bla*$_{OXA-48}$ gene associated with lineages ST101 and ST437 (Figure 6.6). The origin of the *bla*$_{NDM}$ genes at the

HGUV could be elucidated thanks to whole genome analysis. *bla*$_{NDM}$ genes are usually imported from other regions, associated with individual cases and travel [480]. The emergence of these enzymes in Spain was described in a multiregional study [168]. These authors reported two lineages producing *bla*$_{NDM-1}$, ST101 and ST437, that were also present in Valencia. ST101 isolates producing *bla*$_{NDM-1}$ were also reported in Madrid and Catalonia. Epidemiological studies [457] showed that both lineages arrived at the HGUV by the transfer of a patient from a hospital in Madrid in December 2015 to this hospital in Valencia. This was the time when NDM-related episodes increased in the hospital, mainly associated with the dissemination of the ST101 lineage (Figure 6.6). Furthermore, we discovered that the *bla*$_{NDM}$ genes found in the HGUV were carried in two different plasmid backbones, each associated with a different lineage (Figure 6.5, Supplementary Figure 6.6). A more detailed analysis of the spread of the ST437 clone is described in Chapter 7 [456]. It is remarkable how different clones of the same pathogen sharing the same carbapenem resistance mechanisms can have different dissemination dynamics. For instance, lineages ST101 and ST437-NDM were both carriers of NDM enzymes; yet, although ST437-NDM showed higher AMR levels to the antibiotic tested, it was outnumbered by lineage ST101. Indeed, we know from epidemiological data that these lineages entered at the same moment in the same patient, highlighting these differences. Of note, we detected the transfer of pOXA-48 to these lineages, resulting in ST101 and ST437 strains carrying both *bla*$_{OXA-48}$ and *bla*$_{NDM}$ genes, but we did not detect the transfer of the NDM plasmid to other lineages (Supplementary Figure 6.4). This suggests that the dissemination and transfer dynamics of each carbapenemase in the HGUV behaved differently.

We also found 18 strains without any reported resistance mechanism to carbapenems. However, in most cases, the phenotype was probably related to the presence of *bla*$_{CTX-M-15}$, as reported in [481].

This work demonstrates the complexity involved in the initial colonisation of carbapenem-resistant strains in a hospital and emphasises the diversity of the lineages and plasmids involved. Many hospitals have already reported similar CRKp population diversity [151, 156, 476]. Our findings highlight the need to study all carbapenem-resistant strains to understand the overall epidemiology of CRKp even

in a single hospital, as focusing solely on specific CRKp lineages and/or resistance determinants may result in loss of information.

Moreover, we have demonstrated that most relevant lineages at the HGUV were previously present in other Spanish regions, likely being the sources of the initial infections in this hospital. These findings support that screening for colonisation upon hospital admission may help to identify carriers of carbapenem-resistant strains and, therefore, to prevent its subsequent, local transmission to other patients.

8.5 New AMR threats: the NDM-23 carbapenemase

At a time when antimicrobial resistance has become one of the biggest concerns worldwide [8, 10], the emergence of novel alleles and extremely drug resistant plasmids is a serious threat to global public health, more so when they produce carbapenem resistance in one of the most problematic pathogens in clinical settings, such as *K. pneumoniae*. This dissertation has shown that newly emerging AMR lineages play a significant role in the regional AMR picture, as they can spread and become local problems over time. Therefore, understanding how these lineages emerge and disseminate is crucial to combat the spread of AMR.

While searching for AMR genes in the NLSAR collection, we have detected numerous isolates with an unreported variant of the NDM enzyme, denominated NDM-23. In Chapter 7, we have used the data collected in the NLSAR network (including the SKPCV), to select all bla_{NDM}-carrying isolates with similar context to the bla_{NDM23} genes detected to further study its AMR pattern, genetic context and epidemiology in the region.

We found the $bla_{\text{NDM-23}}$ gene in 8 strains that, worrisomely, were non-susceptible to almost all the antimicrobials tested. They were susceptible to only a few last-line antimicrobials with a risk of toxicity or other safety concerns [96]. Furthermore, a high resistance level to the new cephalosporin cefiderocol was detected. Although little is known about the mechanisms of resistance to this antimicrobial, resistance has been reported for *K. pneumoniae* strains with mutations in *cirA* genes when combined with NDM production [115] or the combination of β-lactamases with

mutations in the *baeS* gene (baeSR regulon) [379,465]. In our study, NDM-TOP10 transformants did not show cefiderocol-hydrolysing activity, meaning that neither NDM-1 or NDM-23 hydrolysed cefiderocol by themselves. However, when tested in clinical samples, we observed that 146KP-HG showed a high resistance level (MIC > 256 mg/L). This isolate, as those in clade A, carries a combination of a mutation in the *baeS* gene and other AMR determinants already reported as being related to cefiderocol resistance. However, the MIC reported before was far lower than the one obtained for 146KP-HG (4 versus 256 mg/L) [379]. Moreover, the clone carries a yersiniabactin virulence factor in the chromosome, which encodes an iron-scavenging molecule that enhances its capacity to cause disease associated with bacteraemia and tissue-invasive infections [88]. The combination of the limited number of effective therapeutic options available to treat infections caused by this clone and the increasing number of infections in this region make this clone and its plasmid of particular concern for local and regional public health.

In this chapter, we used genomic epidemiology to describe the emergence and dissemination of this new allele in the CV. We found that it was carried in a multi-drug-resistant plasmid (p146KP-NDM23), with 18 more antimicrobial resistance genes (Figure 7.1), and disseminated through a *K. pneumoniae* ST437 strain (Figure 7.3). The spread of this clone was initiated by a patient transferred from a hospital in Madrid to the HGUV and ended up affecting at least 4 different hospitals in the CV from 2016, until at least 2019 (when our sampling concluded) (Figure 7.4). The findings of this study shed light on the potential for rapid regional transmission through clonal expansion of carbapenemase-imported cases.

We aimed to investigate the gene and plasmid origin to gain understanding of their evolutionary history. When searching for similar plasmids to the p146KP-NDM23 we did not find any similar to it. However, we found it in a genome of a *K. pneumoniae* isolate named PN4 [462]. PN4 is a strain that was involved in a multispecies outbreak in 2010 in Pakistan. Moreover, the phylogenetic analysis of isolates carrying the $bla_{\text{NDM-23}}$ together with all ST437 genomes found in the global collection (Figure 7.3) showed that the closest isolates to cluster A were two samples from Pakistan collected in 2011. Indeed, the estimated tMCRA between the Pakistani and the $bla_{\text{NDM-23}}$ NLSAR samples was estimated only 10 years before, in 1999 (R^2=0.78).

Lastly, the most basal genome of this cluster was the index case, which also was the index case of a multi-clonal and multispecies outbreak detected in the hospital Ramón y Cajal in Madrid [457]. This patient had received healthcare assistance in Pakistan in November 2014 and later, in August 2015, upon return from Pakistan, he was admitted to this hospital in Madrid. In December 2015, the patient was transferred to Valencia and admitted to the HGUV.

The close similarity of the chromosome sequences of isolates in clade A (Figure 7.4) and two isolates from Pakistan, the origin of the p146KP-NDM23, and the epidemiological link between the index case in Spain and that country strongly suggest the origin of cluster A sequences in Pakistan and not in other ST437 isolates previously found in Spain (cluster B) (Figure 7.3).

Again, the use of WGS made evident the spread of this clone through the region, showing its interhospital spread and through the community thus converting into particular concern for local and regional public health authorities. In addition, to caution them on the possibility of finding this clone in their hospitals, the results found in the study could help to guide hospital controls measures. As we found that the *bla*$_{\text{NDM-23}}$ is carried in a non-mobilizable plasmid, and therefore it can only be transmitted by clonal dissemination, hospital control measures will have to focus on the timely detection of *bla*$_{\text{NDM-23}}$ and finding and removing the source of this clone from hospitals.

Moreover, the use of WGS not only provided a better understanding of the emergence of new threats in this hospital setting, but it also allowed the evolutionary analyses that provided valuable insights into the origins and spread of resistance genes and plasmids in bacterial populations [63, 263, 482]. In fact, we were able to trace the evolutionary and epidemiological route of the new allele, the hosting plasmid, and the strain that carried both from Pakistan to Spain.

Finally, this work demonstrates the importance of a well-established surveillance system based on WGS. It also reinforces the fact that global genomic surveillance is needed to track the emergence and spread of new antibiotic resistance genes or MGEs in specific areas of interest.

8.6 Current and future use of genome-based surveillance programs

As previously mentioned, and supported by the findings of this study, genomic surveillance of bacterial pathogens is critical, particularly in the context of the worrisome AMR situation. Genomic surveillance projects enable understanding the prevalence of resistant strains, their genetic determinants, and their transmission patterns. This information is valuable for implementing targeted interventions to prevent the spread of resistant strains. For instance, our observations have significant implications for infection control strategies in the hospitals of the CV. We found that within-hospital spread was the primary driver of the *K. pneumoniae* AMR burden, but inter-hospital transmission also played a role. Furthermore, we identified cases where importation led to complex dissemination of resistant strains. Finally, we also found evidence of patients whose infections may have originated from strains present in their microbiome. These findings provide valuable insights into the patterns of transmission and factors contributing to the spread of resistant strains, which can inform public health policies and interventions.

Additionally, our findings underscore the importance of implementing surveillance strategies based on complete genome sequencing in hospitals. As shown during this dissertation, we observed different lineages of the same ST with distinct dynamics and associated resistance determinants. Traditional methods lack the required level of resolution and, although MLST schemes are frequently employed as relationship markers, they are insufficient to differentiate between lineages of the same ST [251]. Only the access to the genetic variation within whole genomes, could give enough detail to obtain a complete picture of *K. pneumoniae* spread and evolution in healthcare settings. In addition, WGS enables the performance of evolutionary analyses, which aid in understanding the origins and evolution of bacterial populations. Through this approach, we were able to gain valuable insights into the evolution of the worldwide emerging MDR lineage, ST307, and to trace the evolutionary and epidemiological route of the new carbapenemase allele, $bla_{\text{NDM-23}}$.

Thanks to an initiative by the Spanish government through PRAN and the RedLabRA, the implementation of genomic surveillance for AMR pathogens has recently begun

(in 2022). The NLSAR and SKPCV projects have enabled all tertiary hospitals in the CV to perform WGS on pathogens, eliminating the need to rely on other reference laboratories. This initiative marks a significant step towards more efficient and effective surveillance and control of AMR regionally in the CV and nationally in Spain. The WGS capabilities will enable hospitals to rapidly identify and track outbreaks of AMR infections, allowing for timely and targeted interventions to prevent further spread. Additionally, the data generated from the nationwide surveillance system will contribute to global efforts to tackle the AMR global threat.

CHAPTER 9
CONCLUSIONS

Through our work, we have shed light on the genomic characteristics of the *K. pneumoniae* population in the Comunitat Valenciana and how they contribute to the spread of antibiotic resistance. This information is crucial for the development of effective infection control measures to prevent the spread of resistant strains. The major conclusions of this dissertation are:

- While genomics is a powerful tool for studying AMR, data quality is a critical concern to obtain high quality large genomic datasets and ensure the accuracy of downstream analyses.

- *Klebsiella* genomics studies are highly biased towards MDR strains of specific HRCs and mostly restricted to a few countries. Therefore, much *Klebsiella* diversity has not been discovered and documented yet.

- The CV has a high incidence of 3GC-resistant *K. pneumoniae* infections (~20%). Moreover, AMR *K. pneumoniae* infections, especially those resistant to carbapenems, impose a substantial increase on the probability of being hospitalized in the CV.

- The AMR *K. pneumoniae* population differed between the SKPCV hospitals, especially carbapenem-resistant strains. Indeed, the 3GC- and carbapenem-resistant AMR determinants and the corresponding carrier lineages were found to be distinctly distributed in the hospitals, except for lineage ST307 carrying the $bla_{\text{CTX-M-15}}$ gene.

- Most of the *K. pneumoniae* AMR burden in the region was due to within-hospital spread, but there was also evidence of inter-hospital transmission.

- The burden of AMR *K. pneumoniae* in the CV was the result of a diverse set of factors, including singleton lineages likely originating from the community or from the microbiota of patients, as well as a complex interplay between lineage transmission between hospitals and the local proliferation of problematic clones within each hospital.

- Although a high lineage diversity was found, one lineage, ST307, was responsible for most 3GC- and carbapenem-resistant infections and interhospital transmissions in the CV region.

- The emergence of carbapenem-resistance at a hospital can be a highly complex and rapid process, involving many different lineages and carbapenem-resistance mechanisms colonising at the same time.

- The novel carbapenemase NDM variant found, NDM-23, had the same AMR phenotypic effect as its ancestor, NDM-1.

- This gene was found in a non-mobilizable MDR plasmid encoding other 18 AMR genes, which made strains almost pan-drug resistant.

- Global genomic data allowed us to trace the evolutionary and epidemiological route of the new allele, the hosting plasmid, and the strain that carried both from Pakistan to Spain.

- Carbapenemase-imported cases such as the *bla*$_{\text{NDM-23}}$, can rapidly develop into regional transmission through clonal expansion. The new *bla*$_{\text{NDM-23}}$ allele disseminated through several hospitals in CV from 2016 to at least 2019 (end of sample collection).

- Well-established surveillance systems based on WGS can help to identify the AMR determinants and unravel the lineages and MGE that are driving their spread and dissemination. Coupled with global genomic data, they can help tracking the emergence and spread of new antibiotic resistance genes or MGEs in specific areas of interest.

REFERENCES

1. Spagnolo F, Trujillo M, Dennehy JJ. Why Do Antibiotics Exist? MBio. 2021;12: e0196621.

2. Chain E, Florey HW, Gardner AD, Heatley NG, Jennings MA, Orr-Ewing J, et al. Penicillin as a chemotherapeutic agent. Lancet. 1940;236: 226–228.

3. Hawkey PM. The growing burden of antimicrobial resistance. J Antimicrob Chemother. 2008;62: i1–i9.

4. D'Costa VM, King CE, Kalan L, Morar M, Sung WWL, Schwarz C, et al. Antibiotic resistance is ancient. Nature. 2011;477: 457–461.

5. Vandenbroucke-Grauls CMJE, Kluytmans JAJW. Tracing the origins of antibiotic resistance. Nature medicine. 2022. pp. 638–640.

6. Bhullar K, Waglechner N, Pawlowski A, Koteva K, Banks ED, Johnston MD, et al. Antibiotic Resistance Is Prevalent in an Isolated Cave Microbiome. PLoS ONE. 2012. p. e34953. doi:10.1371/journal.pone.0034953

7. Gould K. Antibiotics: from prehistory to the present day. J Antimicrob Chemother. 2016;71: 572–575.

8. 10 global health issues to track in 2021. [cited 14 Dec 2022]. Available: https://www.who.int/news-room/spotlight/10-global-health-issues-to-track-in-2021

9. EClinicalMedicine. Antimicrobial resistance: a top ten global public health threat. EClinicalMedicine. 2021;41: 101221.

10. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399: 629–655.

11. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396: 1204–1222.

12. Antimicrobial resistance in the EU/EEA (EARS-Net) - Annual epidemiological report for 2020. In: European Centre for Disease Prevention and Control [Internet]. 25 Jul 2022 [cited 1 Mar 2023]. Available: https://www.ecdc.europa.eu/en/publications-data/antimicrobial-resistance-eueea-ears-net-annual-epidemiological-report-2020

13. O'Neill J. Tackling drug-resistant infections globally: final report and recommendations. 2016. Available: https://apo.org.au/node/63983

14. Smith R, Coast J. The true cost of antimicrobial resistance. BMJ. 2013;346: f1493.

15. Hofer U. The cost of antimicrobial resistance. Nat Rev Microbiol. 2018;17: 3–3.

16. Laxminarayan R, Duse A, Wattal C, Zaidi AKM, Wertheim HFL, Sumpradit N, et al. Antibiotic resistance-the need for global solutions. Lancet Infect Dis. 2013;13: 1057–1098.

17. Oldenkamp R, Schultsz C, Mancini E, Cappuccio A. Filling the gaps in the global prevalence map of clinical antimicrobial resistance. Proc Natl Acad Sci U S A. 2021;118.

doi:10.1073/pnas.2013515118

18. Pendleton JN, Gorman SP, Gilmore BF. Clinical relevance of the ESKAPE pathogens. Expert Rev Anti Infect Ther. 2013;11: 297–308.

19. Navon-Venezia S, Kondratyeva K, Carattoli A. Klebsiella pneumoniae: a major worldwide source and shuttle for antibiotic resistance. FEMS Microbiol Rev. 2017;41: 252–275.

20. Denissen J, Reyneke B, Waso-Reyneke M, Havenga B, Barnard T, Khan S, et al. Prevalence of ESKAPE pathogens in the environment: Antibiotic resistance status, community-acquired infection and risk to human health. Int J Hyg Environ Health. 2022;244: 114006.

21. WHO publishes list of bacteria for which new antibiotics are urgently needed. [cited 17 Nov 2022]. Available: https://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed

22. Podschun R, Ullmann U. Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clin Microbiol Rev. 1998;11: 589–603.

23. Passet V, Brisse S. Description of Klebsiella grimontii sp. nov. Int J Syst Evol Microbiol. 2018;68: 377–381.

24. Merla C, Rodrigues C, Passet V, Corbella M, Thorpe HA, Kallonen TVS, et al. Description of Klebsiella spallanzanii sp. nov. and of Klebsiella pasteurii sp. nov. Front Microbiol. 2019;10. doi:10.3389/fmicb.2019.02360

25. Hu Y, Wei L, Feng Y, Xie Y, Zong Z. Klebsiella huaxiensis sp. nov., recovered from human urine. Int J Syst Evol Microbiol. 2019;69: 333–336.

26. Bagley ST. Habitat association of Klebsiella species. Infect Control. 1985;6: 52–58.

27. Podschun R, Pietsch S, Höller C, Ullmann U. Incidence of Klebsiella species in surface waters and their expression of virulence factors. Appl Environ Microbiol. 2001;67: 3325–3327.

28. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proc Natl Acad Sci U S A. 2015;112: E3574–81.

29. Estell KE, Young A, Kozikowski T, Swain EA, Byrne BA, Reilly CM, et al. Pneumonia Caused by Klebsiella spp. in 46 Horses. J Vet Intern Med. 2016;30: 314–321.

30. Marques C, Menezes J, Belas A, Aboim C, Cavaco-Silva P, Trigueiro G, et al. Klebsiella pneumoniae causing urinary tract infections in companion animals and humans: population structure, antimicrobial resistance and virulence genes. J Antimicrob Chemother. 2019;74: 594–602.

31. Wyres KL, Lam MMC, Holt KE. Population genomics of Klebsiella pneumoniae. Nat Rev Microbiol. 2020;18: 344–359.

32. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. Nat Commun. 2021;12: 4188.

33. Brisse S, van Himbergen T, Kusters K, Verhoef J. Development of a rapid identification method for Klebsiella pneumoniae phylogenetic groups and analysis of 420 clinical isolates. Clin Microbiol Infect. 2004;10: 942–945.

34. Wyres KL, Holt KE. Klebsiella pneumoniae as a key trafficker of drug resistance genes from environmental to clinically important bacteria. Curr Opin Microbiol. 2018;45: 131–139.

35. Diancourt L, Passet V, Verhoef J, Grimont PAD, Brisse S. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. J Clin Microbiol. 2005;43: 4178–4182.

36. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol. 2012;50: 1355–1361.

37. Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Delannoy-Vieillard A-S, et al. Genomic definition of hypervirulent and multidrug-resistant Klebsiella pneumoniae clonal groups. Emerg Infect Dis. 2014;20: 1812–1820.

38. Zhou H, Liu W, Qin T, Liu C, Ren H. Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of Klebsiella pneumoniae. Front Microbiol. 2017;8: 371.

39. CgMLST.Org nomenclature server. [cited 11 Mar 2023]. Available: https://www.cgmlst.org/ncs

40. Dijkshoorn L, Ursing BM, Ursing JB. Strain, clone and species: comments on three basic concepts of bacteriology. J Med Microbiol. 2000;49: 397–401.

41. Li B, Zhao Y, Liu C, Chen Z, Zhou D. Molecular pathogenesis of Klebsiella pneumoniae. Future Microbiol. 2014;9: 1071–1081.

42. Broberg CA, Palacios M, Miller VL. Klebsiella: a long way to go towards understanding this enigmatic jet-setter. F1000Prime Rep. 2014;6: 64.

43. Dong N, Yang X, Chan EW-C, Zhang R, Chen S. Klebsiella species: Taxonomy, hypervirulence and multidrug resistance. EBioMedicine. 2022;79: 103998.

44. Wu H, Li D, Zhou H, Sun Y, Guo L, Shen D. Bacteremia and other body site infection caused by hypervirulent and classic Klebsiella pneumoniae. Microb Pathog. 2017;104: 254–262.

45. Gonzalez-Ferrer S, Peñaloza HF, Budnick JA, Bain WG, Nordstrom HR, Lee JS, et al. Finding Order in the Chaos: Outstanding Questions in Klebsiella pneumoniae

Pathogenesis. Infect Immun. 2021;89. doi:10.1128/IAI.00693-20

46. Zaidi AKM, Huskins WC, Thaver D, Bhutta ZA, Abbas Z, Goldmann DA. Hospital-acquired neonatal infections in developing countries. Lancet. 2005;365: 1175–1188.

47. Jarvis WR, Munn VP, Highsmith AK, Culver DH, Hughes JM. The epidemiology of nosocomial infections caused by Klebsiella pneumoniae. Infect Control. 1985;6: 68–74.

48. Ben-David D, Kordevani R, Keller N, Tal I, Marzel A, Gal-Mor O, et al. Outcome of carbapenem resistant Klebsiella pneumoniae bloodstream infections. Clin Microbiol Infect. 2012;18: 54–60.

49. Fraenkel-Wandel Y, Raveh-Brawer D, Wiener-Well Y, Yinnon AM, Assous MV. Mortality due to blaKPC Klebsiella pneumoniae bacteraemia. J Antimicrob Chemother. 2016;71: 1083–1087.

50. Chetcuti Zammit S, Azzopardi N, Sant J. Mortality risk score for Klebsiella pneumoniae bacteraemia. Eur J Intern Med. 2014;25: 571–576.

51. Marra AR, Wey SB, Castelo A, Gales AC, Cal RGR, do Carmo Filho JR, et al. Nosocomial bloodstream infections caused by Klebsiella pneumoniae: impact of extended-spectrum β-lactamase (ESBL) production on clinical outcome in a hospital with high ESBL prevalence. BMC Infect Dis. 2006;6. doi:10.1186/1471-2334-6-24

52. Meatherall BL, Gregson D, Ross T, Pitout JDD, Laupland KB. Incidence, risk factors, and outcomes of Klebsiella pneumoniae bacteremia. Am J Med. 2009;122: 866–873.

53. Schwaber MJ, Klarfeld-Lidji S, Navon-Venezia S, Schwartz D, Leavitt A, Carmeli Y. Predictors of carbapenem-resistant Klebsiella pneumoniae acquisition among hospitalized adults and effect of acquisition on mortality. Antimicrob Agents Chemother. 2008;52: 1028–1033.

54. Siu LK, Yeh K-M, Lin J-C, Fung C-P, Chang F-Y. Klebsiella pneumoniae liver abscess: a new invasive syndrome. Lancet Infect Dis. 2012;12: 881–887.

55. Shon AS, Bajwa RPS, Russo TA. Hypervirulent (hypermucoviscous) Klebsiella pneumoniae: a new and dangerous breed. Virulence. 2013;4: 107–118.

56. Prokesch BC, TeKippe M, Kim J, Raj P, TeKippe EM, Greenberg DE. Primary osteomyelitis caused by hypervirulent Klebsiella pneumoniae. Lancet Infect Dis. 2016;16: e190–e195.

57. Li W, Sun G, Yu Y, Li N, Chen M, Jin R, et al. Increasing occurrence of antimicrobial-resistant hypervirulent (hypermucoviscous) Klebsiella pneumoniae isolates in China. Clin Infect Dis. 2014;58: 225–232.

58. Lan Y, Zhou M, Jian Z, Yan Q, Wang S, Liu W. Prevalence of pks gene cluster and characteristics of Klebsiella pneumoniae-induced bloodstream infections. J Clin Lab Anal. 2019;33: e22838.

59. Zhang R, Lin D, Chan EW-C, Gu D, Chen G-X, Chen S. Emergence of Carbapenem-

Resistant Serotype K1 Hypervirulent Klebsiella pneumoniae Strains in China. Antimicrob Agents Chemother. 2016;60: 709–711.

60. Liu C, Shi J, Guo J. High prevalence of hypervirulent infection in the genetic background of elderly patients in two teaching hospitals in China. Infect Drug Resist. 2018;11: 1031–1041.

61. Lederman ER, Crum NF. Pyogenic liver abscess with a focus on Klebsiella pneumoniae as a primary pathogen: an emerging disease with unique clinical characteristics. Am J Gastroenterol. 2005;100: 322–331.

62. Wang J, Yan Y, Xue X, Wang K, Shen D. Comparison of pyogenic liver abscesses caused by hypermucoviscous Klebsiella pneumoniae and non-Klebsiella pneumoniae pathogens in Beijing: a retrospective analysis. J Int Med Res. 2013;41: 1088–1097.

63. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL, et al. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of Klebsiella pneumoniae. PLoS Genet. 2019;15: e1008114.

64. Gu D, Dong N, Zheng Z, Lin D, Huang M, Wang L, et al. A fatal outbreak of ST11 carbapenem-resistant hypervirulent Klebsiella pneumoniae in a Chinese hospital: a molecular epidemiological study. Lancet Infect Dis. 2018;18: 37–46.

65. Decré D, Verdet C, Emirian A, Le Gourrierec T, Petit J-C, Offenstadt G, et al. Emerging severe and fatal infections due to Klebsiella pneumoniae in two university hospitals in France. J Clin Microbiol. 2011;49: 3012–3014.

66. Li J, Ren J, Wang W, Wang G, Gu G, Wu X, et al. Risk factors and clinical outcomes of hypervirulent Klebsiella pneumoniae induced bloodstream infections. Eur J Clin Microbiol Infect Dis. 2018;37: 679–689.

67. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, et al. Gastrointestinal Carriage Is a Major Reservoir of Klebsiella pneumoniae Infection in Intensive Care Patients. Clin Infect Dis. 2017;65: 208–215.

68. Martin RM, Bachman MA. Colonization, Infection, and the Accessory Genome of Klebsiella pneumoniae. Front Cell Infect Microbiol. 2018;8: 4.

69. Sun Y, Patel A, SantaLucia J, Roberts E, Zhao L, Kaye K, et al. Measurement of Klebsiella Intestinal Colonization Density To Assess Infection Risk. mSphere. 2021;6: e0050021.

70. Martin RM, Cao J, Brisse S, Passet V, Wu W, Zhao L, et al. Molecular Epidemiology of Colonizing and Infecting Isolates of. mSphere. 2016;1. doi:10.1128/mSphere.00261-16

71. Lin Y-T, Siu LK, Lin J-C, Chen T-L, Tseng C-P, Yeh K-M, et al. Seroepidemiology of Klebsiella pneumoniae colonizing the intestinal tract of healthy Chinese and overseas Chinese adults in Asian countries. BMC Microbiol. 2012;12: 13.

72. Rao K, Patel A, Sun Y, Vornhagen J, Motyka J, Collingwood A, et al. Risk Factors for Klebsiella Infections among Hospitalized Patients with Preexisting Colonization. mSphere. 2021;6: e0013221.

73. Dorman MJ, Short FL. Genome watch: Klebsiella pneumoniae: when a colonizer turns bad. Nature reviews. Microbiology. 2017. p. 384.

74. Selden R, Lee S, Wang WL, Bennett JV, Eickhoff TC. Nosocomial klebsiella infections: intestinal colonization as a reservoir. Ann Intern Med. 1971;74: 657–664.

75. Paczosa MK, Mecsas J. Klebsiella pneumoniae: Going on the Offense with a Strong Defense. Microbiol Mol Biol Rev. 2016;80: 629–661.

76. Russo TA, Olson R, Fang C-T, Stoesser N, Miller M, MacDonald U, et al. Identification of Biomarkers for Differentiation of Hypervirulent Klebsiella pneumoniae from Classical K. pneumoniae. J Clin Microbiol. 2018;56. doi:10.1128/JCM.00776-18

77. Russo TA, Marr CM. Hypervirulent Klebsiella pneumoniae. Clin Microbiol Rev. 2019;32. doi:10.1128/cmr.00001-19

78. Bengoechea JA, Sa Pessoa J. Klebsiella pneumoniae infection biology: living to counteract host defences. FEMS Microbiol Rev. 2019;43: 123–144.

79. Clegg S, Murphy CN. Epidemiology and Virulence of Klebsiella pneumoniae. Microbiol Spectr. 2016;4. doi:10.1128/microbiolspec.UTI-0005-2012

80. Kawai T. Hypermucoviscosity: An Extremely Sticky Phenotype of Klebsiella pneumoniae Associated with Emerging Destructive Tissue Abscess Syndrome. Clinical Infectious Diseases. 2006;42: 1359–1361.

81. Pan Y-J, Lin T-L, Chen C-T, Chen Y-Y, Hsieh P-F, Hsu C-R, et al. Genetic analysis of capsular polysaccharide synthesis gene clusters in 79 capsular types of Klebsiella spp. Sci Rep. 2015;5: 1–10.

82. Lam MMC, Wick RR, Judd LM, Holt KE, Wyres KL. Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the Klebsiella pneumoniae species complex. Microb Genom. 2022;8. doi:10.1099/mgen.0.000800

83. Follador R, Heinz E, Wyres KL, Ellington MJ, Kowarik M, Holt KE, et al. The diversity of Klebsiella pneumoniae surface polysaccharides. Microb Genom. 2016;2: e000073.

84. Kabha K, Nissimov L, Athamna A, Keisari Y, Parolis H, Parolis LA, et al. Relationships among capsular structure, phagocytosis, and mouse virulence in Klebsiella pneumoniae. Infect Immun. 1995;63: 847–852.

85. Walker KA, Miller VL. The intersection of capsule gene expression, hypermucoviscosity and hypervirulence in Klebsiella pneumoniae. Curr Opin Microbiol. 2020;54: 95–102.

86. Murphy CN, Mortensen MS, Krogfelt KA, Clegg S. Role of Klebsiella pneumoniae type 1 and type 3 fimbriae in colonizing silicone tubes implanted into the bladders of mice as a model of catheter-associated urinary tract infections. Infect Immun. 2013;81: 3009–3017.

87. Behnsen J, Raffatellu M. Siderophores: More than Stealing Iron. mBio. 2016. doi:10.1128/mBio.01906-16

88. Lam MMC, Wick RR, Wyres KL, Gorrie CL, Judd LM, Jenney AWJ, et al. Genetic diversity, mobilisation and spread of the yersiniabactin-encoding mobile element ICEKp in Klebsiella pneumoniae populations. Microb Genom. 2018;4. doi:10.1099/mgen.0.000196

89. Lam MMC, Wyres KL, Judd LM, Wick RR, Jenney A, Brisse S, et al. Tracking key virulence loci encoding aerobactin and salmochelin siderophore synthesis in Klebsiella pneumoniae. Genome Med. 2018;10: 77.

90. Russo TA, Shon AS, Beanan JM, Olson R, MacDonald U, Pomakov AO, et al. Hypervirulent K. pneumoniae secretes more and more active iron-acquisition molecules than "classical" K. pneumoniae thereby enhancing its virulence. PLoS One. 2011;6: e26734.

91. Russo TA, Olson R, Macdonald U, Metzger D, Maltese LM, Drake EJ, et al. Aerobactin mediates virulence and accounts for increased siderophore production under iron-limiting conditions by hypervirulent (hypermucoviscous) Klebsiella pneumoniae. Infect Immun. 2014;82: 2356–2367.

92. Putze J, Hennequin C, Nougayrède J-P, Zhang W, Homburg S, Karch H, et al. Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. Infect Immun. 2009;77: 4696–4703.

93. Huang W-K, Chang JW-C, See L-C, Tu H-T, Chen J-S, Liaw C-C, et al. Higher rate of colorectal cancer among patients with pyogenic liver abscess with Klebsiella pneumoniae than those without: an 11-year follow-up study. Colorectal Dis. 2012;14: e794–801.

94. Viladomiu AS, Pueyo JM, Suñé EL, Zboromyrska Y, Mondejar PL, López JB. Guía de terapéutica antimicrobiana: 2022: Mensa Gatell. Antares; 2022.

95. Bush K, Bradford PA. β-Lactams and β-Lactamase Inhibitors: An Overview. Cold Spring Harb Perspect Med. 2016;6.

96. Villa TG, Veiga Crespo P, editors. Antimicrobial compounds: Current strategies and new alternatives. 2014th ed. Berlin, Germany: Springer; 2013.

97. Giske CG, Monnet DL, Cars O, Carmeli Y, ReAct-Action on Antibiotic Resistance. Clinical and economic impact of common multidrug-resistant gram-negative bacilli. Antimicrob Agents Chemother. 2008;52: 813–821.

98. Paterson DL, Bonomo RA. Extended-spectrum beta-lactamases: a clinical update. Clin Microbiol Rev. 2005;18: 657–686.

99. Arnold RS, Thom KA, Sharma S, Phillips M, Kristie Johnson J, Morgan DJ. Emergence of Klebsiella pneumoniae carbapenemase-producing bacteria. South Med J. 2011;104: 40–45.

100. Cantón R, Novais A, Valverde A, Machado E, Peixe L, Baquero F, et al. Prevalence and spread of extended-spectrum β-lactamase-producing Enterobacteriaceae in Europe.

Clin Microbiol Infect. 2008;14: 144–153.

101. Surveillance Atlas of infectious diseases. [cited 14 Dec 2022]. Available: http://atlas.ecdc.europa.eu/public/index.aspx?Instance

102. Lautenbach E, Patel JB, Bilker WB, Edelstein PH, Fishman NO. Extended-spectrum beta-lactamase-producing Escherichia coli and Klebsiella pneumoniae: risk factors for infection and impact of resistance on outcomes. Clin Infect Dis. 2001;32: 1162–1171.

103. Paterson DL, Ko W-C, Von Gottberg A, Mohapatra S, Casellas JM, Goossens H, et al. Antibiotic therapy for Klebsiella pneumoniae bacteremia: implications of production of extended-spectrum beta-lactamases. Clin Infect Dis. 2004;39: 31–37.

104. Tumbarello M, Spanu T, Sanguinetti M, Citton R, Montuori E, Leone F, et al. Bloodstream infections caused by extended-spectrum-beta-lactamase-producing Klebsiella pneumoniae: risk factors, molecular epidemiology, and clinical outcome. Antimicrob Agents Chemother. 2006;50: 498–504.

105. Cassini A, Högberg LD, Plachouras D, Quattrocchi A, Hoxha A, Simonsen GS, 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. 2019;19: 56–66.

106. Xu L, Sun X, Ma X. Systematic review and meta-analysis of mortality of patients infected with carbapenem-resistant Klebsiella pneumoniae. Ann Clin Microbiol Antimicrob. 2017;16: 18.

107. Durante-Mangoni E, Andini R, Zampino R. Management of carbapenem-resistant Enterobacteriaceae infections. Clin Microbiol Infect. 2019;25: 943–950.

108. Petrosillo N, Giannella M, Lewis R, Viale P. Treatment of carbapenem-resistant Klebsiella pneumoniae: the state of the art. Expert Rev Anti Infect Ther. 2013;11: 159–177.

109. Lee C-R, Lee JH, Park KS, Kim YB, Jeong BC, Lee SH. Global Dissemination of Carbapenemase-Producing Klebsiella pneumoniae: Epidemiology, Genetic Context, Treatment Options, and Detection Methods. Front Microbiol. 2016;7: 895.

110. Logan LK, Weinstein RA. The Epidemiology of Carbapenem-Resistant Enterobacteriaceae: The Impact and Evolution of a Global Menace. J Infect Dis. 2017;215: S28–S36.

111. Yahav D, Giske CG, Grāmatniece A, Abodakpi H, Tam VH, Leibovici L. New β-Lactam-β-Lactamase Inhibitor Combinations. Clin Microbiol Rev. 2020;34. doi:10.1128/CMR.00115-20

112. van Duin D, Bonomo RA. Ceftazidime/Avibactam and Ceftolozane/Tazobactam: Second-generation β-Lactam/β-Lactamase Inhibitor Combinations. Clin Infect Dis. 2016;63: 234–241.

113. Cho JC, Zmarlicka MT, Shaeer KM, Pardo J. Meropenem/Vaborbactam, the First

Carbapenem/β-Lactamase Inhibitor Combination. Ann Pharmacother. 2018;52: 769–779.

114. Hobson CA, Pierrat G, Tenaillon O, Bonacorsi S, Bercot B, Jaouen E, et al. Klebsiella pneumoniae Carbapenemase Variants Resistant to Ceftazidime-Avibactam: an Evolutionary Overview. Antimicrob Agents Chemother. 2022;66: e0044722.

115. Lan P, Lu Y, Chen Z, Wu X, Hua X, Jiang Y, et al. Emergence of High-Level cefiderocol resistance in carbapenem-resistant Klebsiella pneumoniae from bloodstream infections in patients with hematologic malignancies in China. Microbiol Spectr. 2022;10: e0008422.

116. Darby EM, Trampari E, Siasat P, Gaya MS, Alav I, Webber MA, et al. Molecular mechanisms of antibiotic resistance revisited. Nat Rev Microbiol. 2022. doi:10.1038/s41579-022-00820-y

117. Bush K, Jacoby GA. Updated functional classification of beta-lactamases. Antimicrob Agents Chemother. 2010;54: 969–976.

118. Ambler RP. The structure of beta-lactamases. Philos Trans R Soc Lond B Biol Sci. 1980;289: 321–331.

119. Bush K. Past and Present Perspectives on β-Lactamases. Antimicrob Agents Chemother. 2018;62. doi:10.1128/AAC.01076-18

120. Datta N, Kontomichalou P. Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae. Nature. 1965;208: 239–241.

121. Heritage J, M'Zali FH, Gascoyne-Binzi D, Hawkey PM. Evolution and spread of SHV extended-spectrum beta-lactamases in gram-negative bacteria. J Antimicrob Chemother. 1999;44: 309–318.

122. Livermore DM. Defining an extended-spectrum beta-lactamase. Clin Microbiol Infect. 2008;14 Suppl 1: 3–10.

123. Castanheira M, Simner PJ, Bradford PA. Extended-spectrum β-lactamases: an update on their characteristics, epidemiology and detection. JAC Antimicrob Resist. 2021;3: dlab092.

124. Chong Y, Ito Y, Kamimura T. Genetic evolution and clinical impact in extended-spectrum β-lactamase-producing Escherichia coli and Klebsiella pneumoniae. Infect Genet Evol. 2011;11: 1499–1504.

125. Coque TM, Baquero F, Canton R. Increasing prevalence of ESBL-producing Enterobacteriaceae in Europe. Euro Surveill. 2008;13. Available: https://www.ncbi.nlm.nih.gov/pubmed/19021958

126. Bauernfeind A, Jungwirth R, Schweighart S, Theopold M. Antibacterial activity and beta-lactamase stability of eleven oral cephalosporins. Infection. 1990;18 Suppl 3: S155–67.

127. Cantón R, Coque TM. The CTX-M beta-lactamase pandemic. Curr Opin Microbiol.

2006;9: 466–475.

128. Calbo E, Garau J. The changing epidemiology of hospital outbreaks due to ESBL-producing Klebsiella pneumoniae: the CTX-M-15 type consolidation. Future Microbiol. 2015;10: 1063–1075.

129. D'Andrea MM, Arena F, Pallecchi L, Rossolini GM. CTX-M-type β-lactamases: a successful story of antibiotic resistance. Int J Med Microbiol. 2013;303: 305–317.

130. Zhao W-H, Hu Z-Q. Epidemiology and genetics of CTX-M extended-spectrum β-lactamases in Gram-negative bacteria. Crit Rev Microbiol. 2013;39: 79–101.

131. Philippon A, Slama P, Dény P, Labia R. A Structure-Based Classification of Class A β-Lactamases, a Broadly Diverse Family of Enzymes. Clin Microbiol Rev. 2016;29: 29–57.

132. Evans BA, Amyes SGB. OXA β-Lactamases. Clinical Microbiology Reviews. 2014;27: 241–263.

133. Queenan AM, Bush K. Carbapenemases: the versatile beta-lactamases. Clin Microbiol Rev. 2007;20: 440–58, table of contents.

134. Bush K, Bradford PA. Interplay between β-lactamases and new β-lactamase inhibitors. Nat Rev Microbiol. 2019;17: 295–306.

135. Cantón R, Morosini MI, de la Maza OMS, de la Pedrosa EGG. IRT and CMT beta-lactamases and inhibitor resistance. Clin Microbiol Infect. 2008;14 Suppl 1: 53–62.

136. Jacoby GA. AmpC beta-lactamases. Clin Microbiol Rev. 2009;22: 161–82, Table of Contents.

137. Hanson ND, Sanders CC. Regulation of inducible AmpC beta-lactamase expression among Enterobacteriaceae. Curr Pharm Des. 1999;5: 881–894.

138. Philippon A, Arlet G, Jacoby GA. Plasmid-determined AmpC-type β-lactamases. Antimicrob Agents Chemother. 2002;46: 1–11.

139. Dahmen S, Mansour W, Charfi K, Boujaafar N, Arlet G, Bouallègue O. Imipenem resistance in Klebsiella pneumoniae is associated to the combination of plasmid-mediated CMY-4 AmpC β-lactamase and loss of an outer membrane protein. Microb Drug Resist. 2012;18: 479–483.

140. Bradford PA, Urban C, Mariano N, Projan SJ, Rahal JJ, Bush K. Imipenem resistance in Klebsiella pneumoniae is associated with the combination of ACT-1, a plasmid-mediated AmpC beta-lactamase, and the foss of an outer membrane protein. Antimicrob Agents Chemother. 1997;41: 563–569.

141. Yigit H, Queenan AM, Anderson GJ, Domenech-Sanchez A, Biddle JW, Steward CD, et al. Novel carbapenem-hydrolyzing beta-lactamase, KPC-1, from a carbapenem-resistant strain of Klebsiella pneumoniae. Antimicrob Agents Chemother. 2001;45: 1151–1161.

142. Nordmann P, Naas T, Poirel L. Global spread of Carbapenemase-producing

Enterobacteriaceae. Emerg Infect Dis. 2011;17: 1791–1798.

143. Poirel L, Potron A, Nordmann P. OXA-48-like carbapenemases: the phantom menace. J Antimicrob Chemother. 2012;67: 1597–1606.

144. Cantón R, Akóva M, Carmeli Y, Giske CG, Glupczynski Y, Gniadkowski M, et al. Rapid evolution and spread of carbapenemases among Enterobacteriaceae in Europe. Clin Microbiol Infect. 2012;18: 413–431.

145. Nordmann P, Poirel L. The difficult-to-control spread of carbapenemase producers among Enterobacteriaceae worldwide. Clin Microbiol Infect. 2014;20: 821–830.

146. Grundmann H, Glasner C, Albiger B, Aanensen DM, Tomlinson CT, Andrasević AT, et al. Occurrence of carbapenemase-producing Klebsiella pneumoniae and Escherichia coli in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. Lancet Infect Dis. 2017;17: 153–163.

147. Nordmann P, Cuzon G, Naas T. The real threat of Klebsiella pneumoniae carbapenemase-producing bacteria. Lancet Infect Dis. 2009;9: 228–236.

148. Reference gene catalog - pathogen detection - NCBI. [cited 5 Apr 2023]. Available: https://www.ncbi.nlm.nih.gov/pathogens/refgene/

149. Pitout JDD, Nordmann P, Poirel L. Carbapenemase-Producing Klebsiella pneumoniae, a Key Pathogen Set for Global Nosocomial Dominance. Antimicrob Agents Chemother. 2015;59: 5873–5884.

150. Munoz-Price LS, Poirel L, Bonomo RA, Schwaber MJ, Daikos GL, Cormican M, et al. Clinical epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases. Lancet Infect Dis. 2013;13: 785–796.

151. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, et al. Epidemic of carbapenem-resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. Nat Microbiol. 2019;4: 1919–1929.

152. Oteo J, Ortega A, Bartolomé R, Bou G, Conejo C, Fernández-Martínez M, et al. Prospective multicenter study of carbapenemase-producing Enterobacteriaceae from 83 hospitals in Spain reveals high in vitro susceptibility to colistin and meropenem. Antimicrob Agents Chemother. 2015;59: 3406–3412.

153. Poirel L, Héritier C, Tolün V, Nordmann P. Emergence of oxacillinase-mediated resistance to imipenem in Klebsiella pneumoniae. Antimicrob Agents Chemother. 2004;48: 15–22.

154. Mairi A, Pantel A, Sotto A, Lavigne J-P, Touati A. OXA-48-like carbapenemases producing Enterobacteriaceae in different niches. Eur J Clin Microbiol Infect Dis. 2018;37: 587–604.

155. Pitout JDD, Peirano G, Kock MM, Strydom K-A, Matsumura Y. The Global Ascendency of OXA-48-Type Carbapenemases. Clin Microbiol Rev. 2019;33. doi:10.1128/CMR.00102-19

156. Cañada-García JE, Moure Z, Sola-Campoy PJ, Delgado-Valverde M, Cano ME, Gijón D, et al. CARB-ES-19 Multicenter Study of Carbapenemase-Producing Klebsiella pneumoniae and Escherichia coli From All Spanish Provinces Reveals Interregional Spread of High-Risk Clones Such as ST307/OXA-48 and ST512/KPC-3. Front Microbiol. 2022;13: 918362.

157. Walsh TR, Toleman MA, Poirel L, Nordmann P. Metallo-beta-lactamases: the quiet before the storm? Clin Microbiol Rev. 2005;18: 306–325.

158. Kazmierczak KM, Rabine S, Hackel M, McLaughlin RE, Biedenbach DJ, Bouchillon SK, et al. Multiyear, Multinational Survey of the Incidence and Global Distribution of Metallo-β-Lactamase-Producing Enterobacteriaceae and Pseudomonas aeruginosa. Antimicrob Agents Chemother. 2016;60: 1067–1078.

159. Watanabe M, Iyobe S, Inoue M, Mitsuhashi S. Transferable imipenem resistance in Pseudomonas aeruginosa. Antimicrob Agents Chemother. 1991;35: 147–151.

160. Lowe CF, Matic N, Champagne S, Romney MG, Leung V, Ritchie G. The Brief Case: IMP, the Uncommonly Common Carbapenemase. J Clin Microbiol. 2020;58. doi:10.1128/JCM.01094-19

161. Lauretti L, Riccio ML, Mazzariol A, Cornaglia G, Amicosante G, Fontana R, et al. Cloning and characterization of blaVIM, a new integron-borne metallo-beta-lactamase gene from a Pseudomonas aeruginosa clinical isolate. Antimicrob Agents Chemother. 1999;43: 1584–1590.

162. Yong D, Toleman MA, Giske CG, Cho HS, Sundman K, Lee K, et al. Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. Antimicrob Agents Chemother. 2009;53: 5046–5054.

163. Nordmann P, Poirel L, Toleman MA, Walsh TR. Does broad-spectrum β-lactam resistance due to NDM-1 herald the end of the antibiotic era for treatment of infections caused by Gram-negative bacteria? J Antimicrob Chemother. 2011;66: 689–692.

164. Perry JD, Naqvi SH, Mirza IA, Alizai SA, Hussain A, Ghirardi S, et al. Prevalence of faecal carriage of Enterobacteriaceae with NDM-1 carbapenemase at military hospitals in Pakistan, and evaluation of two chromogenic media. J Antimicrob Chemother. 2011;66: 2288–2294.

165. Livermore DM, Walsh TR, Toleman M, Woodford N. Balkan NDM-1: escape or transplant? The Lancet infectious diseases. 2011. p. 164.

166. Dortet L, Poirel L, Nordmann P. Worldwide dissemination of the NDM-type carbapenemases in Gram-negative bacteria. Biomed Res Int. 2014;2014: 249856.

167. Kumarasamy KK, Toleman MA, Walsh TR, Bagaria J, Butt F, Balakrishnan R, et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect Dis. 2010;10: 597–602.

168. Pérez-Vázquez M, Sola Campoy PJ, Ortega A, Bautista V, Monzón S, Ruiz-Carrascoso G, et al. Emergence of NDM-producing Klebsiella pneumoniae and Escherichia coli in Spain: phylogeny, resistome, virulence and plasmids encoding blaNDM-like genes as determined by WGS. J Antimicrob Chemother. 2019;74: 3489–3496.

169. Di Pilato V, Henrici De Angelis L, Aiezza N, Baccani I, Niccolai C, Parisio EM, et al. Resistome and virulome accretion in an NDM-1-producing ST147 sublineage of Klebsiella pneumoniae associated with an outbreak in Tuscany, Italy: a genotypic and phenotypic characterisation. Lancet Microbe. 2022;3: e224–e234.

170. Yu J, Tan K, Rong Z, Wang Y, Chen Z, Zhu X, et al. Nosocomial outbreak of KPC-2- and NDM-1-producing Klebsiella pneumoniae in a neonatal ward: a retrospective study. BMC Infect Dis. 2016;16: 563.

171. Oueslati S, Emeraud C, Grosperrin V, Levy M, Cotellon G, Creton E, et al. Polyclonal Dissemination of NDM-1- and NDM-9-Producing Escherichia coli and Klebsiella pneumoniae in French Polynesia. Antimicrob Agents Chemother. 2021;65. doi:10.1128/AAC.02437-20

172. Wu W, Feng Y, Tang G, Qiao F, McNally A, Zong Z. NDM Metallo-β-Lactamases and Their Bacterial Producers in Health Care Settings. Clin Microbiol Rev. 2019;32. doi:10.1128/CMR.00115-18

173. Pagès J-M, James CE, Winterhalter M. The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. Nat Rev Microbiol. 2008;6: 893–903.

174. Sugawara E, Kojima S, Nikaido H. Klebsiella pneumoniae Major Porins OmpK35 and OmpK36 Allow More Efficient Diffusion of β-Lactams than Their Escherichia coli Homologs OmpF and OmpC. J Bacteriol. 2016;198: 3200–3208.

175. Martínez-Martínez L, Pascual A, Hernández-Allés S, Alvarez-Díaz D, Suárez AI, Tran J, et al. Roles of beta-lactamases and porins in activities of carbapenems and cephalosporins against Klebsiella pneumoniae. Antimicrob Agents Chemother. 1999;43: 1669–1673.

176. Poirel L, Ortiz de la Rosa J-M, Richard A, Aires-de-Sousa M, Nordmann P. CTX-M-33, a CTX-M-15 derivative conferring reduced susceptibility to carbapenems. Antimicrob Agents Chemother. 2019;63. doi:10.1128/AAC.01515-19

177. Nicolas-Chanoine M-H, Mayer N, Guyot K, Dumont E, Pagès J-M. Interplay Between Membrane Permeability and Enzymatic Barrier Leads to Antibiotic-Dependent Resistance in. Front Microbiol. 2018;9: 1422.

178. Ramirez MS, Traglia GM, Lin DL, Tran T, Tolmasky ME. Plasmid-Mediated Antibiotic Resistance and Virulence in Gram-negatives: the Klebsiella pneumoniae Paradigm. Microbiol Spectr. 2014;2: 1–15.

179. Mathers AJ, Peirano G, Pitout JDD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae.

Clin Microbiol Rev. 2015;28: 565–591.

180. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev. 2018;31. doi:10.1128/CMR.00088-17

181. Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids. Microbiol Mol Biol Rev. 2010;74: 434–452.

182. Holmes AH, Moore LSP, Sundsfjord A, Steinbakk M, Regmi S, Karkey A, et al. Understanding the mechanisms and drivers of antimicrobial resistance. Lancet. 2016;387: 176–187.

183. Novick RP. Plasmid incompatibility. Microbiol Rev. 1987;51: 381–395.

184. Couturier M, Bex F, Bergquist PL, Maas WK. Identification and classification of bacterial plasmids. Microbiol Rev. 1988;52: 375–395.

185. Carattoli A, Bertini A, Villa L, Falbo V, Hopkins KL, John Threlfall E. Identification of plasmids by PCR-based replicon typing. J Microbiol Methods. 2005;63: 219–228.

186. Garcillán-Barcia MP, Alvarado A, de la Cruz F. Identification of bacterial plasmids based on mobility and plasmid population biology. FEMS Microbiol Rev. 2011;35: 936–956.

187. Orlek A, Stoesser N, Anjum MF, Doumith M, Ellington MJ, Peto T, et al. Plasmid Classification in an Era of Whole-Genome Sequencing: Application in Studies of Antibiotic Resistance Epidemiology. Front Microbiol. 2017;8. doi:10.3389/fmicb.2017.00182

188. Garcillán-Barcia MP, Francia MV, de la Cruz F. The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev. 2009;33: 657–687.

189. Fricke WF, Welch TJ, McDermott PF, Mammel MK, LeClerc JE, White DG, et al. Comparative genomics of the IncA/C multidrug resistance plasmid family. J Bacteriol. 2009;191: 4750–4757.

190. Lanza VF, de Toro M, Garcillán-Barcia MP, Mora A, Blanco J, Coque TM, et al. Plasmid flux in Escherichia coli ST131 sublineages, analyzed by plasmid constellation network (PLACNET), a new method for plasmid reconstruction from whole genome sequences. PLoS Genet. 2014;10: e1004766.

191. Fondi M, Fani R. The horizontal flow of the plasmid resistome: clues from inter-generic similarity networks. Environ Microbiol. 2010;12: 3228–3242.

192. Tazzyman SJ, Bonhoeffer S. Why There Are No Essential Genes on Plasmids. Mol Biol Evol. 2015;32: 3079–3088.

193. David S, Cohen V, Reuter S, Sheppard AE, Giani T, Parkhill J, et al. Integrated chromosomal and plasmid sequence analyses reveal diverse modes of carbapenemase gene spread among Klebsiella pneumoniae. Proc Natl Acad Sci U S A. 2020;117: 25043–25054.

194. Carattoli A. Resistance plasmid families in Enterobacteriaceae. Antimicrob Agents Chemother. 2009;53: 2227–2238.

195. Acman M, Wang R, van Dorp L, Shaw LP, Wang Q, Luhmann N, et al. Role of mobile genetic elements in the global dissemination of the carbapenem resistance gene blaNDM. Nat Commun. 2022;13: 1131.

196. Naas T, Cuzon G, Villegas M-V, Lartigue M-F, Quinn JP, Nordmann P. Genetic structures at the origin of acquisition of the beta-lactamase bla KPC gene. Antimicrob Agents Chemother. 2008;52: 1257–1263.

197. Leavitt A, Chmelnitsky I, Carmeli Y, Navon-Venezia S. Complete nucleotide sequence of KPC-3-encoding plasmid pKpQIL in the epidemic Klebsiella pneumoniae sequence type 258. Antimicrob Agents Chemother. 2010;54: 4493–4496.

198. Chen L, Mathema B, Chavda KD, DeLeo FR, Bonomo RA, Kreiswirth BN. Carbapenemase-producing Klebsiella pneumoniae: molecular and genetic decoding. Trends Microbiol. 2014;22: 686–696.

199. Poirel L, Bonnin RA, Nordmann P. Genetic features of the widespread plasmid coding for the carbapenemase OXA-48. Antimicrob Agents Chemother. 2012;56: 559–562.

200. Nassif X, Fournier JM, Arondel J, Sansonetti PJ. Mucoid phenotype of Klebsiella pneumoniae is a plasmid-encoded virulence factor. Infect Immun. 1989;57: 546–552.

201. Xu Y, Zhang J, Wang M, Liu M, Liu G, Qu H, et al. Mobilization of the nonconjugative virulence plasmid from hypervirulent Klebsiella pneumoniae. Genome Med. 2021;13: 119.

202. Rodríguez-Beltrán J, DelaFuente J, León-Sampedro R, MacLean RC, San Millán Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. Nat Rev Microbiol. 2021;19: 347–359.

203. Dunn SJ, Connor C, McNally A. The evolution and transmission of multi-drug resistant Escherichia coli and Klebsiella pneumoniae: the complexity of clones and plasmids. Curr Opin Microbiol. 2019;51: 51–56.

204. Boerlin P, Reid-Smith RJ. Antimicrobial resistance: its emergence and transmission. Anim Health Res Rev. 2008;9: 115–126.

205. D'Costa VM, Griffiths E, Wright GD. Expanding the soil antibiotic resistome: exploring environmental diversity. Curr Opin Microbiol. 2007;10: 481–489.

206. Koonin EV, Makarova KS, Aravind L. Horizontal gene transfer in prokaryotes: quantification and classification. Annu Rev Microbiol. 2001;55: 709–742.

207. Carattoli A. Plasmids and the spread of resistance. Int J Med Microbiol. 2013;303: 298–304.

208. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, et al. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. J

Antimicrob Chemother. 2018;73: 1121–1137.

209. San Millan A, Escudero JA, Gifford DR, Mazel D, MacLean RC. Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria. Nat Ecol Evol. 2016;1: 10.

210. San Millan A, Santos-Lopez A, Ortega-Huedo R, Bernabe-Balas C, Kennedy SP, Gonzalez-Zorn B. Small-plasmid-mediated antibiotic resistance is enhanced by increases in plasmid copy number and bacterial fitness. Antimicrob Agents Chemother. 2015;59: 3335–3341.

211. Shen Z, Zhang H, Gao Q, Qin J, Zhang C, Zhu J, et al. Increased Plasmid Copy Number Contributes to the Elevated Carbapenem Resistance in OXA-232-Producing Klebsiella pneumoniae. Microb Drug Resist. 2020;26: 561–568.

212. San Millan A, Toll-Riera M, Qi Q, Betts A, Hopkinson RJ, McCullagh J, et al. Integrative analysis of fitness and metabolic effects of plasmids in Pseudomonas aeruginosa PAO1. ISME J. 2018;12: 3014–3024.

213. Chaves J, Ladona MG, Segura C, Coira A, Reig R, Ampurdanés C. SHV-1 beta-lactamase is mainly a chromosomally encoded species-specific enzyme in Klebsiella pneumoniae. Antimicrob Agents Chemother. 2001;45: 2856–2861.

214. Souli M, Galani I, Giamarellou H. Emergence of extensively drug-resistant and pandrug-resistant Gram-negative bacilli in Europe. Euro Surveill. 2008;13. doi:10.2807/ese.13.47.19045-en

215. Chen L, Kreiswirth BN. Convergence of carbapenem-resistance and hypervirulence in Klebsiella pneumoniae. The Lancet infectious diseases. 2018. pp. 2–3.

216. Lee C-R, Lee JH, Park KS, Jeon JH, Kim YB, Cha C-J, et al. Antimicrobial Resistance of Hypervirulent Klebsiella pneumoniae: Epidemiology, Hypervirulence-Associated Determinants, and Resistance Mechanisms. Front Cell Infect Microbiol. 2017;7: 483.

217. Choby JE, Howard-Anderson J, Weiss DS. Hypervirulent Klebsiella pneumoniae - clinical and molecular perspectives. J Intern Med. 2020;287: 283–300.

218. Han Y-L, Wen X-H, Zhao W, Cao X-S, Wen J-X, Wang J-R, et al. Epidemiological characteristics and molecular evolution mechanisms of carbapenem-resistant hypervirulent Klebsiella pneumoniae. Front Microbiol. 2022;13: 1003783.

219. Tang M, Kong X, Hao J, Liu J. Epidemiological Characteristics and Formation Mechanisms of Multidrug-Resistant Hypervirulent Klebsiella pneumoniae. Front Microbiol. 2020;11: 581543.

220. He J, Du X, Zeng X, Moran RA, van Schaik W, Zou Q, et al. Phenotypic and Genotypic Characterization of a Hypervirulent Carbapenem-Resistant Klebsiella pneumoniae ST17-KL38 Clinical Isolate Harboring the Carbapenemase IMP-4. Microbiol Spectr. 2022;10: e0213421.

221. Heiden SE, Hübner N-O, Bohnert JA, Heidecke C-D, Kramer A, Balau V, et al. A Klebsiella pneumoniae ST307 outbreak clone from Germany demonstrates features of extensive

drug resistance, hypermucoviscosity, and enhanced iron acquisition. Genome Med. 2020;12: 113.

222. Lam MMC, Wyres KL, Wick RR, Judd LM, Fostervold A, Holt KE, et al. Convergence of virulence and MDR in a single plasmid vector in MDR Klebsiella pneumoniae ST15. J Antimicrob Chemother. 2019;74: 1218–1222.

223. Yang X, Wai-Chi Chan E, Zhang R, Chen S. A conjugative plasmid that augments virulence in Klebsiella pneumoniae. Nat Microbiol. 2019;4: 2039–2043.

224. Wyres KL, Holt KE. Klebsiella pneumoniae Population Genomics and Antimicrobial-Resistant Clones. Trends Microbiol. 2016;24: 944–956.

225. Baquero F, Coque TM. Multilevel population genetics in antibiotic resistance. FEMS Microbiol Rev. 2011;35: 705–706.

226. Lowe M, Kock MM, Coetzee J, Hoosien E, Peirano G, Strydom K-A, et al. Klebsiella pneumoniae ST307 with blaOXA-181, South Africa, 2014-2016. Emerg Infect Dis. 2019;25: 739–747.

227. Wyres KL, Nguyen TNT, Lam MMC, Judd LM, van Vinh Chau N, Dance DAB, et al. Genomic surveillance for hypervirulence and multi-drug resistance in invasive Klebsiella pneumoniae from South and Southeast Asia. Genome Med. 2020;12: 11.

228. Bowers JR, Kitchel B, Driebe EM, MacCannell DR, Roe C, Lemmer D, et al. Genomic Analysis of the Emergence and Rapid Global Dissemination of the Clonal Group 258 Klebsiella pneumoniae Pandemic. PLoS One. 2015;10: e0133727.

229. Deleo FR, Chen L, Porcella SF, Martens CA, Kobayashi SD, Porter AR, et al. Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 Klebsiella pneumoniae. Proc Natl Acad Sci U S A. 2014;111: 4988–4993.

230. Marsh JW, Mustapha MM, Griffith MP, Evans DR, Ezeonwuka C, Pasculle AW, et al. Evolution of Outbreak-Causing Carbapenem-Resistant Klebsiella pneumoniae ST258 at a Tertiary Care Hospital over 8 Years. MBio. 2019;10. doi:10.1128/mBio.01945-19

231. Woodford N, Turton JF, Livermore DM. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. FEMS Microbiol Rev. 2011;35: 736–755.

232. Giakoupi P, Maltezou H, Polemis M, Pappa O, Saroglou G, Vatopoulos A, et al. KPC-2-producing Klebsiella pneumoniae infections in Greek hospitals are mainly due to a hyperepidemic clone. Euro Surveill. 2009;14. doi:10.2807/ese.14.21.19218-en

233. Kitchel B, Rasheed JK, Patel JB, Srinivasan A, Navon-Venezia S, Carmeli Y, et al. Molecular epidemiology of KPC-producing Klebsiella pneumoniae isolates in the United States: clonal expansion of multilocus sequence type 258. Antimicrob Agents Chemother. 2009;53: 3365–3370.

234. Chen L, Mathema B, Pitout JDD, DeLeo FR, Kreiswirth BN. Epidemic Klebsiella pneumoniae ST258 is a hybrid strain. MBio. 2014;5: e01355–14.

235. Qi Y, Wei Z, Ji S, Du X, Shen P, Yu Y. ST11, the dominant clone of KPC-producing Klebsiella pneumoniae in China. J Antimicrob Chemother. 2011;66: 307–312.

236. Shen P, Wei Z, Jiang Y, Du X, Ji S, Yu Y, et al. Novel genetic environment of the carbapenem-hydrolyzing beta-lactamase KPC-2 among Enterobacteriaceae in China. Antimicrob Agents Chemother. 2009;53: 4333–4338.

237. Liu Y, Wan L-G, Deng Q, Cao X-W, Yu Y, Xu Q-F. First description of NDM-1-, KPC-2-, VIM-2- and IMP-4-producing Klebsiella pneumoniae strains in a single Chinese teaching hospital. Epidemiol Infect. 2015;143: 376–384.

238. Voulgari E, Gartzonika C, Vrioni G, Politi L, Priavali E, Levidiotou-Stefanou S, et al. The Balkan region: NDM-1-producing Klebsiella pneumoniae ST11 clonal strain causing outbreaks in Greece. J Antimicrob Chemother. 2014;69: 2091–2097.

239. Wang Q, Wang X, Wang J, Ouyang P, Jin C, Wang R, et al. Phenotypic and Genotypic Characterization of Carbapenem-resistant Enterobacteriaceae: Data From a Longitudinal Large-scale CRE Study in China (2012–2016). Clin Infect Dis. 2018;67: S196–S205.

240. Solgi H, Badmasti F, Giske CG, Aghamohammad S, Shahcheraghi F. Molecular epidemiology of NDM-1- and OXA-48-producing Klebsiella pneumoniae in an Iranian hospital: clonal dissemination of ST11 and ST893. J Antimicrob Chemother. 2018;73: 1517–1524.

241. Damjanova I, Tóth A, Pászti J, Hajbel-Vékony G, Jakab M, Berta J, et al. Expansion and countrywide dissemination of ST11, ST15 and ST147 ciprofloxacin-resistant CTX-M-15-type beta-lactamase-producing Klebsiella pneumoniae epidemic clones in Hungary in 2005--the new "MRSAs"? J Antimicrob Chemother. 2008;62: 978–985.

242. Karlowsky JA, Walkty A, Golden AR, Baxter MR, Denisuik AJ, McCracken M, et al. ESBL-positive Escherichia coli and Klebsiella pneumoniae isolates from across Canada: CANWARD surveillance study, 2007-18. J Antimicrob Chemother. 2021;76: 2815–2824.

243. Pereira PS, de Araujo CFM, Seki LM, Zahner V, Carvalho-Assef APD, Asensi MD. Update of the molecular epidemiology of KPC-2-producing Klebsiella pneumoniae in Brazil: spread of clonal complex 11 (ST11, ST437 and ST340). J Antimicrob Chemother. 2013;68: 312–316.

244. Campos-Madueno EI, Moser AI, Jost G, Maffioli C, Bodmer T, Perreten V, et al. Carbapenemase-producing Klebsiella pneumoniae strains in Switzerland: human and non-human settings may share high-risk clones. J Glob Antimicrob Resist. 2022;28: 206–215.

245. Brilhante M, Brawand SG, Endimiani A, Rohrbach H, Kittl S, Willi B, et al. Two high-risk clones of carbapenemase-producing *Klebsiella pneumoniae* that cause infections in pets and are present in the environment of a veterinary referral hospital. J Antimicrob Chemother. 2021;76: 1140–1149.

246. Pulss S, Stolle I, Stamm I, Leidner U, Heydel C, Semmler T, et al. Multispecies and Clonal

Dissemination of OXA-48 Carbapenemase in Enterobacteriaceae From Companion Animals in Germany, 2009—2016. Front Microbiol. 2018;9. doi:10.3389/fmicb.2018.01265

247. Lepuschitz S, Schill S, Stoeger A, Pekard-Amenitsch S, Huhulescu S, Inreiter N, et al. Whole genome sequencing reveals resemblance between ESBL-producing and carbapenem resistant Klebsiella pneumoniae isolates from Austrian rivers and clinical isolates from hospitals. Science of The Total Environment. 2019;662: 227–235.

248. Oteo J, Mencía A, Bautista V, Pastor N, Lara N, González-González F, et al. Colonization with Enterobacteriaceae-Producing ESBLs, AmpCs, and OXA-48 in Wild Avian Species, Spain 2015–2016. Microb Drug Resist. 2018;24: 932–938.

249. Zhao J, Liu C, Liu Y, Zhang Y, Xiong Z, Fan Y, et al. Genomic characteristics of clinically important ST11 Klebsiella pneumoniae strains worldwide. J Glob Antimicrob Resist. 2020;22: 519–526.

250. Dong N, Zhang R, Liu L, Li R, Lin D, Chan EW-C, et al. Genome analysis of clinical multilocus sequence Type 11 Klebsiella pneumoniae from China. Microb Genom. 2018;4. doi:10.1099/mgen.0.000149

251. Liao W, Liu Y, Zhang W. Virulence evolution, molecular mechanisms of resistance and prevalence of ST11 carbapenem-resistant Klebsiella pneumoniae in China: A review over the last 10 years. J Glob Antimicrob Resist. 2020;23: 174–180.

252. Zhou K, Xiao T, David S, Wang Q, Zhou Y, Guo L, et al. Novel Subclone of Carbapenem-Resistant Klebsiella pneumoniae Sequence Type 11 with Enhanced Virulence and Transmissibility, China. Emerg Infect Dis. 2020;26: 289–297.

253. Zhao H, He Z, Li Y, Sun B. Epidemiology of carbapenem-resistant Klebsiella pneumoniae ST15 of producing KPC-2, SHV-106 and CTX-M-15 in Anhui, China. BMC Microbiol. 2022;22: 262.

254. Stoesser N, Giess A, Batty EM, Sheppard AE, Walker AS, Wilson DJ, et al. Genome sequencing of an extended series of NDM-producing Klebsiella pneumoniae isolates from neonatal infections in a Nepali hospital characterizes the extent of community- versus hospital-associated transmission in an endemic setting. Antimicrob Agents Chemother. 2014;58: 7347–7357.

255. Machuca J, López-Cerero L, Fernández-Cuenca F, Mora-Navas L, Mediavilla-Gradolph C, López-Rodríguez I, et al. OXA-48-Like-Producing Klebsiella pneumoniae in Southern Spain in 2014-2015. Antimicrob Agents Chemother. 2019;63. doi:10.1128/AAC.01396-18

256. Shi Q, Han R, Guo Y, Zheng Y, Yang Y, Yin D, et al. Emergence of ST15 Klebsiella pneumoniae Clinical Isolates Producing Plasmids-Mediated RmtF and OXA-232 in China. Infect Drug Resist. 2020;13: 3125–3129.

257. Sánchez-Romero I, Asensio A, Oteo J, Muñoz-Algarra M, Isidoro B, Vindel A, et al. Nosocomial outbreak of VIM-1-producing Klebsiella pneumoniae isolates of multilocus

sequence type 15: molecular basis, clinical risk factors, and outcome. Antimicrob Agents Chemother. 2012;56: 420–427.

258. Vubil D, Figueiredo R, Reis T, Canha C, Boaventura L, DA Silva GJ. Outbreak of KPC-3-producing ST15 and ST348 Klebsiella pneumoniae in a Portuguese hospital. Epidemiol Infect. 2017;145: 595–599.

259. Sacco F, Raponi G, Oliva A, Bibbolino G, Mauro V, Di Lella FM, et al. An outbreak sustained by ST15 Klebsiella pneumoniae carrying 16S rRNA methyltransferases and blaNDM: evaluation of the global dissemination of these resistance determinants. Int J Antimicrob Agents. 2022;60: 106615.

260. Rad ZR, Rad ZR, Goudarzi H, Goudarzi M, Alizade H, Mazraeh FN, et al. Detection of NDM-1 producing Klebsiella pneumoniae ST15 and ST147 in Iran during 2019-2020. Acta Microbiol Immunol Hung. 2021. doi:10.1556/030.2021.01381

261. Karampatakis T, Antachopoulos C, Iosifidis E, Tsakris A, Roilides E. Molecular epidemiology of carbapenem-resistant Klebsiella pneumoniae in Greece. Future Microbiol. 2016;11: 809–823.

262. Afolayan AO, Oaikhena AO, Aboderin AO, Olabisi OF, Amupitan AA, Abiri OV, et al. Clones and Clusters of Antimicrobial-Resistant Klebsiella From Southwestern Nigeria. Clin Infect Dis. 2021;73: S308–S315.

263. Moradigaravand D, Martin V, Peacock SJ, Parkhill J. Evolution and Epidemiology of Multidrug-Resistant Klebsiella pneumoniae in the United Kingdom and Ireland. MBio. 2017;8. doi:10.1128/mBio.01976-16

264. Quiñones D, Valverde A, Rodríguez-Baños M, Kobayashi N, Zayaz A, Abreu M, et al. High clonal diversity in a non-outbreak situation of clinical ESBL-producing Klebsiella pneumoniae isolates in the first national surveillance program in Cuba. Microb Drug Resist. 2014;20: 45–51.

265. Breurec S, Guessennd N, Timinouni M, Le TAH, Cao V, Ngandjio A, et al. Klebsiella pneumoniae resistant to third-generation cephalosporins in five African and two Vietnamese major towns: multiclonal population structure with two major international clonal groups, CG15 and CG258. Clin Microbiol Infect. 2013;19: 349–355.

266. Roberts LW, Hoi LT, Khokhar FA, Hoa NT, Van Giang T, Bui C, et al. Genomic characterisation of multidrug-resistant Escherichia coli, Klebsiella pneumoniae, and Acinetobacter baumannii in two intensive care units in Hanoi, Viet Nam: a prospective observational cohort study. Lancet Microbe. 2022;3: e857–e866.

267. Hu Y, Liu C, Shen Z, Zhou H, Cao J, Chen S, et al. Prevalence, risk factors and molecular epidemiology of carbapenem-resistant Klebsiella pneumoniae in patients from Zhejiang, China, 2008-2018. Emerg Microbes Infect. 2020;9: 1771–1779.

268. Zhang Y, Chen C, Wu J, Jin J, Xu T, Zhou Y, et al. Sequence-Based Genomic Analysis Reveals Transmission of Antibiotic Resistance and Virulence among Carbapenemase-Producing Klebsiella pneumoniae Strains. mSphere. 2022;7: e0014322.

269. Qin X, Wu S, Hao M, Zhu J, Ding B, Yang Y, et al. The Colonization of Carbapenem-Resistant Klebsiella pneumoniae: Epidemiology, Resistance Mechanisms, and Risk Factors in Patients Admitted to Intensive Care Units in China. J Infect Dis. 2020;221: S206–S214.

270. Chen J, Hu C, Wang R, Li F, Sun G, Yang M, et al. Shift in the Dominant Sequence Type of Carbapenem-Resistant Klebsiella pneumoniae Bloodstream Infection from ST11 to ST15 at a Medical Center in Northeast China, 2015-2020. Infect Drug Resist. 2021;14: 1855–1863.

271. Ewers C, Stamm I, Pfeifer Y, Wieler LH, Kopp PA, Schønning K, et al. Clonal spread of highly successful ST15-CTX-M-15 Klebsiella pneumoniae in companion animals and horses. J Antimicrob Chemother. 2014;69: 2676–2680.

272. Zhou K, Lokate M, Deurenberg RH, Tepper M, Arends JP, Raangs EGC, et al. Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum β-lactamase producing ST15 Klebsiella pneumoniae. Sci Rep. 2016;6: 20840.

273. Al Fadhli AH, Jamal WY, Rotimi VO. Elucidating the virulence genes harboured by carbapenemase- and non-carbapenemase-producing carbapenem-resistant Klebsiella pneumoniae rectal isolates from patients admitted to intensive care units using whole-genome sequencing in Kuwait. J Med Microbiol. 2022;71. doi:10.1099/jmm.0.001554

274. Ocampo AM, Chen L, Cienfuegos AV, Roncancio G, Chavda KD, Kreiswirth BN, et al. A Two-Year Surveillance in Five Colombian Tertiary Care Hospitals Reveals High Frequency of Non-CG258 Clones of Carbapenem-Resistant Klebsiella pneumoniae with Distinct Clinical Characteristics. Antimicrob Agents Chemother. 2016;60: 332–342.

275. Park DJ, Yu JK, Park KG, Park Y-J. Genotypes of Ciprofloxacin-Resistant Klebsiella pneumoniae in Korea and Their Characteristics According to the Genetic Lineages. Microb Drug Resist. 2015;21: 622–630.

276. Harada K, Shimizu T, Mukai Y, Kuwajima K, Sato T, Usui M, et al. Phenotypic and Molecular Characterization of Antimicrobial Resistance in Klebsiella spp. Isolates from Companion Animals in Japan: Clonal Dissemination of Multidrug-Resistant Extended-Spectrum β-Lactamase-Producing Klebsiella pneumoniae. Front Microbiol. 2016;7: 1021.

277. Castanheira M, Farrell SE, Wanger A, Rolston KV, Jones RN, Mendes RE. Rapid expansion of KPC-2-producing Klebsiella pneumoniae isolates in two Texas hospitals due to clonal spread of ST258 and ST307 lineages. Microb Drug Resist. 2013;19: 295–297.

278. Habeeb MA, Haque A, Nematzadeh S, Iversen A, Giske CG. High prevalence of 16S rRNA methylase RmtB among CTX-M extended-spectrum β-lactamase-producing Klebsiella pneumoniae from Islamabad, Pakistan. Int J Antimicrob Agents. 2013;41: 524–526.

279. Wyres KL, Hawkey J, Hetland MAK, Fostervold A, Wick RR, Judd LM, et al. Emergence and rapid global dissemination of CTX-M-15-associated Klebsiella pneumoniae strain ST307. J Antimicrob Chemother. 2019;74: 577–581.

280. Villa L, Feudi C, Fortini D, Brisse S, Passet V, Bonura C, et al. Diversity, virulence, and antimicrobial resistance of the KPC-producing Klebsiella pneumoniae ST307 clone. Microb Genom. 2017;3: e000110.

281. Strydom KA, Chen L, Kock MM, Stoltz AC, Peirano G, Nobrega DB, et al. Klebsiella pneumoniae ST307 with OXA-181: threat of a high-risk clone and promiscuous plasmid in a resource-constrained healthcare setting. J Antimicrob Chemother. 2020;75: 896–902.

282. Bocanegra-Ibarias P, Garza-González E, Morfín-Otero R, Barrios H, Villarreal-Treviño L, Rodríguez-Noriega E, et al. Molecular and microbiological report of a hospital outbreak of NDM-1-carrying Enterobacteriaceae in Mexico. PLoS One. 2017;12: e0179651.

283. Piazza A, Comandatore F, Romeri F, Brilli M, Dichirico B, Ridolfo A, et al. Identification of blaVIM-1 Gene in ST307 and ST661 Klebsiella pneumoniae Clones in Italy: Old Acquaintances for New Combinations. Microb Drug Resist. 2019;25: 787–790.

284. Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis JJ, et al. Population Genomic Analysis of 1,777 Extended-Spectrum Beta-Lactamase-Producing Klebsiella pneumoniae Isolates, Houston, Texas: Unexpected Abundance of Clonal Group 307. MBio. 2017;8. doi:10.1128/mBio.00489-17

285. Bonura C, Giuffrè M, Aleo A, Fasciana T, Di Bernardo F, Stampone T, et al. An Update of the Evolving Epidemic of blaKPC Carrying Klebsiella pneumoniae in Sicily, Italy, 2014: Emergence of Multiple Non-ST258 Clones. PLoS One. 2015;10: e0132936.

286. Cejas D, Elena A, Guevara Nuñez D, Sevillano Platero P, De Paulis A, Magariños F, et al. Changing epidemiology of KPC-producing Klebsiella pneumoniae in Argentina: Emergence of hypermucoviscous ST25 and high-risk clone ST307. J Glob Antimicrob Resist. 2019;18: 238–242.

287. Oteo-Iglesias J, Pérez-Vázquez M, Sola Campoy P, Moure Z, Sánchez Romero I, Sánchez Benito R, et al. Emergence of blood infections caused by carbapenemase-producing Klebsiella pneumoniae ST307 in Spain. J Antimicrob Chemother. 2020;75: 3402–3405.

288. Loconsole D, Accogli M, De Robertis AL, Capozzi L, Bianco A, Morea A, et al. Emerging high-risk ST101 and ST307 carbapenem-resistant Klebsiella pneumoniae clones from bloodstream infections in Southern Italy. Ann Clin Microbiol Antimicrob. 2020;19: 24.

289. Peirano G, Chen L, Kreiswirth BN, Pitout JDD. Emerging Antimicrobial-Resistant High-Risk Klebsiella pneumoniae Clones ST307 and ST147. Antimicrob Agents Chemother. 2020;64. doi:10.1128/AAC.01148-20

290. Davies YM, Cunha MPV, Dropa M, Lincopan N, Gomes VTM, Moreno LZ, et al. Pandemic Clones of CTX-M-15 Producing Klebsiella pneumoniae ST15, ST147, and ST307 in Companion Parrots. Microorganisms. 2022;10. doi:10.3390/microorganisms10071412

291. Savin M, Bierbaum G, Mutters NT, Schmithausen RM, Kreyenschmidt J, García-Meniño I, et al. Genetic Characterization of Carbapenem-Resistant Klebsiella spp. from Municipal and Slaughterhouse Wastewater. Antibiotics (Basel). 2022;11. doi:10.3390/antibiotics11040435

292. Rodrigues C, Desai S, Passet V, Gajjar D, Brisse S. Genomic evolution of the globally disseminated multidrug-resistant Klebsiella pneumoniae clonal group 147. Microb Genom. 2022;8. doi:10.1099/mgen.0.000737

293. Coelho A, Mirelis B, Alonso-Tarrés C, Nieves Larrosa M, Miró E, Clivillé Abad R, et al. Detection of three stable genetic clones of CTX-M-15-producing Klebsiella pneumoniae in the Barcelona metropolitan area, Spain. J Antimicrob Chemother. 2009;64: 862–864.

294. Tavoschi L, Forni S, Porretta A, Righi L, Pieralli F, Menichetti F, et al. Prolonged outbreak of New Delhi metallo-beta-lactamase-producing carbapenem-resistant Enterobacterales (NDM-CRE), Tuscany, Italy, 2018 to 2019. Euro Surveill. 2020;25. doi:10.2807/1560-7917.ES.2020.25.6.2000085

295. Peirano G, Lascols C, Hackel M, Hoban DJ, Pitout JDD. Molecular epidemiology of Enterobacteriaceae that produce VIMs and IMPs from the SMART surveillance program. Diagn Microbiol Infect Dis. 2014;78: 277–281.

296. Biedrzycka M, Urbanowicz P, Guzek A, Brisse S, Gniadkowski M, Izdebski R. Dissemination of Klebsiella pneumoniae ST147 NDM-1 in Poland, 2015-19. J Antimicrob Chemother. 2021;76: 2538–2545.

297. Di Pilato V, Errico G, Monaco M, Giani T, Del Grosso M, Antonelli A, et al. The changing epidemiology of carbapenemase-producing Klebsiella pneumoniae in Italy: toward polyclonal evolution with emergence of high-risk lineages. J Antimicrob Chemother. 2021;76: 355–361.

298. Mammina C, Bonura C, Aleo A, Fasciana T, Brunelli T, Pesavento G, et al. Sequence type 101 (ST101) as the predominant carbapenem-non-susceptible Klebsiella pneumoniae clone in an acute general hospital in Italy. Int J Antimicrob Agents. 2012;39: 543–545.

299. Roe CC, Vazquez AJ, Esposito EP, Zarrilli R, Sahl JW. Diversity, Virulence, and Antimicrobial Resistance in Isolates From the Newly Emerging Klebsiella pneumoniae ST101 Lineage. Front Microbiol. 2019;10: 542.

300. Oteo J, Pérez-Vázquez M, Bautista V, Ortega A, Zamarrón P, Saez D, et al. The spread of KPC-producing Enterobacteriaceae in Spain: WGS analysis of the emerging high-risk clones of Klebsiella pneumoniae ST11/KPC-2, ST101/KPC-2 and ST512/KPC-3. J Antimicrob Chemother. 2016;71: 3392–3399.

301. Shankar C, Shankar BA, Manesh A, Veeraraghavan B. KPC-2 producing ST101 Klebsiella pneumoniae from bloodstream infection in India. J Med Microbiol. 2018;67: 927–930.

302. Avgoulea K, Di Pilato V, Zarkotou O, Sennati S, Politi L, Cannatelli A, et al. Characterization of Extensively Drug-Resistant or Pandrug-Resistant Sequence Type 147 and 101 OXA-48-Producing Klebsiella pneumoniae Causing Bloodstream Infections

in Patients in an Intensive Care Unit. Antimicrob Agents Chemother. 2018;62. doi:10.1128/AAC.02457-17

303. Palmieri M, D'Andrea MM, Pelegrin AC, Mirande C, Brkic S, Cirkovic I, et al. Genomic Epidemiology of Carbapenem- and Colistin-Resistant Klebsiella pneumoniae Isolates From Serbia: Predominance of ST101 Strains Carrying a Novel OXA-48 Plasmid. Front Microbiol. 2020;11: 294.

304. Hazırolan G, Karagöz A. Emergence of carbapenemase-producing and colistin resistant Klebsiella pneumoniae ST101 high-risk clone in Turkey. Acta Microbiol Immunol Hung. 2020;67: 216–221.

305. Brahmia S, Lalaoui R, Nedjai S, Djahmi N, Chettibi S, Rolain J-M, et al. First Clinical Cases of KPC-2-Producing Klebsiella pneumoniae ST258 in Algeria and Outbreak of Klebsiella pneumoniae ST101 Harboring blaOXA-48 Gene in the Urology Department of Annaba Hospital. Microb Drug Resist. 2021;27: 652–659.

306. Messaoudi A, Haenni M, Bouallègue O, Saras E, Chatre P, Chaouch C, et al. Dynamics and molecular features of OXA-48-like-producing Klebsiella pneumoniae lineages in a Tunisian hospital. J Glob Antimicrob Resist. 2020;20: 87–93.

307. Gonçalves GB, Furlan JPR, Vespero EC, Pelisson M, Stehling EG, Pitondo-Silva A. Spread of multidrug-resistant high-risk Klebsiella pneumoniae clones in a tertiary hospital from southern Brazil. Infect Genet Evol. 2017;56: 1–7.

308. De Koster S, Rodriguez Ruiz JP, Rajakani SG, Lammens C, Glupczynski Y, Goossens H, et al. Diversity in the characteristics of Klebsiella pneumoniae ST101 of human, environmental, and animal origin. Front Microbiol. 2022;13: 838207.

309. Brisse S, Fevre C, Passet V, Issenhuth-Jeanjean S, Tournebize R, Diancourt L, et al. Virulent clones of Klebsiella pneumoniae: identification and evolutionary scenario based on genomic and phenotypic characterization. PLoS One. 2009;4: e4982.

310. Luo Y, Wang Y, Ye L, Yang J. Molecular epidemiology and virulence factors of pyogenic liver abscess causing Klebsiella pneumoniae in China. Clin Microbiol Infect. 2014;20: O818–24.

311. Qu T-T, Zhou J-C, Jiang Y, Shi K-R, Li B, Shen P, et al. Clinical and microbiological characteristics of Klebsiella pneumoniae liver abscess in East China. BMC Infect Dis. 2015;15: 161.

312. Lam MMC, Wyres KL, Duchêne S, Wick RR, Judd LM, Gan Y-H, et al. Population genomics of hypervirulent Klebsiella pneumoniae clonal-group 23 reveals early emergence and rapid global dissemination. Nat Commun. 2018;9: 2703.

313. Struve C, Roe CC, Stegger M, Stahlhut SG, Hansen DS, Engelthaler DM, et al. Mapping the Evolution of Hypervirulent Klebsiella pneumoniae. MBio. 2015;6: e00630.

314. Risk Assessment: Emergence of hypervirulent Klebsiella pneumoniae ST23 carrying carbapenemase genes in EU/EEA countries. In: European Centre for Disease

Prevention and Control [Internet]. 17 Mar 2021 [cited 9 Jan 2023]. Available: https://www.ecdc.europa.eu/en/publications-data/risk-assessment-emergence-hypervirulent-klebsiella-pneumoniae-eu-eea

315. Lan P, Jiang Y, Zhou J, Yu Y. A global perspective on the convergence of hypervirulence and carbapenem resistance in Klebsiella pneumoniae. J Glob Antimicrob Resist. 2021;25: 26–34.

316. Balloux F, Brønstad Brynildsrud O, van Dorp L, Shaw LP, Chen H, Harris KA, et al. From Theory to Practice: Translating Whole-Genome Sequencing (WGS) into the Clinic. Trends Microbiol. 2018;26: 1035–1048.

317. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl J m., Laurent F, et al. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill. 2013;18: 20380.

318. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. Pathology. 2015;47: 199–210.

319. Price V, Ngwira LG, Lewis JM, Baker KS, Peacock SJ, Jauneikaite E, et al. A systematic review of economic evaluations of whole-genome sequencing for the surveillance of bacterial pathogens. Microb Genom. 2023;9. doi:10.1099/mgen.0.000947

320. Lewis T, Loman NJ, Bingle L, Jumaa P, Weinstock GM, Mortiboy D, et al. High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates from a hospital outbreak. J Hosp Infect. 2010;75: 37–41.

321. Köser CU, Holden MTG, Ellington MJ, Cartwright EJP, Brown NM, Ogilvy-Stuart AL, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med. 2012;366: 2267–2275.

322. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, NISC Comparative Sequencing Program Group, Henderson DK, et al. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med. 2012;4: 148ra116.

323. Popovich KJ, Snitkin ES. Whole Genome Sequencing—Implications for Infection Prevention and Outbreak Investigations. Curr Infect Dis Rep. 2017;19: 15.

324. Hawken SE, Snitkin ES. Genomic epidemiology of multidrug-resistant Gram-negative organisms. Ann N Y Acad Sci. 2019;1435: 39–56.

325. Sundermann AJ, Chen J, Miller JK, Martin EM, Snyder GM, Van Tyne D, et al. Whole-genome sequencing surveillance and machine learning for healthcare outbreak detection and investigation: A systematic review and summary. Antimicrob Steward Healthc Epidemiol. 2022;2: e91.

326. Stimson J, Gardy J, Mathema B, Crudu V, Cohen T, Colijn C. Beyond the SNP Threshold: Identifying Outbreak Clusters Using Inferred Transmissions. Mol Biol Evol. 2019;36: 587–603.

327. Ludden C, Coll F, Gouliouris T, Restif O, Blane B, Blackwell GA, et al. Defining nosocomial transmission of Escherichia coli and antimicrobial resistance genes: a genomic surveillance study. Lancet Microbe. 2021;2: e472–e480.

328. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis. 2013;13: 137–146.

329. Sherry NL, Lane CR, Kwong JC, Schultz M, Sait M, Stevens K, et al. Genomics for Molecular Epidemiology and Detecting Transmission of Carbapenemase-Producing Enterobacterales in Victoria, Australia, 2012 to 2016. J Clin Microbiol. 2019;57. doi:10.1128/JCM.00573-19

330. Thorpe HA, Booton R, Kallonen T, Gibbon MJ, Couto N, Passet V, et al. A large-scale genomic snapshot of Klebsiella spp. isolates in Northern Italy reveals limited transmission between clinical and non-clinical settings. Nat Microbiol. 2022;7: 2054–2067.

331. Gorrie CL, Da Silva AG, Ingle DJ, Higgs C, Seemann T, Stinear TP, et al. Key parameters for genomics-based real-time detection and tracking of multidrug-resistant bacteria: a systematic analysis. Lancet Microbe. 2021;2: e575–e583.

332. Bush SJ, Foster D, Eyre DW, Clark EL, De Maio N, Shaw LP, et al. Genomic diversity affects the accuracy of bacterial single-nucleotide polymorphism–calling pipelines. Gigascience. 2020;9: giaa007.

333. Valiente-Mullor C, Beamud B, Ansari I, Francés-Cuesta C, García-González N, Mejía L, et al. One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. PLoS Comput Biol. 2021;17: e1008678.

334. De Maio N, Wu C-H, Wilson DJ. SCOTTI: Efficient Reconstruction of Transmission within Outbreaks with the Structured Coalescent. PLoS Comput Biol. 2016;12: e1005130.

335. Didelot X, Fraser C, Gardy J, Colijn C. Genomic Infectious Disease Epidemiology in Partially Sampled and Ongoing Outbreaks. Mol Biol Evol. 2017;34: 997–1007.

336. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nat Rev Microbiol. 2016;14: 150–162.

337. Hall MD, Woolhouse MEJ, Rambaut A. Using genomics data to reconstruct transmission trees during disease outbreaks. Rev Sci Tech. 2016;35: 287–296.

338. Fricke WF, Rasko DA. Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. Nat Rev Genet. 2014;15: 49–55.

339. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med. 2006;34: 1589–1596.

340. van Belkum A, Dunne WM Jr. Next-generation antimicrobial susceptibility testing. J Clin Microbiol. 2013;51: 2018–2024.

341. van Belkum A, Bachmann TT, Lüdke G, Lisby JG, Kahlmeter G, Mohess A, et al. Developmental roadmap for antimicrobial susceptibility testing systems. Nat Rev Microbiol. 2019;17: 51–62.

342. Jorgensen JH, Ferraro MJ. Antimicrobial susceptibility testing: a review of general principles and contemporary practices. Clin Infect Dis. 2009;49: 1749–1755.

343. Su M, Satola SW, Read TD. Genome-Based Prediction of Bacterial Antibiotic Resistance. J Clin Microbiol. 2019;57. doi:10.1128/JCM.01405-18

344. McDermott PF, Tyson GH, Kabera C, Chen Y, Li C, Folster JP, et al. Whole-Genome Sequencing for Detecting Antimicrobial Resistance in Nontyphoidal Salmonella. Antimicrob Agents Chemother. 2016;60: 5515–5520.

345. Eyre DW, De Silva D, Cole K, Peters J, Cole MJ, Grad YH, et al. WGS to predict antibiotic MICs for Neisseria gonorrhoeae. J Antimicrob Chemother. 2017;72: 1937–1947.

346. Hendriksen RS, Bortolaia V, Tate H, Tyson GH, Aarestrup FM, McDermott PF. Using Genomics to Track Global Antimicrobial Resistance. Front Public Health. 2019;7: 242.

347. Nunez-Garcia J, AbuOun M, Storey N, Brouwer MS, Delgado-Blas JF, Mo SS, et al. Harmonisation of in-silico next-generation sequencing based methods for diagnostics and surveillance. Sci Rep. 2022;12: 14372.

348. Stoesser N, Batty EM, Eyre DW, Morgan M, Wyllie DH, Del Ojo Elias C, et al. Predicting antimicrobial susceptibilities for Escherichia coli and Klebsiella pneumoniae isolates using whole genomic sequence data. J Antimicrob Chemother. 2013;68: 2234–2244.

349. Bauer MJ, Peri AM, Lüftinger L, Beisken S, Bergh H, Forde BM, et al. Optimized Method for Bacterial Nucleic Acid Extraction from Positive Blood Culture Broth for Whole-Genome Sequencing, Resistance Phenotype Prediction, and Downstream Molecular Applications. J Clin Microbiol. 2022;60: e0101222.

350. Kim JI, Maguire F, Tsang KK, Gouliouris T, Peacock SJ, McAllister TA, et al. Machine Learning for Antimicrobial Resistance Prediction: Current Practice, Limitations, and Clinical Perspective. Clin Microbiol Rev. 2022;35: e0017921.

351. Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, et al. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. Sci Rep. 2018;8: 421.

352. Ellington MJ, Ekelund O, Aarestrup FM, Canton R, Doumith M, Giske C, et al. The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee. Clin Microbiol Infect. 2017;23: 2–22.

353. Peacock SJ, Parkhill J, Brown NM. Changing the paradigm for hospital outbreak detection by leading with genomic surveillance of nosocomial pathogens. Microbiology. 2018;164: 1213–1219.

354. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, et al. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of

verotoxigenic Escherichia coli. J Clin Microbiol. 2014;52: 1501–1510.

355. Parcell BJ, Gillespie SH, Pettigrew KA, Holden MTG. Clinical perspectives in integrating whole-genome sequencing into the investigation of healthcare and public health outbreaks - hype or help? J Hosp Infect. 2021;109: 1–9.

356. Argimón S, Masim MAL, Gayeta JM, Lagrada ML, Macaranas PKV, Cohen V, et al. Integrating whole-genome sequencing within the National Antimicrobial Resistance Surveillance Program in the Philippines. Nat Commun. 2020;11: 2719.

357. Hodcroft EB, Zuber M, Nadeau S, Vaughan TG. SeqCOVID-SPAIN consortium. Stadler, T.

358. Revez J, Espinosa L, Albiger B, Leitmeyer KC, Struelens MJ, ECDC National Microbiology Focal Points and Experts Group. Survey on the Use of Whole-Genome Sequencing for Infectious Diseases Surveillance: Rapid Expansion of European National Capacities, 2015-2016. Front Public Health. 2017;5: 347.

359. Argimón S, David S, Underwood A, Abrudan M, Wheeler NE, Kekre M, et al. Rapid Genomic Characterization and Global Surveillance of Klebsiella Using Pathogenwatch. Clin Infect Dis. 2021;73: S325–S335.

360. Sherry NL, Horan KA, Ballard SA, Gonçalves da Silva A, Gorrie CL, Schultz MB, et al. An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. Nat Commun. 2023;14: 60.

361. NIHR Global Health Research Unit on Genomic Surveillance of AMR. Whole-genome sequencing as part of national and international surveillance programmes for antimicrobial resistance: a roadmap. BMJ Glob Health. 2020;5. doi:10.1136/bmjgh-2019-002244

362. Doyle RM, O'Sullivan DM, Aller SD, Bruchmann S, Clark T, Coello Pelegrin A, et al. Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. Microb Genom. 2020;6. doi:10.1099/mgen.0.000335

363. Cañada-García JE, Pérez-Vázquez M, Oteo-Iglesias J. RedLabRA; a Spanish Network of Microbiology Laboratories for the Surveillance of Antibiotic Resistant Microorganisms. Rev Esp Quimioter. 2021;34 Suppl 1: 12–14.

364. Miro E, Rossen JWA, Chlebowicz MA, Harmsen D, Brisse S, Passet V, et al. Core/Whole Genome Multilocus Sequence Typing and Core Genome SNP-Based Typing of OXA-48-Producing Klebsiella pneumoniae Clinical Isolates From Spain. Front Microbiol. 2019;10: 2961.

365. Miró E, Agüero J, Larrosa MN, Fernández A, Conejo MC, Bou G, et al. Prevalence and molecular epidemiology of acquired AmpC β-lactamases and carbapenemases in Enterobacteriaceae isolates from 35 hospitals in Spain. Eur J Clin Microbiol Infect Dis. 2013;32: 253–259.

366. Pérez-Vázquez M, Oteo J, García-Cobos S, Aracil B, Harris SR, Ortega A, et al. Phylogeny,

resistome and mobile genetic elements of emergent OXA-48 and OXA-245 Klebsiella pneumoniae clones circulating in Spain. J Antimicrob Chemother. 2016;71: 887–896.

367. Cañada-García JE, Grippo N, de Arellano ER, Bautista V, Lara N, Navarro AM, et al. Phenotypic and molecular characterization of IMP-producing Enterobacterales in Spain: Predominance of IMP-8 in Klebsiella pneumoniae and IMP-22 in Enterobacter roggenkampii. Front Microbiol. 2022;13: 1000787.

368. Muñoz I, Vanaclocha H, Martín-Sierra M, González F. [Microbiological Surveillance Network in the Valencian community]. Enferm Infecc Microbiol Clin. 2008;26: 77–81.

369. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer Science & Business Media; 2009.

370. Magiorakos A-P, Srinivasan A, Carey RB, Carmeli Y, Falagas ME, Giske CG, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect. 2012;18: 268–281.

371. Pechlivanis I, Wawrzyniak S, Engelhardt M, Schmieder K. Evidence level in the treatment of meningioma with focus on the comparison between surgery versus radiotherapy. A review. J Neurosurg Sci. 2011;55: 319–328.

372. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLOS Computational Biology. 2017. p. e1005595. doi:10.1371/journal.pcbi.1005595

373. Wick R. Guide to bacterial genome assembly. Github; Available: https://github.com/rrwick/Trycycler/wiki/Guide-to-bacterial-genome-assembly

374. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30: 2068–2069.

375. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, et al. Identification of Klebsiella capsule synthesis loci from whole genome data. Microb Genom. 2016;2: e000102.

376. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep. 2021;11: 12728.

377. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother. 2020;75: 3491–3500.

378. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421.

379. Simner PJ, Beisken S, Bergman Y, Ante M, Posch AE, Tamma PD. Defining Baseline Mechanisms of Cefiderocol Resistance in the Enterobacterales. Microb Drug Resist.

2022;28: 161–170.

380. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother. 2014;58: 3895–3903.

381. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44: W16–21.

382. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20: 257.

383. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29: 1072–1075.

384. Smits THM. The importance of genome sequence quality to microbial comparative genomics. BMC Genomics. 2019;20: 662.

385. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25: 1754–1760.

386. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011;27: 2987–2993.

387. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10. doi:10.1093/gigascience/giab008

388. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17: 132.

389. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35: 526–528.

390. Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A. 1973;70: 3321–3323.

391. Simpson EH. Measurement of Diversity. Nature. 1949;163: 688–688.

392. Pham MH, Hoi LT, Beale MA, Khokhar FA, Hoa NT, Musicha P, et al. Evidence of widespread endemic populations of highly multidrug resistant Klebsiella pneumoniae in hospital settings in Hanoi, Vietnam: a prospective cohort study. Lancet Microbe. 2023. doi:10.1016/S2666-5247(22)00338-X

393. Comandatore F, Sassera D, Bayliss SC, Scaltriti E, Gaiarsa S, Cao X, et al. Gene Composition as a Potential Barrier to Large Recombinations in the Bacterial Pathogen Klebsiella pneumoniae. Genome Biol Evol. 2019;11: 3240–3251.

394. Castillo-Ramírez S. Beyond microbial core genomic epidemiology: towards pan genomic epidemiology. Lancet Microbe. 2022;3: e244–e245.

395. Perrin A, Rocha EPC. PanACoTA: a modular tool for massive microbial comparative genomics. NAR Genom Bioinform. 2021;3: lqaa106.

396. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020;21: 180.

397. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30: 3059–3066.

398. Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016;4: e1660.

399. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25: 1972–1973.

400. Steenwyk JL, Buida TJ 3rd, Li Y, Shen X-X, Rokas A. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. 2020;18: e3001007.

401. Filipski A, Murillo O, Freydenzon A, Tamura K, Kumar S. Prospects for building large timetrees using molecular data with incomplete gene coverage among species. Mol Biol Evol. 2014;31: 2542–2550.

402. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microbial Genomics. 2016. doi:10.1099/mgen.0.000056

403. tseemann. GitHub - tseemann/snippy: Rapid haploid variant calling and core genome alignment. In: GitHub [Internet]. [cited 28 Mar 2022]. Available: https://github.com/tseemann/snippy

404. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26: 841–842.

405. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32: 268–274.

406. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 2013;30: 1188–1195.

407. Xu S, Li L, Luo X, Chen M, Tang W, Zhan L, et al. *Ggtree* : A serialized data object for visualization of a phylogenetic tree and annotation data. Imeta. 2022;1. doi:10.1002/imt2.56

408. Hennart M, Guglielmini J, Bridel S, Maiden MCJ, Jolley KA, Criscuolo A, et al. A Dual Barcoding Approach to Bacterial Strain Nomenclature: Genomic Taxonomy of Klebsiella pneumoniae Strains. Mol Biol Evol. 2022;39. doi:10.1093/molbev/msac135

409. Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, Santos S, et al. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. Microb Genom. 2018;4. doi:10.1099/mgen.0.000166

410. Klebsiella PasteurMLST. [cited 15 Jan 2023]. Available: https://bigsdb.pasteur.fr/cgi-bin/bigsdb/bigsdb.pl?db=pubmlst_klebsiella_isolates

411. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res. 2019;47: 5539–5549.

412. Gouliouris T, Coll F, Ludden C, Blane B, Raven KE, Naydenova P, et al. Quantifying acquisition and transmission of Enterococcus faecium using genomic surveillance. Nat Microbiol. 2021;6: 103–111.

413. Sherry NL, Gorrie CL, Kwong JC, Higgs C, Stuart RL, Marshall C, et al. Multi-site implementation of whole genome sequencing for hospital infection control: A prospective genomic epidemiological analysis. The Lancet Regional Health - Western Pacific. 2022;23: 100446.

414. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol. 2016;2: vew007.

415. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. Microb Genom. 2017;3: e000128.

416. Arredondo-Alonso S, Rogers MRC, Braat JC, Verschuuren TD, Top J, Corander J, et al. mlplasmids: a user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. Microb Genom. 2018;4. doi:10.1099/mgen.0.000224

417. Cury J, Touchon M, Rocha EPC. Integrative and conjugative elements and their hosts: composition, distribution and organization. Nucleic Acids Res. 2017;45: 8943–8956.

418. Pfeifer E, Moura de Sousa JA, Touchon M, Rocha EPC. Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. Nucleic Acids Res. 2021;49: 2655–2673.

419. Pfeifer E, Bonnin RA, Rocha EPC. Phage-Plasmids Spread Antibiotic Resistance Genes through Infection and Lysogenic Conversion. MBio. 2022;13: e0185122.

420. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microb Genom. 2018;4. doi:10.1099/mgen.0.000206

421. Galata V, Fehlmann T, Backes C, Keller A. PLSDB: a resource of complete bacterial plasmids. Nucleic Acids Res. 2019;47: D195–D202.

422. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 2008;36: W181–4.

423. Guy L, Kultima JR, Andersson SGE. genoPlotR: comparative gene and genome visualization in R. Bioinformatics. 2010;26: 2334–2335.

424. Nordmann P, Boulanger AE, Poirel L. NDM-4 Metallo-β-Lactamase with Increased Carbapenemase Activity from Escherichia coli. Antimicrobial Agents and Chemotherapy. 2012. pp. 2184–2186. doi:10.1128/aac.05961-11

425. Poirel L, Walsh TR, Cuvillier V, Nordmann P. Multiplex PCR for detection of acquired carbapenemase genes. Diagn Microbiol Infect Dis. 2011;70: 119–123.

426. Viehweger A, Blumenscheit C, Lippmann N, Wyres KL, Brandt C, Hans JB, et al. Context-aware genomic surveillance reveals hidden transmission of a carbapenemase-producing Klebsiella pneumoniae. Microb Genom. 2021;7. doi:10.1099/mgen.0.000741

427. Low AJ, Koziol AG, Manninger PA, Blais B, Carrillo CD. ConFindr: rapid detection of intraspecies and cross-species contamination in bacterial whole-genome sequence data. PeerJ. 2019;7: e6995.

428. Goig GA, Blanco S, Garcia-Basteiro AL, Comas I. Contaminant DNA in bacterial sequencing experiments is a major source of false genetic variability. BMC Biol. 2020;18: 24.

429. Hamilton AL, Kamm MA, Ng SC, Morrison M. Proteus spp. as Putative Gastrointestinal Pathogens. Clin Microbiol Rev. 2018;31. doi:10.1128/CMR.00085-17

430. Baliga P, Shekar M, Kallappa GS. Genome-Wide Identification and Analysis of Chromosomally Integrated Putative Prophages Associated with Clinical Klebsiella pneumoniae Strains. Curr Microbiol. 2021;78: 2015–2024.

431. Grundmann H, Glasner C, Albiger B, Aanensen DM, Tomlinson CT, Andrasević AT, et al. Occurrence of carbapenemase-producing Klebsiella pneumoniae and Escherichia coli in the European survey of carbapenemase-producing Enterobacteriaceae (EuSCAPE): a prospective, multinational study. Lancet Infect Dis. 2017;17: 153–163.

432. Yu F, Hu L, Zhong Q, Hang Y, Liu Y, Hu X, et al. Dissemination of Klebsiella pneumoniae ST11 isolates with carbapenem resistance in integrated and emergency intensive care units in a Chinese tertiary hospital. J Med Microbiol. 2019;68: 882–889.

433. He Z, Xu W, Zhao H, Li W, Dai Y, Lu H, et al. Epidemiological characteristics an outbreak of ST11 multidrug-resistant and hypervirulent Klebsiella pneumoniae in Anhui, China. Front Microbiol. 2022;13: 996753.

434. Liu S, Ding Y, Xu Y, Li Z, Zeng Z, Liu J. An outbreak of extensively drug-resistant and hypervirulent Klebsiella pneumoniae in an intensive care unit of a teaching hospital in Southwest China. Front Cell Infect Microbiol. 2022;12: 979219.

435. Kochan TJ, Nozick SH, Medernach RL, Cheung BH, Gatesy SWM, Lebrun-Corbin M, et al. Genomic surveillance for multidrug-resistant or hypervirulent Klebsiella pneumoniae among United States bloodstream isolates. BMC Infect Dis. 2022;22: 603.

436. Hernández M, López-Urrutia L, Abad D, De Frutos Serna M, Ocampo-Sosa AA, Eiros JM.

First Report of an Extensively Drug-Resistant ST23 Klebsiella pneumoniae of Capsular Serotype K1 Co-Producing CTX-M-15, OXA-48 and ArmA in Spain. Antibiotics (Basel). 2021;10. doi:10.3390/antibiotics10020157

437. Hawkey J, Wyres KL, Judd LM, Harshegyi T, Blakeway L, Wick RR, et al. ESBL plasmids in Klebsiella pneumoniae: diversity, transmission and contribution to infection burden in the hospital setting. Genome Med. 2022;14: 97.

438. Rivera-Izquierdo M, Láinez-Ramos-Bossini AJ, Rivera-Izquierdo C, López-Gómez J, Fernández-Martínez NF, Redruello-Guerrero P, et al. OXA-48 Carbapenemase-Producing Enterobacterales in Spanish Hospitals: An Updated Comprehensive Review on a Rising Antimicrobial Resistance. Antibiotics (Basel). 2021;10. doi:10.3390/antibiotics10010089

439. Urbanowicz P, Bitar I, Izdebski R, Baraniak A, Literacka E, Hrabák J, et al. Epidemic Territorial Spread of IncP-2-Type VIM-2 Carbapenemase-Encoding Megaplasmids in Nosocomial Pseudomonas aeruginosa Populations. Antimicrob Agents Chemother. 2021;65. doi:10.1128/AAC.02122-20

440. Lopes R, Furlan JPR, Stehling EG. Acquisition of blaIMP-13 on a novel IncP-7 plasmid in XDR VIM-2-positive Pseudomonas aeruginosa belonging to the global high-risk clone ST235 in an agricultural ecosystem. J Glob Antimicrob Resist. 2022;30: 403–405.

441. Gijón D, Curiao T, Baquero F, Coque TM, Cantón R. Fecal carriage of carbapenemase-producing Enterobacteriaceae: a hidden reservoir in hospitalized and nonhospitalized patients. J Clin Microbiol. 2012;50: 1558–1563.

442. Loli A, Tzouvelekis LS, Tzelepi E, Carattoli A, Vatopoulos AC, Tassios PT, et al. Sources of diversity of carbapenem resistance levels in Klebsiella pneumoniae carrying blaVIM-1. J Antimicrob Chemother. 2006;58: 669–672.

443. Navon-Venezia S, Leavitt A, Schwaber MJ, Rasheed JK, Srinivasan A, Patel JB, et al. First report on a hyperepidemic clone of KPC-3-producing Klebsiella pneumoniae in Israel genetically related to a strain causing outbreaks in the United States. Antimicrob Agents Chemother. 2009;53: 818–820.

444. European Centre for Disease Prevention and Control. Antimicrobial resistance in the EU/EEA (EARS-Net) - Annual Epidemiological Report 2020. ECDC; 2022 Jul. Available: https://www.ecdc.europa.eu/sites/default/files/documents/AER-EARS-Net-2020.pdf

445. Asensio A, Oliver A, González-Diego P, Baquero F, Pérez-Díaz JC, Ros P, et al. Outbreak of a multiresistant Klebsiella pneumoniae strain in an intensive care unit: antibiotic use as risk factor for colonization and infection. Clin Infect Dis. 2000;30: 55–60.

446. Tórtola MT, Lavilla S, Miró E, González JJ, Larrosa N, Sabaté M, et al. First detection of a carbapenem-hydrolyzing metalloenzyme in two enterobacteriaceae isolates in Spain. Antimicrob Agents Chemother. 2005;49: 3492–3494.

447. Livermore DM, Canton R, Gniadkowski M, Nordmann P, Rossolini GM, Arlet G, et al. CTX-M: changing the face of ESBLs in Europe. J Antimicrob Chemother. 2007;59: 165–

174.

448. Ruiz de Alegría C, Rodríguez-Baño J, Cano ME, Hernández-Bello JR, Calvo J, Román E, et al. Klebsiella pneumoniae strains producing extended-spectrum beta-lactamases in Spain: microbiological and clinical features. J Clin Microbiol. 2011;49: 1134–1136.

449. Gijón D, Tedim AP, Valverde A, Rodríguez I, Morosini M-I, Coque TM, et al. Early OXA-48-Producing Isolates Recovered in a Spanish Hospital Reveal a Complex Introduction Dominated by Sequence Type 11 (ST11) and ST405 Klebsiella pneumoniae Clones. mSphere. 2020;5. doi:10.1128/mSphere.00080-20

450. Babouee B, Widmer AF, Dubuis O, Ciardo D, Droz S, Betsch BY, et al. Emergence of four cases of KPC-2 and KPC-3-carrying Klebsiella pneumoniae introduced to Switzerland, 2009-10. Euro Surveill. 2011;16. doi:10.2807/ese.16.11.19817-en

451. Politi L, Gartzonika K, Spanakis N, Zarkotou O, Poulou A, Skoura L, et al. Emergence of NDM-1-producing Klebsiella pneumoniae in Greece: evidence of a widespread clonal outbreak. J Antimicrob Chemother. 2019;74: 2197–2202.

452. Fuster B, Salvador C, Tormo N, García-González N, Gimeno C, González-Candelas F. Molecular epidemiology and drug-resistance mechanisms in carbapenem-resistant Klebsiella pneumoniae isolated in patients from a tertiary hospital in Valencia, Spain. J Glob Antimicrob Resist. 2020;22: 718–725.

453. Fajardo-Lubián A, Ben Zakour NL, Agyekum A, Qi Q, Iredell JR. Host adaptation and convergent evolution increases antibiotic resistance without loss of virulence in a major human pathogen. PLoS Pathog. 2019;15: e1007218.

454. Codjoe FS, Donkor ES. Carbapenem Resistance: A Review. Med Sci (Basel). 2017;6. doi:10.3390/medsci6010001

455. Fuster B, Tormo N, Salvador C, Gimeno C. Detection of two simultaneous outbreaks of Klebsiella pneumoniae coproducing OXA-48 and NDM-1 carbapenemases in a tertiary-care hospital in Valencia, Spain. New Microbes and New Infections. 2020. p. 100660. doi:10.1016/j.nmni.2020.100660

456. García-González N, Beamud B, Fuster B, Giner S, Domínguez MV, Sánchez A, et al. Tracking the Emergence and Dissemination of a blaNDM-23 Gene in a Multidrug Resistance Plasmid of Klebsiella pneumoniae. Microbiol Spectr. 2023; e0258522.

457. Hernández-García M, Pérez-Viso B, León-Sampedro R, Navarro-San Francisco C, López-Fresneña N, Díaz-Agero C, et al. Outbreak of NDM-1+CTX-M-15+DHA-1-producing Klebsiella pneumoniae high-risk clone in Spain owing to an undetectable colonised patient from Pakistan. Int J Antimicrob Agents. 2019;54: 233–239.

458. Nabeela Farhat AUK. Evolving trends of New Delhi Metallo-betalactamse (NDM) variants: A threat to antimicrobial resistance. Infect Genet Evol. 2020;86: 104588.

459. Li B, Hu Y, Wang Q, Yi Y, Woo PCY, Jing H, et al. Structural diversity of class 1 integrons and their associated gene cassettes in Klebsiella pneumoniae isolates from a hospital

in China. PLoS One. 2013;8: e75805.

460. Boyd DA, Tyler S, Christianson S, McGeer A, Muller MP, Willey BM, et al. Complete nucleotide sequence of a 92-kilobase plasmid harboring the CTX-M-15 extended-spectrum beta-lactamase involved in an outbreak in long-term-care facilities in Toronto, Canada. Antimicrob Agents Chemother. 2004;48: 3758–3764.

461. Chen C-J, Wu T-L, Lu P-L, Chen Y-T, Fung C-P, Chuang Y-C, et al. Closely related NDM-1-encoding plasmids from Escherichia coli and Klebsiella pneumoniae in Taiwan. PLoS One. 2014;9: e104899.

462. Wailan AM, Sartor AL, Zowawi HM, Perry JD, Paterson DL, Sidjabat HE. Genetic Contexts of blaNDM-1 in Patients Carrying Multiple NDM-Producing Strains. Antimicrob Agents Chemother. 2015;59: 7405–7410.

463. Toleman MA, Bennett PM, Walsh TR. ISCR elements: novel gene-capturing systems of the 21st century? Microbiol Mol Biol Rev. 2006;70: 296–316.

464. Dolejska M, Villa L, Dobiasova H, Fortini D, Feudi C, Carattoli A. Plasmid content of a clinically relevant Klebsiella pneumoniae clone from the Czech Republic producing CTX-M-15 and QnrB1. Antimicrob Agents Chemother. 2013;57: 1073–1076.

465. Bao J, Xie L, Ma Y, An R, Gu B, Wang C. Proteomic and Transcriptomic Analyses Indicate Reduced Biofilm-Forming Abilities in Cefiderocol-Resistant. Front Microbiol. 2021;12: 778190.

466. Gorrie CL, Mirčeta M, Wick RR, Judd LM, Lam MMC, Gomi R, et al. Genomic dissection of Klebsiella pneumoniae infections in hospital patients reveals insights into an opportunistic pathogen. Nat Commun. 2022;13: 3017.

467. Keynan Y, Rubinstein E. The changing face of Klebsiella pneumoniae infections in the community. Int J Antimicrob Agents. 2007;30: 385–389.

468. Chong Y, Shimoda S, Yakushiji H, Ito Y, Miyamoto T, Kamimura T, et al. Community spread of extended-spectrum β-lactamase-producing Escherichia coli, Klebsiella pneumoniae and Proteus mirabilis: a long-term study in Japan. J Med Microbiol. 2013;62: 1038–1043.

469. Marimuthu K, Venkatachalam I, Koh V, Harbarth S, Perencevich E, Cherng BPZ, et al. Whole genome sequencing reveals hidden transmission of carbapenemase-producing Enterobacterales. Nat Commun. 2022;13: 3052.

470. Ludden C, Moradigaravand D, Jamrozy D, Gouliouris T, Blane B, Naydenova P, et al. A One Health Study of the Genetic Relatedness of Klebsiella pneumoniae and Their Mobile Elements in the East of England. Clin Infect Dis. 2020;70: 219–226.

471. Sequence Set Browser :: NCBI. [cited 25 Apr 2023]. Available: https://www.ncbi.nlm.nih.gov/Traces/wgs/?page=1&view=wgs

472. Merchant S, Wood DE, Salzberg SL. Unexpected cross-species contamination in genome sequencing projects. PeerJ. 2014;2: e675.

473. Lu J, Salzberg SL. Removing contaminants from databases of draft genomes. PLoS Comput Biol. 2018;14: e1006277.

474. Spadar A, Phelan J, Elias R, Modesto A, Caneiras C, Marques C, et al. Genomic epidemiological analysis of Klebsiella pneumoniae from Portuguese hospitals reveals insights into circulating antimicrobial resistance. Sci Rep. 2022;12: 13791.

475. Yang Y, Yang Y, Chen G, Lin M, Chen Y, He R, et al. Molecular characterization of carbapenem-resistant and virulent plasmids in Klebsiella pneumoniae from patients with bloodstream infections in China. Emerg Microbes Infect. 2021;10: 700–709.

476. Teo JQ-M, Tang CY, Tan SH, Chang HY, Ong SM, Lee SJ-Y, et al. Genomic Surveillance of Carbapenem-Resistant Klebsiella pneumoniae from a Major Public Health Hospital in Singapore. Microbiol Spectr. 2022;10: e0095722.

477. Antimicrobial resistance surveillance in Europe 2023 - 2021 data. In: European Centre for Disease Prevention and Control [Internet]. 14 Apr 2023 [cited 3 May 2023]. Available: https://www.ecdc.europa.eu/en/publications-data/antimicrobial-resistance-surveillance-europe-2023-2021-data

478. Hernández-García M, Pérez-Viso B, Francisco CN-S, Baquero F, Morosini MI, Ruiz-Garbajosa P, et al. Intestinal co-colonization with different carbapenemase-producing Enterobacterales isolates is not a rare event in an OXA-48 endemic area. EClinicalMedicine. 2019. pp. 72–79. doi:10.1016/j.eclinm.2019.09.005

479. León-Sampedro R, DelaFuente J, Díaz-Agero C, Crellen T, Musicha P, Rodríguez-Beltrán J, et al. Pervasive transmission of a carbapenem resistance plasmid in the gut microbiota of hospitalized patients. Nat Microbiol. 2021;6: 606–616.

480. Nordmann P, Poirel L, Walsh TR, Livermore DM. The emerging NDM carbapenemases. Trends Microbiol. 2011;19: 588–595.

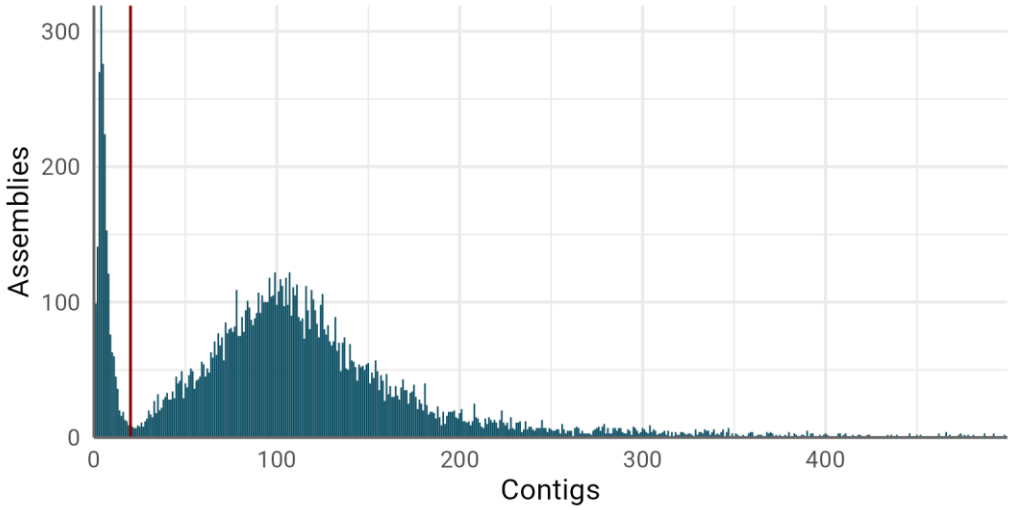
481. Adler M, Anjum M, Andersson DI, Sandegren L. Influence of acquired β-lactamases on the evolution of spontaneous carbapenem resistance in Escherichia coli. J Antimicrob Chemother. 2013;68: 51–59.

482. Baker S, Thomson N, Weill F-X, Holt KE. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. Science. 2018;360: 733–738.

SUPPLEMENTARY DATA

CHAPTER 3

Supplementary figures



Supplementary Figure 3.1. Distribution of the number of contigs per assembly of the whole collection (13,133). The threshold used of 20 contigs to include or not the genome for the custom plasmid database is marked by a red line.

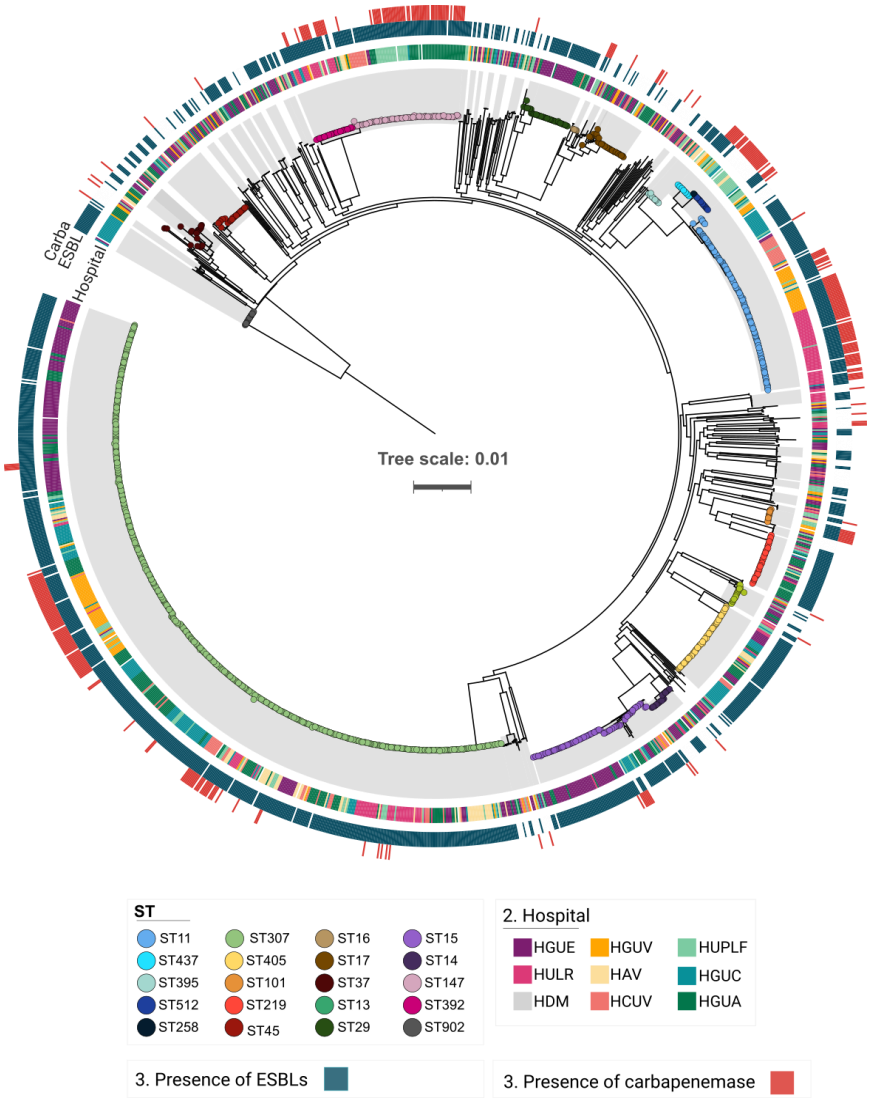
CHAPTER 4

Supplementary tables

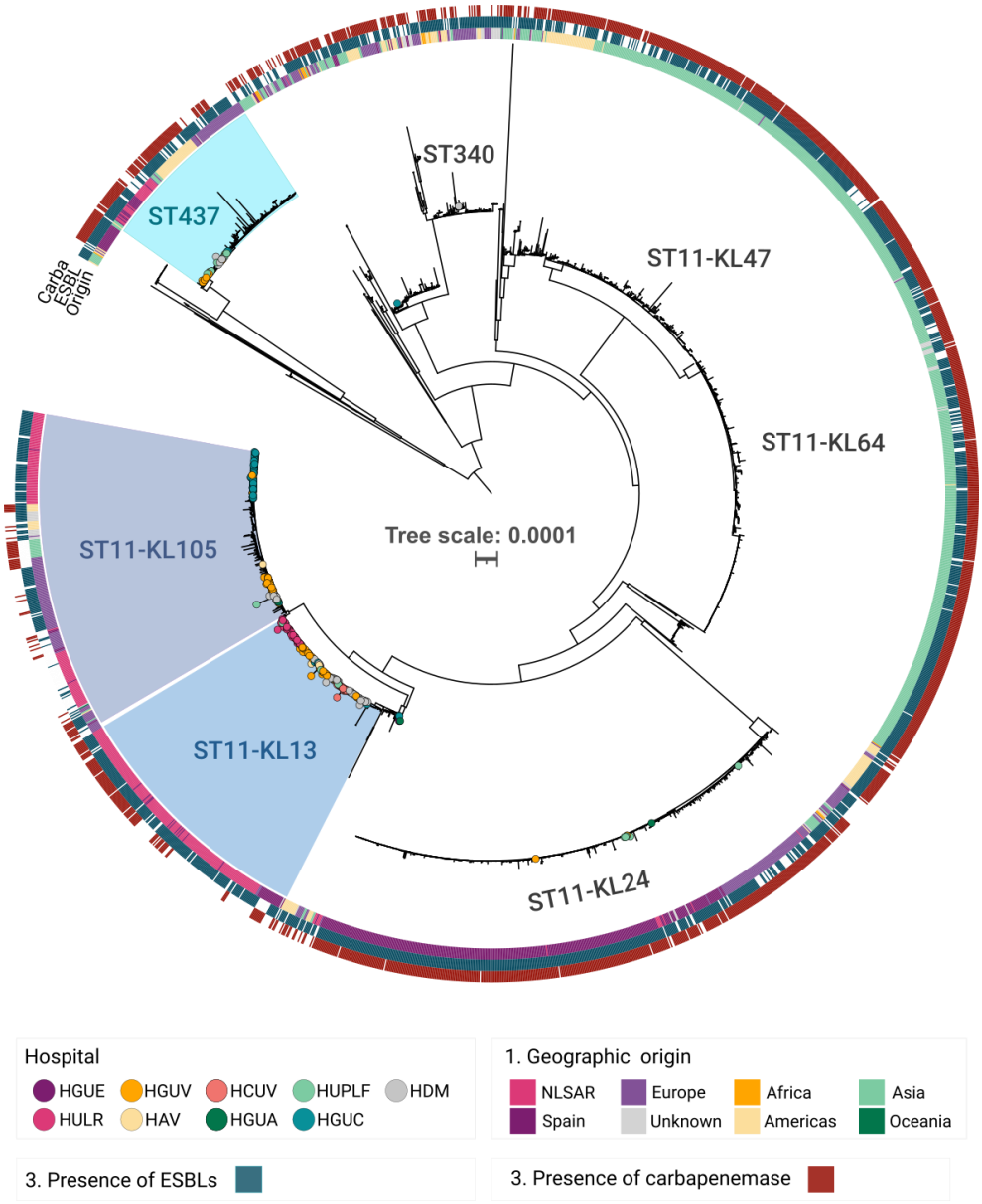
Supplementary Table 4.1. Information of all genomes included in the initial analysis, both from the database and sequenced in this project (SKPCV and NLSAR). This table contains information about the accession and biosample numbers, quality control statistics, such as the number of reads, taxonomy assignment by Kraken and assembly quality, heterozygous positions, mash distances and the Kleborate results. doi: https://doi.org/10.6084/m9.figshare.22637356.v1

CHAPTER 5

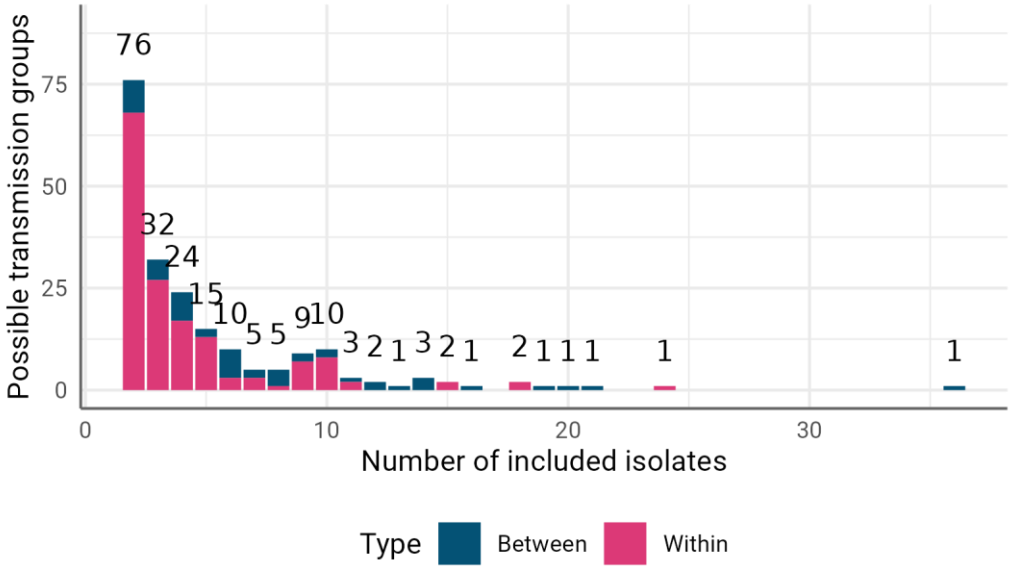
Supplementary figures



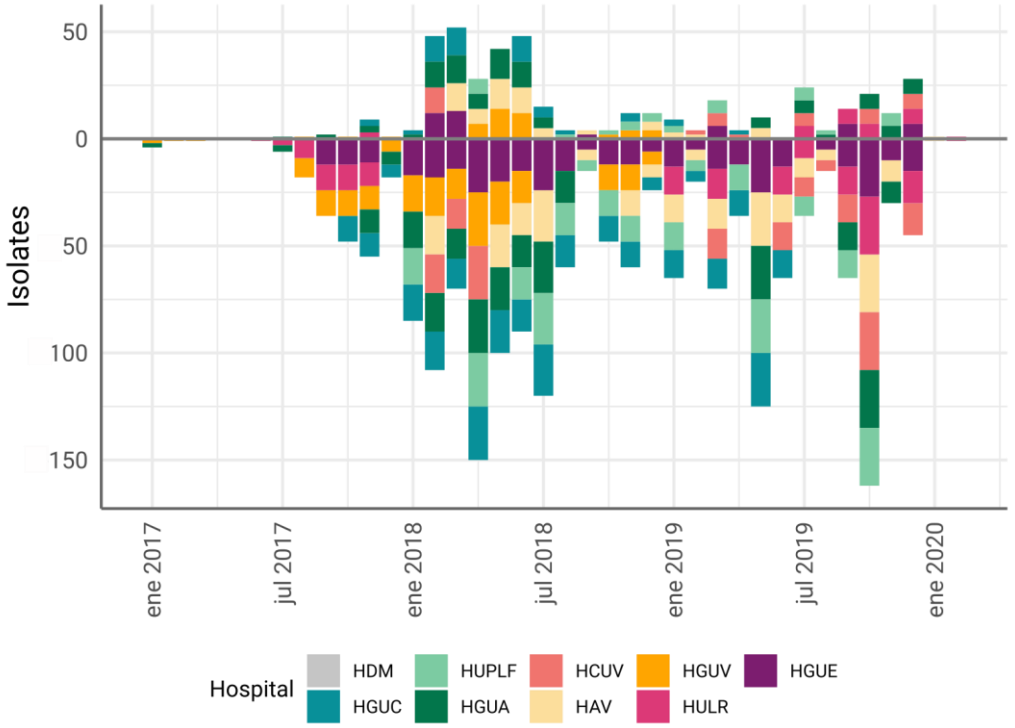
Supplementary Figure 5.1. Maximum likelihood tree of the 1,604 genomes collected under the SKPCV surveillance program. The tree was made using the core-genome alignment (90%, 4,034 genes, 3,928,610 bp). Sequence types are marked with different colours in dots at the end of each tip. Grey highlights indicate Sublineages (SLs). The inner circle indicates the hospital whereas the two outer circles indicate the presence of ESBLs and carbapenemases.



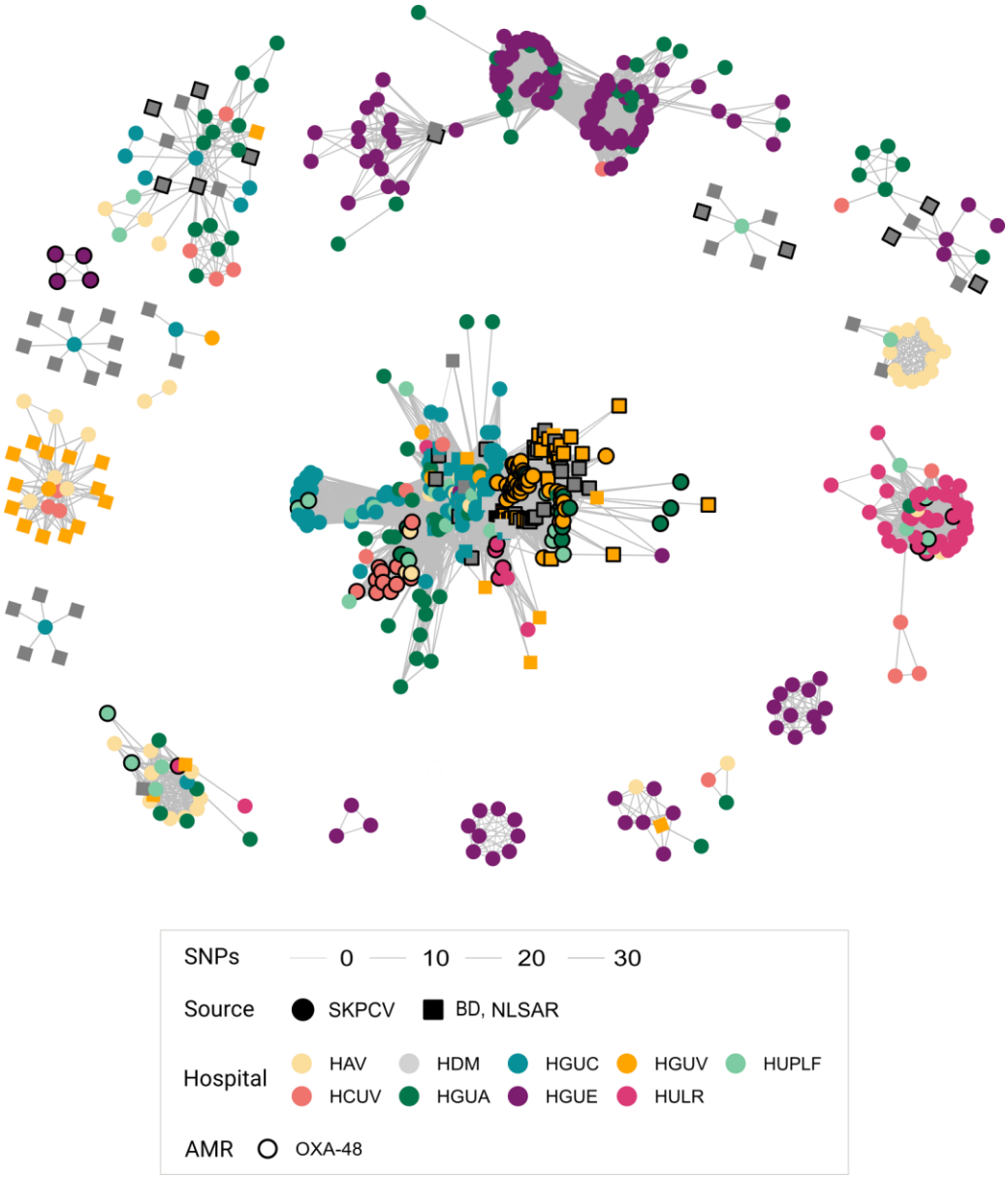
Supplementary Figure 5.2. Global phylogenetic tree of the sublineage that includes ST11 (SL3). The phylogeny with the core genome at 90% of 2,327 isolates and a total alignment length of 2,681,784 bp. Colours in dots at the end of each tip indicate the hospital. The inner circle indicates the geographic origin of each isolate whereas the two outer circles indicate the presence of ESBLs and carbapenemases.



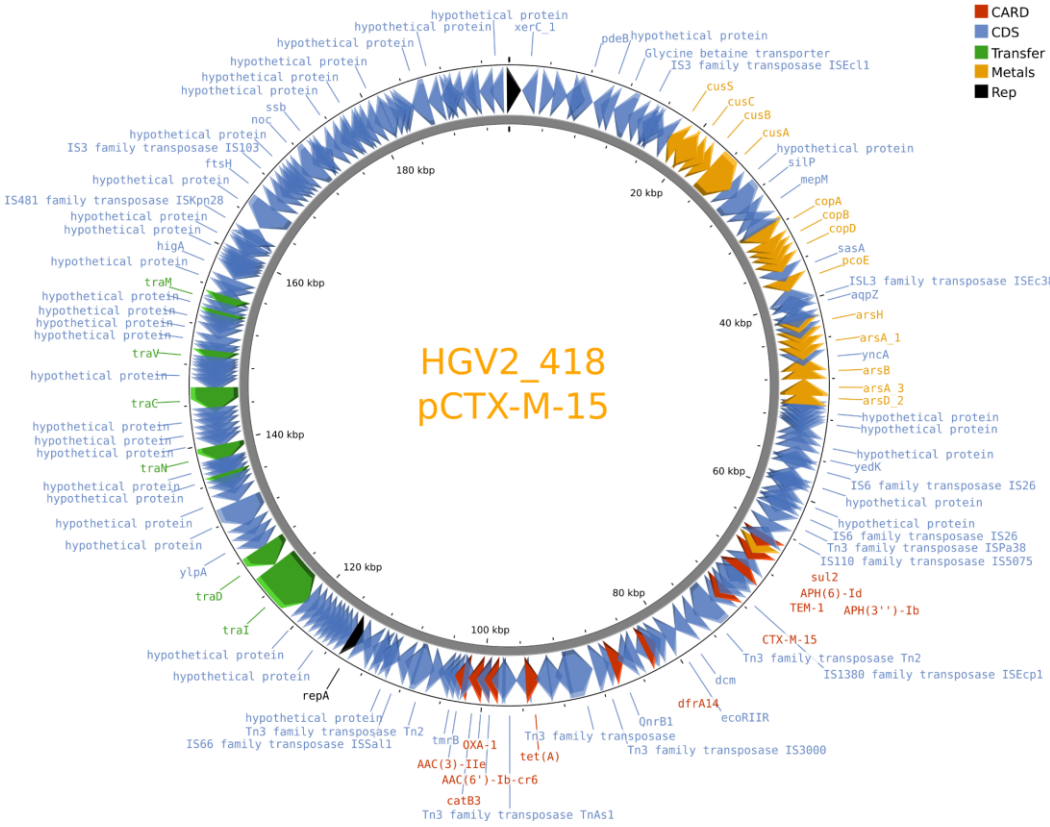
Supplementary Figure 5.3. Distribution of the sizes of possible transmission groups (PTG), where the y-axis represents the count of the PTG and the x-axis represents the size of the groups. The bars in the histogram are coloured to differentiate groups in which all isolates were collected in the same hospital (within) and groups with isolates from different hospitals (between).



Supplementary Figure 5.4. Histogram showing the number of ST307 isolates collected during the SKPCV. In the upper part isolates not included in any possible transmission chain are represented, whereas the lower part includes those that were included in possible transmission groups.



Supplementary Figure 5.5. Graphs representing relationships among ST307 genomes. Only pairwise relationships under 35 SNPs were considered. Isolates collected under the NLSAR project are coloured by hospitals. The shape indicates whether they were collected under the SKPCV project or not. Framed shapes indicate that the isolates carry a *bla*$_{\text{OXA-48}}$ gene.



Supplementary Figure 5.6. Plasmid carrying the *bla*$_{\text{CTX-M-15}}$ gene extracted from the long-read assembly of the HGV2_418 isolate. CDS are coloured in black for replicon genes, yellow if they are related to metal genes, red if they are AMR genes or green for transfer genes.

Supplementary tables

Supplementary Table 5.1. The number of episodes obtained from the RedMIVA database for each province, year, and resistance group.

Id/Type	Group	Susceptible	Resistant to 3G C	Resistant to 3GC + Fluoroquinolones + Aminglycosides	Carbapenem resitant	Resistant to 3GC + Fluoroquinolones + Aminoglycosides + Carbapenems	Total
Hospitalization group	Not hospitalized	40,268 (84.3%)	2,609 (5.5%)	4,607 (9.6%)	41 (0.1%)	235 (0.5%)	47,760 (100.0%)
Hospitalization group	Hospitalized	59,702 (71.6%)	6,366 (7.6%)	15,023 (18.0%)	363 (0.4%)	1,906 (2.3%)	83,360 (100.0%)
Province	ALC	37,252 (78.7%)	2,975 (6.3%)	6,110 (12.9%)	96 (0.2%)	901 (1.9%)	47,334 (100.0%)
Province	CAS	8,654 (77.5%)	640 (5.7%)	1,792 (16.0%)	17 (0.2%)	64 (0.6%)	11,167 (100.0%)
Province	VLC	54,063 (74.4%)	5,360 (7.4%)	11,728 (16.2%)	291 (0.4%)	1,176 (1.6%)	72,618 (100.0%)
Request year	2010	4,169 (87.7%)	151 (3.2%)	402 (8.5%)	5 (0.1%)	27 (0.6%)	4,754 (100.0%)
Request year	2011	5,165 (89.2%)	212 (3.7%)	399 (6.9%)	2 (0.0%)	12 (0.2%)	5,790 (100.0%)
Request year	2012	5,744 (87.7%)	290 (4.4%)	492 (7.5%)	9 (0.1%)	12 (0.2%)	6,547 (100.0%)
Request year	2013	5,782 (84.3%)	382 (5.6%)	656 (9.6%)	8 (0.1%)	30 (0.4%)	6,858 (100.0%)
Request year	2014	6,689 (78.8%)	505 (6.0%)	1,197 (14.1%)	27 (0.3%)	66 (0.8%)	8,484 (100.0%)
Request year	2015	7,735 (77.7%)	648 (6.5%)	1,417 (14.2%)	11 (0.1%)	143 (1.4%)	9,954 (100.0%)
Request year	2016	7,853 (75.6%)	641 (6.2%)	1,591 (15.3%)	39 (0.4%)	270 (2.6%)	10,394 (100.0%)
Request year	2017	8,139 (72.6%)	742 (6.6%)	2,072 (18.5%)	47 (0.4%)	218 (1.9%)	11,218 (100.0%)
Request year	2018	8,565 (69.5%)	971 (7.9%)	2,482 (20.1%)	54 (0.4%)	255 (2.1%)	12,327 (100.0%)
Request year	2019	9,502 (73.4%)	1,059 (8.2%)	2,139 (16.5%)	41 (0.3%)	202 (1.6%)	12,943 (100.0%)
Request year	2020	8,613 (73.0%)	1,010 (8.6%)	1,943 (16.5%)	32 (0.3%)	204 (1.7%)	11,802 (100.0%)
Request year	2021	10,217 (73.6%)	999 (7.2%)	2,289 (16.5%)	55 (0.4%)	323 (2.3%)	13,883 (100.0%)
Request year	2022	11,797 (73.0%)	1,365 (8.4%)	2,551 (15.8%)	74 (0.5%)	379 (2.3%)	16,166 (100.0%)

Supplementary Table 5.2. Estimated odd ratios of being hospitalised for each susceptible episode and AMR type studied.

Type	Odds ratio	Confidence interval	p-value	Significance
Susceptible	0.47	0.46 - 0.48	< 2.2e-16	***
Resistant to 3G C	1.43	1.36 - 1.5	< 2.2e-16	***
Resistant to 3GC + Fluoroquinolones + Aminglycosides	2.06	1.99 - 2.13	< 2.2e-16	***
Carbapenem resitant	5.09	3.68 - 7.22	< 2.2e-16	***
Resistant to 3GC + Fluoroquinolones + Aminoglycosides + Carbapenems	4.73	4.13 - 5.45	< 2.2e-16	***

Supplementary Table 5.3. Genomic and clinical information of the 1,604 isolates collected during the SKPCV. doi: https://doi.org/10.6084/m9.figshare.22779971.v1

Supplementary Table 5.4. Population diversity estimated for the whole population. In the lower hemi matrix nucleotide divergence (Dxy) per site between hospitals, is shown. The diagonal is the nucleotide diversity (π) per site within hospitals. The upper hemi matrix is the net divergence (Da) estimated between hospitals.

	HAV	HCUV	HGUA	HGUC	HGUE	HGUV	HULR	HUPLF
HAV	**0,004613**	0,000136	0,000101	0,000115	0,000163	0,000095	0,000233	0,000254
HCUV	0,005082	**0,005279**	0,000096	0,000173	0,000170	0,000065	0,000048	0,000022
HGUA	0,004952	0,005280	**0,005089**	0,000184	0,000061	0,000225	0,000288	0,000141
HGUC	0,005316	0,005707	0,005622	**0,005788**	0,000191	0,000163	0,000256	0,000272
HGUE	0,005074	0,005414	0,005210	0,005689	**0,005209**	0,000249	0,000375	0,000241
HGUV	0,004774	0,005077	0,005142	0,005430	0,005225	**0,004745**	0,000061	0,000174
HULR	0,004998	0,005146	0,005291	0,005608	0,005438	0,004892	**0,004917**	0,000108
HUPLF	0,005365	0,005466	0,005490	0,005971	0,005650	0,005352	0,005371	**0,005609**

Supplementary Table 5.5. Population diversity estimated for the whole population after removing the *bla*$_{CTX\text{-}m\text{-}15}$ gene carriers. The lower hemi matrix shows the nucleotide diversity (π) per site between hospitals. The diagonal is the nucleotide diversity (π) per site within hospitals. The upper hemi matrix shows the net divergence (Da) estimated between hospitals.

	HAV	HCUV	HGUA	HGUC	HGUE	HGUV	HULR	HUPLF
HAV	**0,005809**	0,000057	0,000174	0,000297	0,000322	0,000195	0,000241	0,000050
HCUV	0,005749	**0,005576**	0,000177	0,000367	0,000339	0,000122	0,000128	0,000021
HGUA	0,005927	0,005814	**0,005698**	0,000492	0,000148	0,000569	0,000589	0,000137
HGUC	0,006620	0,006573	0,006759	**0,006836**	0,000494	0,000460	0,000519	0,000396
HGUE	0,006110	0,006010	0,005880	0,006796	**0,005767**	0,000614	0,000780	0,000341
HGUV	0,005650	0,005460	0,005968	0,006428	0,006048	**0,005100**	0,000080	0,000204
HULR	0,005569	0,005340	0,005862	0,006361	0,006087	0,005054	**0,004848**	0,000209
HUPLF	0,005823	0,005678	0,005855	0,006683	0,006094	0,005624	0,005502	**0,005739**

Supplementary Table 5.6. Core genomes were obtained for each of the sublineages.

Sublineage	LIMS name	STs included	Number of isolates	Core 90%	Alignment length	SNP alignment
SL1	SL2258;SL307;SL322	ST307	1110	4393	4268326	16936
SL2	SL10008;SL322;SL405	ST405	277	4750	4560668	16131
SL3	SL10009;SL14;SL258	ST11 + ST437	2327	2726	2681784	50486
SL5	SL14;SL147;SL889	ST147 + ST392	694	4077	3926877	35933
SL6	SL101;SL3157	ST101	403	3941	3806777	11922
SL7	SL13;SL416	ST13	108	4352	4178830	19807
SL8	SL3010;SL33;SL3321	ST33	16	4490	4315360	3783
SL10	SL17;SL337	ST17	144	4463	4307046	8411
SL11	SL15	ST15	825	4088	3968878	39327
SL13	SL1107;SL14;SL1878	ST902	29	4501	4301273	25801
SL15	SL14;SL29;SL34	ST29	98	4428	4261070	11031
SL22	SL14	ST14	244	4344	4207205	25488
SL49	SL1245;SL14;SL45	ST45	220	4461	4301061	30142
SL63	SL10009;SL14;SL258	ST258 + ST512	1846	3320	3225844	38875
SL65	SL10009;SL14;SL395	ST395	109	4562	4348317	10712
SL67	SL107;SL3157	ST219	66	4572	4388698	16849

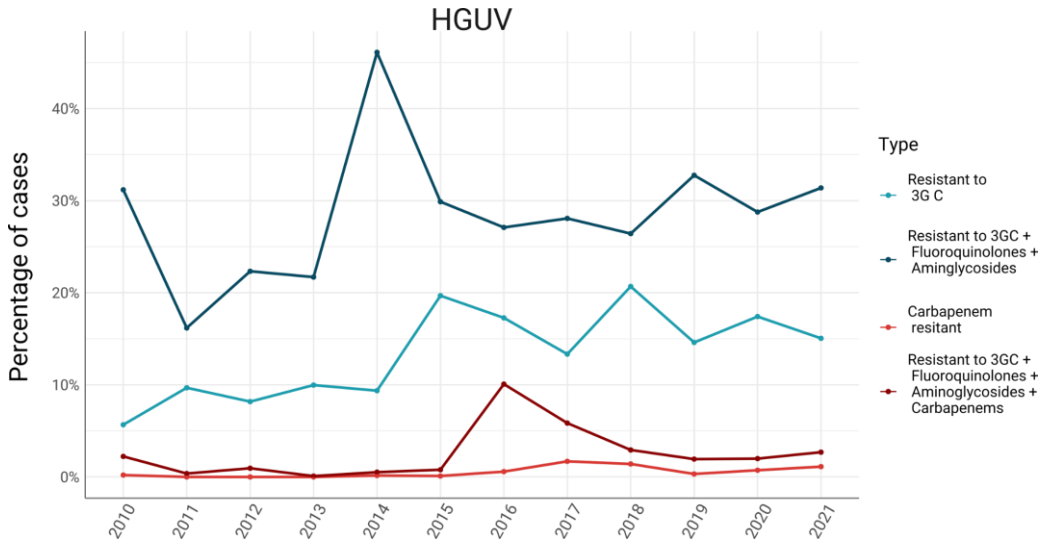
Supplementary Table 5.7. Sublineages clusters with 5 or more SKPCV isolates. Total number, SKPCV, NLSAR and database isolates per cluster. The number of genes, total alignment and SNP alignment values obtained from the core-genome at 90% of each cluster.

Sublineages clusters	Total	SCVKP	NLSAR	DB	Core 90%	Alignment length	SNP alignment length	PTG SNP cut-off
SL1	1110	552	107	451	4393	4268326	16936	21.3
SL10-2	38	17	3	18	4670	4489154	1476	22.4
SL11-4	237	5	2	230	4366	4209917	4278	21.0
SL11-5	84	14	1	69	4319	4167924	2190	20.8
SL11-6	61	5	2	54	4418	4272074	20585	21.4
SL11-8	353	69	14	270	4426	4282434	9152	21.4
SL13-2	25	19	2	4	4756	4539913	180	22.7
SL15-2	57	30	1	26	4584	4399345	4121	22.0
SL187-2	8	8	0	0	4716	4483222	162	22.4
SL19-2	66	5	1	60	4519	4331252	3297	21.7
SL2-2	246	57	14	175	4913	4697802	3916	23.5
SL22-1	32	7	1	24	4566	4415012	5707	22.1
SL22-6	133	11	1	121	4497	4349731	11493	21.7
SL266-1	10	6	0	4	4796	4554997	289	22.8
SL3-1	145	13	14	118	4085	3901660	2706	19.5
SL3-11	489	5	3	481	4181	3995023	6778	20.0
SL3-5	243	104	77	62	4631	4411021	5641	22.1
SL3-7	284	30	94	160	3624	3458277	3891	17.3
SL408-2	10	8	0	2	4558	4406745	1189	22.0
SL49-1	23	6	0	17	4642	4473521	1392	22.4
SL49-6	107	20	1	86	4554	4381864	6330	21.9
SL5-1	128	31	1	96	4671	4448039	3733	22.2
SL5-6	453	84	9	360	4202	4043603	16393	20.2
SL6-5	66	7	33	26	4967	4686554	2024	23.4
SL63-4	697	15	2	680	3694	3505000	10875	17.5
SL65-5	44	11	1	32	4668	4450419	2170	22.3
SL657-2	5	5	0	0	4686	4476612	2615	22.4
SL659-2	5	5	0	0	4566	4338263	66	21.7
SL67-2	60	39	1	20	4742	4534305	1579	22.7
SL7-4	77	11	1	65	4395	4225184	2933	21.1
SL8-2	10	5	4	1	4849	4554535	4803	22.8

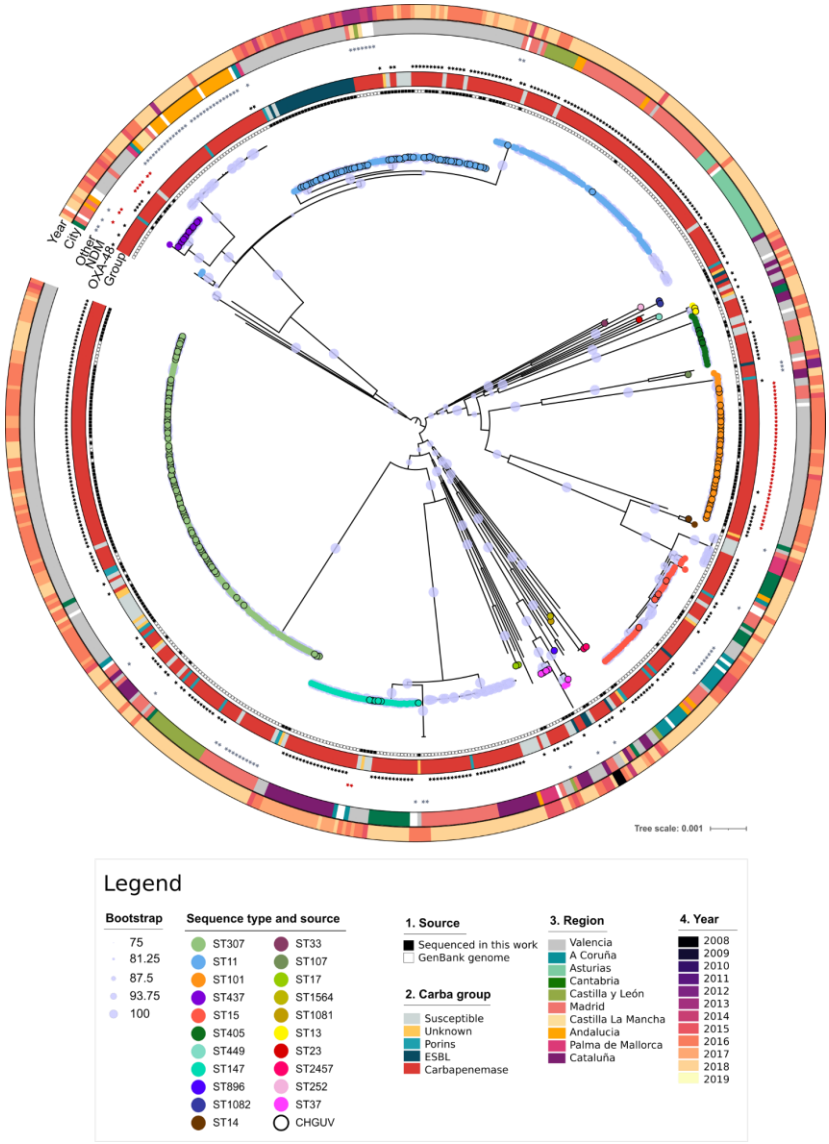
Supplementary Table 5.8. Possible Transmission groups including at least 1 SKPCV isolate. doi: https://doi.org/10.6084/m9.figshare.22779971.v1

CHAPTER 6

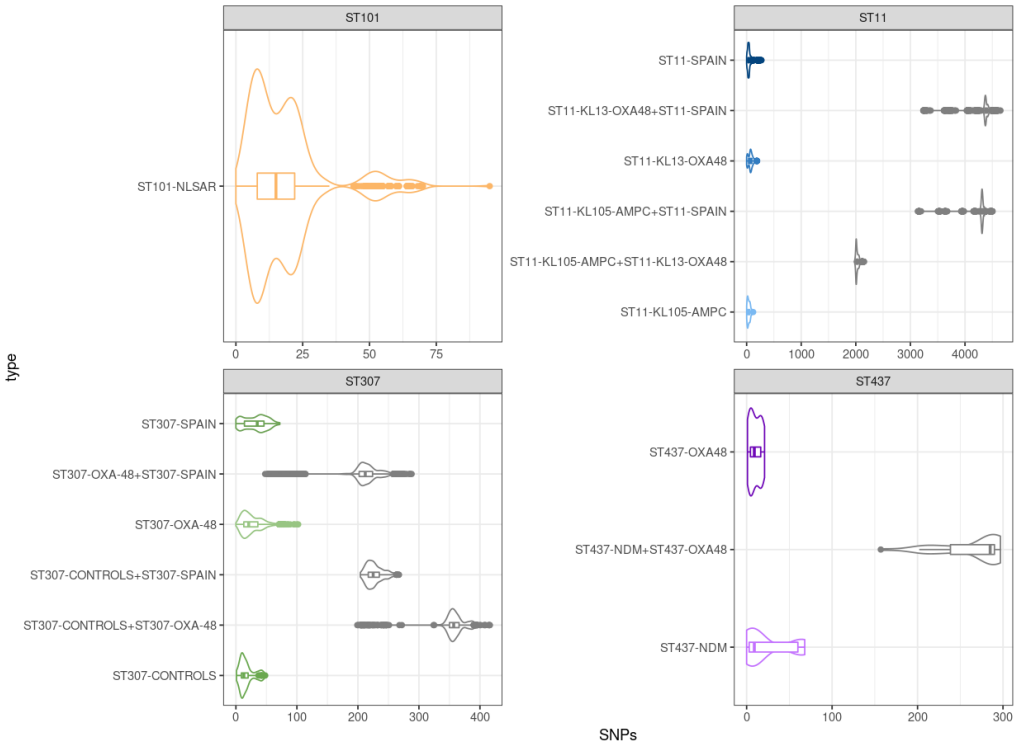
Supplementary figures



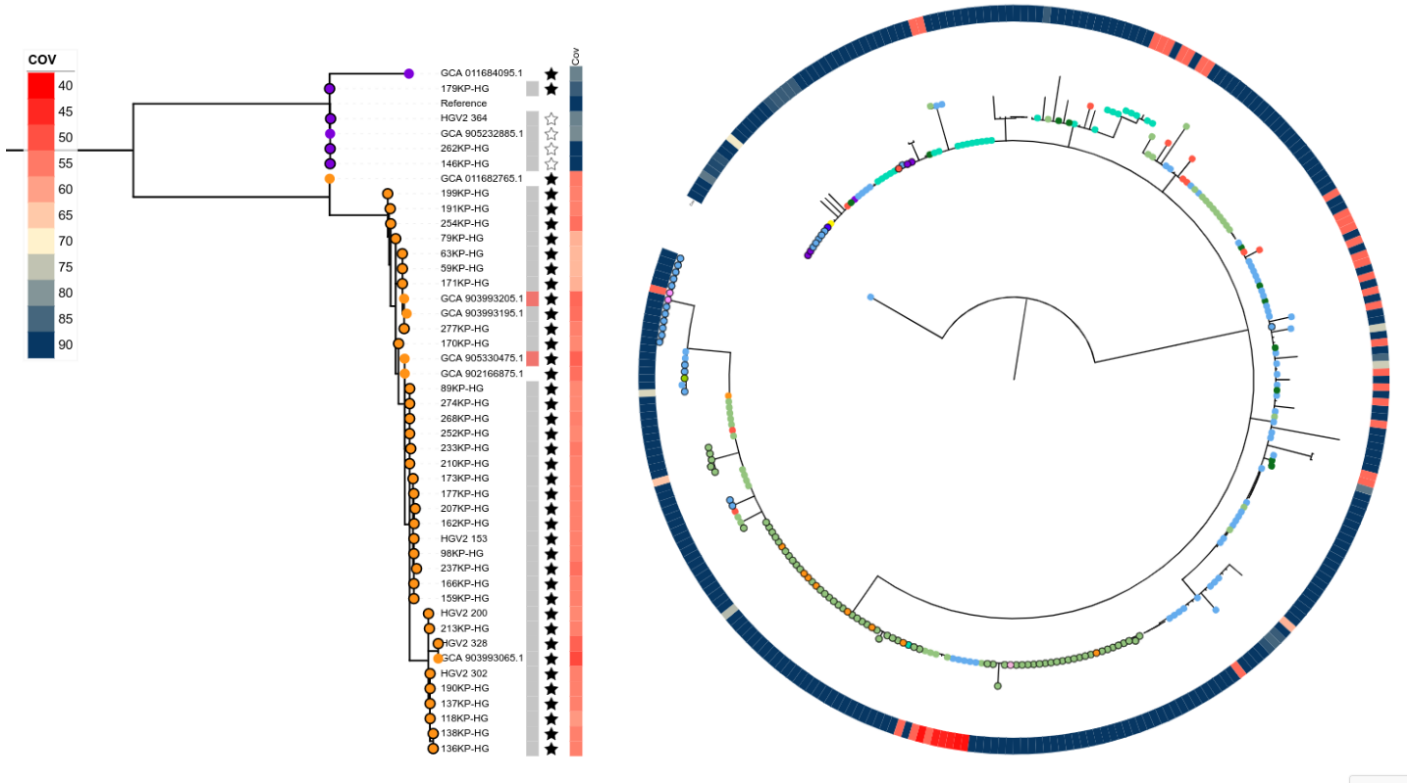
Supplementary Figure 6.1. Relative incidence of cephalosporin- and carbapenem-resistant K. pneumoniae isolates in the HGUV during the period 2010 to 2021. Data was obtained from the Microbiological Network of the Comunitat Valenciana (RedMIVA).



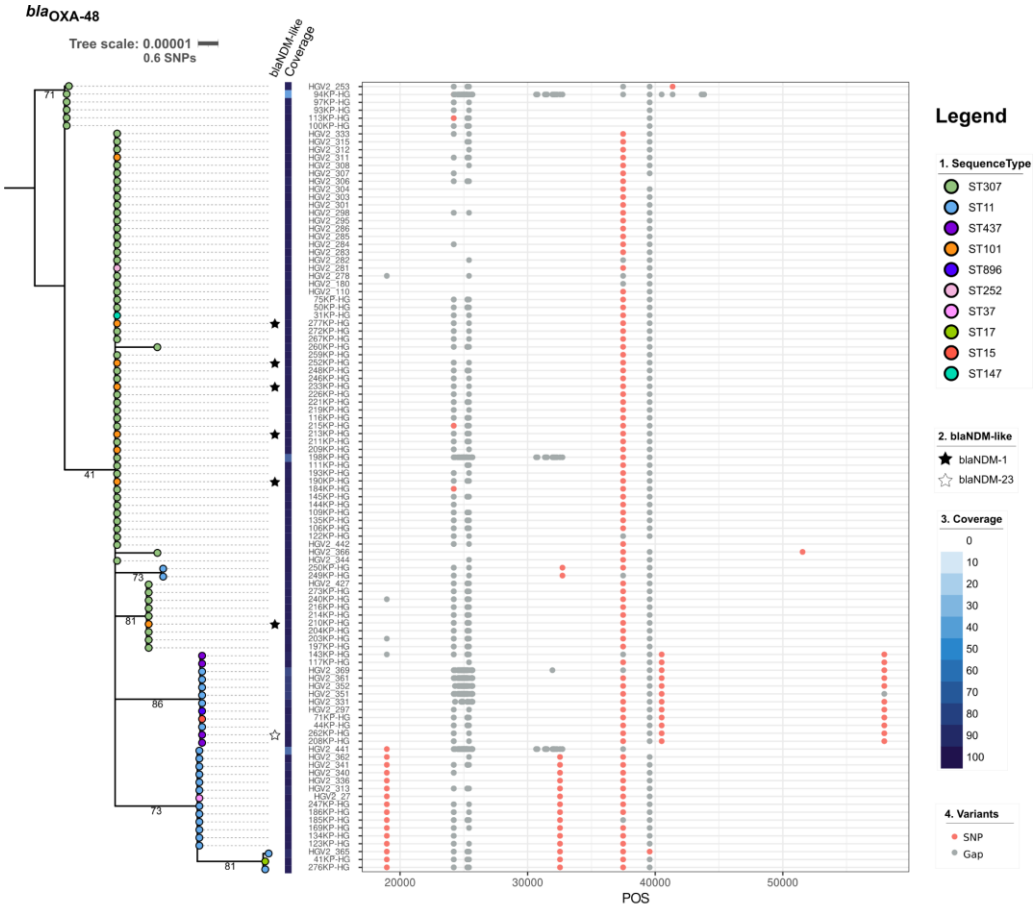
Supplementary Figure 6.2. ML tree of the core genome of the 224 isolates collected in the HGUV and the 360 Spanish isolates downloaded from GenBank. Dots in the tree indicate the ST. Only STs found in the HGUV are shown, and those isolates collected in the HGUV are encircled in black. Blue dots on the branches represent bootstrap support values. Only values above 75% are indicated. The innermost ring indicates the source of the isolates, filled in black if they were collected in this work and in grey if obtained from public databases. The next ring indicates the carbapenemase groups, whereas the third and the outermost rings indicate the city and year of the collection, respectively.



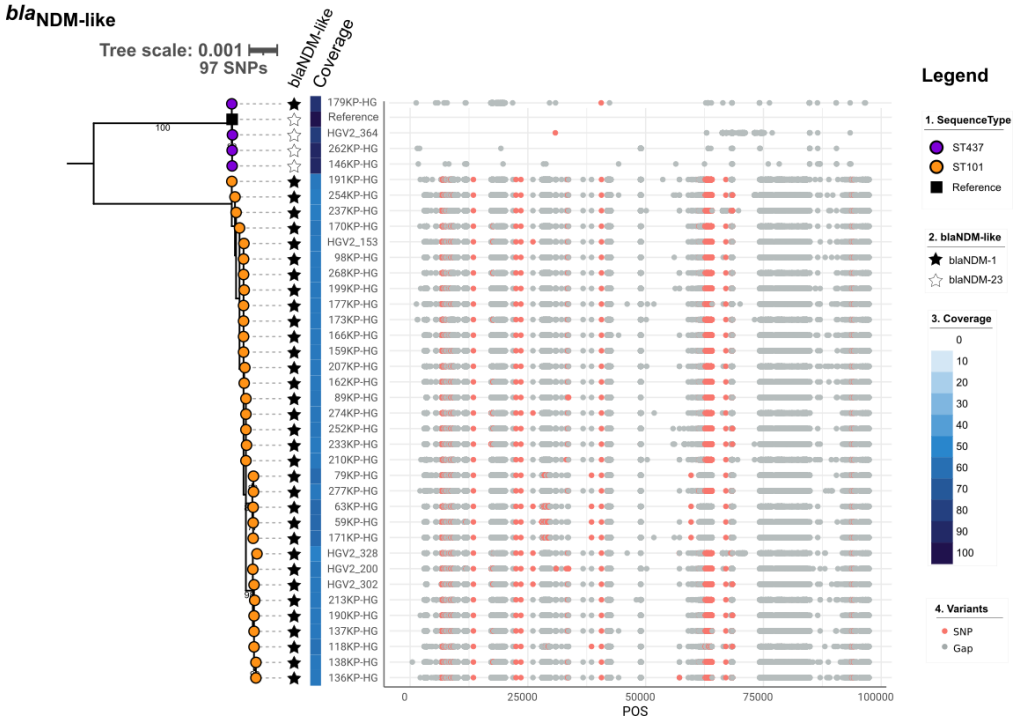
Supplementary Figure 6.3. Distributions of SNP differences within and between lineages of the major STs. The SNP distances were obtained from the core genome alignment of each ST.



Supplementary Figure 6.4. Phylogenetic trees of A) NDM-like and B) OXA-48 plasmids, including the Spanish downloaded genomes. The outer ring represents the coverage value according to the scale at the left side.



Supplementary Figure 6.5. Maximum likelihood tree of pOXA48 plasmids (61,881 bp) with detail of the SNPs (red dots) in each isolate. Grey dots indicate missing positions in the corresponding sample.



Supplementary Figure 6.6. Maximum likelihood tree of NDM plasmids (97,784 bp) with detail of the SNPs (red dots) in each isolate. Grey dots indicate missing positions in the corresponding sample.

Supplementary tables

Supplementary Table 6.1. Clinical information of the isolates collected and sequenced in this study. The table includes patient data such as gender, age, the strain collection date, and the antimicrobial resistance profile for the antibiotics tested. doi: https://doi.org/10.6084/m9.figshare.22639348.v1

Supplementary Table 6.2. Genome information of the sequenced isolates, including assembly statistics, carbapenem resistance group, sequence type, capsular type, antimicrobial resistance, and virulence genes. This table includes the accession number of the raw data at SRA. doi: https://doi.org/10.6084/m9.figshare.22639348.v1

Supplementary Table 6.3. Genome information and metadata for the GenBank Spanish genomes used in this study. It includes information about the isolation source, location, and collection year. doi: https://doi.org/10.6084/m9.figshare.22639348.v1

Supplementary Table 6.4. Number of gene families included in each core genome and alignment length used to derive the ML phylogenetic trees. doi: https://doi.org/10.6084/m9.figshare.22639348.v1

Supplementary Table 6.5. Number of SNPs found in ST11, within and between lineages. doi: https://doi.org/10.6084/m9.figshare.22639348.v1

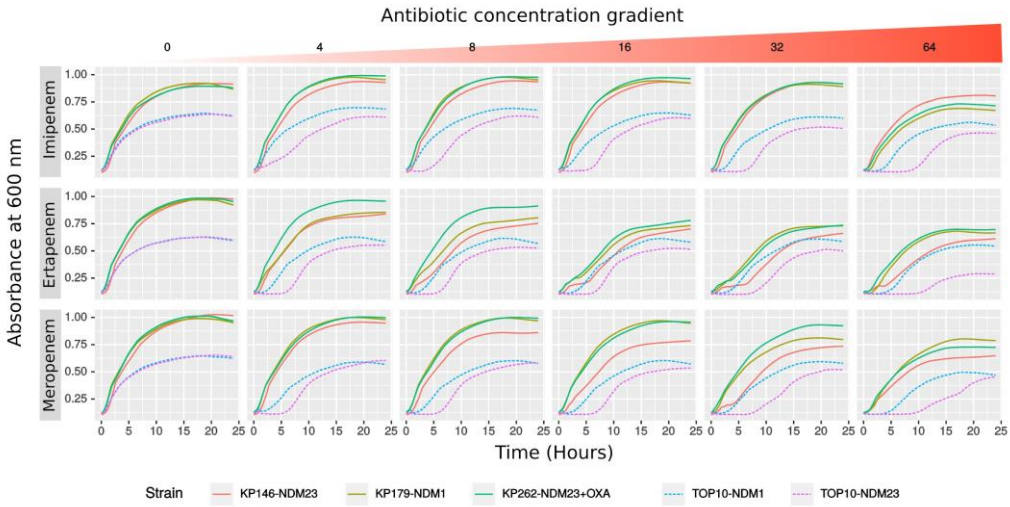
Supplementary Table 6.6. Number of SNPs found in ST101, within and between lineages doi: https://doi.org/10.6084/m9.figshare.22639348.v1

Supplementary Table 6.7. Number of SNPs found in ST437, within and between lineages. doi: https://doi.org/10.6084/m9.figshare.22639348.v1

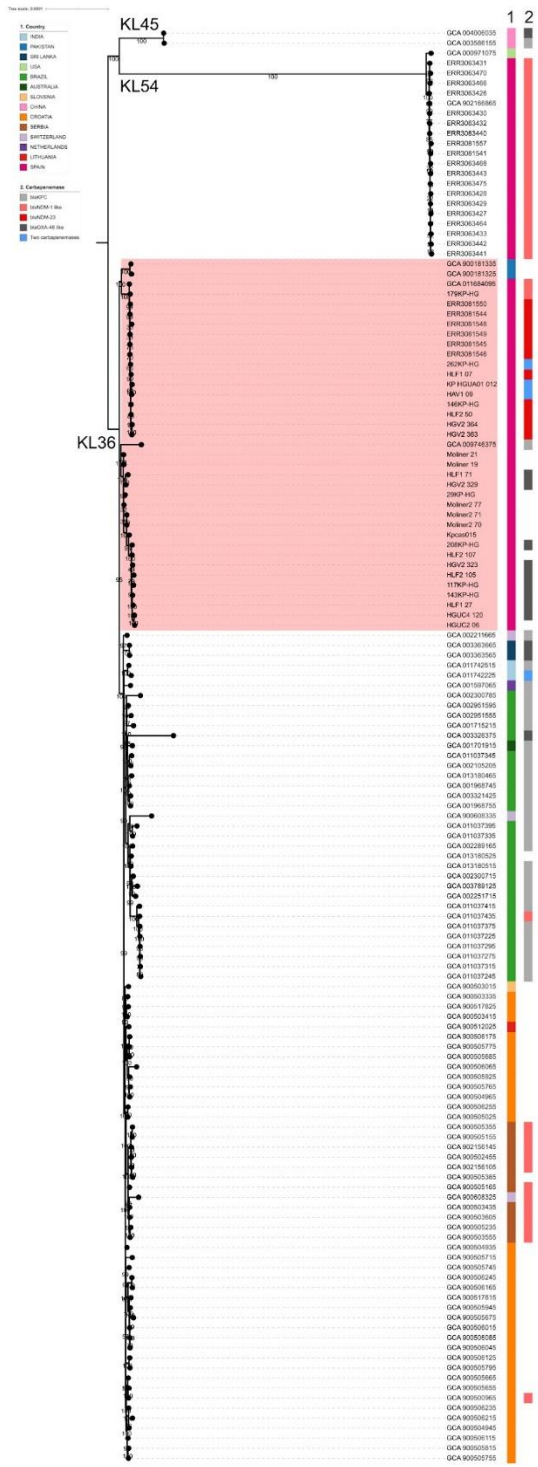
Supplementary Table 6.8. Number of SNPs found in ST307, within and between lineages. doi: https://doi.org/10.6084/m9.figshare.22639348.v1

CHAPTER 7

Supplementary figures



Supplementary Figure 7.1. Comparative growth dynamics under different carbapenem antibiotics concentrations for 3 clinical isolates and 2 transformed E. coli TOP10 strains with NDM-1 and NDM-23.



Supplementary Figure 7.2. Maximum Likelihood phylogenetic tree from complete genome sequences of ST437 isolates included in the analysis.

Supplementary tables

Supplementary Table 7.1. Carbapenemase-production tests used to detect colonisation at patient admission. doi: https://doi.org/10.6084/m9.figshare.22639351.v1

Supplementary Table 7.2. BLAST similarity search results for the repA gene in p146KP-NDM23. doi: https://doi.org/10.6084/m9.figshare.22639351.v1

Supplementary Table 7.3. Summary of BLAST similarity search results for p146KP-NDM23 backbone (region 1 - 12175 + region 73023 - 97321). Only sequences with a query coverage above 90% are reported. doi: https://doi.org/10.6084/m9.figshare.22639351.v1

Supplementary Table 7.4. BLAST similarity search results for p146KP-NDM23 Class 1 integron carrying *bla*$_{\text{NDM-23}}$ (region 34416-44583). Only sequences with a query coverage above 85% are reported. doi: https://doi.org/10.6084/m9.figshare.22639351.v1

Supplementary Table 7.5. Lists of isolates and database genomes included in this study. Accession number, bioproject, biosample, strain name, centre name, country, isolation date, and genetic relevant information as virulence and resistance genes are reported. doi: https://doi.org/10.6084/m9.figshare.22639351.v1

Supplementary Table 7.6. Main features of phages detected by PHASTER in isolate 146KP-HG. Intact phages were removed from the complete genome alignment for phylogenetic analyses. doi: https://doi.org/10.6084/m9.figshare.22639351.v1

Supplementary Table 7.7. Mapping statistics of reads aligned to the chromosome of the 146KP-HG isolate. Reference length, number of aligned positions to the reference, coverage, number of unaligned positions, heterozygous positions, and low depth positions. doi: https://doi.org/10.6084/m9.figshare.22639351.v1

Supplementary Table 7.8. Pairwise distance matrix based on the number of differences obtained from the whole genome alignment (5.3 Mb). doi: https://doi.org/10.6084/m9.figshare.22639351.v1

Supplementary Table 7.9. Coverage of plasmids found in 146KP-HG isolate for each genome and number of CDS found in the remaining accessory genome. doi: https://doi.org/10.6084/m9.figshare.22639351.v1

RESUMEN EN CASTELLANO

Capítulo 1. Introducción

La resistencia a los antimicrobianos como amenaza urgente

En 2019, la Organización Mundial de la Salud (OMS) declaró la resistencia a los antimicrobianos (RAM) como una de las diez principales amenazas para la salud pública mundial, junto con el cambio climático y otras enfermedades infecciosas como el ébola, el sida y el dengue. En este mismo año, un estudio mundial estimó que la RAM causó directamente 1,27 millones de muertes y que 4,95 millones de muertes estuvieron directamente asociadas a ella. Además del coste en vidas humanas, la RAM también ha supuesto un incremento en el coste económico de la asistencia sanitaria. Cada año, sólo en la UE, se gastan entre 1.100 y 1.500 millones de euros adicionales en costes sanitarios debido a la RAM.

Klebsiella pneumoniae como patógeno prioritario de la RAM

Klebsiella pneumoniae es una bacteria Gram-negativa que representa uno de los patógenos bacterianos más preocupantes. Este patógeno causa alrededor de un tercio de todas las infecciones por gram negativos en centros sanitarios, donde se asocia con la RAM y con infecciones nosocomiales, afectando principalmente a pacientes hospitalizados con condiciones médicas subyacentes. Junto con su alta prevalencia, *K. pneumoniae* ha ganado notoriedad como patógeno debido a su potencial para adquirir y acumular RAM rápidamente. En las últimas décadas, las tasas de prevalencia de *K. pneumoniae* no susceptible a los antibióticos más utilizados han aumentado de forma alarmante. En 2017, la OMS reconoció a *K. pneumoniae* como uno de los patógenos prioritarios más críticos que requieren urgentemente el desarrollo de nuevos antimicrobianos. De hecho, en 2019, *K. pneumoniae* se clasificó como la segunda causa principal de muertes atribuidas a la RAM.

Existen dos patotipos de *K. pneumoniae*: el patotipo "clásico" y el "hipervirulento". No hay una definición clara que los diferencie, pero generalmente se pueden distinguir mediante varias características epidemiológicas, clínicas y genéticas. Las cepas "clásicas" de *K. pneumoniae* (cKp) son patógenos oportunistas que causan

principalmente infecciones hospitalarias en pacientes hospitalizados e inmunodeprimidos, como neumonía, infecciones del tracto urinario y de tejidos blandos, que pueden complicarse en una bacteriemia. En las últimas décadas, la adquisición de RAM ha dado lugar a la aparición de linajes multirresistentes. Las infecciones por cepas multirresistentes suelen empeorar debido a la falta de opciones terapéuticas, lo que termina aumentado las tasas de mortalidad.

Por el contrario, las cepas "hipervirulentas" de *K. pneumoniae* (hvKp) suelen actuar como verdaderos patógenos, causando infecciones adquiridas en la comunidad en huéspedes sanos. HvKp causa infecciones altamente invasivas, especialmente abscesos hepáticos piógenos que pueden metastatizar y complicarse en meningitis, endoftalmitis y fascitis necrotizante. Estas infecciones se han notificado en todo el mundo, pero la mayoría de los casos se producen en Asia o en individuos con contacto reciente con la región asiática.

Tratamientos antimicrobianos y RAM

Los antibióticos de primera línea y los más utilizados para tratar infecciones por *K. pneumoniae* son los β-lactámicos. Dependiendo del lugar o la gravedad de la infección, también se utilizan otras clases de antibióticos como los aminoglucósidos, la fosfomicina, la tigeciclina o las fluoroquinolonas. La aparición de cepas RAM ha comprometido la eficacia de estos tratamientos de primera línea, haciendo que la elección de un régimen antibiótico adecuado sea un reto para los médicos. De hecho, es el alarmante aumento en la adquisición de RAM, lo que ha convertido a cKp en un patógeno clínicamente relevante. Por el contrario, hvKp tiende a no presentar RAM, por lo que su relevancia como patógeno RAM sigue siendo baja. Es por ello que este trabajo se centra más en las cepas del patotipo clásico.

Las cepas que suponen un mayor problema y preocupación son aquellas que han adquirido resistencia a las cefalosporinas de tercera generación (C3G) y a los carbapenémicos, ya que son el tratamiento por elección de primera y última línea, respectivamente. En la actualidad, la prevalencia mundial de *K. pneumoniae* resistente a C3G es muy alta y tiende al alza. En Europa, en 2020, el 33,9% de las *K. pneumoniae* notificadas a la EARS-Net eran resistentes a este grupo de antibióticos.

La alta prevalencia de cepas resistentes a C3G ha llevado a un mayor uso de carbapenémicos como terapia, causando una gran presión selectiva hacia organismos resistentes a carbapenémicos. Es por ello que se han notificado aumentos alarmantes de la prevalencia de cepas de *K. pneumoniae* resistentes a los carbapenémicos (CRKp) en todo el mundo. Aunque, actualmente, la prevalencia de las infecciones por CRKp sigue siendo relativamente baja, su impacto es enorme. Esto se debe en gran medida a la elevada mortalidad atribuible y morbilidad de estas infecciones. Se ha estimado que la mortalidad asociada a las infecciones por CRKp adquirida en hospitales oscila entre el 48% y el 65%.

Como consecuencia del aumento en la prevalencia de CRKp, nos enfrentamos a una creciente escasez de tratamientos eficaces. Las infecciones leves por *K. pneumoniae* se han vuelto muy complicadas de tratar y las infecciones más graves, como la neumonía y la bacteriemia, han pasado a ser potencialmente mortales. Esta situación afecta no solo a las personas inmunodeficientes, sino también a las sanas.

Determinantes de la RAM en *K. pneumoniae*

La resistencia a los antimicrobianos en *K. pneumoniae* se debe principalmente a la adquisición de nuevos genes de resistencia. El número de genes de RAM adquiridos en esta especie es mayor (>400) que en cualquier otro patógeno RAM como otras Enterobacterias o *Pseudomonas aeruginosa*. El principal mecanismo de RAM de esta especie es la acción de las β-lactamasas. Estas son enzimas capaces de inactivar los antibióticos β-lactámicos. Se han descrito miles de β-lactamasas, siendo las más relevantes las β-lactamasas de espectro extendido (BLEE), las β-lactamasas tipo AmpC mediadas por plásmidos y las carbapenemasas. Además, estos determinantes RAM suelen combinarse con otros que confieren resistencia a aminoglucósidos, quinolonas, polimixinas, tigeciclina y fosfomicina. Estos son algunos de los tratamientos indicados después de los β-lactámicos. La presencia y acumulación de distintos determinantes de RAM en las cepas produce que estas sean resistentes a la mayoría e incluso a todos los antibióticos disponibles para su eliminación.

En *K. pneumoniae*, estos determinantes RAM están codificados principalmente en elementos genéticos móviles (EGM), como plásmidos y transposones. Los plásmidos desempeñan varias funciones clave en la evolución de la RAM ya que aumentan la movilidad de los genes de RAM y, por tanto, la probabilidad de su acumulación en una sola célula, produciendo fenotipos multirresistentes e incluso pan-resistente.

Linajes de alto riesgo

K. pneumoniae se caracteriza por una amplia diversidad poblacional, compuesta por más de 250 linajes filogenéticos. Aunque cientos de linajes pueden causar infecciones en humanos, sólo una pequeña fracción contribuye a la mayoría de ellas. Estos linajes se han denominado "clones de alto riesgo" o "problemas globales", ya que están distribuidos por todo el mundo y son muy prevalentes en los entornos sanitarios.

Los clones de alto riesgo relacionados con la RAM incluyen el bien estudiado CG258, junto con los CG307, CG147, CG14/15 y CG101, entre otros. En muchos entornos hospitalarios de todo el mundo, como América, Europa y Asia, estos linajes representan la mayoría de las infecciones resistentes a los C3G y a los carbapenémicos. Sin embargo, la prevalencia y distribución en muchos otros países, en su mayoría países en vías de desarrollo, está poco estudiada, ya que sólo se conoce la epidemiología de alguna ciudad grande u hospitales terciarios. En consecuencia, nuevos estudios pueden cambiar nuestra comprensión de estos linajes como "problemas globales".

La población de *K. pneumoniae* hipervirulenta es menos diversa y la mayoría de las infecciones correspondientes se deben a un pequeño número de linajes y serotipos, especialmente los CG23, CG65, ST66, ST86, ST65 y CG38.

Hasta hace poco, los patotipos RAM y hvKp se consideraban poblaciones no superpuestas, ya que pertenecen a ramas distintas y no relacionadas en la filogenia de la especie, y presentan genes de RAM y factores de virulencia distintos. No obstante, en los últimos años, se han notificado cepas que muestran características tanto de multirresistencia como de hipervirulencia. Este tipo de patógenos plantea

una importante preocupación para la salud pública, ya que pueden producir infecciones graves con opciones terapéuticas limitadas.

La genómica y la epidemiología

La secuenciación de genomas completos (WGS, por las siglas en inglés de *Whole Genome Sequencing*) ha desempeñado un papel importante en el desentrañamiento de la taxonomía, la ecología y la evolución de *K. pneumoniae*. Además, se ha convertido en una herramienta importante en las tareas de microbiología clínica.

Con la disminución del coste de la WGS y del tiempo requerido para obtener resultados, la WGS se está implantando entre las tareas rutinarias de los laboratorios de microbiología clínica apoyando, e incluso sustituyendo, algunas técnicas tradicionales. Entre sus aplicaciones más comunes se incluyen la caracterización de aislados, la predicción del fenotipo de resistencia y el rastreo de la propagación de agentes infecciosos entre pacientes y las fuentes de infección. Toda esta información puede combinarse para llevar a cabo la vigilancia de patógenos, una de las aplicaciones clave de la WGS.

Las estrategias de vigilancia son necesarias para controlar la aparición y diseminación local de clones problemáticos y con RAM y, por tanto, controlar su propagación en entornos hospitalarios. Además, la vigilancia puede ayudar a optimizar la administración de antimicrobianos e informar las decisiones sobre tratamientos.

Implementación de WGS en hospitales españoles y el proyecto NLSAR

Actualmente, el uso de la WGS no está implementado en la rutina de microbiología clínica de los hospitales españoles. No obstante, el Plan Nacional contra la Resistencia a los Antibióticos (PRAN) incluye medidas para adoptar e integrar en un futuro la WGS en la práctica clínica rutinaria. El PRAN tiene como objetivo reducir el riesgo de selección y propagación de la RAM y, por consiguiente, su impacto clínico y en la salud pública. Como parte de este plan, se creó la Red Española de Laboratorios para la Vigilancia de Microorganismos Resistentes (RedLabRA), una red

nacional de laboratorios de microbiología coordinados que tiene como objetivo llevar a cabo estudios moleculares de enfermedades infecciosas. Una de las acciones futuras de RedLabRA es promover la integración de la WGS en la práctica clínica rutinaria entre 2022 y 2024.

En 2018, el grupo de Epidemiología Molecular de FISABIO y la Universitat de València encabezó una iniciativa para establecer un marco para llevar a cabo la vigilancia de la resistencia a los antimicrobianos en la región, conocida como la Red de Laboratorios para la Vigilancia de la Resistencia a los Antimicrobianos (NLSAR, por sus siglas en inglés).

La NLSAR está coordinada por un nodo central en el grupo de Epidemiología Molecular de FISABIO. Su objetivo principal es servir como primer paso hacia la integración de la WGS en la vigilancia de patógenos con resistencias a los antimicrobianos en la Comunitat Valenciana, pero también generar información que pueda ayudar a comprender las rutas de transmisión, los orígenes y la evolución de los patógenos en los entornos clínicos regionales.

Capítulo 2. Objetivos

El objetivo principal de esta tesis es investigar la epidemiología genómica de la población de *K. pneumoniae e*n la Comunitat Valenciana mediante la aplicación de la WGS y el análisis genómico, con especial énfasis en los principales linajes y determinantes de RAM.

Los objetivos específicos desarrollados para cumplir con el objetivo principal son:

- Establecer las bases para el análisis de la epidemiología genómica de *Klebsiella pneumoniae* en la Comunitat Valenciana.
- Estudiar la epidemiología genómica de *Klebsiella pneumoniae* RAM en la Comunitat Valenciana.
- Comprender cómo ocurre la emergencia inicial y la colonización posterior de determinantes resistentes a carbapenémicos en hospitales.

- Utilizar las ventajas de la WGS para detectar la aparición de nuevas amenazas en entornos hospitalarios.

Capítulo 3. Material y métodos

La NLSAR y el proyecto SKPCV.

Para estudiar la población de *K. pneumoniae* productoras de BLEEs y carbapenemasas en la CV, se creó el proyecto Vigilancia de *K. pneumoniae* en la Comunitat Valenciana (SKPCV, de las siglas en inglés).

El SKPCV se llevó a cabo entre 2017 y 2019 a escala regional en 8 de los hospitales de la NLSAR recientemente desarrollado. Los hospitales incluidos son: Consorcio Hospital General Universitario de Valencia (HGUV), Hospital Clínico Universitario de Valencia (HCUV), Hospital Universitario y Politécnico La Fe de Valencia (HUPLF), Hospital Arnau de Vilanova de Valencia (HAV), Hospital de la Ribera de Alzira (HULR), Hospital General Universitario de Castellón (HGUC), Hospital General Universitario de Alicante (HGUA), y Hospital General Universitario de Elx (HGUE). Todos estos hospitales son de nivel terciario, excepto el HAV y el HULR, que son de nivel secundario. Los laboratorios de microbiología clínica de estos hospitales sirven como laboratorios de referencia para los hospitales de primer y segundo nivel de la CV. En total, atienden al 45,5% de la población de la CV.

El único criterio establecido para el proyecto fue incluir, para cada hospital y trimestre, los 30 primeros aislados de *K. pneumoniae* productores de BLEEs y/o carbapenemasas, junto con los 10 primeros aislados clínicos susceptibles a C3G, incluyendo un único aislado por paciente. Se tuvo en cuenta tanto las infecciones como las colonizaciones. Se recogieron los datos clínicos pertinentes, como la fecha de toma de muestra, el servicio del hospital, el sexo y la edad del paciente y las pruebas de sensibilidad a los antimicrobianos.

Para algunos análisis incluimos aislados de *K. pneumoniae* recogidos en los hospitales de la NLSAR recogidos para otros proyectos. Estos incluyeron cepas productoras de BLEEs recogidas durante 2014 en el HGUC, una colección no susceptible a carbapenémicos del HGUV de 2015 a 2017, y muestras históricas del

Hospital Dr. Moliner (no en el NLSAR). Estas cepas se utilizaron para proporcionar un contexto histórico para las cepas del SKPCV en los capítulos 4 y 5, mientras que todo el conjunto de datos del HGUV se analiza con más detalle en el capítulo 6.

Aislamiento, identificación de especie, antibiogramas y WGS

Todas las cepas recogidas por los laboratorios de la NLSAR se sometieron a los mismos procedimientos de aislamiento, antibiograma y secuenciación, independientemente de su inclusión en el proyecto SKPCV o no. La recogida de aislados, la identificación de especies y los antibiogramas se realizaron en los laboratorios de Microbiología Clínica de cada hospital como parte de sus tareas rutinarias. Las cepas se cultivaron durante toda la noche y se sometieron a extracción de ADN con extractores automatizados. Todas las muestras se secuenciaron utilizando la tecnología de lecturas cortas de Illumina. La secuenciación se realizó en una plataforma Illumina NextSeq 500 y con el kit de preparación de librerías Nextera XT. Además, algunos aislados también fueron secuenciados mediante tecnología de secuenciación de lecturas largas (PacBio y MinION) por el Servicio de Secuenciación y Bioinformática de FISABIO.

Genomas de la base de datos

Se incluyeron genomas de *K. pneumoniae* obtenidos de la base de datos de secuencias GenBank, la colección de secuencias de referencia (RefSeq) y el archivo de lecturas de secuencias (SRA). En el caso de RefSeq, se incluyeron todos los ensamblados disponibles el 13/6/2020. La obtención de genomas del GenBank y del SRA se incluyeron las lecturas del SRA y los ensamblados del GenBank etiquetados con "*Klebsiella pneumoniae*" AND "Spain" (acceso el 4 de abril de 2022).

Reconstrucción de los genomas

La calidad de la secuenciación de cada muestra se evaluó con **FastQC** v0.11.9 y **MultiQC** v1.11 y se usó **Prinseq-lite** v0.20.4 para filtrar y recortar las lecturas. Las muestras secuenciadas por lecturas cortas y las de Nanopore se ensamblaron *de novo* utilizando **Unicycler** v0.4.8. En el caso de Nanopore, se hicieron ensamblados híbridos con las lecturas cortas de cada muestra. Las muestras secuenciadas por

PacBIO se ensamblaron con **HGAP** o **Unicycler**, dependiendo de la profundidad obtenida para cada muestra. Todos los ensamblados de PacBio resultantes se pulieron con **Pilon** y con las lecturas cortas con hasta 10 rondas. Finalmente, los genes de todos los ensamblados se anotaron utilizando Prokka v1.14.6.

Secuencio Tipo, resistencia a antimicrobianos, determinantes de virulencia y plásmidos

Para determinar el Secuenciotipo (ST) e identificar los genes de virulencia y de resistencia a los antimicrobianos se utilizó Kleborate. Para identificar y verificar los genes de resistencia antimicrobiana, se utilizaron **AMRFinderPlus**, **ResFinder**, **staramr** y **BLAST**. Los mecanismos que contribuyen a la resistencia al cefiderocol se caracterizaron comparando los aislados con el genoma de referencia de *K. pneumoniae* ATCC 13883 e identificando mutaciones genéticas relacionadas previamente con resistencia al cefiderocol. Para el tipado de replicones plasmídicos, utilizamos **PlasmidFinder**, mientras que para los fagos utilizamos **PHASTER**.

Filtrado de ensamblados

Para evaluar los genomas obtenidos en el NLSAR y los genomas descargados, desarrollamos y aplicamos un método de filtrado en cinco pasos. Se evaluó la clasificación taxonómica de las lecturas con **Kraken2** y la asignación taxonómica general del genoma completo con **Kleborate**. La calidad de los genomas reconstruidos se evaluó con **QUAST**. Para eliminar posibles muestras con mezcla de diferentes cepas de *Klebsiella*, se estimó el número de genes afectados por lecturas heterocigotas. Para ello, las lecturas cortas de cada aislado se mapearon contra el genoma de referencia, NTUH K2204 (AP006725) utilizando **BWA** con la opción mem. Para determinar la similitud entre todos los aislados de la colección, y así eliminar posibles genomas muy distantes, estimamos distancias mash con **Mash** y el módulo dist.

Diversidad y curvas de rarefacción

La diversidad de cada población se estimó utilizando STs y el índice de diversidad de Simpson implementado en el paquete de R **vegan**. Además, se realizaron curvas

de rarefacción utilizando el paquete **vegan** y sus funciones rarefy y rarecurve, que se representaron utilizando el paquete de R **ggplot2**.

Análisis genómico

Comparación entre aislados

En los capítulos 4, 5 y 6, para comparar genéticamente los aislados, se estimaron los genes del "core", es decir, aquellos que están presentes en la mayoría de los aislados. En cada capítulo se utilizó una herramienta de análisis diferente. Para los capítulos 4 y 5 utilizamos **PanACota**, mientras que en el capítulo 6 utilizamos **panaroo**. Tras obtener las familias de genes presentes en el genoma "core" y para obtener el alineamiento de cada gen se realizó externamente utilizando **mafft** v7.490 con la opción --adjust-direction y el software **AMAS.py**.

Durante este trabajo, nos dimos cuenta de que el software de anotación de genes introducía una gran acumulación de SNPs falsos. Para resolver este problema, desarrollamos e implementamos un algoritmo que produce alineamientos de genes libres de errores. El algoritmo se basa en encontrar ramas largas en el árbol de cada gen y comprobar si se deben a la acumulación de SNPs seguidos únicos en el alineamiento.

En el caso de **panaroo**, estos los errores no fueron tan significativos ya que la cantidad de muestras a analizar era mucho menor. Por lo tanto, los errores se eliminaron con **ClipKit** con el modo kpic-gappy. Para ambos programas, la concatenación de los alineamientos resultantes se realizó con **AMAS.py concat**. Las posiciones variables se extrajeron utilizando **snp-sites**.

En algunos casos, la comparación de genomas se hizo mediante mapeo de las lecturas contra una referencia. Para ello se utilizó **Snippy**.

Los árboles filogenéticos de máxima verosimilitud se realizaron con **IQTREE** y se representaron con Interactive Tree OnLine y en R con **ggtree**.

Definición de linajes

Para definir linajes dentro de *K. pneumoniae*, se utilizó el esquema cgMLST de 694 genes propuesto por Bialek-Davenet *et al*. Para inferir el perfil de genes, utilizamos el programa **Chewbbaca** y la secuencia de los genes que se descargaron desde el sitio web PasteurMLST. Las distancias entre los perfiles obtenidos para cada muestra se obtuvieron mediante **cgml-dists** y se utilizaron definir sublinajes siguiendo el umbral propuesto por Hennart *et al.* Los grupos se obtuvieron con las funciones hclust y cut tree implementadas en el paquete de R **stats**. A su vez, dentro de los sublinajes, se definieron clusters con el paquete de R **fastBAPS**.

Definición e identificación de transmisiones

Definimos como posibles grupos de transmisión (PTG) aquellos aislados con distancias de pares de SNPs iguales o inferiores a 5 SNPs por Mbp comparados. Para determinar los PTGs, utilizamos las distancias de SNPs obtenidas por **snp-dist** seguido de un agrupamiento mediante las funciones hclust y cuttree de R **stats** v4.1.2. Para construir las redes de transmisión, utilizamos el paquete **network** v1.18.1 de R y para representarlas utilizamos las librerías **ggplot2** v3.4.2 y **ggnetwork** v0.5.10.

Análisis de plásmidos

Similitud entre plásmidos

Para la comparación de plásmidos, creamos una base de datos de plásmidos cerrados a partir de los con menos de 20 contigs. Para identificar las secuencias plasmídicas y poder extraerlas, utilizamos **mlplasmids**. Para enriquecer la base de datos con posibles plásmidos cerrados de la colección NLSAR, interrogamos todos los ensamblados de lecturas cortas que no se habían tenido en cuenta para añadir a todos los contigs que se predijo que eran plásmidos y estaban circularizados. Las relaciones entre los plásmidos de la base de datos se calcularon con la puntuación ponderada de relación de repertorios genéticos (**wGRR**), como se describe en Pfeifer *et al*. (2021).

Identificación y clasificación de plásmidos

Los plásmidos de la base de datos se caracterizaron mediante distintas herramientas. Se utilizó **PlasmidFinder**, **Mob-typer** y **AMRfinderplus**. Finalmente, los plásmidos se anotaron utilizando **Prokka**.

Para identificar plásmidos similares publicados anteriormente, utilizamos **BLAST** con la colección de nucleótidos estándar del NCBI y PLSDB. Los plásmidos se representaron utilizando el servidor **CGview** y la biblioteca **genoplotR**.

Experimentos de transformación y conjugación con *bla*$_{\text{NDM-23}}$

Para evaluar la susceptibilidad a antimicrobianos de los aislados portadores de genes *bla*$_{\text{NDM}}$ relevantes, realizamos ensayos de clonación y transformación. Las secuencias completas de los genes *bla*$_{\text{NDM}}$ se amplificaron y los fragmentos se clonaron utilizando el kit de clonación TOPO TA y se transformaron en una cepa *E. coli* TOP10. Las pruebas de susceptibilidad a antibióticos se hicieron utilizando el sistema automatizado MicroScan WalkAway. Para los carbapenémicos, se comprobó la concentración mínima inhibitoria (CMI) mediante el método de microdilución en caldo, mientras que el cefiderocol se comprobó mediante un kit comercial de tiras reactivas.

Para estudiar la movilidad del plásmido portador del gen *bla*$_{\text{NDM}}$, realizamos ensayos de conjugación. Para ello utilizamos la *E. coli* J53 resistente a la azida como receptora. La movilización de los genes de la carbapenemasa en *E. coli* J53 se confirmó mediante CHROMagar y PCR. Se realizaron réplicas técnicas y biológicas para cada aislado.

Resultados y discusión

En esta tesis, hemos investigado la epidemiología de aislados de *K. pneumoniae* resistentes a C3G y carbapenémicos en varios hospitales de la Comunitat Valenciana, utilizando datos genómicos. Este estudio ha realizado contribuciones significativas en: i) la creación de una colección de genomas para la vigilancia genómica de *K. pneumoniae* en la CV; ii) desentrañar la estructura poblacional y la

diversidad genética de *K. pneumoniae* en entornos clínicos de la CV, así como la evolución del linaje de alto riesgo ST307; iii) una comprensión más profunda de la compleja colonización de los entornos hospitalarios por cepas de *K. pneumoniae* resistente a carbapenémicos; iv) dilucidar la emergencia de gen $bla_{\text{NDM-23}}$, y demostrar cómo los casos importados de cepas productoras de carbapenemasas pueden diseminarse rápidamente a nivel regional a través de la expansión clonal.

Capítulo 4. De la genómica regional de *K. pneumoniae* a la global

Los proyectos de vigilancia genómica como el desarrollado en esta tesis, el SKPCV, requieren la integración de genomas "contextuales" a escala local y global para entender su situación epidemiológica, ya que a menudo la diseminación de linajes cruza fronteras geográficas. Además, los genomas de las bases de datos pueden ayudar a descifrar el origen temporal y geográfico de linajes y genes de RAM emergentes.

Bajo el programa SKPCV recogimos 1.774 aislados de *K. pneumoniae.* Además, incluimos 414 aislados retrospectivos recogidos de tres hospitales de la NLSAR, lo que resultó en un total de 2.188 aislados. Para proporcionar contexto incluimos datos externos. En el momento en que interrogamos las bases de datos, pudimos descargar un total de 11.553 genomas de *K. pneumoniae.*

Los grandes conjuntos de datos genómicos son un recurso valioso, pero también pueden convertirse en un obstáculo si no se manejan adecuadamente, como se muestra en esta tesis. En el capítulo 4 describimos la colección final de genomas tanto del NLSAR/SKPCV como de las bases de datos. Para asegurar la calidad y fiabilidad de los resultados, diseñamos un filtro con diferentes pasos que nos permitió eliminar genomas que correspondían a otras especies, estaban contaminados, eran de mala calidad o tenían un efecto negativo en los análisis posteriores. Además, en este capítulo ofrecemos una visión general de las dos colecciones resultantes (NLSAR y bases de datos) que se explotarán en los capítulos siguientes

Finalmente, de los 13.741 genomas considerados inicialmente, 13.133 (95,6%) genomas pasaron el filtro de control de calidad: 1.999 (91,4%) ensamblados de NLSAR y 11.134 (96,3%) ensamblados procedentes de bases de datos.

La colección NLSAR tuvo un número muy alto de aislados contaminantes o mal identificados. Sorprendentemente, 99 genomas de las bases de datos también estaban mal identificados o contaminados.

Por ello, al diseñar protocolos rutinarios de WGS en los laboratorios de microbiología clínica, ha de considerarse tanto la calidad como la contaminación. Además, los flujos de trabajo rutinarios de epidemiología genómica deben implementar estos filtros de control o similares. Somos conscientes de que la complejidad y el número de pasos escalonados en nuestro flujo de trabajo pueden limitar su implementación en flujos de trabajo de vigilancia genómica más generales y rápidos. Por lo tanto, se deben establecer esfuerzos para estandarizar pasos clave y garantizar resultados precisos y reproducibles en todos los laboratorios.

La colección de la NLSAR en un contexto global

En el capítulo 4, también caracterizamos la colección de genomas y evaluamos la similitud de los genomas de la base de datos con los recolectados en los hospitales de la NLSAR.

La colección final de la NLSAR consta de 1.604 aislados secuenciados recogidos en el marco del proyecto SKPCV en 8 hospitales diferentes de la CV y 395 aislados de otros proyectos retrospectivos. Esta colección de genomas de *K. pneumoniae* es uno de los conjuntos de datos más grandes disponibles para comprender la epidemiología genómica de este patógeno en una región específica. Además de los de la NLSAR, incluimos 11.134 genomas de bases de datos en la colección global. Esta colección incorpora aislados de al menos 94 países tomados al menos a lo largo de 109 años.

Es importante tener en cuenta que tener más genomas no equivale necesariamente a tener más información o diversidad. A pesar del gran número de genomas recogidos en las bases de datos, todavía hay muchas regiones y países que siguen

estando subrepresentados o no representados en absoluto. Vale la pena señalar que algunos de los países que siguen estando subrepresentados ya han sido identificados como reservorios de cepas resistentes a carbapenémicos. La distribución en y dentro de los continentes dista mucho de ser uniforme. La mayoría de los aislados se recogieron en Europa, especialmente en Italia y el Reino Unido. España estaba poco representada en RefSeq, con sólo 274 ensamblados. Al incluir los genomas españoles disponibles en las bases de datos GenBank y SRA, mejoramos la representación española con 1.153 ensamblados en total. Los genomas asiáticos procedían principalmente de China, mientras que los de América se recogieron sobre todo en EE.UU.

Además, es importante considerar que la baja diversidad observada en regiones como América y Europa puede no ser completamente representativa de la situación real, ya que sus colecciones están altamente sesgadas hacia determinantes específicos de la resistencia a los antimicrobianos y clones de alto riesgo.

En cuanto a los linajes, identificamos 168 ST diferentes en la colección NLSAR. En general, los principales linajes encontrados en la NLSAR también estaban presentes en las colecciones globales y son reconocidos como linajes RAM de alto riesgo. La colección NLSAR mostró una composición de linajes similar a la de la española, ya que ambas estaban compuestas mayoritariamente por los ST307 y ST11. A pesar de ser el segundo y quinto linajes más frecuentes en toda la colección, los ST258 y ST512, eran muy poco frecuentes en el resto de las regiones, incluyendo España y el NLSAR.

Los determinantes de la RAM y plásmidos encontrados en la colección NLSAR ya se habían descrito en otros lugares. Sin embargo, encontramos algunas excepciones: la enzima NDM-23 y el plásmido que la portaba sólo se encontraron en la colección NLSAR o en genomas españoles, y el linaje ST11 co-portador de las enzimas OXA-48 y CTX-M-1, del que apenas se ha informado fuera de la NLSAR o España. Además, no encontramos plásmidos similares en la base de datos de plásmidos personalizada a los que portan VIM y un KPC.

Esta colección de genomas cuidadosamente curada no sólo mejorará nuestra comprensión de la evolución y transmisión de la resistencia a los antibióticos en

K. pneumoniae en la Comunitat Valenciana, sino que también será un recurso valioso para investigar posibles vínculos y eventos de transmisión dentro y entre otras regiones en todo el mundo.

Capítulo 5. Análisis genómico de *K. pneumoniae* RAM in la CV

España ha sido identificada como una de las regiones europeas con tasas más altas de RAM en *K. pneumoniae*, especialmente a C3G y carbapenémicos. Para obtener información sobre la dinámica y evolución de la población RAM de *K. pneumoniae* en la Comunitat Valenciana, se estableció el proyecto SKPCV y se recolectaron 1.604 aislados de *K. pneumoniae* de ocho hospitales diferentes. Además, la información obtenida de más de 11.000 genomas en bases de datos públicas y 400 genomas recolectados anteriormente en los mismos hospitales (NLSAR) ayudó a dilucidar posibles orígenes y rastrear brotes globales.

Primero estudiamos la situación epidemiológica actual de la CV. En general los datos obtenidos de la prevalencia de cepas resistentes a C3G y carbapenémicos fue muy similar a la reportada en España y en Europa. La mayoría de estos hospitales tenían una alta prevalencia de episodios de resistencia a las C3G, mientras que mantenían tasas relativamente bajas de resistencia a los carbapenémicos.

Para investigar los determinantes genéticos de la resistencia a las 3GC y carbapenémicos, así como sus modos de adquisición y diseminación, analizamos los datos genómicos de la colección SKPCV. Nuestro estudio reveló una gran diversidad poblacional en general y dentro de cada hospital. Esta diversidad es similar a la encontrada en otros programas de vigilancia a nivel hospitalario o nacional. Además, los linajes y determinantes de la RAM más frecuentes en la CV eran también los más abundantes en España y Europa.

Los análisis realizados de diversidad genética dentro y entre hospitales y la detección de posibles grupos de transmisión sugieren que la diseminación clonal dentro de los hospitales es el mecanismo principal que impulsa la propagación de la RAM en la región, en línea con los hallazgos de otros países. Sin embargo, nuestros resultados también sugieren que la transmisión entre hospitales es significativa e implica rutas de transmisión más complejas, con transmisiones

intermedias no detectadas, lo que destaca la importancia de las infecciones adquiridas en la comunidad.

Un aspecto innovador de nuestro estudio es el uso del análisis de alineamiento del genoma "core", en lugar del mapeo del genoma, para detectar posibles grupos de transmisión. Los alineamientos del genoma "core" generalmente no se utilizan para este propósito. En nuestro estudio, hemos desarrollado un algoritmo que elimina estos errores, lo que nos permite realizar análisis de transmisión utilizando el mismo enfoque para 31 STs diferentes, incluidos linajes raros. Esta metodología proporciona una ventaja significativa para la vigilancia genómica de patógenos, ya que elimina el sesgo de seleccionar una referencia estrechamente relacionada para cada uno de los STs y permite un enfoque estandarizado para cada patógeno y sublinaje.

El linaje ST307 portador de *bla*$_{\text{CTX-M-15}}$

El ST307 es el linaje más prevalente de la colección SKPCV, con 559 aislados (34.8%). Además, como se muestra en el análisis de diversidad y transmisión, este ST fue responsable de la mayoría de las transmisiones dentro y entre hospitales. De hecho, casi la mayoría de todos los PTGs se atribuyeron al grupo ST307, incluyendo aquellos involucrados en transmisión interhospitalaria.

Para realizar un análisis más detallado, investigamos las relaciones entre los aislados ST307 recogidos en la base de datos global. La filogenia global del linaje indicó que los aislados de SKPCV no se agrupaban en un solo clado, sino que estaban dispersos en diferentes clados. Esto sugiere que la distribución actual de ST307 en la CV se debe a múltiples introducciones del linaje en la región, seguido de una posterior diseminación y diversificación. Sin embargo, para comprender completamente la evolución de este clado en la CV, se necesitan análisis adicionales.

Capítulo 6. La complejidad detrás de la colonización de CRKp

Para reducir la RAM es necesario entender los factores que conducen a la emergencia y diseminación de cepas resistentes en ambientes hospitalarios. En este capítulo, evaluamos la colonización inicial y la posterior evolución de *K. pneumoniae*

resistente a carbapenémicos en un hospital de Valencia, el hospital terciario HGUV. Para ello, analizamos los primeros aislados no susceptibles a carbapenémicos detectados en este hospital (enero de 2015) y extendimos el análisis hasta 4 años después (diciembre de 2018).

Sorprendentemente, los casos iniciales de CRKp en el HGUV se debieron principalmente a cepas de *K. pneumoniae* que no producían carbapenemasas, a pesar de que estas últimas son el mecanismo más comúnmente encontrado. Aunque la mayoría de los estudios suelen centrarse en estudiar cepas productoras de carbapenemasas, en este trabajo hemos demostrado que existen otros mecanismos asociados con el fenotipo de resistencia y que también necesitan ser controlados. No obstante, este linaje fue reemplazado rápidamente por aislados productores de carbapenemasas. Se encontraron dos tipos diferentes de carbapenemasas en este hospital, OXA-48 y NDM, diseminadas por 6 linajes diferentes.

Este trabajo demuestra la complejidad involucrada en la colonización inicial de cepas resistentes a carbapenémicos en un hospital, y enfatiza la diversidad de linajes, detergentes de la RAM y plásmidos involucrados.

En muchos hospitales se ha reportado una diversidad poblacional de CRKp similar a la de este trabajo, pero una vez ya establecidas las poblaciones, sin conocer cómo apareció este patógeno en el hospital. Nuestros hallazgos resaltan la necesidad de estudiar todas las cepas resistentes a carbapenémicos para comprender la epidemiología general de CRKp, incluso en un solo hospital, ya que centrarse únicamente en linajes específicos de CRKp y/o determinantes de resistencia puede resultar en la pérdida de información.

Además, hemos demostrado que los linajes más relevantes en el HGUV estaban previamente presentes en otras regiones españolas, siendo probablemente las fuentes de las infecciones iniciales en este hospital. Estos hallazgos respaldan que la detección de colonización al momento del ingreso al hospital puede ayudar a identificar portadores de cepas resistentes a carbapenémicos y, por lo tanto, prevenir su posterior transmisión local a otros pacientes.

Capítulo 7. Nuevas amenazas RAM: la carbapenemasa NDM-23

Uno de los objetivos esenciales de la vigilancia genómica es informar sobre nuevas amenazas y facilitar la mejora de las estrategias de intervención para controlar su propagación.

Mientras buscábamos genes RAM en la colección NLSAR, detectamos numerosos aislados con una variante no reportada de la enzima NDM, denominada NDM-23. En ese momento, esta variante no estaba disponible en las bases de datos de genes de resistencia, lo que impulsó una investigación más profunda. Hasta la fecha, se han reportado 50 variantes diferentes de la NDM, pero solo se han publicado las características de 21 de ellas. Cabe destacar que algunas variantes de la enzima muestran mayor capacidad de hidrolizar carbapenémicos. En ese momento, el alelo *bla*$_{\text{NDM-23}}$ había sido depositado como una secuencia completa de CDS en GenBank (acceso MH450214.1). Sin embargo, no se había publicado ninguna cepa portadora o información sobre el contexto genético del gen, los linajes hospedadores o el fenotipo que produce.

En este capítulo, utilizamos la epidemiología genómica para describir la emergencia y diseminación del *bla*$_{\text{NDM-23}}$, un nuevo alelo codificante de la carbapenemasa NDM-23 en la CV.

La secuencia del *bla*$_{\text{NDM-23}}$ difiere del *bla*$_{\text{NDM-1}}$ en una sustitución no sinónima en el codón 101 (I101L). Para comprobar si esta sustitución aminoacídica cambiaba el perfil de susceptibilidad a los antimicrobianos, seleccionamos dos aislados de la colección con antecedentes clínicos y genéticos altamente similares, pero con diferentes alelos del gen: el aislado 179KP-HG llevaba *bla*$_{\text{NDM-1}}$ mientras que el aislado 146KP-HG llevaba *bla*$_{\text{NDM-23}}$. Nuestros resultados mostraron que los genes *bla*$_{\text{NDM}}$ producen enzimas con el mismo perfil de resistencia y que eran responsables de gran parte del perfil de resistencia a β-lactámicos de las cepas clínicas; pero no conferían resistencia a otras clases de antimicrobianos. Tampoco mostraron capacidad hidrolítica del cefiderocol, un antibiótico nuevo.

Descubrimos que el gen *bla*$_{\text{NDM-23}}$ estaba en un plásmido multirresistente (p146KP-NDM23), con 18 genes RAM más y que dentro del plásmido, estaba en un

integrón de clase 1 complejo asociado con el ISCR1. Se han descrito contextos genéticos similares previamente con otras variantes de *bla*$_{NDM}$. Para conocer la movilidad del plásmido, buscamos marcadores *in silico* y experimentos *in vitro*. No pudimos detectar ninguna relaxasa ni genes de transferencia en la secuencia del p146KP-NDM23. Por lo tanto, la movilización de p146KP-NDM23 solo podría ocurrir en trans, con la ayuda de plásmidos co-residentes. Los ensayos de conjugación mostraron que el plásmido no se transfería a las células receptoras. Aunque estos ensayos fueron negativos, los análisis bioinformáticos mostraron la presencia del complejo integrón de clase 1 en diferentes plásmidos en la familia *Enterobacteriaceae*. Lo que sugiere que este integrón podría tener capacidad para transponerse.

El plásmido con el gen *bla*$_{NDM-23}$ se diseminó a través de una cepa de *K. pneumoniae* del ST437. La propagación de este clon se inició con un paciente trasladado desde un hospital de Madrid al HGUV. El caso índice de la diseminación de este clon, correspondiente al genoma GCA_011684095, también fue el caso índice de un brote multiclonal y multispecie detectado en el Hospital Ramón y Cajal en Madrid. Este paciente era un hombre de 36 años, originario de Pakistán y residente en Valencia. El paciente recibió atención médica en Pakistán después de sufrir un accidente de tráfico en noviembre de 2014. El 26 de agosto de 2015, al regresar de Pakistán, fue ingresado en la Unidad de Neurocirugía del HRyC en Madrid. En diciembre de 2015, el paciente fue trasladado a Valencia y admitido en el hospital HGUV. Unos meses después, este paciente pasó por la Unidad de Urgencias en el HUPLF y fue admitido en la Unidad General. El brote terminó afectando al menos a 4 hospitales diferentes en la CV desde 2016, hasta al menos 2019 (cuando concluyó nuestro muestreo). Esto explica la rápida propagación del linaje a través de la región, pero también apunta a Pakistán como posible origen de la cepa que causó este evento de diseminación. Los hallazgos de este estudio arrojan luz sobre el potencial de rápida transmisión regional a través de la expansión clonal de casos importados de cepas productoras de carbapenemasas.

Una de las principales ventajas de utilizar la WGS en la vigilancia de la RAM es la capacidad de rastrear las rutas evolutivas y epidemiológicas de los genes y plásmidos de resistencia, proporcionando una mejor y más completa comprensión

de la propagación global de bacterias resistentes. Además, esta información puede ayudar a informar las intervenciones de salud pública para prevenir su propagación adicional.

En este estudio, el uso de WGS permitió la identificación del origen y las rutas de transmisión del nuevo alelo y su plásmido asociado. Esta información se puede utilizar para desarrollar intervenciones específicas, como restricciones de viaje o protocolos de detección, para prevenir la importación y propagación de estas cepas en entornos de atención médica. En general, este estudio destaca la importancia de continuar la vigilancia utilizando WGS para monitorear la aparición y propagación de amenazas RAM e informar las acciones de salud pública.

Presente y futuro de los sistemas de vigilancia basados en WGS

La vigilancia genómica de patógenos es crítica, especialmente en el contexto de la RAM. Los proyectos de vigilancia genómica permiten comprender la prevalencia de cepas resistentes, sus determinantes genéticos y sus patrones de transmisión. Esta información es valiosa para implementar intervenciones dirigidas para prevenir la propagación de cepas resistentes. Por ejemplo, en este estudio encontramos que la diseminación intrahospitalaria fue la principal causa de transmisión de *K. pneumoniae* resistente en la CV, y que la transmisión inter-hospitalaria también desempeñó un papel. Además, identificamos casos de RAM importados que condujeron a una diseminación regional. Finalmente, también encontramos pacientes cuyas infecciones pueden haberse originado a partir de cepas presentes en su microbioma. Estos resultados proporcionan información valiosa sobre los patrones de transmisión y los factores que contribuyen a la propagación de cepas resistentes.

Además, nuestros resultados destacan la importancia de implementar estrategias de vigilancia basadas en la secuenciación completa del genoma y no sobre los métodos tradicionales de los hospitales. Como se demuestra en esta tesis, observamos diferentes linajes del mismo ST con dinámicas distintas y determinantes de resistencia asociados. Los métodos tradicionales carecen del nivel de resolución necesario. Solo el acceso a la variación genética de genomas

completos puede proporcionar suficiente detalle para obtener una imagen completa de la propagación y evolución de *K. pneumoniae* en entornos sanitarios. Además, la WGS permite realizar análisis evolutivos, que ayudan a comprender los orígenes y la evolución de las poblaciones bacterianas.

Gracias a una iniciativa del gobierno español a través de PRAN y la RedLabRA, la implementación de la vigilancia genómica de patógenos RAM se ha puesto en marcha recientemente (2022). Además, gracias a los proyectos NLSAR y SKPCV, todos los hospitales terciarios de la CV tienen la infraestructura necesaria para realizar WGS de patógenos, eliminando la necesidad de depender de otros laboratorios de referencia. Esta iniciativa marca un paso significativo hacia una vigilancia y control más eficientes y efectivos de la RAM a nivel regional y nacional.

Capítulo 9. Conclusiones

Las principales conclusiones de este trabajo son:

- Si bien la genómica es una herramienta poderosa para estudiar la RAM, asegurar la calidad de los datos es crítico para obtener bases de datos de genomas completos grandes y de gran calidad y además garantizar la calidad de los análisis posteriores.

- Los estudios de genómica en *K. pneumoniae* están altamente sesgados hacia cepas multirresistentes o clones de alto riesgo específicos y principalmente restringidos a países concretos. Por lo tanto, todavía queda mucha diversidad de esta especie sin estudiar.

- La CV tiene una alta incidencia de infecciones por *K. pneumoniae* resistentes a C3G (~20%). Además, las infecciones por *K. pneumoniae* resistentes a antibióticos, especialmente aquellas resistentes a los carbapenémicos imponen un aumento de la probabilidad de ser hospitalizado.

- La población de *K. pneumoniae* resistente es distinta entre los hospitales de la NLSAR, especialmente las cepas resistentes a los carbapenémicos. De hecho, los determinantes de resistencia a C3G y carbapenémicos y los

correspondientes linajes portadores están distribuidos de manera diferente en los hospitales, excepto por el linaje ST307 portador del gen *bla*$_{\text{CTX-M-15}}$.

- La mayoría de la RAM en *K. pneumoniae* en la CV se debió a
- la propagación dentro del hospital, pero también hubo evidencia de transmisión entre hospitales.
- La actual situación de *K. pneumoniae* RAM en la CV es el resultado de un conjunto diverso de factores, que incluyen linajes que probablemente se originaron en la comunidad o en la microbiota de los pacientes, así como una interacción compleja entre la transmisión de linajes entre hospitales y la proliferación local de clones problemáticos dentro de cada hospital.
- Aunque se encontró una alta diversidad de linajes, un linaje, el ST307, fue responsable de la mayoría de las infecciones resistentes a C3G y carbapenémicos y de las transmisiones entre hospitales en la región.
- La aparición de resistencia a los carbapenémicos en un hospital puede ser un proceso altamente complejo y rápido, que involucra muchos linajes diferentes y mecanismos de resistencia a los carbapenémicos que colonizan al mismo tiempo.
- La nueva variante de carbapenemasa NDM encontrada, NDM-23, tuvo el mismo efecto fenotípico de RAM que su ancestro, NDM-1.
- Este gen se encontró en un plásmido MDR no movilizable que codificaba otros 18 genes de RAM, lo que hizo que las cepas fueran casi completamente resistentes a los medicamentos.