THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Exploiting *Klebsiella pneumoniae* genomics for its significance in One Health

Jiayu Wang

Doctor of Philosophy

THE UNIVERSITY OF EDINBURGH

2024

Abstract

This thesis explores the genomics and epidemiological dynamics of *Klebsiella variicola (K. var)* and *Klebsiella pneumoniae (K. pne)*, focusing on their adaptation to human hosts, resistance profiles, and virulence traits.

We begin by investigating the population structure and niche adaptation of *K. var* through a comprehensive genomic dataset. Our analysis reveals how this pathogen, traditionally associated with plants, has transitioned into the human niche through the acquisition of genes either due to antibiotic overuse or proximity to *K. pne*, highlighting its genetic variability and potential public health threat. To understand its behaviour in clinical infections, we conducted a case study in Scotland involving 119 clinical strains collected in 2023. This study identified multidrug-resistant and virulent strains carrying the yersiniabactin gene, underscoring the resistance and virulence of *K. var* in Scotland and emphasizing the need for ongoing surveillance.

We next shifted our focus to *K. pne*, a close relative of *K. var*, examining epidemic clones such as ST11, ST15, and ST258. We analysed their global dissemination and resistance/virulence profiles, with particular attention to ST11, which has caused significant outbreaks in China. Our genomic analysis revealed distinct evolutionary paths in the ST11 lineage, particularly within the Asian cluster, where resistance to extended-spectrum beta-lactams and virulence traits, such as aerobactin and yersiniabactin, were prevalent. Furthermore, we identified key resistance mechanisms linked to fluoroquinolone and colistin resistance, particularly in a hypervirulent clonal expansion cluster of ST11.

Overall, this thesis advances the understanding of how these *Klebsiella* species evolve, adapt, and contribute to significant public health challenges due to their resistance and virulence traits. These findings underscore the need for continued surveillance and control measures to mitigate the risks posed by these emerging pathogens under a One Health framework.

Lay Summary

Bacteria from the *Klebsiella pneumoniae* species complex are found all over the world in humans, animals, plants, and the environment. Two important members are *Klebsiella variicola* (*K. var*) and *Klebsiella pneumoniae* (*K. pne*). These bacteria can live harmlessly in the gut or on plants, but under certain conditions, they can cause serious infections such as pneumonia, bloodstream infections, and urinary tract infections.

Until recently, *K. var* was thought to be mainly an environmental bacterium, found in plants, soil, and water. However, researchers are now seeing it more often in hospital patients, which is worrying because it can be hard to detect and may have traits that make infections more severe. This thesis investigates how *K. var* has been adapting to humans, including how it has acquired antibiotic resistance and virulence (disease-causing) traits. By analysing genetic data from hundreds of *K. var* samples from around the world, we found evidence that it is becoming better suited to living in humans. In some cases, it appears to have acquired resistance and virulence genes from its close relative *K. pne*. A detailed study of *K. var* samples collected in Scottish hospitals in 2023 showed that some strains are resistant to many antibiotics and carry genes that could make infections harder to treat.

We also studied *K. pne*, a major cause of hospital outbreaks worldwide. Some lineages, such as one called ST11, have spread widely and become resistant to multiple antibiotics. We focused on ST11 strains from Asia and found that they often combine high levels of resistance—including to last-resort antibiotics like colistin—with genes that make them more virulent. These traits make infections caused by these strains particularly difficult to treat and control.

Overall, this research shows how *K. var* and *K. pne* are evolving to survive in human hosts, becoming more resistant to treatment and more capable of causing severe disease. The findings highlight the importance of close monitoring, rapid detection, and strong infection control measures to prevent the spread of these dangerous bacteria.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my principal supervisor, Dr. Thamarai Schneiders, for her patient support and guidance throughout the past four years. Coming from a mainly bioinformatics background, I greatly appreciate her exceptional teaching in microbiology, her patient supervision, and her innovative ideas that kept my research on track. I am thankful for her mentorship, which has shaped my growth as a researcher.

I would also like to extend my heartfelt thanks to my second supervisor, Prof. David Gally, and my supervisors in the Netherlands, including Dr. Els Wessels, Prof. Ed Kuiper, and Dr. Eelco Franz, for their invaluable ideas and support. A special mention would go to Dr. Claudia Swart-Coipan, who has been almost like a bioinformatics supervisor to me. Despite never meeting her in person, I owe her a lot debt of gratitude for her weekly debugging sessions and her continuous guidance. I also wish to thank all the co-authors of our paper for their insightful ideas and constructive feedback.

The support of our group members have been super important in my PhD journey. I would like to thank Nur Afifah Izzati Binti Saiful Bahari and Kyriakos Vidalis for their laboratory assistance, unwavering support, and daily encouragement. Practising English with both of you and exchanging research ideas has been an absolute joy every day. I believe you are both incredibly talented and destined for remarkable achievements. I would also like to thank former PhD students, Dr. Charlie Metcalf and Dr. Martin Vandeputte, for setting an inspiring example of dedication and excellence.

Finally, I am deeply grateful to my parents and my husband for their unconditional love and support. I would like to thank my parents for raising me and standing behind every decision I have made. To my husband, thank you for being my rock during stressful times, for transforming my worries into love, and for always being there for me. I could not have achieved this without you.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Jiayu Wang

Contents

Abstract ii

Lay Summary iii

Acknowledgements iv

Declaration v

1 Introduction 1

1.1 *Klebsiella pneumoniae* Species Complex . . . 1

1.1.1 Introduction and identification of KpSC . . . 1

1.1.2 Species Characteristics within KpSC . . . 2

1.1.3 Genetic Diversity within KpSC . . . 5

1.2 Antimicrobial Resistance . . . 5

1.2.1 Mechanisms of Intrinsic Antimicrobial Resistance . . . 6

1.2.2 Importance of ESBL, Carbapenem and Colistin Resistance . . . 8

1.2.3 Role of Plasmids and Mobile Genetic Elements in AMR Spread . . . 12

1.2.4 Comparative Analysis of AMR Profiles Between KpSC Species . . . 15

1.2.5 Impact of AMR on Clinical Outcomes and Relevance to One Health Framework . . . 16

1.3 Virulence Factors . . . 17

1.3.1 Overview of Virulence . . . 17

1.3.2 Capsule Polysaccharide . . . 18

1.3.3 Siderophores . . . 21

1.3.4 Fimbriae and Adhesins . . . 24

1.3.5 Pathotypes and Epidemiology of Hypervirulent *Klebsiella pneumoniae* and *Klebsiella variicola* . . . 26

1.3.6 Clinical Implications of Virulence Factors . . . 28

1.4 Convergence of Resistance and Virulence . . . 29

1.4.1 Convergence in *K. pne* . . . 29

1.4.2 Convergence of resistance and virulence in *K. var* . . . 35

1.5 Population Structure, Epidemiology, and Evolution . . . 36

1.5.1 Studies in *K. pne* . . . 36

1.5.2 Studies in *K. var* . . . 39

1.6 Knowledge Gap and Summary of Findings . . . 41

2 Material and Methods **44**
2.1 Genomic DNA purification . . . 44
2.2 Genome Assembly . . . 44
2.2.1 Quality Assessment of Sequencing Reads . . . 44
2.2.2 Trimming the Reads . . . 45
2.2.3 Genome Assembly . . . 45
2.2.4 Quality Assessment for Genome Assemblies . . . 45
2.3 Building Databases . . . 45
2.3.1 *Klebsiella variicola* database . . . 45
2.3.2 *Klebsiella pneumoniae* ST11, ST15 and ST258 database . . . 46
2.3.3 Scotland *Klebsiella variicola* database . . . 47
2.4 Metadata Statistics Tests . . . 47
2.5 Annotation . . . 47
2.6 Population structure . . . 47
2.6.1 Core genome phylogeny . . . 47
2.6.2 Multi-Locus Sequence Typing (MLST) . . . 48
2.6.3 Accessory genome phylogeny . . . 48
2.6.4 Hierarchical clustering . . . 48
2.6.5 SNP distance . . . 48
2.7 Pan-genome analysis . . . 48
2.7.1 Type strain . . . 49
2.7.2 Average Nucleotide Identity . . . 49
2.8 Evolutionary analysis . . . 49
2.8.1 Ancestral state tracing . . . 49
2.8.2 Time-measured phylogeny . . . 50
2.8.3 Recombination prediction . . . 51
2.9 Resistance profile . . . 51
2.9.1 Statistical tests of resistance genes . . . 52
2.9.2 Genomic location of resistance genes . . . 52
2.10 Virulence profile . . . 52
2.10.1 Capsule locus . . . 52
2.10.2 Siderophores and genotoxins . . . 53
2.11 Other specific genes . . . 53
2.12 Genome-wide association studies (GWAS) . . . 53
2.12.1 Gene presence or absence based GWAS . . . 53

3 *Klebsiella variicola* population structure and niche adaptation **54**
3.1 Introduction . . . 54
3.2 *Klebsiella variicola* database . . . 56

3.3 Population Structure and Evolution Dynamics . . . 58
3.4 Adaptation of Non-Human *Kv* into Human Host . . . 60
3.5 Genome-Wide Association Analyses . . . 63
3.6 Resistance . . . 66
3.7 Virulence . . . 71
3.8 Discussion . . . 74

4 Scotland *Klebsiella variicola* case study 77
4.1 Introduction . . . 77
4.2 Construction of the Scotland *Klebsiella variicola* database . . . 78
4.3 Population Structure . . . 79
4.4 Resistance and Virulence . . . 81
4.5 Pan-genome and GWAS analyses . . . 85
4.6 Discussion . . . 85

5 A Large-scale genomic analysis on the epidemiology, evolution, resistance, and virulence of *Klebsiella pneumoniae* epidemic lineage ST11 87
5.1 Introduction . . . 87
5.2 Population structure of ST11, ST15 and ST258 . . . 88
5.3 Resistance and virulence score of ST11, ST15 and ST258 . . . 93
5.4 Resistance and virulence profile of ST11 . . . 95
5.5 Resistance and virulence profile of ST15 and ST258 . . . 103
5.6 Membrane permeability proteins . . . 105
5.7 Core genome and hierarchical clustering of ST11 . . . 107
5.8 Evolution and transmission . . . 110
5.9 Genome Wide Association Studies . . . 111
5.10 Discussion . . . 119

6 Discussion 121
6.1 General Discussion . . . 121
6.2 The versatility of *K. var* . . . 122
6.3 The Importance of *K. pne* being a human pathogen . . . 124
6.4 The Future of This Field . . . 125
6.4.1 Building Unbiased Datasets . . . 125
6.4.2 Dry and Wet Lab . . . 126
6.4.3 Antibiotic Overuse and Sampling . . . 127
6.4.4 Short and Long Read Sequencing . . . 127

Appendices

A Supplementary Data for Chapter 3 128

B Supplementary Data for Chapter 4 **134**

C Supplementary Data for Chapter 5 **139**

Bibliography **142**

Chapter 1

Introduction

This introduction chapter will provide context to the *Klebsiella* species and lineages of interest by reviewing their known characteristics, including species composition, antibiotic resistance, virulence factors, population structure, epidemiology, and evolutionary dynamics.

1.1 *Klebsiella pneumoniae* Species Complex

This section will review the *Klebsiella pneumoniae* Species Complex (KpSC), including its species composition, identification methods, characteristics of each species, and their genomic composition covering the core and accessory genome.

1.1.1 Introduction and identification of KpSC

Klebsiella is a genus of Gram-negative, facultatively anaerobic, rod-shaped bacteria within the family *Enterobacteriaceae* (Ryan, Ray, & Sherris, 2010). The KpSC comprises a group of closely related *Klebsiella* species that are related to both clinical and environmental health (Holt et al., 2015; Ryan et al., 2010). As is shown in Figure 1.1, the KpSC consists of *Klebsiella pneumoniae (K. pne)* (KpI), *Klebsiella quasipneumoniae (K. quasi)* (KpII), and *Klebsiella variicola (K. var)* (KpIII), which are genetically distinct species (Holt et al., 2015) but share similar morphological and biochemical characteristics (Long et al., 2017).

Accurate identification of *K. var*, *K. pne*, and *K. quasi* is crucial for understanding their distinct pathogenic roles, emerging resistance and virulence profiles, and epidemiological patterns, enabling clinical management and infection control. Current identification methods include microbiological methods, Polymerase Chain Reaction (PCR)-based methods, Matrix-Assisted Laser Desorption/Ionisation-Time of Flight (MALDI-TOF) method and Whole Genome Sequencing (WGS):

Microbiological methods: Microbiological methods rely on biochemical reactions, metabolic profiles, and phenotypic traits to classify microorganisms. Microbiological automatic systems including Phoenix, Vitek 2XL, or MicroScan cannot distinguish between species within KpSC due to the high similarity in biochemical and phenotypic characteristics among *K. pne*, *K. var*, and *K. quasi*, and will invariably categorise all KpSC members as *K. pne* (Rodríguez-Medina, Barrios-Camacho, Duran-Bedolla, & Garza-Ramos, 2019).

PCR-based methods: PCR-based methods amplify and identify chromosomal species-specific markers (Bej, Mahbubani, & Atlas, 1991) such as bla_{SHV} for *K. pne*, bla_{LEN} for *K. var*, and bla_{OKP} for *K. quasi*. This method is not always reliable and is only suitable for rough identification (Long et al., 2017), as *bla* genes can be found in multiple species. For example, our analyses showed that bla_{SHV} and bla_{OKP} also occur in a minority of *K. var* (Long et al., 2017; J. Wang et al., 2024). Recently, a triplex real-time PCR assay (ZKIR-T) was developed to simultaneously detect all KpSC members while specifically identifying *K. pne* sensu stricto, achieving 98% concordance with whole-genome sequencing (Ligowska-Marzęta et al., 2024).

MALDI-TOF: MALDI-TOF mass spectrometry identifies bacteria by identifying their proteins with a laser and analysing the mass-to-charge ratio of the resulting ions to generate a species-specific spectral fingerprint (Hillenkamp, Karas, Beavis, & Chait, 1991). Since the update to the MALDI-TOF (specifically Bruker reference library version 6.0.0.0), *K. var* and *K. pne* can be differenciated through specific mass spectrometry peaks (Rodrigues, Passet, Rakotondrasoa, & Brisse, 2018). However, this method remains limited because not all clinical diagnostic labs utilise access to MALDI-TOF.

WGS: WGS determines the complete DNA sequence of an organism's genome, providing comprehensive insights into its genetic composition and variation. WGS offers the most accurate identification of KpSC members, though it is not the most cost-effective approach. WGS is often followed by the use of Kleborate (M. Lam et al., 2021), a widely utilised tool for studying the KpSC. Kleborate identifies the species, detects resistance and virulence genes, and determines the strain's sequence type (ST) (M. Lam et al., 2021).

1.1.2 Species Characteristics within KpSC

Members of the KpSC are found not only in human hosts but also in animals, environmental, and plant sources (Holt et al., 2015). The clinical relevance of KpSC has grown increasingly significant, as these organisms are opportunistic pathogens, particularly in healthcare settings (Gorrie et al., 2022).

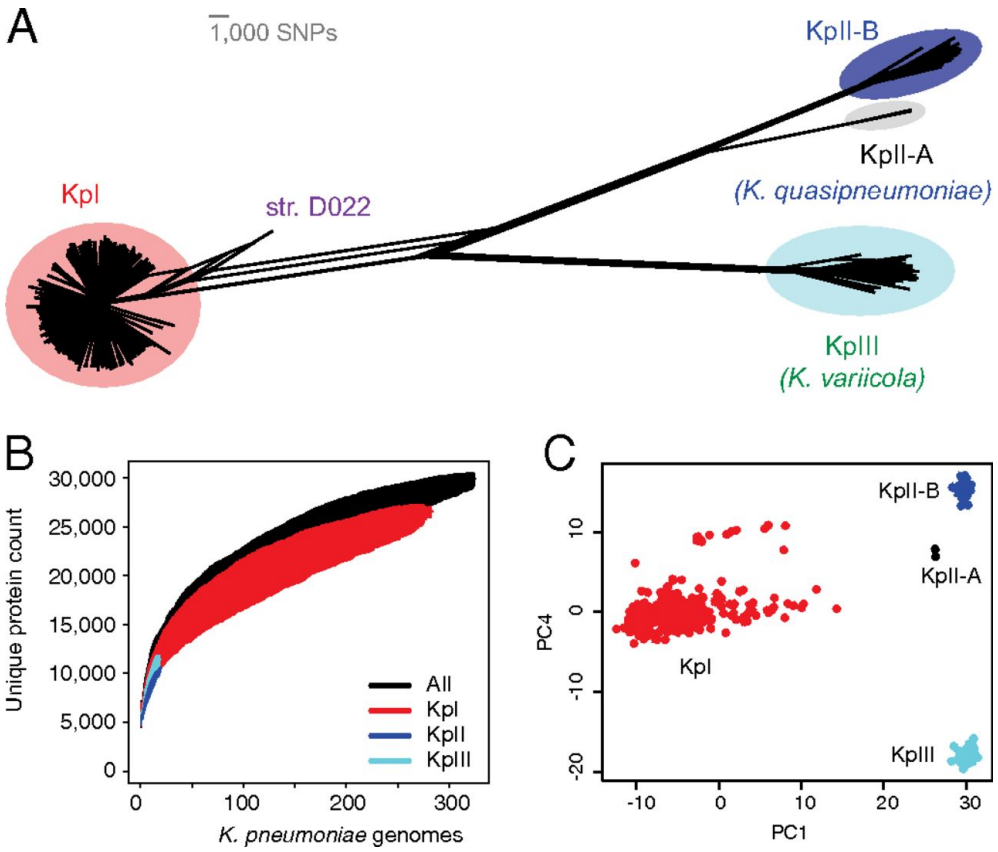


Figure 1.1: The Phylogroups and Pangenome of *Klebsiella pneumoniae* Species Complex. **A**: Split network of 328 KpSC genomes with phylogroups highlighted. **B**: Pangenome accumulation curves. **C**: PCA analysis based on the presence of common (5–95% prevalence) accessory genes. (Holt et al., 2015)

K. pne

Among KpSC species, *K. pne* (KpI) is the more prevalent cause of nosocomial infections, responsible for conditions such as pneumonia, bloodstream infections, and urinary tract infections (Holt et al., 2015; Mohd Asri et al., 2021). These infections are challenging to treat because this specie is capable of rapidly acquiring antimicrobial resistance (AMR) mechanisms, which will lead to resistance against many last-line antibiotics, including colistin and carbapenems (Y. Liu et al., 2022; Rojas et al., 2017). In Greece, over 66% of *K. pne* isolates were identified as Carbapenem-Resistant (CR) (Galani et al., 2018; “Surveillance atlas of infectious diseases”, 2021). It is estimated that more than 3% of the *K. pne* population is resistant to colistin (Azam et al., 2021; Uzairue et al., 2022). Another study reported that 13% of CR *K. pne* isolates are also colistin resistant (Rojas et al., 2017).

In addition to classical *K. pne* (cKP), hypervirulent *K. pne* (hvKP) strains, which are typically less resistant, can cause more severe diseases, such as liver abscesses and meningitis (Iwasaki et al., 2017). HvKP often carries a virulence plasmid containing siderophores and the *rmp* genes which typically produces a thick capsular polysaccharide (CPS). This is related to the hypermucoviscosity (HMV) phenotype, which distinguish it from cKP (Y.-T. Chen et al., 2004; Walker, Treat, Sepúlveda, & Miller, 2020; X. Yang et al., 2023).

K. quasi

K. quasi (KpII) is also related to the human host. Although HMV strains were reported before (Arena et al., 2015), *K. quasi* is generally less prevalent, less resistant and virulent than *K. pne* (Holt et al., 2015; Mathers et al., 2019; Octavia et al., 2019).

K. var

In contrast to *K. pne* and *K. pne* which is strongly related to the humans, *Klebsiella variicola* (*K. var*) was initially inferred to be associated with plants, as it was first isolated from banana, rice, sugar cane, and maize (Rodríguez-Medina et al., 2019; Rosenblueth, Martínez, Silva, & Martínez-Romero, 2004). However, in recent years, *K. var* has become increasingly recognised as a pathogen in human infections, largely due to prior misidentification as *K. pne* before the update to the MALDI-TOF system (Rodrigues et al., 2018). Notably, *K. var* has shown higher virulence rates than *K. pne* in certain human infections, particularly in bloodstream infections (Farzana et al., 2019; Imai et al., 2019; Maatallah et al., 2014). Certain reports have described hypervirulent *Klebsiella variicoa* (hvKv) strains (Matsuda et al., 2023; Morales-León et al., 2021). These findings suggest that the pathogenicity of *K. var* may be underestimated.

Genetically, *K. pne*, *K. quasi*, and *K. var* exhibit notable distinctions at the core and accessory genome level, which lead to phylogenetic clustering and differences in accessory genome content (Holt et al., 2015; Rodríguez-Medina et al., 2024). Despite these distinctions, all three species are capable of horizontal gene transfer (HGT) between and within species (Mathers et al., 2019; Morgado, Fonseca, & Vicente, 2022; Pedersen et al., 2020; J. Wang et al., 2024). The HGT events facilitate the exchange of resistance, virulence, or any gene that helps adaptation to flow, which could enhance their adaptability in different sources and pathogenic potential.

1.1.3 Genetic Diversity within KpSC

Species in KpSC exhibit extensive genomic diversity, primarily due to their large accessory genome that contributes to its adaptability across diverse hosts and habitats as is shown in Figure 1.1 (Holt et al., 2015; R. M. Martin & Bachman, 2018). This genomic diversity is critical to KpSC's evolutionary success, which could facilitate the adaptation to various selective pressures including antimicrobial exposure and immune system interactions (Mills et al., 2024; Thorpe et al., 2022).

The KpSC pangenome, comprising both core and accessory genes, reveals a complex interplay between conserved and variable genetic elements. Core genes, shared by all KpSC members, encode proteins that are essential for bacterial cellular functions. In contrast, accessory genes are often acquired through horizontal gene transfer, which could enhance the functional versatility of strains or species (Rocha, Henriques, Gomila, & Manaia, 2022; Segerman, 2012). Unlike species such as *Mycobacterium tuberculosis* or *Bacillus anthracis*, which maintain distinct species boundaries and only have minimal gene flow between strains (Segerman, 2012), *Klebsiella* exhibits extensive accessory genome variability, as genetic material can cross species lines (Pedersen et al., 2020; J. Wang et al., 2024). This dynamic pangenome structure allows KpSC members to rapidly acquire traits such as AMR mechanisms or virulence factors, which could strengthen their ecological fitness and pathogenic potential. The flexibility of the accessory genome underscores its importance in KpSC's evolutionary route, enhancing the adaptability of KpSC to diverse sources and host interactions.

Capsule variation and exchange in *Klebsiella* species represent an additional, naturally occurring through recombination, can radically alter phage susceptibility and plasmid conjugation rates, with large-volume capsules shielding against phages but hindering plasmid transfer, and smaller capsules enabling more efficient horizontal gene acquisition (Haudiquet et al., 2024).

Among the KpSC species, *K. pne*, a strict human pathogen, and *K. var*, initially recognised as a plant pathogen but increasingly emerging as a human pathogen, have garnered significant interest.

1.2 Antimicrobial Resistance

Studying the AMR profile in *Klebsiella* species is essential for understanding the drivers of resistance, informing treatment strategies, and mitigating the spread of multidrug-resistant strains. This section examines AMR mechanisms in *Klebsiella*, distinguishing between intrinsic and extrinsic resistance, and explores the pathways facilitating resistance, their dissemination, clinical implications, and a comparative analysis across species within the KpSC.

1.2.1 Mechanisms of Intrinsic Antimicrobial Resistance

Intrinsic AMR mechanisms are inherent to *Klebsiella* species and include resistance genes, membrane permeability, efflux and influx genes, and specific mutations. This section will provide an overview of the genetic and functional basis of the intrinsic mechanisms.

Intrinsic resistance genes: Certain AMR genes, such as *fosA* and various *bla* genes, are chromosomally present in members in KpSC (bla_{SHV}, bla_{OKP}, and bla_{LEN} are chromosomally located in *K. pne*, *K. quasi* and *K. var* respectively (Holt et al., 2015)) and contribute to baseline resistance against specific antibiotics such as fosfomycin and β-lactams (Y. Li, Kumar, Zhang, Wu, & Wu, 2023). *Klebsiella* isolates carrying *fosA* tend to have higher Minimum Inhibitory Concentrations (MICs) of fosfomycin, indicating reduced susceptibility (Elliott et al., 2019; Ito et al., 2017), while *bla* genes encode β-lactamases, enzymes that could degrade β-lactam antibiotics (Y. Li et al., 2023).

Porins: Permeability genes, particularly outer membrane porin coding genes like *ompK35* and *ompK36* (homologous to *ompF* and *ompC* in *Escherichia coli*), regulate the entrance of molecules across the bacterial cell membrane. Modifications (David et al., 2022) or loss of function (Tsai et al., 2011) in these proteins can limit the entry of antibiotics, which will thereby reduce antibiotic susceptibility. It was found that the GD insertion in loop 3 of OmpK36 could restrict the entry of carbapenems into the bacterial cell, which is widely found in Clonal Group 258 (CG258) in *K. pne* (David et al., 2022).

Pumps: Efflux pumps including *oqxAB* and *acrAB* play an important role in antimicrobial resistance by actively transporting a variety of antibiotics out of the bacterial cell, which could reduce the antibiotic concentrations (Veleba, Higgins, Gonzalez, Seifert, & Schneiders, 2012) (J. Li et al., 2019) (Padilla et al., 2010). This efflux mechanism enhances resistance to multiple antibiotic classes, including fluoroquinolones, tetracyclines, and certain β-lactams. The *acrAB* efflux pump belongs to the Resistance-Nodulation-Division (RND) family transporters (Nikaido & Takatsuka, 2009). It forms a complex with the outer membrane protein TolC and creates a tripartite pump which is capable of expelling a broad range of toxic compounds from the bacteria cell. The regulatory gene *acrR* serves as a repressor for *acrAB* expression (Padilla et al., 2010; Schneiders, Amyes, & Levy, 2003). Mutations or disruptions in *acrR* can lead to overexpression of the *acrAB* pump, thereby increasing the bacterial resistance to antibiotics that would otherwise accumulate in the cell. The *oqxAB* pump contributes to resistance in various bacteria by helping export antibiotics out of bacterial cells, particularly quinolone, tigecycline, chloramphenicol and nitrofurantoin (J. Li et al., 2019). In addition to being chromosomally located, *oqxAB* genes are also found on the plasmids in *E. coli*, but are intrinsically present in *Klebsiella* and *Raoultella* species (Kim et al., 2009).

Regulatory genes: *ompK35* and *ompK36* are regulated by *ompR*, which a regulatory role in maintaining outer membrane integrity and contributes to antibiotic susceptibility (Yoshida, Qin, Egger, & Inouye, 2006) (Janssen et al., 2024). *oqxAB* expression is regulated by the repressor gene *oqxR*, which modulates pump activity. Mutations in *oqxR* can overexpress *oqxAB*, resulting in higher levels of pump expression and thereby contributing to increased resistance. *oqxAB* could also be up-regulated by the intrinsic factors RarA and RamA (J. Li et al., 2019). Regulatory genes such as *ramA* and *ramR* present broader control over efflux systems, especially within the RND family (Rosenblum, Khan, Gonzalez, Hasan, & Schneiders, 2011) (De Majumdar et al., 2015) (Veleba & Schneiders, 2012). RamA is a global activator that can upregulate both *acrAB* and *oqxAB* efflux pumps, leading to increased antibiotic expulsion (Q. Xu et al., 2021). RamR acts as a local repressor of *ramA*; mutations in *ramR* result in derepression of *ramA* (Rosenblum et al., 2011), which subsequently enhances the activity of both *acrAB* and *oqxAB* efflux pumps. The *ram* locus is a key factor in the multidrug-resistant phenotype observed in *Klebsiella* strains (De Majumdar et al., 2015). The *soxS* and *soxR* genes are critical regulators within *Klebsiella*, especially under oxidative stress conditions, which could also contribute to antimicrobial resistance mechanisms (Anes et al., 2021; Jiménez-Castellanos et al., 2016). SoxR functions as a sensor-regulator that responds to oxidative stress and redox-active compounds , which will thereby activate *soxS* (Anes et al., 2021). *soxS* activation has been shown to upregulate the *acrAB* efflux pump, which increases resistance tetracycline (Anes et al., 2021). This *soxS-soxR* regulatory system allows *Klebsiella* species to respond to oxidative and antibiotic pressures. By linking efflux pump activity with oxidative stress responses, *soxS* and *soxR* play a key role in the development of multidrug-resistant phenotypes (Jiménez-Castellanos et al., 2016), supporting the survival and persistence of *Klebsiella* in challenging clinical settings.

Mutations: Specific chromosomal mutations can further contribute to intrinsic resistance in *Klebsiella* in addition to the permeability proteins and the regulators. The GD insertion in *ompK36* loop 3 is related to carbapenem resistance by restricting the size of the pore (David et al., 2022). Mutations in *gyrA* and *parC* could confer quinolone resistance by altering the antibiotic's target site (Kareem et al., 2021) (Zeng et al., 2020) (Rezaei, Tajbakhsh, Naeimi, & Yousefi, 2024). Similarly, modifications in *pmrB* and *phoQ* are linked to colistin resistance (Jayol et al., 2014; X. Liu et al., 2022; Mirshekar, Zadeh, Moghadam, Shahbazi, & Jazi, 2024). Together, these mutations could further strengthen bacterial capacity for a broad range of antibiotics.

Taken together, the intrinsic AMR genes, the interplay between the above membrane permeability proteins and their regulatory genes, and the specific mutations form a complex and robust resistance mechanism, allowing *Klebsiella* species to dynamically response to antimicrobial pressures. This adaptability enhances their ability to survive in environments with high antibiotic concentrations and contributes to their clinical significance as MultiDrug-Resistant (MDR) pathogens.

1.2.2 Importance of ESBL, Carbapenem and Colistin Resistance

Apart from the intrinsic genes, extrinsic AMR mechanisms involve the acquisition of resistance genes via HGT. HGT enables *Klebsiella* species to acquire additional AMR genes, significantly broadening their resistance profiles. Importantly, the acquisition of Extended-spectrum β-lactamases (ESBLs), carbapenem, and colistin resistance genes (Azam et al., 2021; Boralli et al., 2023; de Oliveira Alves et al., 2022) confers high levels of resistance to currently available antibiotics (Figure 1.2). This ability to acquire resistance genes through HGT enhances bacterial adaptability within the human host and presents significant challenges for clinical treatment. In this section, we focus on key extrinsic AMR genes in *Klebsiella*.

The focus on resistance to ESBLs, carbapenems, and colistin is crucial as these antibiotics are considered essential in treating severe infections caused by MDR organisms, specifically in *Klebsiella* (Semet et al., 2024). ESBL genes, such as $bla_{\text{CTX-M}}$, compromise the efficacy of many β-lactam antibiotics, leaving carbapenems as a primary treatment option (Kiratisin, Apisarnthanarak, Laesripa, & Saifon, 2008). However, the emergence of carbapenem resistance genes, such as bla_{KPC} and bla_{NDM}, has made carbapenems ineffective in many cases (Solomkin et al., 2010). Colistin, which is a last-line therapy for carbapenem-resistant strains, has also been undermined by the emergence of *mcr* genes (Kaye et al., 2022). These resistance mechanisms severely limit therapeutic options, emphasising their critical importance in clinical settings and the necessity of understanding their prevalence and impact.

The prevalence of ESBL, carbapenemase, and colistin-resistance genes within *Klebsiella* has become a major public health concern, contributing to the increasing emergence of MDR strains and limited treatment options in clinical settings (Holt et al., 2015). This section provides an overview of reported resistance cases across different regions, variations in resistance levels among *Klebsiella* species, and the corresponding clinical implications.

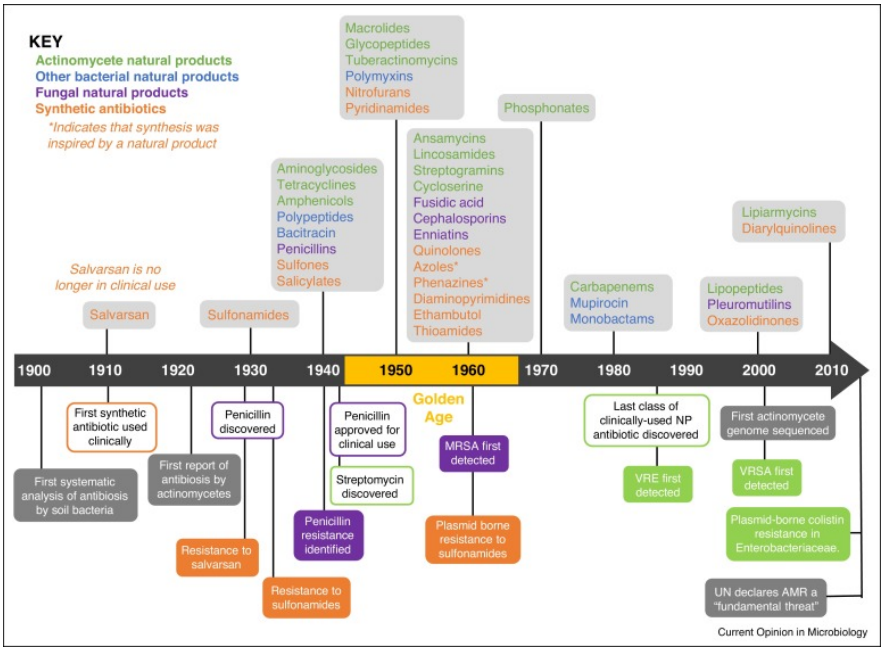


Figure 1.2: Timeline of new classes of antibiotics reaching the clinic and key dates of antimicrobial resistance. (Hutchings et al., 2019)

ESBL

β-lactam antibiotics, such as penicillins and cephalosporins, are widely used in treating infectious disease (Bush & Bradford, 2016). ESBL belongs to β-lactamase which is capable of hydrolysing the third-generation β-lactam antibiotics (Eisler, 2012; Mora-Ochomogo & Lohans, 2021). β-lactamases are widely present among *Enterobacteriaceae*, especially in *Klebsiella* and *Escherichia* species (Breilh, Texier-Maugein, Allaouchiche, Saux, & Boselli, 2013). ESBLs are predominantly encoded by genes such as $bla_{\text{CTX-M}}$, bla_{KPC} and bla_{OXA} (Kiratisin et al., 2008). ESBL-producing strains can degrade commonly used third generation β-lactam antibiotics (Castanheira, Simner, & Bradford, 2021).

Among *Klebsiella* species, *K. pne* presents a greater concern regarding the global spread of MDR strains compared to *K. quasi* and *K. var*. A study focused on *K. pne* strains collected from a Portuguese hospital identified $bla_{\text{CTX-M-15}}$ and certain bla_{SHV} variants (notably $bla_{\text{SHV-106}}$ and $bla_{\text{SHV-27}}$) as the most prevalent ESBL genes (Carvalho et al., 2021). These strains exhibited resistance to multiple antibiotics, including ciprofloxacin, aminoglycosides, and tetracyclines. Significantly, two of the ESBL-producing strains also carried carbapenemase genes ($bla_{\text{KPC-2,3}}$ and $bla_{\text{OXA-48}}$), and the study identified high-risk clonal lineages, specifically ST11 and ST15, among these *K. pne* strains (Carvalho et al., 2021). Similarly, a study on *K. pne* strains collected from Ghana revealed that over 70% of ESBL-producing strains were MDR, with resistance to tetracycline and sulfamethoxazole-trimethoprim being widespread. The study also reported $bla_{\text{CTX-M-15}}$ as the most frequently detected ESBL gene (Akenten et al., 2023).

Notably, analyses of specific *K. pne* lineages revealed that ST11 (Y. Liu et al., 2022), ST15 (Zhao, He, Li, & Sun, 2022), ST147 (Gondal, Saleem, Jahan, Choudhry, & Yasmin, 2020), ST307 (Boonstra et al., 2020), ST14 (Mshana et al., 2013), ST348 (Pons et al., 2020), and ST258 (Marsh et al., 2019) are significantly enriched in ESBL genes compared to other lineages. In contrast, data from clinical strains in Japan showed that ESBL production is less prevalent in *K. quasi* (12 out of 74, 16.2%) and *K. var* (1 out of 74, 1.4%) compared to *K. pne* (61 out of 74, 82.4%) (Sakai & Maesaki, 2022). Similarly, a study from China identified only 2 ESBL-producing *K. quasi* strains and 1 ESBL-producing *K. var* strain among 184 isolates collected from community-onset infections, while the majority of the ESBL-producing strains belonged to *K. pne* (J. Zhang et al., 2016). These data highlight the dominant role of *K. pne* in the dissemination of MDR and ESBL-producing strains globally, underscoring its critical impact on public health compared to other *Klebsiella* species.

Carbapenem

The widespread resistance to penicillins and cephalosporins has accelerated the use of carbapenems—critical last-line β-lactam antibiotics such as meropenem, doripenem, and imipenem—for the treatment of severe *Klebsiella* infections (Solomkin et al., 2010). Carbapenem resistance (CR) in *Klebsiella* is increasingly reported worldwide, with particularly high prevalence in Asia, Southern Europe, and the Middle East, where it poses major challenges for infection control and patient management.

In Asia, China has emerged as a hotspot for CR *K. pne*, with the ST11 lineage recognised as the principal driver. One nationwide study found that ST11 accounted for 88.5% (154/174) of CR isolates (Y. Liu et al., 2022), while another reported hypervirulent CR ST11 strains causing fatal pneumonia in all affected patients despite antibiotic therapy (Gu et al., 2018). Longitudinal surveillance in North China revealed a rapid rise in bla_{NDM}-carrying CR strains, which replaced the previously dominant bla_{KPC} types, with ST11 remaining the predominant clone, followed by ST15 (Duan et al., 2022). In Southeast Asia, CR strains were detected in 9 of 11 surveyed countries, with Indonesia, the Philippines, Thailand, and Vietnam all reporting resistance rates exceeding 5% (Malchione, Torres, Hartley, Koch, & Goodman, 2019).

In Europe, ST258—a close relative of ST11—has been a major CR clone. In the United States, 25% of ST258 isolates with GD insertions in OmpK36 were carbapenem-resistant (Wong et al., 2019). Southern Europe reports some of the highest CR rates globally, including Romania (52.8%), Greece (69.7%), Bulgaria (56.8%), and Cyprus (33.6%) (“Surveillance atlas of infectious diseases”, 2021). In Greece, bla_{KPC} and bla_{NDM} were the most prevalent carbapenemases (66.5% and 13.7%, respectively) (Galani et al., 2018). In the Middle East, $bla_{\text{OXA-48}}$ dominates in Saudi Arabia, detected in 81.5% of *K. pne* isolates alongside high rates of multidrug resistance (Al-Zahrani & Alasiri, 2018).

Although less common, *K. var* can also acquire carbapenem resistance through horizontal gene transfer from *K. pne*, such as the plasmid-mediated acquisition of $bla_{\text{KPC-2}}$ (L. Huang et al., 2021). These patterns highlight the central role of high-risk clones—particularly ST11, ST15, and ST258—in the global dissemination of carbapenem resistance.

Colistin

Resistance to carbapenems significantly restricts treatment options to less effective and more toxic alternatives, such as colistin (Kaye et al., 2022). Colistin serves as another last-resort antibiotic for treating carbapenem-resistant infections.

Colistin resistance is regulated by the PmrAB system, which modifies lipid A and the LPS core phosphates, enhancing bacterial survival in hostile environments (H. D. Chen & Groisman, 2013). In low Mg^{2+} conditions, PhoPQ can upregulate *pmrA* via the PmrD protein. Alternatively, mutations in *pmrB* may independently activate PmrA, resulting in increased colistin resistance without the need for PhoPQ mutations (H. D. Chen & Groisman, 2013). In addition to the PmrAB system, plasmid-mediated *mcr* genes can also lead to colistin resistance (Abozahra, Gaballah, & Abdelhamid, 2023). *mcr* encodes a phosphoethanolamine transferase enzyme, which modifies the outer membrane lipopolysaccharides by adding phosphoethanolamine to the phosphate groups in lipid A. This modification reduces the net negative charge on the bacterial cell surface, and consequently, the binding affinity of polymyxins to the bacterial cell wall decreases, leading to colistin resistance (Karki et al., 2021).

Although less prevalent than carbapenem resistance, colistin resistance rates have been rising globally, increasing from less than 2% to 9% (Petrosillo, Taglietti, & Granata, 2019). Of particular concern is the emerging trend of resistance to both carbapenems and colistin. Recent studies report colistin resistance rates as high as 20% among carbapenem-resistant *K.pne* isolates in Asia (Azam et al., 2021). This dual resistance of carbapenems and colistin is increasingly observed in the last decade, where carbapenem-resistant isolates also exhibit colistin resistance, further narrowing treatment options (Galani et al., 2018). For instance, in Greece, colistin was effective against only 59.6% of carbapenem-resistant isolates (Galani et al., 2018), while in Italy, over 43% of carbapenem-resistant isolates were also resistant to colistin (Monaco et al., 2014).

Geographical and clinical variations in carbapenem and colistin resistance prevalence reflect differences in local antibiotic usage policies, infection control practices, and the genetic composition of *Klebsiella* strains. Higher resistance rates are typically observed in intensive care units (ICUs) and other healthcare environments where antibiotic pressure is high (Qin et al., 2020; Uzairue et al., 2022). Surveillance data indicate that *K. pne* remains the predominant carrier of carbapenem and colistin resistance (“Surveillance atlas of infectious diseases”, 2021), but emerging evidence suggests that *K. var* may also be developing resistance at a notable rate (J. Wang et al., 2024).

1.2.3 Role of Plasmids and Mobile Genetic Elements in AMR Spread

Mobile genetic elements (MGEs) play a critical role in the dissemination of AMR within *Klebsiella* (Acman et al., 2022; Xanthopoulou et al., 2020). These genetic platforms facilitate the HGT of resistance genes, allowing bacteria to rapidly acquire and disseminate traits that enhance survival in the presence of antimicrobial agents. Unlike chromosomal mutations, which occur relatively slowly and are limited to vertical inheritance (estimated at approximately 18.4 mutations per genome per year in *K. pne*) (Stoesser et al., 2014), the transmission of MGEs significantly accelerates the spread of AMR by mediating HGT, and could happen across diverse bacterial species.

Plasmids, as mobile genetic elements, serve as efficient vectors for AMR genes, often carrying multiple resistance determinants within a single replicon (Karaman, Çiçek, Şemen, & Şaban Beriş, 2024). The widespread distribution of conjugative plasmids has been particularly significant in the global emergence of MDR *Klebsiella* pathogens.

Other mobile genetic elements, including transposons, insertion sequences (IS), and integrons, further enhance this process by integrating and mobilising resistance genes within plasmids or chromosomal DNA (Acman et al., 2022). The modularity of MGEs allows bacteria to adapt rapidly to selective pressures, driving the evolution of resistance phenotypes.

HGT is the principal mechanism by which MGEs spread AMR genes. HGT occurs through processes such as conjugation, transformation, and transduction, with conjugation being the most important for the dissemination of AMR determinants (Burmeister, 2015). This transfer enables the rapid emergence of resistant strains in clinical, agricultural, and environmental settings, raising the challenges associated with controlling AMR (Burmeister, 2015). The interplay between MGEs and HGT highlights the need for a comprehensive understanding of these mechanisms to devise effective strategies to combat the spread of AMR.

Plasmids

It was found that in *K. pne*, over 86% of AMR genes and virulence genes were located on plasmids (J. Zhang et al., 2023). Plasmids are usually categorised into three main classes based on their ability to transfer between bacterial cells: (1) conjugative, (2) mobilisable, and (3) non-mobilisable plasmids (Smillie, Garcillán-Barcia, Francia, Rocha, & de la Cruz, 2010), and each class contributes differently to the dissemination of AMR genes:

Conjugative plasmids encode the complete set of genes required for their transfer between bacterial cells through conjugation (Smillie et al., 2010). Conjugative plasmids in *Klebsiella* are key players in horizontal gene transfer, contributing to the spread of antimicrobial resistance (e.g., ESBL and carbapenem resistant genes) (Aghamohammad, Khazani Asforooshani, Malek Mohammadi, Sholeh, & Badmasti, 2023). These plasmids often harbour essential conjugative modules such as relaxases, type IV coupling proteins, and type IV secretion systems,

enabling self-transmission or assisting in mobilizing other genetic elements (Smillie et al., 2010). Successful *K. pne* epidemic clones, particularly ST11, ST14, and ST16, act as key strain types for resistance dissemination (Aghamohammad et al., 2023). Conjugative plasmids in *Klebsiella* predominantly belong to IncFII, IncR, IncM, IncX, and IncL replicon types. A study focusing on an outbreak of ST1049-KL5 CR *K. pne* strains in a hospital in China showed that an conjugative IncM1 plasmid, pKPC_NDM, carrying $bla_{\text{KPC-2}}$, $bla_{\text{NDM-1}}$ and *qnrS1*, was responsible for the plasmid mediated CR transmission (H. Liu et al., 2024). It was shown that strains carrying this plasmid are resistant to nearly all β-lactams, carbapenems, and β-lactamase inhibitor combinations, with susceptibility only to polymyxin B and tigecycline.

Mobilisable plasmids lack the complete conjugation machinery but possess specific origins of transfer genes (such as the *oriT* gene) and genes that enable them to be mobilised by co-residing conjugative plasmids (such as *traI*, *traD*, *mobA*, and *mobC* genes) (Smillie et al., 2010). Mobilisable plasmids played a significant role in the hospital outbreak of colistin resistance *K. pne* in China, as they facilitated the horizontal transfer of the *mcr-1* gene (Tian et al., 2017; R. Wang et al., 2018). These study found that *mcr-1* was located on IncX4 or IncFII plasmids in all isolates, which also harboured additional resistance genes like $bla_{\text{KPC-2}}$, $bla_{\text{TEM-1}}$ and $bla_{\text{CTX-M-3}}$. This highlights the importance of mobilisable plasmids as vectors for spreading critical resistance genes in healthcare environments (Tian et al., 2017; R. Wang et al., 2018). These plasmids can efficiently acquire and disseminate AMR genes by borrowing the transfer machinery of conjugative plasmids, contributing to the horizontal transfer of resistance determinants especially for the *mcr-1* gene.

Non-mobilisable plasmids in the *Klebsiella* are defined by the absence of essential elements required for self-transfer or assisted mobilisation, such as relaxase genes, type IV coupling proteins, and the origin of transfer (*oriT*) regions. Different from conjugating and mobilisable plasmids, non-mobilisable plasmids rely exclusively on vertical transmission for their maintenance within bacterial cells and often exhibit stability due to selective pressures exerted by the antibiotic resistance genes they carry. For example, previous studies demonstrated that in the p12181-KPC plasmid, the disruption of critical conjugal transfer genes, such as *traV*, rendered the plasmid non-conjugative despite carrying a broad range of resistance determinants, including $bla_{\text{KPC-2}}$, *mph(A)*, and *strAB* (J. Feng et al., 2017). Interestingly, a study on plasmids in *Escherichia coli* and *K. pne* isolates from chronic wounds in Ghana identified 8 conjugative plasmids (4 carrying $bla_{\text{CTX-M-15}}$), 11 mobilisable plasmids (2 carrying $bla_{\text{CTX-M-15}}$), and 9 non-mobilisable plasmids (1 carrying $bla_{\text{CTX-M-15}}$) (Pankok et al., 2022). Non-mobilisable plasmids can still significantly contribute to the dissemination of resistance through the clonal expansion of their host strain. This underscores their epidemiological importance in hospital-acquired infections and the persistence of clonal *Klebsiella* lineages.

Understanding the role of these plasmid classes is essential for studying the mechanisms underlying the spread of AMR and developing strategies to mitigate its dissemination.

Other Mobile Genetic Elements

Other MGEs such as integrons, transposons and IS elements are critical drivers of AMR dissemination, facilitating the horizontal transfer of resistance genes within and between bacteria populations (Acman et al., 2022; Xanthopoulou et al., 2020), each playing an important role in the acquisition, integration, and mobilisation of resistance determinants:

Integrons function as genetic platforms that capture gene cassettes through site-specific recombination, often encoding multiple resistance genes within a single structure (Domingues, da Silva, & Nielsen, 2012). For instance, the class 1 integron carrying *dfrA1* and *aadA1* gene cassettes in *Klebsiella* showed their role in the transfer of trimethoprim and aminoglycoside resistance across environments, in the hospital wastewater and clinical settings (Rolbiecki et al., 2024).

Transposons, are DNA segments capable of excising and re-inserting into different genomic locations, enabling the mobilisation of resistance genes between chromosomes and plasmids (Pray, 2008). A previous example is the IS26-flanked multidrug resistance region in *K. pne* plasmids, which includes $bla_{\text{CTX-M-15}}$ and multiple other resistance determinants, highlighting the mobility and stability of such elements (Pajand et al., 2023).

IS elements, apart from integrons and transposons, are the smallest type of MGE, act by disrupting genomic regions or facilitating the mobilisation of adjacent resistance genes through their transposase activity (Mahillon & Chandler, 1998). For instance, ISEcp1 has been shown to mediate the mobilisation of $bla_{\text{CTX-M-15}}$ in *Klebsiella* strains from both clinical and environmental sources, emphasising their widespread impact (Pajand et al., 2023; Rolbiecki et al., 2024).

Taken together, these elements contribute to the adaptability of *Klebsiella*, enhancing their ability to survive in environments with higher antimicrobial selection pressures, such as hospitals settings, specifically intensive care units, and also wastewater, where carbapenem-resistant clones like ST147 and ST392 could survive (Pajand et al., 2023; Rolbiecki et al., 2024). The interplay among integrons, transposons, and IS elements underscores their collective importance in the evolution and spread of MDR pathogens, as demonstrated by the dissemination of $bla_{\text{NDM-1}}$ and $bla_{\text{OXA-48}}$ among clonal complexes in diverse settings (Pajand et al., 2023; Rolbiecki et al., 2024).

1.2.4 Comparative Analysis of AMR Profiles Between KpSC Species

The KpSC members demonstrate a remarkable capacity to acquire and disseminate AMR genes, making them a critical target for surveillance and intervention (Rodríguez-Medina et al., 2024). Within KpSC, *K. pne*, *K. var*, and *K. quasi* exhibit significant diversity influenced by intrinsic resistance mechanisms, MGEs, and ecological adaptation. Recent analyses (Calland et al., 2023; Potter et al., 2018; Rodríguez-Medina et al., 2024; J. Wang et al., 2024) provide detailed insights into these species' AMR patterns, highlighting their varying capacities to acquire, maintain, and disseminate resistance genes:

K. pne: *K.pne* consisted of the majority of KpSC isolates (95%) and displays the broadest AMR profile, with high phenotypic and genotypic diversity (Rodríguez-Medina et al., 2024). The intrinsic AMR carried by *K. pne* is bla_{SHV} (Holt et al., 2015). It frequently harbours the extrinsic ESBL genes ($bla_{\text{CTX-M-15}}$, bla_{KPC} and bla_{OXA} (Akenten et al., 2023; Carvalho et al., 2021; Castanheira et al., 2021; Kiratisin et al., 2008; Y. Liu et al., 2022)) and carbapenemases (bla_{NDM}, $bla_{\text{OXA-48}}$ and bla_{KPC} (Al-Zahrani & Alasiri, 2018; Duan et al., 2022; Galani et al., 2018; Gu et al., 2018; Malchione et al., 2019)), alongside with intrinsic colistin resistance mechanisms (mutations in *pmrAB* or *phoPQ*) (Abozahra et al., 2023; H. D. Chen & Groisman, 2013; Karki et al., 2021). The prevalence of MDR strains (64.4%) is largely attributed to plasmid-mediated resistance genes (bla_{SHV}, bla_{TEM}) and the abundance of MGEs, particularly IncF plasmids, which drive horizontal gene transfer (Rodríguez-Medina et al., 2024). These traits underscore *K. pne*'s clinical dominance and adaptability, particularly in hospital environments.

K. quasi: *K. quasi* represents a small proportion of KpSC isolates (Akinola & Dahunsi, 2024; Rodríguez-Medina et al., 2024) and exhibits intrinsic resistance through the chromosomal bla_{OKP} gene (Fevre, Passet, Weill, Grimont, & Brisse, 2005). Extended-spectrum β-lactamase (ESBL) production appears uncommon, although $bla_{\text{CTX-M}}$ genes can occasionally be acquired (Rodríguez-Medina et al., 2024). Carbapenem resistance is also rarely reported, typically associated with bla_{NDM} or $bla_{\text{OXA-48}}$ in combination with outer membrane porin mutations such as *ompK35* and *ompK36* (Arabaghian et al., 2019). Colistin resistance remains infrequent and has been linked primarily to the acquisition of *mcr* genes (Hatrongjit et al., 2024).

K. var: *K. var*, though less prevalent (2.5%) compared to *K. pne* (95%) in clinical isolates, exhibits intrinsic resistance levels similar to *K. quasi* (Rodríguez-Medina et al., 2024). While reports of ESBL-producing *K. var* strains are increasing, comprehensive data on the overall rates of resistance remain limited. A previous analysis revealed that nearly half (13 out of 23, percentage = 56.5%) of the *K. var* strains tested were ESBL producers (Garza-Ramos et al., 2015). The rate of carbapenem resistance in *K. var* is comparatively low, with only 10 out of 145 strains (6.7%) reported as carbapenem-resistant (Potter et al., 2018). Colistin resistance

has also been documented in *K. var*, primarily associated with mutations in the *phoQ* gene (Lu, Feng, McNally, & Zong, 2018). Overall, the available data on the resistance profiles of *K. var* remain limited and sparse, and further large-scale studies are necessary to better understand its antimicrobial resistance patterns and their clinical implications.

In summary, the comparative analysis of KpSC species highlights key distinctions in their AMR profiles:

- *K. pne*: demonstrates the broadest resistance spectrum, driven by its extensive repertoire of resistance mechanisms and the high prevalence of clinically significant resistance genes such as bla_{NDM} and $bla_{\text{CTX-M}}$.
- *K. quasi*: has a more limited AMR profile but can acquire resistance genes from *K. pne*.
- *K. var*: the overall resistance data remains limited and needs further investigation.

The phenotypic heterogeneity and potential transmission of AMR genes among KpSC members highlight significant public health implications, particularly in the context of hospital-acquired infections. Understanding these differences is crucial for developing targeted surveillance systems and effective control strategies to mitigate the spread of resistance within the KpSC.

1.2.5 Impact of AMR on Clinical Outcomes and Relevance to One Health Framework

The rise of carbapenem- and colistin-resistant *Klebsiella* strains underscores the urgent need for targeted antimicrobial control and strengthened surveillance systems to monitor and contain the spread of these high-risk pathogens. The clinical impact of AMR in *Klebsiella* is huge, affecting treatment outcomes and complicating patient management in hospital settings. MDR *Klebsiella* strains present significant treatment challenges, as the limited effectiveness of last-line antibiotics, such as carbapenems, colistin, and tigecycline, severely restricts therapeutic options. These strains often require combination therapy or the use of novel or less-tested drugs, which may not be optimal and have increased toxicity (Kaye et al., 2022).

The prevalence of carbapenem-resistant *K. pne* and emerging dual resistance to carbapenems and colistin further makes clinical decisions difficult. Infections caused by these strains are associated with higher mortality rates, prolonged hospital stays, and increased healthcare costs (Duan et al., 2022). For example, carbapenem-resistant strains linked to hypervirulent *K. pne* ST11 have shown poor treatment responses, resulting in fatal outcomes (Gu et al., 2018). Similarly, colistin-resistant strains, driven by *mcr* genes often leave clinicians with few effective treatment options, necessitating the development of alternative therapeutic approaches (Azam et al., 2021; Galani et al., 2018).

Emerging therapies such as bacteriophage therapy (Ichikawa et al., 2023), antimicrobial peptides (Fleeman, Macias, Brodbelt, & Davies, 2020), and combination therapies (Medeiros, Rigatto, Falci, & Zavascki, 2019) are being explored as potential solutions. However, these approaches are still in developmental stages and face significant barriers to widespread implementation, as phages require tailored cocktails for different bacterial strains (Ichikawa et al., 2023) and peptides often face barriers like bacterial capsules or resistance mechanisms that reduce their efficacy (Fleeman et al., 2020).

From a One Health perspective, the environmental reservoirs of KpSC, particularly *K. var*, could potentially play a pivotal role in the dissemination of AMR. The ability of *Klebsiella* species to acquire resistance genes from each other, specifically *K. pne*, through plasmid exchange and chromosomal recombination highlights the connection of human, animal, and environmental health. Studies have identified *K. var* in plant, animal, and environments (Duran-Bedolla, Garza-Ramos, Rodríguez-Medina, Aguilar Vera, & Barrios-Camacho, 2021; Martínez-Romero et al., 2018; Rodríguez-Medina et al., 2019; Rosenblueth et al., 2004), emphasising its potential for zoonotic and environmental transmission. These findings underscore the importance of integrating environmental and animal surveillance and management into AMR control strategies. Addressing AMR in *Klebsiella* requires collaborated actions, including improved antibiotic stewardship, enhanced infection control measures, and strengthened surveillance systems across not only human, but also veterinary and environmental health. Such integrated efforts are important to control the clinical and ecological impact of AMR in *Klebsiella* and reduce its global AMR burden.

1.3 Virulence Factors

In addition to resistance mechanisms, virulence factors play a crucial role in *Klebsiella* infections. This section will review the major virulence factors of *Klebsiella*, discussing the pathotypes and epidemiology of hypervirulent strains and the clinical implications of these virulence factors.

1.3.1 Overview of Virulence

Virulence is a critical factor in terms of pathogenicity in *Klebsiella*, driving the ability of these species to cause severe infections in both hospital and community settings (Monteiro, Cordeiro, & Reis, 2024; Riwu, Effendi, Rantam, Khairullah, & Widodo, 2022). Virulence factors enable *Klebsiella* species to survive from host immune responses, establish infections, and disseminate within the host (Bengoechea & Sa Pessoa, 2019). The clinical relevance of KpSC virulence is underscored by its association with a wide range of infections, including respiratory system infection, blood stream infection, urinary tract infection, meningitis and liver abscesses, many of which exhibit high morbidity and mortality rates (Monteiro et al., 2024).

Klebsiella species showed diverse virulence phenotypes that can be broadly categorised as classical, hypervirulent, and hypermucoviscous (Monteiro et al., 2024). Classical strains are typically associated with nosocomial infections and often exhibit MDR, while hypervirulent strains are both prevalent in community-acquired and nosocomial infections and are characterised by promoted virulence traits such as capsule over-production and siderophore synthesis (Iwasaki et al., 2017). Hypervirulent strains usually carry the siderophores and the *rmp* gene cluster (Iwasaki et al., 2017). Hypermucoviscous strains, distinguished by their ability to form viscous phenotype, is often related to hypervirulent phenotypes and are strongly associated with invasive infections, including liver abscesses and meningitis (Walker & Miller, 2020).

Among KpSC members, *K. pne* is the most clinically significant, which showed both classical and hypervirulent traits. *K. var* and *K. quasi*, though less studied, are increasingly recognised for their emerging virulence potential these days. Strains of *K. var* isolated from BSI in human hosts have demonstrated virulence traits comparable to those of *K. pne* (Farzana et al., 2019; Maatallah et al., 2014), while certain *K. quasi* strains exhibit hypermucoviscous phenotypes (Arena et al., 2015).

Understanding the virulence factors in *Klebsiella* is critical for both clinical and public health contexts, as it offers valuable insights into disease pathogenesis, informs the development of therapeutic strategies, and guides infection control measures. The interplay between virulence and antimicrobial resistance further complicates treatment, underscoring the importance of monitoring the transmission of virulence traits within *Klebsiella* populations. Key virulence factors in *Klebsiella* include the capsule locus, siderophores, fimbriae, and toxins. This section reviews these factors in detail.

1.3.2 Capsule Polysaccharide

Capsule polysaccharide (CPS) is a key virulence factor of the *Klebsiella*, which plays a pivotal role in immune evasion, serum resistance, and biofilm formation (Rendueles, 2020; Struve & Krogfelt, 2003). The CPS forms a thick and protective layer surrounding the bacterial cell, shielding it from host immune responses and antimicrobial agents (Rendueles, 2020).

The protective properties of CPS contribute to the pathogenicity by preventing opsonisation and membrane attack complex formation (Monteiro et al., 2024; Rendueles, 2020). Additionally, the capsule inhibits phagocytosis by macrophages and neutrophils, enabling the bacteria to evade clearance by the innate immune system (March et al., 2013). Furthermore, CPS acts as a barrier against antimicrobial peptides, reducing their effectiveness and contributing to the persistence of *Klebsiella* strains in hostile environments (Fleeman et al., 2020). The capsule locus from genome assembly data could be recognised by the Kaptive software (M. M. Lam, Wick, Judd, Holt, & Wyres, 2022), which is widely used for screening the capsule genes and loci.

Synthesis of CPS

The capsule is composed of repeating polysaccharide subunits, the synthesis of which is regulated by genes encoded in the capsule locus. The capsule locus is present in each strain and represents an intrinsic virulence factor of *Klebsiella* (Monteiro et al., 2024). Key intrinsic genes involved in capsule biosynthesis, transport, and polymerisation in *K. pne* include *wcaJ*, *wbaP*, *wzi*, *wzc*, *wzx*, *wzy*, and *wza*, each playing a distinct and important role in synthesising the protective polysaccharide capsule (Rendueles, 2020).

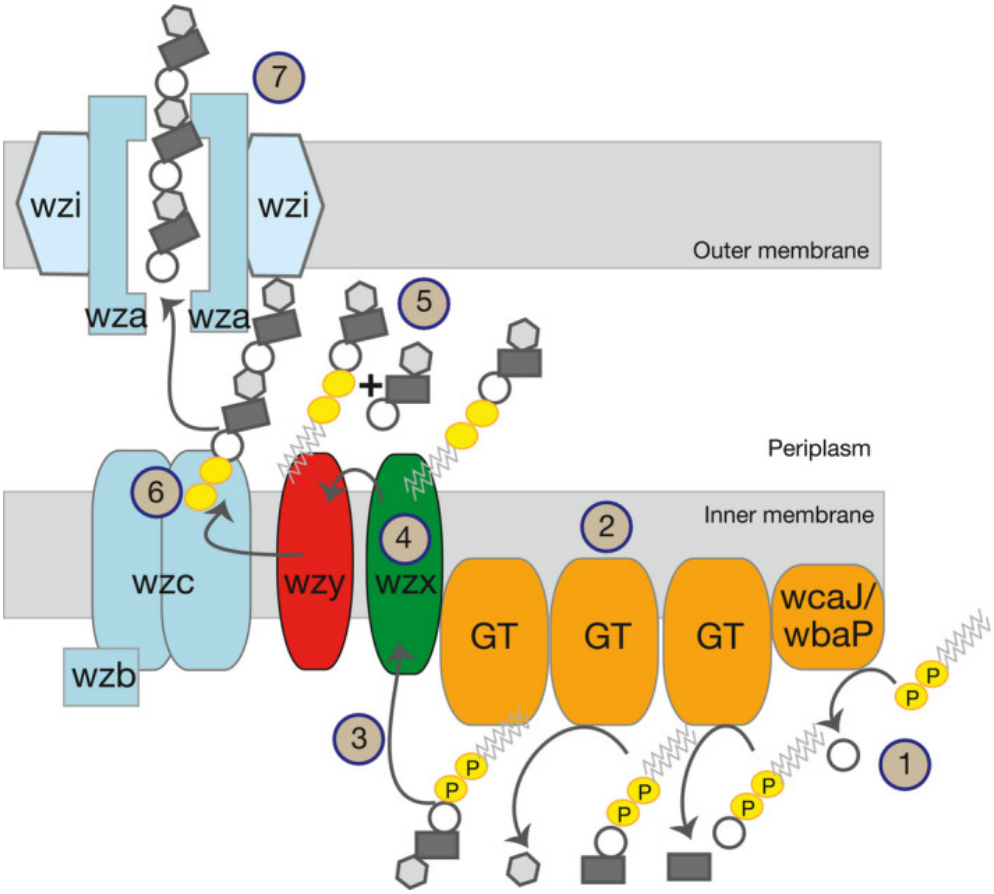


Figure 1.3: Diagram of the capsule biosynthesis pathway in *K. pne*. (Rendueles, 2020)

The first step of the biosynthesis of the capsule is the activity of glycosyltransferases encoded by *wcaJ* or *wbaP*, which could initiate the process by linking the first sugar moiety (glucose or galactose) to an undecaprenyl phosphate lipid carrier, which forms the foundation for subsequent oligosaccharide addition by other glycosyltransferases, enabling the synthesis of diverse capsule structures (Rendueles, 2020) (Figure 1.3). Importantly, studies have shown that the deletion of *wcaJ* will stop capsule production in both classical and hypervirulent strains of *K. pne* (Bain et al., 2024; Farooq et al., 2020; Y. H. Tan, Chen, Chu, Sham, & Gan, 2020).

Next, the polymerisation and translocation progress of these polysaccharides is mediated by *wzx* and *wzy* genes.*wzx* could flip the lipid-linked sugars across the inner membrane, while *wzy* could polymerise the oligosaccharide units into longer chains. The capsule length is regulated by *wzc*, which is a tyrosine autokinase that controls polymerisation activity (Rendueles, 2020) (Figure 1.3).

Finally, the assembled capsule is exported to the cell surface by *wza*, which is a protein that facilitates the transfer and secretin of the capsule from the periplasm to the extracellular environment. The capsule is then anchored to the bacterial surface through *wzi*, which is a protein that could interact with outer membrane structures and thus stabilise the capsule (Rendueles, 2020) (Figure 1.3).

These key capsule genes work together to contribute to the protective functions of CPS. However, the capsule locus often includes additional genes to form various capsule K-types that expand its functional capabilities, such as the *rmlABCD* genes, which contribute to the structural diversity of capsular polysaccharides by providing rhamnose for incorporation into the capsule, and the *manBC* genes, which encode enzymes involved in the conversion of fructose-6-phosphate into GDP-D-mannose (Shu et al., 2009). These additional genes enhance the CPS by supporting virulence through mechanisms such as immune evasion, serum resistance, and antiphagocytic properties (Shu et al., 2009). The diversity of capsule types (K-types), determined by the genetic variability of the biosynthetic loci, is associated with different levels of virulence and host specificity (Rendueles, 2020).

Variations within the capsule genes lead to diverse capsule K-types, which contribute to the virulence and adaptability of *Klebsiella* strains. To date, over 100 K-types have been identified (M. M. Lam et al., 2022), which shows the genetic and functional diversity of CPS within *Klebsiella*.

Specific K-types

Specific K-type, such as KL1 and KL2, characterised by particular arrangements of glycosyltransferase genes, are strongly associated with hypervirulent *K. pne*, particularly in invasive infections such as liver abscesses (Al-Busaidi et al., 2024). These hypervirulent strains showed enhanced serum resistance and a greater capacity for tissue invasion (Al-Busaidi et al., 2024; Remya, Shanthi, & Sekar, 2018). The association between KL1 and KL2 and hypervirulence presents the critical role of CPS in host-pathogen interactions, particularly in severe invasive infections such as liver abscesses and meningitis.

Other K-types, such as KL64, associated with the CR *K. pne* ST11 lineage, have played a pivotal role in the epidemic success of CR *K. pne* in China (K. Zhou et al., 2023). KL64 has replaced the previously dominant KL47, with KL64-carrying strains enriched in hypervirulence factors, including the *rmp* genes (K. Zhou et al., 2023).

CPS and vaccines

K-antigen-based vaccines are among the most prevalent strategies for combating *Klebsiella* infections, as targeting the capsule is highly effective due to its role as a strong virulence factor (Dangor et al., 2024; Feldman et al., 2019). Vaccines targeting capsule polysaccharides (CPS) aim to induce specific antibodies that enhance opsonophagocytosis and complement-mediated killing, thereby preventing bacterial colonisation and infection (Dangor et al., 2024).

Current vaccine approaches include those targeting a single capsule locus (Feldman et al., 2019) and multivalent formulations (Granström, Wretlind, Markman, & Cryz, 1988; Kolls, Norton, & Chen, 2022). Single-locus targeted vaccines have shown promise in inducing immunity, particularly against hypervirulent strains expressing KL1 and KL2 capsular types, which are responsible for some of the most severe infections (Feldman et al., 2019). Multivalent formulations, on the other hand, are designed to provide protective immunity against 24 capsular types, including KL2, which are associated with invasive diseases such as neonatal sepsis and infections in immunocompromised individuals (Granström et al., 1988; Kolls et al., 2022).

These vaccine strategies are being developed to protect neonates and other vulnerable populations, such as hospitalised or immunosuppressed patients, to reduce the burden of *Klebsiella* infections and associated healthcare costs (Dangor et al., 2024).

In summary, CPS is a key virulence factor in *Klebsiella*, underpinning immune evasion, resistance to host defences, and enhanced pathogenicity. The genetic diversity of capsule types and their association with hypervirulent phenotypes highlight the importance of CPS in the clinical success of KpSC strains.

1.3.3 Siderophores

Siderophores are essential for iron acquisition in *Klebsiella*, particularly under iron-restricted conditions in host environments (Kumar et al., 2024). Iron is vital for bacterial metabolism, and *Klebsiella* synthesises siderophores to compete in environments where free iron availability is limited (Monteiro et al., 2024). Four key siderophores are variably produced: enterobactin (*ent*), yersiniabactin (*ybt*), salmochelin (*iro*), and aerobactin (*iuc*) (Kumar et al., 2024). These siderophores significantly enhance *Klebsiella*'s capacity to colonise host tissues, thereby contributing to its virulence and pathogenicity (Kumar et al., 2024).

Enterobactin (*ent* genes)

Enterobactin (*ent*) is a chromosomally encoded siderophore conserved across all *Klebsiella* strains (Han et al., 2022; M. M. Lam et al., 2018). The biosynthesis of enterobactin is mediated by the *entABCDEF* genes. Enterobactin plays a fundamental role in iron acquisition but is often inadequate in human hosts due to immune system countermeasures (Han et al., 2022). Specifically, its effectiveness is limited by lipocalin-2, a host protein that binds to iron-loaded enterobactin, preventing bacterial iron uptake (Monteiro et al., 2024).

This limitation makes enterobactin less effective in overcoming host iron-sequestration defences compared to other siderophores, such as aerobactin, salmochelin, and yersiniabactin, which either evade lipocalin-2 binding or utilise distinct mechanisms for iron acquisition (Han et al., 2022; Monteiro et al., 2024). While enterobactin serves a foundational role in iron metabolism, its contribution to virulence is less pronounced compared to these more specialised siderophores, particularly in hypervirulent or MDR strains.

Yersiniabactin (*ybt* genes)

Yersiniabactin (*ybt*) is the second most prevalent siderophore in *Klebsiella*, found in approximately half of classic strains and nearly all hypervirulent strains (Holt et al., 2015; Yan, Zhou, Zou, & Liu, 2016). Its biosynthesis is mediated by the *ybtAEPQSTUX* and *irp* genes, with the receptor encoded by *fyuA* (Z. Zhu et al., 2021). These genes are widely used as markers in studies investigating *ybt*-associated virulence (Monteiro et al., 2024). Yersiniabactin often coexists with other virulence factors, contributing to the pathogenesis of various infections (Monteiro et al., 2024).

Yersiniabactin plays a critical role in evading the host's innate immune protein lipocalin-2, particularly during pulmonary infections (Bachman et al., 2011). Its prevalence is significantly higher in isolates of RSI compared to blood and urine samples. For example, in India, *ybt* was present in 60.5% of RTI isolates, and was the second most common siderophore after enterobactin (Remya, Shanthi, & Sekar, 2019). Similarly, in ventilator-associated pneumonia isolates from China, the prevalence of *ybtS* was 83.6%, with no significant differences between classical strains (88. 6%) and hypervirulent strains (71. 6%) (Yan et al., 2016).

In BSI, the prevalence of yersiniabactin exhibits geographical variability. *ybtS* was detected in 63.3% of isolates in India (Remya et al., 2019), 55.8% in studies in China (M. Zhou et al., 2020), and 29.5% in Japan (Namikawa et al., 2019). These findings suggest a potential association between yersiniabactin and BSI, although further studies outside Asia are needed to confirm this correlation. Taken together, the widespread presence of yersiniabactin, particularly in hypervirulent strains, underscores its importance as a key virulence factor in *Klebsiella*.

Salmochelin (*iro* genes)

The *iro* locus in *Klebsiella* encodes salmochelin, a glycosylated derivative of enterobactin that enables the bacteria to evade host lipocalin-2-mediated iron sequestration (M. M. Lam et al., 2018). Salmochelin is crucial in iron-limited environments and is strongly associated with hypervirulent strains, particularly those causing pyogenic liver abscesses (Monteiro et al., 2024). A previous study reported a 100% prevalence of *iro* genes in hypervirulent strains isolated from liver abscess cases (Yan et al., 2016), where they frequently co-occur with other virulence factors such as *iuc* (aerobactin) and *rmpA* (hypermucoviscosity regulator) (M. M. Lam et al., 2018; Monteiro et al., 2024).

While *iro* significantly enhances virulence in systemic infections, its role in respiratory infections has also been highlighted, with evidence suggesting that it could worsen disease outcomes in pneumonia models compared to strains lacking this siderophore (Yan et al., 2016). The co-expression of *iro* with other virulence genes underscores its importance in multidimensional pathogenicity. However, the *iro* locus is generally less common in classical strains, confirming its strong association with hypervirulence and invasive diseases (M. M. Lam et al., 2018).

Aerobactin (*iuc* genes)

Aerobactin is a critical virulence factor predominantly found in hypervirulent *Klebsiella* strains and is rare in classical strains (Monteiro et al., 2024). It was found to be a reliable marker for identifying hypervirulent strains, with diagnostic precision as high as 96% (Russo et al., 2018). The biosynthesis of aerobactin is regulated by the *iucABCD* gene cluster, while the *iutA* gene encodes its outer membrane receptor (Bailey et al., 2018). Both of these genes are frequently targeted in genetic studies of virulence (Bailey et al., 2018).

Previous studies about prevalence data showed significant geographical variability. For example, aerobactin was detected in 50.9% of *K. pne* bloodstream isolates in China (M. Zhou et al., 2020), compared to just 6.6% in India (Remya et al., 2019). In hypermucoviscous strains, aerobactin was universally present and co-occured with the *rmpA* gene, which could drive their hypermucoviscous phenotype (M. M. Lam et al., 2018). In Singapore, aerobactin was the most prevalent virulence factor in isolates from bacteremia cases (T. Y. Tan, Ong, Cheng, & Ng, 2019). Similarly, in a Chinese population, aerobactin was the virulence factor most frequently detected in liver abscess isolates, which reflected the iron-rich nature of the liver and blood environments that prefers its role in pathogenesis (Luo, Wang, Ye, & Yang, 2014). In contrast, its prevalence is markedly lower in urinary tract infections, with only 1.76% of isolates exhibiting the *iucA* gene in an Indian study (Remya et al., 2019). Aerobactin also plays an important role in infections of the central nervous system. In Taiwan, it was detected

in 66.7% of meningitis isolates and was more strongly associated with primary meningitis than with cases after craniotomy (Ku et al., 2017). The strong association of aerobactin with hypermucoviscosity and invasive infections underscores its importance as a key virulence determinant in *Klebsiella*.

Other iron uptake genes

Unlike siderophores, where small molecules secreted by bacteria to bind iron from host sources and subsequently transport the iron-siderophore complex back into the bacterial cell via specific receptors, the *kfu* and *fec* genes in *Klesielle* directly facilitate the uptake of specific iron (Daoud, Al-Marzooq, Moubareck, Ghazawi, & Collyns, 2022). The *fec* system is responsible for the uptake of ferric citrate, while the *kfu* system enables the acquisition of free or protein-bound iron, bypassing the need for extracellular chelation (Daoud et al., 2022). By facilitating the acquisition of essential nutrients like iron, the *kfu* and *fec* systems play crucial roles in the virulence of *Klebsiella*, enabling it to establish and sustain infections within the host.

In summary, these genes play a critical role in *Klebsiella* virulence by enabling efficient iron acquisition, which is a vital process for bacterial metabolism and survival in iron-restricted host environments (Monteiro et al., 2024). The production of multiple siderophores, including enterobactin, salmochelin, aerobactin, and yersiniabactin, could help in iron acquisition, and will enhance the bacteria's ability to evade host defences, colonise tissues, and establish infections (Han et al., 2022; M. M. Lam et al., 2018). These siderophores are differentially associated with hypervirulent and classical strains, with aerobactin and salmochelin showing strong links to hypervirulence (Russo et al., 2018; Yan et al., 2016), while enterobactin and yersiniabactin are more widely spread (Holt et al., 2015; M. M. Lam et al., 2018; Yan et al., 2016). Clinically, the siderophore arsenal contributes significantly to the pathogenicity of *Klebsiella*, driving invasive diseases such as pyogenic liver abscesses, bloodstream infections, and meningitis (Ku et al., 2017; Luo et al., 2014; Remya et al., 2019; Yan et al., 2016).

1.3.4 Fimbriae and Adhesins

Fimbriae are hair-like structures on the surface of *Klebsiella* that play a critical role in adhesion to host tissues and surfaces (Wu, Huang, Fung, & Peng, 2010). This adhesion is mediated by fimbrial adhesins, proteins located at the tips of fimbriae, which specifically recognise and bind to receptors on host cells (Wu et al., 2010). *Klebsiella* is thought to possess ten types of fimbriae, with type 1 and type 3 being the most significant (Y.-J. Huang, Liao, Wu, & Peng, 2009; Struve, Bojer, & Krogfelt, 2008).

Type 1 Fimbriae

Type 1 fimbriae is encoded by the *fim* gene cluster (*fimABCDEFGHIK*) (Sauer et al., 2016). FimH adhesin is the primary functional component, which specifically binds to mannose-rich glycoproteins, such as uroplakin Ia in the urothelium (Sauer et al., 2016). This interaction is critical for urinary tract colonisation, making type 1 fimbriae a key virulence factor (Sauer et al., 2016). Nearly all *Klebsiella* strains express FimH, contributing to their ability to infect diverse sites, including urinary track and blood stream infections (El Fertas-Aissani, Messai, Alouache, & Bakour, 2013; Struve, Bojer, & Krogfelt, 2009). Mutant strains lacking functional *fim* clusters showed markedly reduced urovirulence, which highlights the critical role of type 1 fimbriae in infection (Struve et al., 2009).

Type 3 Fimbriae

Type 3 fimbriae, encoded by the *mrk* gene cluster (*mrkABCDF*) (Stahlhut et al., 2013). Among these genes, *mrkD* plays a pivotal role in biofilm formation, and is the adhesin responsible for binding to extracellular matrix components, such as collagen and fibrinogen, as well as abiotic surfaces (Stahlhut et al., 2013). Biofilms provide *Klebsiella* with enhanced protection against host immune responses (Struve et al., 2009) and antibiotics (Bandeira, Almeida Carvalho, Duarte, & Jordao, 2014). Notably, biofilm-associated bacteria exhibit resistance to antibiotics that could be 10 to 1000-fold higher than the planktonic counterparts (Bandeira et al., 2014). Biofilm could also help *Klebsiella* to colonise medical devices such as catheters and endotracheal tubes, which is a major factor in healthcare-associated infections (Bellifa et al., 2013; Murphy, Mortensen, Krogfelt, & Clegg, 2013; Singhai, Malik, Shahid, Malik, & Goyal, 2012). Notably, *mrkD* alone has been shown to drive biofilm formation, independent of *fimH* (Bandeira et al., 2014; Fu et al., 2018).

While type 3 fimbriae are not as crucial as type 1 fimbriae in the context of urinary tract infections, they contribute significantly to long-term infections and the persistence of *Klebsiella* in hospital environment (Bandeira et al., 2014; Stahlhut et al., 2013). Their role in biofilm formation could also enhance *Klebsiella*'s ability to colonise and survive on medical devices, which will increase the chance of nosocomial infections (Bellifa et al., 2013; Murphy et al., 2013; Singhai et al., 2012). Additionally, the co-existence of type 1 and type 3 fimbriae in some strains showed a synergistic effect (Struve et al., 2009), enabling *Klebsiella* to adapt to diverse niches within the host and external environments.

Taken together, fimbriae and adhesins enable *Klebsiella* to adhere to host tissues, evade immune responses, and persist in the host, contributing to its ability to cause infections and persist in clinical devices (Wu et al., 2010). This underscores their critical role in *Klebsiella*'s virulence and epidemiology.

1.3.5 Pathotypes and Epidemiology of Hypervirulent *Klebsiella pneumoniae* and *Klebsiella variicola*

Among the *Klebsiella* species, *K. pne* is the most frequently associated with severe human infections. The cKp strains are well-known opportunistic pathogens that primarily cause nosocomial infections such as pneumonia, bloodstream infections, and urinary tract infections, particularly in immunocompromised patients (Monteiro et al., 2024). However, within *K. pne*, a distinct subset known as hvKp has emerged, which usually carries siderophores and *rmp* genes, and is capable of causing invasive diseases in healthy individuals with high mortality rates. These include life-threatening conditions such as liver abscesses, meningitis, and endophthalmitis (Monteiro et al., 2024). The clinical significance of hvKp has grown with its rising prevalence globally, presenting a unique challenge due to its combination of hypervirulence and, increasingly, multidrug resistance (Russo & Marr, 2019).

Emerging evidence suggests that *K. var*, traditionally classified as an environmental or plant-associated bacterium (Rosenblueth et al., 2004), is evolving into an emerging human pathogen (Rodríguez-Medina et al., 2019). Although *K. var* infections were historically underreported, likely due to misidentification as *K. pne*, advancements in genomic and proteomic tools, such as MALDI-TOF, have revealed its growing clinical presence. Hypervirulent *K. var* (hvKv) strains have now been documented, with reports suggesting that infections caused by hvKv may exhibit even higher mortality rates than those caused by hvKp (Farzana et al., 2019; Maatallah et al., 2014; Matsuda et al., 2023; Morales-León et al., 2021).

Given the clinical relevance and rising prevalence of both hvKp and hvKv, understanding their pathotypes, resistance profiles, and epidemiology is critical. Such insights will inform the development of effective diagnostic tools, targeted treatment strategies, and infection control measures to mitigate their increasing impact on global public health.

Pathotypes

hvKp and hvKv represent distinct pathotypes within their respective species, characterised by enhanced virulence, invasive capabilities, and unique genetic traits.

hvKp is distinguished from cKp by its ability to cause severe invasive infections, including pyogenic liver abscesses, community-acquired pneumonia, meningitis, and endophthalmitis (Choby, Howard-Anderson, & Weiss, 2020). These infections often occur in healthy individuals and exhibit a high mortality rate (Choby et al., 2020). The characteristic of hvKp pathogenesis is the overproduction of CPS, driven by genes such as *rmpA* and *rmpA2*, which could enhance capsule production and hypermucoviscosity (Russo & Marr, 2019). Key hvKp-specific virulence factors include large virulence plasmids carrying siderophore systems (*iucABCD* for aerobactin and *iro* for salmochelin), and genomic islands like ICEKp10 encoding colibactin and yersiniabactin (Y.-H. Huang et al., 2018; Lai et al., 2014).

hvKv shares several phenotypic and genotypic traits with hvKp but exhibits distinct characteristics shaped by its evolutionary divergence. hvKv strains are associated with invasive infections, including BSIs and UTIs, often with high mortality rates (Morales-León et al., 2021). Notably, hvKv has been reported to harbour a large virulence plasmid (pVir_030666) encoding *rmpA*, *rmpA2*, and siderophore systems such as aerobactin, salmochelin, and yersiniabactin (Dong, Lin, Zhang, Chan, & Chen, 2018). Despite a frameshift mutation in *rmpA2* leading to the absence of the hypermucoviscous phenotype, these strains displayed significant virulence (Dong, Lin, et al., 2018). A significant outbreak of hvKv was reported in Bangladesh, where hvKv strains caused a fatal neonatal sepsis outbreak with high mortality rates (Farzana et al., 2019). These strains lacked *rmpA* and *rmpA2*, yet demonstrated virulence through iron acquisition systems and fimbriae. Additionally, they carried $bla_{\text{NDM-1}}$ and $bla_{\text{CTX-M-15}}$, illustrating the convergence of MDR and hypervirulence, further emphasizing the clinical importance of hvKv.

Epidemiology

hvKp is especially enriched in Asia, especially in Taiwan, South Korea and Japan, where it is a leading cause of liver abscess (Choby et al., 2020; Matsuda et al., 2023). However, hvKp is increasingly reported globally, with infections documented in North America (Nadasy, Domiati-Saad, & Tribble, 2007; Russo & Marr, 2019), Europe (Assessment, 2021), and South America (Morales-León et al., 2021). Gastrointestinal colonisation is a critical reservoir for hvKp, with studies showing high carriage rates among healthy individuals (J. Yang et al., 2022). Transmission occurs via contaminated food, water, and close contact, with healthcare environments emerging as significant reservoirs due to the evolution of extensively resistant hvKp strains from classical strains through plasmid acquisition (Russo & Marr, 2019).

hvKv is globally distributed but frequently misidentified as hvKp, leading to underreporting of its clinical prevalence (Rodríguez-Medina et al., 2019). It has been implicated in healthcare-associated infections such as BSIs, UTIs, and pneumonia, with documented outbreaks in South America (Morales-León et al., 2021) and Asia (Farzana et al., 2019; Matsuda et al., 2023). hvKv are less common than hvKp, but current studies would underscore its pathogenic potential. *K. var* has been increasingly recognised as human pathogen these days (Farzana et al., 2019; Lu et al., 2018; Morales-León et al., 2021). The convergence of hypervirulence and antimicrobial resistance in hvKv, as observed in outbreaks involving $bla_{\text{CTX-M-15}}$ and $bla_{\text{NDM-1}}$, underscores its growing clinical relevance (Farzana et al., 2019).

In summary, the hypervirulent *K. pne* and *K. var* represent emerging threats to global health, characterised by their invasive pathotypes and expanding epidemiological footprints (Rodríguez-Medina et al., 2019; Russo & Marr, 2019). While hvKp remains the predominant cause of hypervirulent infections worldwide, hvKv is an under-recognised pathogen with unique ecological and clinical features. Accurate identification, surveillance, and targeted interventions are critical to mitigating the impact of these hypervirulent *Klebsiella* species in both community and healthcare settings.

1.3.6 Clinical Implications of Virulence Factors

The virulence factors of the *Klebsiella* species significantly impact host-pathogen interactions, disease severity, treatment outcomes, and clinical management (Gomez-Simmonds & Uhlemann, 2017).

Host Interactions and Immune Evasion

Klebsiella species showed a range of strategies to modulate host immune responses, enabling evasion and systemic dissemination (Bengoechea & Sa Pessoa, 2019). Virulence determinants, such as the CPS, inhibit phagocytosis, suppress complement activation, and neutralise antimicrobial peptides (Rendueles, 2020; Struve & Krogfelt, 2003). Furthermore, KpSC species actively suppress pro-inflammatory cytokines, including TNF-α, and modulate adaptive immune responses by dampening Th17 cell activity, which is crucial for bacterial clearance (Bengoechea & Sa Pessoa, 2019; Clemente et al., 2017). These immune evasion mechanisms not only promote survival within the host but also facilitate the establishment of infections in diverse niches (Bengoechea & Sa Pessoa, 2019).

The interplay between virulence factors and host/habitat adaptation is critical for *Klebsiella* pathogenesis. In human hosts, the enhanced capsule production and siderophore-mediated iron acquisition systems of hypervirulent strains support survival in nutrient-restricted environments like the bloodstream and liver (Russo & Marr, 2019). Conversely, in non-human hosts/habitats, certain virulence traits, such as colibactin production, may provide competitive advantages, enabling adaptation to various reservoirs (Addington, Sandalli, & Roe, 2024; M. M. Lam et al., 2018).

Impact on Treatment Outcomes and Disease

Virulence factors are directly linked to disease severity and treatment outcomes. Hypervirulent strains, particularly those encoding the K1 capsule type, are associated with invasive syndromes such as pyogenic liver abscesses, metastatic infections, and high morbidity (Al-Busaidi et al., 2024; Remya et al., 2018). The expression of siderophores, including aerobactin

and salmochelin, enhances bacterial fitness under iron-limited conditions, increasing the likelihood of systemic dissemination and poor clinical outcomes (Kumar et al., 2024; Monteiro et al., 2024; Russo & Marr, 2019). Moreover, the hypermucoviscosity phenotype is related to increased resistance to neutrophil killing and antimicrobial agents, which will make treatment more complicated (March et al., 2013).

1.4 Convergence of Resistance and Virulence

The convergence of resistance and virulence in *Klebsiella* species represents a critical public health concern, as it will give rise to strains that are both difficult to treat and more capable of causing severe infections. There are three primary mechanisms that lead to convergence (Arcari & Carattoli, 2023): MDR isolates acquiring hypervirulent plasmids; hypervirulent isolates acquiring MDR plasmids (e.g., carbapenemase genes like bla_{KPC}, bla_{NDM}, bla_{OXA}); or acquisition of hybrid plasmids combining both resistance and virulence traits.

These processes are driven by mobile genetic elements such as IncF, IncX, and IncHI1 plasmids and integrative conjugative elements (M. M. Lam et al., 2019).

1.4.1 Convergence in *K. pne*

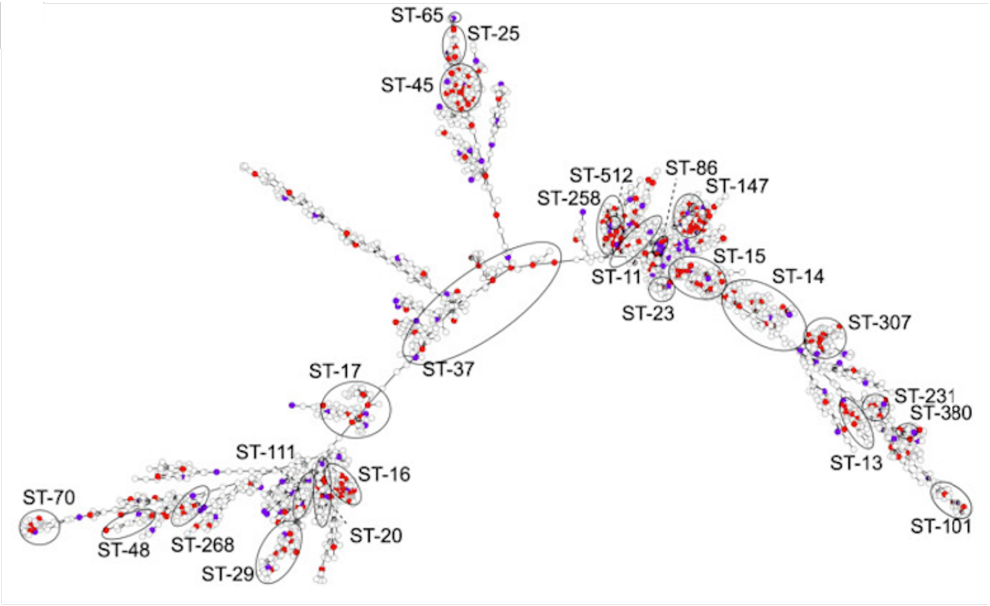


Figure 1.4: Minimum Spanning Tree of 3123 *K. pne* to show the phylogenetic relationship of different STs. (M. J. Martin et al., 2023)

K. pne ST11 lineage

Clonal group 258 (CG258) is one of the most successful *K. pne* lineages worldwide, characterised by extensive multidrug resistance and strong epidemic potential. CG258 includes globally epidemic sequence types ST11, ST258, and ST512, representing a highly disseminated MDR *K. pne* group (C. Liu et al., 2022) (Figure 1.4). This group is a frequent carrier of the carbapenemase gene bla_{KPC}, typically found on plasmids with a pKpQIL backbone (C. Liu et al., 2022). ST11 is now the predominant bla_{KPC}-producing clone in China, and the bla_{KPC} gene was frequently harboured on plasmids (Duan et al., 2022; Qi et al., 2011). Multiple ST11 outbreaks occurred in China in recent years:

Back in 2015, a study of 28 ST11 CR *K. pne* infections across nine cities in China identified a hypervirulent and CR *K. pne* strain from a blood sample, marking it a significant public health threat (Y. Zhang et al., 2015). This is a convergence case of plasmid-mediated acquisition of carbapenemase genes (e.g., $bla_{\text{KPC-2}}$) into virulent K2 strains. This hypervirulent and CR *K. pne* strain exhibited hypermucoviscosity, carbapenem resistance, and increased serum resistance and mouse lethality compared to other CRKP isolates. Molecular analysis revealed decreased expression of OmpK35/36, production of ESBLs, and the presence of virulence genes *rmpA*, aerobactin, *entB*, and *mrkD*. Capsular genotyping identified the strain as KL2 serotype. The findings highlight the need for heightened clinical awareness and robust epidemiologic surveillance to mitigate the spread of hypervirulent carbapenem-resistant strains.

In 2015, another study in a Beijing tertiary hospital investigated CR hvKp isolates and compared it with carbapenem-resistant classic *K. pne* (Yao et al., 2015). Among 33 cases, seven CR hvKp isolates were identified, belonging to ST25, ST65, and ST11, which showed KL2 and non-typeable capsular serotypes. Most CR hypervirulent strains carried the $bla_{\text{KPC-2}}$ gene and pLVPK-related virulence loci, including *iroN*. The CR hvKp infections were nosocomially acquired and showed no significant differences in clinical outcomes compared to CR cKP cases, although having enhanced virulence and resistance profiles. The study highlights the bi-directional evolution of hypervirulent strains (ST65 and ST25) acquiring carbapenem resistance and carbapenem-resistant strains (ST11) gaining hypervirulent traits, indicating a major challenge in managing hospital-acquired *K. pne* infections.

Later on in 2018, in Zhejiang, China, a study described the convergence of carbapenem resistance and hypervirulence in ST11 *K. pne* strains associated with a fatal outbreak (Dong, Yang, Zhang, Chan, & Chen, 2018). The strains showed both MDR and hypervirulence, which was caused by the coexistence of a conjugative $bla_{\text{KPC-2}}$-bearing MDR plasmid and a non-conjugative pLVPK-like virulence plasmid. Interestingly, the integration of an ISKpn18 insertion sequence disrupted the *mgrB* gene, which could confer colistin resistance and making the

strains resistant to colistin. In the study, they showed that homologous regions between the MDR and virulence plasmids suggest co-integration as a possible mechanism for transferring virulence plasmids from ST23 hypervirulent strains to ST11 CR strains, which highlights the plasmid-mediated evolution of this “superbug.”

More recent studies identified the emergence of a novel subclone of *K. pne* ST11 carrying KL64 capsule locus (ST11-KL64) in a tertiary hospital in China, characterized by a concerning convergence of carbapenem resistance and hypervirulence in Zhejiang, China (K. Zhou et al., 2020). The previously dominant ST11-KL47 was progressively replaced by ST11-KL64 during 2013–2017, driven by capsular recombination and acquisition of pLVPK-like virulence plasmids carrying *rmpA, rmpA2*, and aerobactin genes. ST11-KL64 strains demonstrated enhanced survival rates, biofilm production, and resistance to human neutrophils, and also showed increased transmissibility and higher mortality rates compared to ST11-KL47. This subclone’s evolution highlights a more dangerous convergence of antimicrobial resistance and virulence, which underscores the need for urgent surveillance and infection control measures to control its dissemination.

In 2020, another study showed the convergence of resistance and virulence in *K. pne* ST11, particularly in hypervirulent and CR strains, isolated from diarrheal stool samples of inpatients in Zhejiang Province, China (Zheng et al., 2020). The $bla_{\text{KPC-2}}$ gene and virulence-associated plasmids, including those harbouring *rmpA* and *rmpA2*, were found in a significant proportion of isolates, with diverse capsular types (KL105, KL47, KL64). These strains showed MDR and high genetic diversity, which suggests complex evolutionary processes involving multiple plasmids and lineages. The findings underscore the gut’s role as a reservoir for ST11 CR hvKp, and showed its dissemination and potential for nosocomial outbreaks in China. Enhanced surveillance in clinical settings, particularly ICUs, is critical for controlling the spread of the KL64 ST11 high-risk clones.

A study in 2020 across 11 cities in China identified the emergence of K47 and K64 hypervirulent carbapenem-resistant ST11 *K. pne* strains, which combine multidrug resistance with hypervirulence, posing significant clinical threats (Q. Yang et al., 2020). These strains harboured the resistance genes including $bla_{\text{KPC-2}}$, $bla_{\text{CTX-M-65}}$, and $bla_{\text{TEM-1}}$, often located on unique plasmids like pR16-Hv-CRKp1, pR16-Hv-CRKp2 and pR16-Hv-CRKp3, which also carry predicted virulence genes. Despite that these strains lacked classical virulence genes such as *rmpA/rmpA2*, some strains showed high virulence in infection models, attributed to alternative virulence factors. The convergence of resistance and virulence in these strains underscores the urgent need for enhanced surveillance and control measures across China.

In Nanchang, China, a study in 2021 showed the convergence of resistance and virulence in ST11 *K. pne* isolated from bloodstream infections (Du et al., 2021). These strains co-harbour $bla_{\text{KPC-2}}$-carrying resistance plasmids and pLVPK-like virulence plasmids, with some of the strains possessing hybrid plasmids encoding both of the resistance and virulence genes. The

isolates exhibit MDR, enhanced virulence, and high plasmid transferability, with key virulence genes (*rmpA, rmpA2, iutA, ybtS*) contributing to their pathogenicity. *Galleria mellonella* infection models confirmed higher virulence compared to cKp. These strains also carried diverse capsular types, with KL47 and KL64 being the predominant ones. The dominance of ST11 in these infections, combined with high mortality (66.7%), is posing significant public health threats due to their potential for rapid dissemination and limited treatment options.

The analysis focusing on the evolution and emergence of ST11 CR hvKP showed that ST11 strains, originally known for their resistance to carbapenem, have evolved into hypervirulent and CR strains through the acquisition of a conjugative plasmid, p17ZR-91-Vir-KPC (Xie et al., 2021). This plasmid, which combines resistance and virulence genes, can transfer from hypervirulent K1/K2 *K. pne* strains to non-K1/K2 capsular types like ST11. The fusion of a non-conjugative virulence plasmid with a conjugative resistance plasmid facilitates the HGT, and enables ST11 to have both high virulence and multidrug resistance. This evolutionary trajectory represents a critical public health concern, as ST11 CR hvKp strains have become prevalent in clinical settings, particularly in China, and are associated with severe infections and high mortality. The study underscores the adaptability of ST11 and its capacity to become a dominant and highly pathogenic lineage.

Another study focusing on the evolution of KL64 ST11 examined the genetic and phenotypic evolution of CR ST11 in China from 2014 to 2019 (K. Zhou et al., 2023). This study showed that the replacement of the previously dominant subclone KL47 by KL64, which exhibited enhanced virulence and resistance. They reported that the KL64 strains emerged through recombination in CPS locus, which was driven by a unique *recC* mutation (His935Arg) that increased recombination proficiency and enabled rapid diversification into the Kl64 subclone. This subclone accumulated MGEs such as plasmids, prophages, and insertion sequences, including virulence plasmids carrying *rmpA* and *iucABCD* contributing to hypervirulence traits and resistance to phagocytosis and antibiotics like tetracycline and sulfamethoxazole trimethoprim. Phylogenetic analysis suggests that the KL64 isolates originated in China around 2006, evolving independently of other global ST11 subclones.

The convergence of resistance and virulence in *K. pne* has been extensively documented in the globally dominant ST11 lineage. Studies across China reveal the rapid evolution of CR hvKp strains, which combine MDR with hypervirulence through the acquisition of virulence plasmids (*rmpA, iucABCD, ybtS*) and resistance plasmids ($bla_{\text{KPC-2}}$, $bla_{\text{CTX-M-15}}$) (Du et al., 2021; Xie et al., 2021; Q. Yang et al., 2020; Zheng et al., 2020; K. Zhou et al., 2023; M. Zhou et al., 2020). These strains are associated with diverse capsular types, particularly KL64, which has replaced the previously dominant KL47 due to enhanced recombination proficiency and capsular locus diversification (Xie et al., 2021; K. Zhou et al., 2023). Plasmid-mediated HGT is a key mechanism driving this convergence, as demonstrated by hybrid plasmids carrying both resistance and virulence determinants. The emergence of these strains underscores the

need for enhanced surveillance, particularly in high-risk settings such as intensive care units, to control their spread. The following subsections will explore similar evolutionary trends in other high-risk lineages, such as ST15 and ST258, further illustrating the global impact of convergence in *K. pne*.

K. pne ST15 lineage

ST15 represent another high-risk clone of *K. pne* with distinct yet overlapping epidemiological and genomic characteristics with ST11 lineage (Figure 1.4). As described above, ST11 is predominantly associated with CR, driven by the acquisition of carbapenemase genes such as bla_{KPC}, and has been implicated in high-mortality outbreaks of CR hvKp, particularly in China. In contrast, ST15 is a globally disseminated MDR clone frequently linked to nosocomial outbreaks and commonly harbouring ESBL genes like $bla_{\text{CTX-M}}$ (Y. Hu et al., 2024).

A study reported the genetic characterisation of two *K. pne* ST15 isolates from Norwegian hospitals, which showed a convergence of MDR and hypervirulence determinants within a single plasmid (M. M. Lam et al., 2019). The isolates harboured large hv-MDR plasmids (pKp104014_1 and pKp112126_1), which combine sequences typical of KpVP-1 virulence plasmids (encoding *iuc* and *rmpA2*) and IncFIIK MDR plasmids (carrying $bla_{\text{CTX-M-15}}$ and additional AMR genes). These plasmids showed to have formed through homologous recombination. One isolate's plasmid contains eight AMR genes ($bla_{\text{CTX-M-15}}$, bla_{TEM}, *aac3'-IIa, dfrA1, satA2*, bla_{SHV}, *sul1* and *aadA1*), while the other harbours remnants of AMR loci (bla_{TEM} and *aac3'-IIa*, and $bla_{\text{CTX-M-15}}$). Both isolates are linked epidemiologically and genomically to Romania. The study highlights the ability of such plasmids to simultaneously transmit resistance and virulence determinants.

A study in Shanghai, China identified five *K. pne* isolates belonging to ST15, and two isolates showed convergence in resistance and MDR (Y. Li et al., 2021). These ST15 isolates, along with others, exhibited multidrug resistance, with all carbapenem-resistant strains carrying the $bla_{\text{KPC-2}}$ gene. The ST15 strains in the study carried virulence genes such as *entB*, which encodes for the siderophore enterobactin. Additional virulence determinants, including genes for type 1 fimbriae (*fimH*) and type 3 fimbriae (*mrkD*), were also present in these strains. Hypervirulent-associated genes, such as *rmpA* and *rmpA2*, were also identified in some ST15 isolates. These genes collectively contribute to the strains' enhanced capacity for adherence, evasion of host defences, and virulence.

In a study using strains collected from three U.S. hospitals, the authors identified *K. pne* ST15 isolates as part of the emerging convergent strains that combine features of multidrug resistance and hypervirulence (Kochan et al., 2023). These ST15 isolates were found to harbour virulence genes typically associated with hypervirulent strains, such as *iuc* (aerobactin biosynthesis genes), *iro* (salmochelin biosynthesis genes), and occasionally *rmpA* or *rmpA*. Despite possessing these genes, the ST15 isolates did not exhibit enhanced virulence compared to

cKp strains in a murine pneumonia model. This attenuation was attributed to potential mutations in key virulence genes or the acquisition of these genes on plasmids that were not fully functional. The findings highlight the complexity of resistance and virulence convergence in *K. pne* and suggest that the presence of hypervirulence-associated genes does not guarantee hypervirulent phenotypes.

In summary, unlike ST11, which is predominantly associated with CR and hypervirulence, ST15's emergence highlights a concerning global concern, which poses a significant challenge to public health through its broad geographical distribution and plasmid versatility.

K. pne ST258 lineage

The convergence of resistance and virulence was also found in the ST258 lineage, which is more related to North America.

The study using strains collected from the U.S. hospitals also found the convergence in ST258 *K. pne* strains, which were known for their diverse resistance determinants, including carbapenemase genes like bla_{KPC} (Kochan et al., 2023). The ST258 isolates has also acquired hvKp-associated genes such as *iuc, iro, rmpA*, and *rmpA2* genes. The *iuc* gene were commonly detected, but their contribution to siderophore production and virulence varied among isolates. The absence of hypermucoviscosity in most ST258 isolates was predicted to be correlated with diminished virulence in animal models. The acquisition of virulence traits in ST258 was referred to be primarily plasmid-mediated. However, plasmid incompatibilities and mutations in recipient strains could restrict the full expression of hypervirulence phenotypes. Despite the potential clinical risks, these convergent isolates largely demonstrate diminished virulence, which suggests continued surveillance and functional studies.

The other study investigated the convergence of resistance and virulence in *K. pne* ST258 in the United States and Europe showed that the convergence was driven by adaptive evolution involving mutations in capsule biosynthesis genes (*wzc, wbaP*) (Ernst et al., 2020). These mutations resulted in hypercapsule-producing strains, enhancing phagocytosis resistance, dissemination, and virulence in bloodstream infections. These mutations in capsule-deficient strains could also facilitate biofilm formation and epithelial invasion in urinary tract infections. While hyper-capsule production significantly increased lethality in animal models, capsule-deficient mutants exhibited persistence and drug tolerance. Notably, these adaptations arose independently across strains, highlighting their role in ST258's pathogenic success and its ability to adapt to diverse host environments. This dual evolution of hypervirulence and persistence underscores the complexity of ST258 as a high-risk clone.

1.4.2 Convergence of resistance and virulence in *K. var*

The convergence of resistance and virulence in *Klebsiella* species is a critical evolutionary development with important clinical implications. While traditionally associated with *K. pne*, emerging evidence highlights this convergence in *K. var*, a species originally linked to plant and environmental niches but increasingly recognised as a human pathogen (Holt et al., 2015; Rosenblueth et al., 2004). This section presents key case studies documenting the convergence of resistance and virulence in *K. var*.

An important study describing the convergence in *K. var* showed a clonal outbreak of *K. var* in a neonatal intensive care unit in Bangladesh, associated with high mortality (54.5%) (Farzana et al., 2019). The outbreak spanned October 2016 to January 2017, during which the *K. var* strain acquired an MDR plasmid harbouring $bla_{\text{NDM-1}}$ and $bla_{\text{CTX-M-15}}$ resistance genes, likely driven by high ceftriaxone and amikacin usage in the hospital. These genes were present alongside 21 resistance determinants, including *aph(3), aadA1*, and $_{\text{aac(6)}}$, leading to extensive antibiotic resistance. WGS confirmed a unique capsular locus and virulence factors such as *ent* (iron acquisition), *kfu* (iron uptake), *fim* and *mrk* (adhesins), and *traT* (complement resistance), though the strain lacked hypermucoid markers like *rmpA*. The mortality rate was higher in neonates infected with $bla_{\text{NDM-1}}$-negative strains, correlating with increased virulence observed in Galleria mellonella models. The findings highlight *K. var* virulence and adaptability, raising concerns about the spread of MDR pathogens in resource-limited settings with suboptimal infection control and empirical antibiotic usage.

A study characterised a colistin-resistant, hypervirulent *K. var* strain (WCHKV030666) isolated from the blood of a cholangitis patient in China (Lu et al., 2018). The strain harboured key hypervirulence genes, including *rmpA* and *rmpA2* (mucoid phenotype regulators), siderophore systems (*iucABCD, iutA, iroBCDN, ybtAEPQSTUX*), and additional metabolic, and toxin-antitoxin (*ccdA, ccdB*) genes on a large virulence plasmid (pVir_030666). This plasmid, resembling but more enriched than pLVPK, was non-self-transmissible due to truncations in conjugation-related genes. The strain exhibited colistin resistance and carried other resistance determinants, including $bla_{\text{LEN-16}}$, *oqxAB*, and *fosA*. Phenotypically mucoid and demonstrated hypervirulence in a Galleria mellonella model, with 0% larval survival within 24 hours.

The study in Chile identified a hypermucoviscous *K. var* isolate, UCO-495, with significant virulence and MDR traits (Morales-León et al., 2021). Confirmed through whole-genome sequencing and Galleria mellonella virulence assays, this *K. var* strain showed resistance to broad-spectrum cephalosporins and colistin, driven by $bla_{\text{LEN-25}}$, $bla_{\text{SHV-12}}$, and $bla_{\text{TEM-18}}$ genes. Its virulence profile included the aerobactin (*iutA*), fimbrial genes (*fimABCDFEGH* and *mrkABCDF*), and biofilm-associated genes (*treC* and *sugE*), contributing to its persistence and higher virulence compared to *K. pne* UCO-494. UCO-495 also demonstrated medium biofilm formation, resistance to human serum bactericidal activity, and was classified

as a novel sequence type. The hypervirulence phenotype of *K. var* appears to result from a combination of siderophore systems, biofilm formation, and resistance mutations, posing significant public health challenges and necessitating further molecular and epidemiological investigations.

In summary, although originally identified in plants, *K. var* is increasingly recognised as a significant human pathogen, which showed convergence of resistance and virulence similar to *K. pne*. Previous reports highlighted its potential severity, with strains harbouring MDR determinants such as bla_{NDM-1} and $bla_{CTX-M-15}$, alongside virulence factors like *rmpA, rmpA2*, and siderophore genes. Notable cases, including hypervirulent, colistin-resistant isolates in China and hypermucoviscous strains in Chile, demonstrate its pathogenicity for invasive infections and persistence. These findings underscore *K. var*'s potential to compete with *K. pne* in clinical impact, necessitating enhanced surveillance and intervention strategies.

The convergence of MDR and hypervirulence in *K. pne* and *K. var* represents a growing public health threat. The boundaries between MDR and hypervirulent pathotypes are blurring. Thus, enhanced surveillance, genomic diagnostics, and novel therapeutic strategies are critical for controlling the dissemination of convergent high-risk clones.

1.5 Population Structure, Epidemiology, and Evolution

1.5.1 Studies in *K. pne*

K. pne is a well-studied species due to its strong relation to human infections. Population structure and epidemiological analyses have demonstrated that *K. pne* lineages can show significant variation in resistance and virulence traits (Wyres, Lam, & Holt, 2020). The evolution of *K. pne* occurs through multiple mechanisms, including HGT, capsule variation, and the convergence of resistance and virulence (Wyres et al., 2020).

A recent paper provides a detailed discussion on the population structure, epidemiology, and evolution of *K. pne* (Figure 1.5), emphasising the interplay between its genetic diversity and its role as a significant pathogen (Wyres et al., 2020). They showed that the population of *K. pne* is highly diverse. Genomic analyses using core-genome multilocus sequence typing (cgMLST) and WGS have identified hundreds of distinct lineages, and each of them differs by approximately 0.5% in nucleotide divergence (Wyres et al., 2020). The accessory genome is dynamic, with most genes being rare across the population, reflecting ongoing HGT events. Chromosomal recombination and plasmid-mediated HGT drive genetic exchange, particularly in traits such as capsule biosynthesis and antimicrobial resistance (Wyres et al., 2020). The population is organised into well-defined clades, aligning with distinct epidemiological and clinical features. Global problem clones such as the MDR lineage CG258 dominate specific clinical settings.

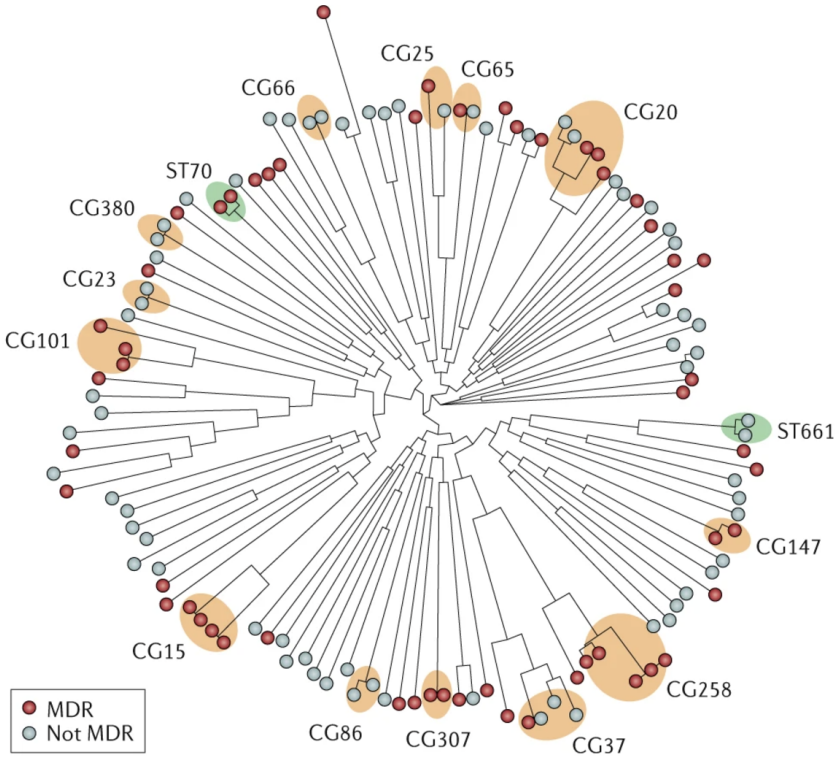


Figure 1.5: Population structure of *K. pne* genomes with tips coloured to indicate individual isolates that are multidrug resistant (Wyres et al., 2020)

.

The epidemiology of *K. pne* is closely linked to its association with human environments (Wyres et al., 2020). It is a leading cause of both healthcare-associated infections (HAIs) and community-acquired infections (CAIs). HAIs are primarily caused by MDR clones, including CG258, CG15, CG101, and CG307, which carry large plasmid-borne antimicrobial resistance determinants and are associated with severe infections such as pneumonia, urinary tract infections, and bloodstream infections (Wyres et al., 2020). On the other hand, CAIs are dominated by hypervirulent strains such as ST23, which are associated with invasive diseases like liver abscesses and endophthalmitis due to virulence plasmids encoding siderophores and capsule synthesis genes (Wyres et al., 2020). The geographical distribution of clones also shows variability, with CG258 being prevalent in the Americas and Southern Europe, while ST11, a lineage under CG258, is the predominant strain in China (Wyres et al., 2020). Surveillance data showed an increasing overlap between the traditional MDR and hypervirulent clones, especially in Asia, raising concerns about the emergence of hybrid strains.

The evolution of *K. pne* is characterised by extensive HGT, recombination, and selective pressures from antimicrobial overuse (Wyres et al., 2020). The species frequently acquire resistance genes, virulence loci, and capsule types through plasmids, integrative conjugative elements, or transposons. Specific clonal groups, such as CG23 in hypervirulent CAIs and CG258 in MDR HAIs, have evolved to specialise in particular environments. Capsule loci undergo frequent recombination, contributing to immune evasion and niche adaptation, with over 138 distinct K-loci identified (Wyres et al., 2020). Convergence of AMR and virulence has been documented in cases where MDR clones acquire virulence plasmids or hybrid plasmids carrying both resistance and virulence determinants, highlighting a critical evolutionary trajectory that could create highly virulent and resistant pathogens (Wyres et al., 2020).

Table 1.1: The proportion (%) of CR *K. pne*, hv *K. pne* and CR and hv *K. pne* in different areas (Heng et al., 2024).

Areas	**CR *K. pne***	**hv *K. pne***	**CR and hv *K. pne***
Latin America and Caribbean	68.97	0.82	0
Africa	33.09	1.85	1.15
Asia-Pacific (China excluded)	51.15	14.58	10.22
China	65.72	38.39	28.01
North America (mainly USA)	50.74	1.57	1.05
Europe and Central Asia	50.60	5.87	4.93
Not recorded	41.41	24.30	22.88
Overall	52.21	11.99	12.98

A comprehensive genomic study of 16,797 *K. pne* strains from over 100 countries, spanning 1937 to 2021, provided critical insights into the pathogen's global population structure and evolution (Heng et al., 2024). Using Uniform Manifold Approximation and Projection (UMAP) to analyse core genome distances, the study identified 60 high-density genetic clusters representing major epidemic strains, with diverse geographic origins, isolation years, and gene profiles (Heng et al., 2024). This study traced evolutionary routes and documented the global convergence of CR and hvKp phenotypes, with CR strains comprising 52.21% of the dataset and a notable rise in CR and hvKp strains, particularly in China but also increasingly in other regions (Heng et al., 2024). Key genetic markers such as *acrR* and *ompK35* were linked to resistance and virulence, offering potential targets for new typing methods to distinguish high- and low-risk strains (Heng et al., 2024). Temporal and geographic analyses revealed the historical and ongoing dissemination patterns of CR and CR-hypervirulent strains, highlighting their stabilisation as major public health threats. This study compiled the current largest *K. pne* dataset and underscored the urgent need to address the convergence of resistance and hypervirulence in *K. pne* (Heng et al., 2024).

In summary, *K. pne* exhibits extensive genomic diversity, with hundreds of lineages shaped by HGT, chromosomal recombination, and selective pressures. Distinct clades, such as hypervirulent CG23 and MDR CG258, dominate clinical settings, driving both HAI and CAIs (Wyres et al., 2020). Increasing overlap between MDR and hypervirulent clones, especially in Asia, raises concerns about hybrid strains. Recent analysis of 16,797 strains from 1937–2021 underscores the global convergence of CR and hvKp phenotypes, posing major public health risks and emphasising the need for enhanced genomic surveillance (Heng et al., 2024).

1.5.2 Studies in *K. var*

Limited knowledge are known about the population structure and epidemiology of *K. var*. However, two key studies have significantly contributed to understanding its population structure and molecular epidemiology. This section reviews these studies to summarise the current knowledge about *K. var*.

A population structure study provided comprehensive insights into the genomic and phenotypic traits of *K. var*, emphasising its distinctiveness within the *Klebsiella* genus (Potter et al., 2018) (Figure 1.6). Phylogenomic analysis of 145 *K. var* genomes revealed two distinct lineages which were characterised by minimal recombination and a diverse, star-like population structure. This showed *K. var*'s broad distribution across geographic regions and host environments (Potter et al., 2018).

The study highlighted *K. var*'s antimicrobial resistance profile, which showed near-universal conservation of the β-lactamase gene bla_{LEN} (Potter et al., 2018). Although carbapenemase genes such as bla_{KPC} and bla_{NDM} were rare, they were globally distributed across clinical and environmental isolates. Efflux pump genes like *oqxAB* were prevalent as is shown in *K. pne*. Overall, *K. var* clinical isolates exhibited greater antibiotic susceptibility compared to *K. pne*, which showed significant differences in resistance profiles (Potter et al., 2018).

In terms of virulence factors, a murine UTI model showed strain-specific uropathogenicity. Certain *K. var* strain, such as WUSM_KV_39, showed higher bladder bacterial loads than the uropathogenic *K. pne* TOP52. Type 1 pili, regulated by the *fim* operon, were critical for bladder colonisation, while mutations in *fimD* influenced it's UTI pathogenic potential (Potter et al., 2018).

Another previous study investigated the molecular epidemiology of *K. var* (Barrios-Camacho et al., 2019) (Figure 1.7). Although a universal MLST scheme for all *Klebsiella* strains would be ideal, the existing scheme did not adequately resolve *K. var* diversity, prompting the development of a species-specific MLST scheme. Using this *K. var*-specific scheme, the study analysed 254 genomes, including 226 publicly available sequences and 28 newly isolated

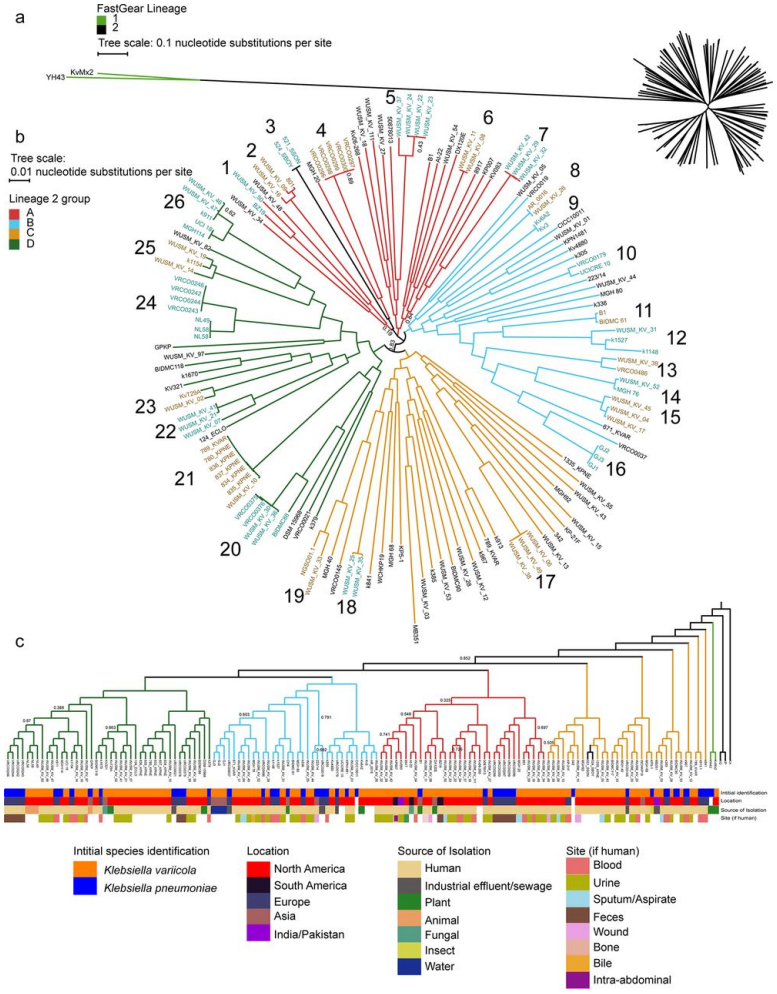


Figure 1.6: Population structure of *K. var* genomes. **a**: Core genome phylogeny of 145 *K. var* genomes. **b**: Recombination-free parSNP tree of lineage 2 genomes. **c**: Relevant metadata for initial identification, geographic location, source of isolation, and body site are adjacent to the assembly names (Potter et al., 2018)

.

strains. The findings revealed remarkable genomic diversity and identified 166 distinct STs, 39 of which were shared among multiple isolates. Clonal Complex 1 (CC1) was globally distributed and predominantly associated with human infections, with ST10 as the founder ST in CC1.

Phylogenetic analyses showed no strict boundaries between isolates from humans, plants, and other sources, indicating a shared genetic pool across diverse environments (Barrios-Camacho et al., 2019). Antimicrobial resistance was a significant concern, as several isolates carried ESBLs and carbapenemases, including $bla_{\text{KPC-2}}$, $bla_{\text{NDM-1}}$, and $bla_{\text{OXA-181}}$, highlighting the clinical threat posed by multidrug-resistant *K. var* (Barrios-Camacho et al., 2019). The study also raised the probability of *K. var* to undergo cross-kingdom transmission, particularly from plants to humans (Barrios-Camacho et al., 2019).

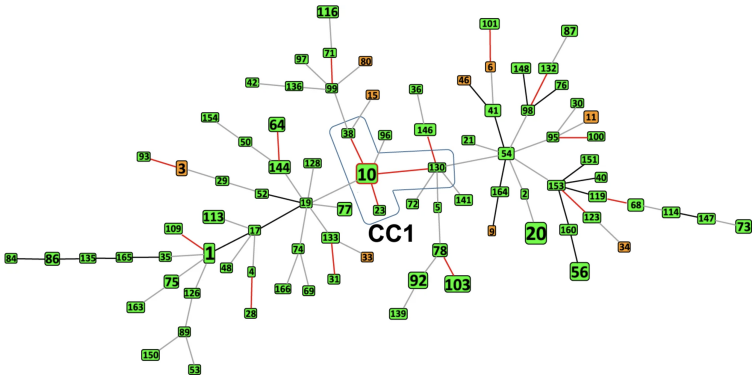


Figure 1.7: Minimum Spanning Tree of 254 *K. var* genomes. The founder ST10 and ST38, ST23 and ST130 correspond to Clonal Complex 1. The isolates from humans and plants are coloured in green and orange, respectively (Barrios-Camacho et al., 2019).

.

Clinically, *K. var* was found in severe infections, including bloodstream infections, and showed increasing virulence (Barrios-Camacho et al., 2019). Beyond clinical settings, *K. var* was also identified in various sources, which showed its various ecological adaptability. This paper underscores the importance of specialised MLST schemes for studying *K. var* epidemiology to track the potential pandemic clones (Barrios-Camacho et al., 2019).

In summary, although these studies provided unprecedented insights into the population structure and epidemiology of *K. var*, the datasets used are not comparable in scale to those available for *K. pne*. This limitation could affect the findings, which may change as the *K. var* genome database expands in terms of its' diversity. Currently, a comprehensive study on *K. var* is lacking. Additionally, the origin of *K. var* remains unknown, as no evolutionary analyses have been conducted to trace its evolutionary trajectory—an urgent gap that needs to be addressed.

1.6 Knowledge Gap and Summary of Findings

Despite significant progress, many knowledge gaps persist in the study of *K. var* and *K. pne* ST11 epidemic lineage. For *K. var*, current genomic analyses need to be re-visited as the dataset continues to expand with advancements, such as updates to the MALDI-TOF system. Unlike *K.pne*, the key aspects of this species, including its origin, evolutionary trajectory, and underlying genetic basis, resistance and virulence profile on a border dataset remain undefined and require further investigation.

For *K. pne* ST11, while extensive research has focused on strains from China, this lineage is a globally distributed epidemic, and greater efforts are needed to include and study strains from outside China. Additionally, comparative analyses between ST11, ST15, and ST258 remain limited. The genetic factors underlying ST11's greater success compared to ST15 and ST258, as well as the specific factors contributing to its dominance in China, remain largely unexplored.

This thesis explores the genomics of *K. var* and *K. pne* while answering these questions. Chapters 3 and 4 focus on *K. var*, while Chapter 5 examines *K. pne*. Beyond the introduction and methodology chapters, the final chapter (Chapter 6) summarises the findings and provides a blueprint for the future of this field.

In Chapter 3, we conducted a large-scale genomic analysis on a global multisource dataset of *K. var*, uncovering the genetic basis of its niche adaptation. A key distinction of our study from previous literature lies in the use of the largest dataset to date, enabling us to comprehensively build the population structure, resistance and virulence profiles, and evolutionary trajectory of *K. var*. The main differences between our work and prior studies are summarised in Table 1.2. From our study, we highlight the following key findings: (1) the remarkable diversity and plasticity of *K. var*, (2) its adaptation driven by the attrition of cold-shock genes and the acquisition of resistance genes, (3) horizontal gene transfer from *K. pne* to *K. var*, and (4) distinct differences in capsule profiles between *K. var* and *K. pne*.

Table 1.2: Comparison of previous and our work in terms of *Klebsiella variicola*.

K. var	Previous Work	Our Work
No. of Genomes Used for Core Genome Phylogeny	145 (Potter et al., 2018)	715
No. of Genomes Used for MLST	226 (Barrios-Camacho et al., 2019)	715
No. of Genomes Used for Resistance Profile	574 (Osei Sekyere & Reta, 2021)	715
No. of Genomes Used for Virulence Profile	145 (Potter et al., 2018)	715
No. of Genomes Used for GWAS	NA	689
Ancestral State Reconstruction	NA	Non-Human
Transition Rate Between Sources	NA	0.3545
Adaptation	NA	AMR
No. of Novel Capsule Loci	1 (Farzana et al., 2019)	28

In Chapter 4, we conducted a local case study on Scottish *K. var*. This chapter highlighted the following findings: (1) the diversity and plasticity of *K. var*, demonstrated by the absence of genetic differences between strains causing different diseases or collected from different cities; (2) the effectiveness of resistance control in Scotland, reflected by the lower carriage of AMR genes; and (3) the association between the acquisition of bla_{SHV} and the capsule locus KL123.

In Chapter 5, we analysed the currently available genomes of ST11, ST15, and ST258 *K. pne*, with a particular focus on ST11. The main differences between our work and prior studies are summarised in Table 1.3. Our key findings include: (1) ST11 is a bigger threat to public health than ST15 and ST258 and is the dominant clone in China, (2) ST11 originated outside of Asia, (3) upon transmission to Asia, KL47 ST11 evolved into KL64 ST11 under antibiotic pressure, and (4) the evolution of ST11 was driven by key genes, including resistance genes and iron acquisition genes.

Table 1.3: Comparison of previous and our work in terms of *Klebsiella pneumoniae* ST11.

K. pne ST11	Previous Work	Our Work
Hierarchical Clustering	NA	1157
No. of Genomes Used for Evolutionary Analyses	794 (K. Zhou et al., 2023)	1157
No. of Chinese Genomes	523 (F. Hu et al., 2024)	621
The Evolution of ST11	730 (J. Wang, Feng, & Zong, 2023)	1157
No. of Genomes used for GWAS	393 (Pei et al., 2022)	1157
Comparative Analysis with ST15 and ST258	NA	1157, 848 and 1193

Chapter 2

Material and Methods

2.1 Genomic DNA purification

Genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA). Briefly, 1 ml of overnight bacterial culture was pelleted, followed by protein precipitation and DNA purification using isopropanol and ethanol. The quality of the extracted DNA was then assessed using Invitrogen Qubit Fluorometers (Thermo Fisher Scientific, Waltham, MA, USA). The DNA samples were then dissolved in 50 μ of distilled water and sent to Novogene (Novogene Europe, Cambridge, UK) for short-read sequencing.

2.2 Genome Assembly

2.2.1 Quality Assessment of Sequencing Reads

The integrity and quality of the short-read sequencing data were assessed using FastQC (Andrews et al., 2010), a widely recognized tool for quality control of sequence reads. The criteria for deeming a fastq file as acceptable were defined as follows: (1) The lower quartile of per-base sequence quality scores for all bases must exceed 5, and the median score should exceed 20, ensuring high-quality reads throughout. (2) The most frequently observed mean quality score should present a singular peak, indicating high quality across the majority of reads. (3) A balanced nucleotide composition is required, with the bias between Adenine (A) and Thymine (T), or Guanine (G) and Cytosine (C), remaining within 20% at any position. (4) The GC content at each base must be similar to the expected genomic profile, deviating no more than 10% from the average of *Klebsiella*, which is 57.56% (M. Chen et al., 2016). (5) The GC content distribution across each sequence should present a single peak, and the sum of deviations from a normal distribution should not exceed 30%. (6) The proportion of undetermined nucleotide bases (N) at any position should remain below 20%. (7) Sequence length uniformity is essential, with all sequence lengths exceeding 1 and exhibiting a similar distribution. (8) Duplicated sequences should not exceed 50% of the total reads. (9) Overrepresented sequences should comprise no more than 1% of the total reads.

Once these quality checks are successfully passed, the read data are considered suitable for proceeding to the next phase of analysis, which is read trimming. If the quality criteria are not met, the genomic DNA sample is re-extracted and re-sequenced to ensure the acceptable quality of sequencing data.

2.2.2 Trimming the Reads

According to 'Per Base Sequence Quality', calls of poor quality were trimmed using Trimmomatic (Version 0.39) (Bolger, Lohse, & Usadel, 2014). When applying Trimmomatic, we adhered to the following rules to ensure having high-quality reads: (1) Removed bases from the start if below a quality score of 20 (LEADING:20). (2) Removed bases from the end if below a quality score of 20 (TRAILING:20). (3) Scanned the read with a 4-base sliding window, cutting when the average quality per base dropped below 15 (SLIDINGWINDOW:4:15). (4) Dropped reads shorter than 36 bases (MINLEN:36).

After trimming the reads, we applied FastQC again to ensure the trimmed reads still met the standards listed in Section 2.2.1 (Andrews et al., 2010).

2.2.3 Genome Assembly

We applied SPAdes (Bankevich et al., 2012) for genome assembly, using single paired-end libraries in gzipped format after the trimmed reads passed the quality check (Andrews et al., 2010).

2.2.4 Quality Assessment for Genome Assemblies

To assess the quality of genome assemblies, we applied Quality Assessment Tool for Genome Assemblies (QUAST) (Gurevich, Saveliev, Vyahhi, & Tesler, 2013). We evaluated key metrics such as N50, genome size, GC content, and the number of contigs to check the concordance with the reference genome of *K. var* (GCF_009648975.1) to ensure that the genomic data used for downstream analyses met high-quality standards.

2.3 Building Databases

2.3.1 *Klebsiella variicola* database

We utilized a total of 882 *K. var* genome sequences. This collection includes 584 strains accessible via the National Center for Biotechnology Information (NCBI) in September 2022, 279 strains provided by Edward Feil (Thorpe et al., 2022), 3 unpublished genomes from Refath Farzana's study in 2019 except from those published on NCBI (Farzana et al., 2019), 3 strains

from Sands et al.'s study (Sands et al., 2021), and 13 strains sequenced using Illumina short-read sequencing provided by Paul Higgins. The corresponding metadata, including ecological host/habitate, geographic location, and collection date, was sourced from the NCBI BioSample database. Geographic locations were classified into continent, and collection dates were categorized by year. Additionally, to complement our *K. var* data, we built a representative dataset of *K. pne* for comparative analysis. This was achieved by randomly selecting 1,000 genome assemblies from the publicly available *K. pne* assemblies on the NCBI database.

2.3.2 *Klebsiella pneumoniae* ST11, ST15 and ST258 database

All publicly available *Klebsiella pneumoniae* (Kp) genomic data were systematically retrieved from the NCBI RefSeq database as of May 5, 2023 using Batch Entrez (RRID:SCR_016634) which include a total of 16,917 *K. pne* assemblies. To identify the sequence type (ST) of each strain, Kleborate (M. Lam et al., 2021) was employed, filtering restricted to strains classified as ST11, ST15, and ST258. This refinement resulted in a dataset comprising 2,470 ST11, 920 ST15, and 1,446 ST258 strains. Metadata attributes, including strain, collection date, geographic location, host, host disease, isolation source, and environmental context, were extracted from the corresponding BioSample entries on NCBI using the XML2 package (Wickham, Hester, & Ooms, 2018). These attributes were then systematically organized into a structured database with columns titled 'Strain', 'Collection Date', 'Geographic Location', 'Source', and 'Host Disease'. The 'Collection Date' was categorized by year, and 'Geographic Location' was grouped by continent. The 'Source' metadata includes categories of Human, Animal, Plant, Environment, and Missing, while 'Host Disease' included classifications like Bloodstream Infection (BSI), Urinary Tract Infection (UTI), Respiratory System Infection (RSI), Gastro-Intestinal Tract Infection (GITI), Other Diseases, along with a category for Missing data.

Next, strains lacking metadata information were excluded from our database. This filtration resulted in a refined dataset comprising 2,350 ST11, 852 ST15, and 1,221 ST258 strains. The dataset specific to ST11 needed further scrutiny due to its considerable size, which exceeded the computational capacities of our current resource, EDDIE. Consequently, ST11 strains lacking disease annotations and categorized within the 'Missing' group, were additionally excluded. This step resulted in a more manageable subset of 1,157 ST11 strains. In the final stage of database optimization, duplicate assemblies were meticulously identified and removed. The resultant *K. pne* database thus comprised 1,157 ST11 strains, 848 ST15 strains, and 1,193 ST258 strains.

2.3.3 Scotland *Klebsiella variicola* database

We collaborated with hospitals in Edinburgh, Glasgow, Dumfries, and Galloway, collecting *K. var* samples from January 2023 to January 2024. Genomic DNA was purified using the methods described in Section 2.1.

We received 129 clinical samples from hospitals in Edinburgh, Glasgow and Dumfries and Galloway, where there are 3 no-growth samples, 5 mixed samples, 1 *Klebsiella pneumoniae* strain, and 1 *Klebsiella quasipneumoniae* strain. A total of 119 samples were sent to Novogene for Microbial Whole Genome Sequencing, and genome assembly was performed as described in Section 2.2.

2.4 Metadata Statistics Tests

To test if the metadata of *K. var* were independent of each other, we conducted a Generalized Linear Model (GLM) analysis in R, specifically focusing on source and continent annotations. Due to the bias in our dataset, where we have unequal isolates from each source and continent, the source category, which originally contained human, animal, plant, and environment, was simplified into 'human' and 'non-human'. Similarly, the continent category, which originally included Africa, Asia, Europe, North America, and South America, was simplified into 'Europe' and 'rest of the world'. We applied the GLM to both the simplified and the original metadata sets, and checked whether their correlation was significantly different from zero.

2.5 Annotation

All *Klebsiella* genomic assemblies were systematically annotated using PROKKA (Version 1.14.6) (Seemann, 2014), predominantly with default parameters, except for setting the genus as *Klebsiella*. This ensured a genus-specific approach to annotation and a more accurate annotation output. The comprehensive annotation process was conducted on the EDDIE computational platform.

2.6 Population structure

2.6.1 Core genome phylogeny

For the Single Nucleotide Polymorphism (SNP)-based core genome phylogeny, we applied ParSNP (Treangen, Ondov, Koren, & Phillippy, 2014) for recombination-free phylogeny construction. The recombination-free core genome alignment from ParSNP was converted into fasta format using Harvesttools (Treangen et al., 2014), and Seqret (Rice, Rice, Bleasby, & Ison, 2011) was used to generate the Nexus (.nex) alignment file.

The corresponding metadata including source, continent, host disease, resistance and virulence profile and hierarchical groups are visualised on the core genome tree using iTol (Letunic & Bork, 2007)

2.6.2 Multi-Locus Sequence Typing (MLST)

The Multi Locus Sequence Typing (MLST) profile was generated using either Pasteur scheme (Diancourt, Passet, Verhoef, Grimont, & Brisse, 2005) or *K. var*-specific MLST scheme (Barrios-Camacho et al., 2019). PHYLOViZ 2.0 (Francisco et al., 2012) was then applied for goeBURST analysis and visualization of Minimum Spanning Tree (MST). To assess the convergence between the core genome phylogeny and the MLST, we calculated Baker's Gamma Index (Baker, 1974) using R.

2.6.3 Accessory genome phylogeny

The gene presence and absence-based Neighbour-Joining tree of the accessory genome was constructed using FastTree (Price, Dehal, & Arkin, 2010). The source and continent metadata were visualized on the accessory genome tree using iTOL (Letunic & Bork, 2007).

2.6.4 Hierarchical clustering

We applied RhierBAPS, which utilizes hierarchical clustering, to group the sub-lineages from the core genome tree (Tonkin-Hill, Lees, Bentley, Frost, & Corander, 2018). The hierarchical groups were visualized on the core genome tree using iTOL (Letunic & Bork, 2007). The SNP distances between different groups of strains are calculated by snp-dists (Seemann, 2018)

2.6.5 SNP distance

To calculate the SNP distances between strains, we calculated the pairwise SNP distance from core genome alignment using Seemann (2018). The strains were then clustered based on specific phenotypes, such as geographic location, source or disease, to calculate the average SNP distance within or between groups.

2.7 Pan-genome analysis

For building the pan-genome of *K. var*, we applied Roary (Page et al., 2015) using the Generic Feature Format (GFF) files annotated by PROKKA (Version 1.14.6) (Seemann, 2014). We opted to prevent paralog splitting (-s) and increased the cluster number setting (-g) to 60,000 if needed. This adjustment was necessary to accommodate our dataset, which exceeded

Roary's default cluster number limit of 50,000. The accessory genes binary profile was used to build the NJ accessory genome tree in Section 2.6.3. The gene presence and absence table was used for GWAS analysis in Section 2.11. Protein sequences from the required ortholog groups were extracted using the 'query_pan_genome' function of Roary.

2.7.1 Type strain

We used the gene presence and absence matrix generated by Roary (Page et al., 2015) to calculate Euclidean pairwise distances using the SciPy library (Virtanen et al., 2020) within Python. These pairwise distances for the *K. var* dataset served as the basis for Multi-Dimensional Scaling (MDS) analysis applied in R using the `cmdscale` function (J. Wang & Wang, 2012). Within this MDS framework, the type *K. var* isolate is defined as the one nearest to the centroid of all *K. var* isolates (Dorado-Morales, Vilanova, P. Garay, Martí, & Porcar, 2015).

2.7.2 Average Nucleotide Identity

The Average Nucleotide Identity (ANI) was determined using FastANI to analyze the genomic similarities within *Klebsiella* strains (Jain, Rodriguez-R, Phillippy, Konstantinidis, & Aluru, 2018). To calculate ANI, we utilized the NCBI reference genomes for *Klebsiella pneumoniae* (GCF_000240185.1) and *Klebsiella variicola* (GCF_009648975.1) to ensure the reliability of our ANI calculations. Our focus was on comparing ANI values between *K. var* strains with and without the bla_{SHV} gene. To reduce the imbalance in the dataset, we employed random over- and under-sampling for strains with and without the bla_{SHV} gene in our comparison. To reduce the influence of random variation in this method, we repeated the sampling process 10 times. Next, the Welch unequal variance *t* test was applied to assess the statistical significance of the ANI differences observed in these re-sampled groups.

2.8 Evolutionary analysis

2.8.1 Ancestral state tracing

In source ancestral state reconstruction in *K. var*, we included only those 594 *K. var* strains with source annotation. We grouped all strains from animal, plant, and environmental origins under the 'non-human' category to reduce the over-representation of strains from human sources. We utilized BayesTraits V3.0.5 (Pagel & Meade, 2013) to trace ancestral states, employing the Maximum Likelihood model in two separate analyses.

The initial model did not incorporate any fitted parameters, e.g. κ, λ, or δ. Conversely, the second model incorporated the parameter κ to assess whether trait evolution is associated with the phylogeny. The log-likelihood for the first model is -839.314049, while for the second model it is -415.12330. We performed a χ^2 test on the Likelihood Ratio (LR) to determine if including κ significantly improved the model fit, using a degree of freedom (df) of 1 for a one-sided test:

$$LR = 2 \times \big((-415.12330) - (-839.314049)\big) = 848.381492 \gg 3.841 \tag{2.1}$$

This LR value is significantly higher than the critical value of 3.841, indicating that the second model's likelihood is significantly greater. Therefore, the κ parameter offers a more accurate representation of our data, and the estimation of κ is robust. Thus, we chose the κ-inclusive model for subsequent analyses.

2.8.2 Time-measured phylogeny

To build a time-measured phylogeny in *K. var*, we established a smaller dataset by randomly selecting 45 strains, which are collectively represented across every continent, source, and year of collection. This dataset was designed to comprehensively represent the diversity of our *K. var* strain collection.

In the *K. var* project, we employed BEAST (V2.6.7) (Suchard et al., 2018) for time-measured phylogenetic analysis on this subset. Given that the clustering of *K. var* collection dates largely falls within the recent five years, we chose the Piqmee package (Boskova & Stadler, 2020). We used the default Hasegawa–Kishino–Yano model for nucleotide evolution and implemented a strict molecular clock model. The model parameters, including posterior, likelihood, and prior, stabilized after 1,000,000 iterations. Nevertheless, we continued the analysis up to 93,665,000 states to confirm the model's optimal fit. To avoid the risk of local optima, we replicated the analysis three times, and the parameters for the other two runs stabilized after 1,000,000 iterations as well.

Next, we consolidated the log and tree files using LogCombiner (Drummond & Rambaut, 2007). Tracer (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) was applied to examine the log file parameters. The maximum clade credibility tree was constructed using TreeAnnotator (Drummond & Rambaut, 2007), and finally, FigTree (Rambaut, 2023) was employed for the visualization and annotation of the time-measured phylogenetic tree.

In the *K. pne* project, we utilized BactDating (Didelot, Croucher, Bentley, Harris, & Wilson, 2018a) to analyze the evolutionary timelines of the strains. Due to computational constraints, we randomly sampled a subset of the *K. pne* ST11 database, selecting 200 strains from the Asian cluster and 200 strains from the other clusters. We constructed the core genome phylogeny of this subset using ParSNP (Treangen et al., 2014). BactDating was then employed for Bayesian dating of the phylogenetic nodes and to predict the root location (Didelot et al., 2018a).

2.8.3 Recombination prediction

For recombination analysis, we firstly constructed a core-genome alignment with ParSNP (Treangen et al., 2014), specifically opting not to exclude recombination events. This alignment was then converted into FASTA format using Harvesttools (Treangen et al., 2014). Subsequently, we employed Gubbins (Croucher et al., 2015) to analyze the alignment, utilizing the NCBI reference genomes for *Klebsiella pneumoniae* (*K. pne*, GCF_000240185.1) and *Klebsiella variicola* (*K. var*, GCF_009648975.1) to identify recombinant sequences. We quantified the recombination rate of *K. var* as previously described for *K. pne* using the formula (Wyres et al., 2019):

$$\mathscr{R} = \frac{L_r}{L_g} \tag{2.2}$$

where $\mathscr{R}$ denotes recombination rate; L_r denotes the total length of all recombination events predicted; L_g denotes genome length. This approach allowed us to calculate the proportion of the genome predicted to be recombination.

2.9 Resistance profile

To build the AMR carriage profiles for each strain, we either used AMRFinderPlus (Version 2.1.39) (Feldgarden et al., 2021) or Abricate (Seemann, 2019) and then merged the columns representing allelic variations in AMR genes for visualization. These profiles were then converted into a binary format indicating gene presence or absence and integrated with the core genome phylogeny using iTol (Letunic & Bork, 2007) for visualisation. The diversity within the AMR profiles was quantified and graphically represented through the iNEXT package (Hsieh, Ma, & Chao, 2016) in R. To check the impact of variations in AMR sequences, we utilised PolyPhen-2 (Adzhubei et al., 2010) to determine whether these genetic alterations in the AMR

genes are likely to be neutral or deleterious. We utilized PlasmidFinder (V2.0) (Carattoli et al., 2014) to analyze the plasmid carriage profiles within each *K. var* isolate. To investigate the association between all plasmids and their corresponding AMR carriage, we applied Welch unequal variance *t* tests.

2.9.1 Statistical tests of resistance genes

We recorded all variations in amino acid sequences associated with intrinsic AMR genes. The occurrence of these variations was then statistically compared against source classifications using the Fisher's exact test to determine any significant associations in R (Team et al., 2013). Similarly, for extrinsic AMRs, Fisher's exact tests were employed to evaluate the relationship between the presence or absence of these genes and the 'human' or 'non-human' source classifications.

2.9.2 Genomic location of resistance genes

The genomic location of the genes in pan-genome was determined based on the proportion of core ($\geq$ 95%) and accessory ($\leq$ 95%) genes present on the same contig. If a contig predominantly consists of core genes, we infer that the resistance gene is located on the core genome. Conversely, if most genes on the contig are accessory genes, the resistance gene is inferred to be part of the accessory genome.

2.10 Virulence profile

2.10.1 Capsule locus

The capsule locus for each strain was characterized using Kaptive (V3.0) (Stanton, Hetland, Löhr, Holt, & Wyres, 2025). Loci that did not match any known types in the Kaptive database and were identified as 'None' in the output were inferred as novel. The visualization of these capsule loci was done with BioRender.com. Additionally, we extracted the genes in the capsule loci (between the flanking genes *galF* and *gnd*) of KL34 variant *K. var* strains and annotated them using Pannzer2 (V2.0) (Törönen, Medlar, & Holm, 2018) to find potential novel genes.

To ensure that the capsule profiles obtained from short read data were not artefacts of misassembly or GC-content bias, we mapped capsule genes back to their genome assemblies and confirmed that none were located within 500 bp of contig edges, where assembly artefacts are likely to occur.

2.10.2 Siderophores and genotoxins

We determined the presence of siderophores and genotoxins by local BLASTP searches against each genome assembly (Altschul, Gish, Miller, Myers, & Lipman, 1990), with a threshold of 95% for identification. The siderophore profiles, including gene presence or absence, were mapped onto the core genome phylogeny and visualized using iTol (Letunic & Bork, 2007).

2.11 Other specific genes

Other specific genes of interest were mapped using BLASTP (Altschul et al., 1990) and extracted from the pan-genome built from Roary (Page et al., 2015). The alignments were built using MAFFT (Katoh & Standley, 2013) and the phylogenies were built using FastTree (Price et al., 2010).

2.12 Genome-wide association studies (GWAS)

2.12.1 Gene presence or absence based GWAS

We conducted Genome-Wide Association Studies (GWAS) to study the association between the genotype and phenotype at the pan-genome gene presence or absence level. Our approach included two GWAS methodologies: pan-genome GWAS and phylogenetic-corrected GWAS.

For the pan-genome GWAS, we applied Scoary (Brynildsrud, Bohlin, Scheffer, & Eldholm, 2016) to the Roary-generated gene presence/absence table to calculate associations between all genes in the pan-genome and the phenotype. We furthermore applied both Bonferroni correction and False Discovery Rate (FDR) (or Benjamini Hochberg) correction to the p values and used a threshold of 0.05 for adjusted p values.

In the phylogenetic-corrected GWAS, we applied treeWAS (et al., 2017) in R to assess the statistical association between the phenotype and genotype while correcting the recombination and clonal expansion effect in the core genome phylogeny. The core genome phylogeny constructed by ParSNP (Treangen et al., 2014) and the pan-genome data generated by Roary (Page et al., 2015) were used as the input for identifying significant associations. We employed both Correlation analysis and Fisher's exact test, adjusting the p values with both FDR and Bonferroni corrections. We used 0.05 as the threshold for adjusted p values.

Chapter 3

Klebsiella variicola population structure and niche adaptation

The content of this chapter corresponds to the preprint paper:

Genome Analyses of A Worldwide Multisource Collection of *Klebsiella variicola* Reveal Adaptation Relevant for Human Infection

Jiayu Wang, Claudia E. Coipan, Eelco Franz, Ed Kuiper, Els Wessels, David L. Gally, Karthick Vasudevan, Balaji Veerarhagavan, Refath Farzana, Nicola Holden, Ruth Zadoks, Paul G. Higgins, Thamarai Schneiders*, BioRxiv, 2024

Link: https://www.biorxiv.org/content/10.1101/2024.10.28.617863v1

3.1 Introduction

Klebsiella variicola (*K. var*), part of the *Klebsiella pneumoniae* species complex (KpSc), is an emerging human pathogen long associated with plants, animals, and other environmental sources. Given the broad dissemination of *Klebsiella variicola* strains, transmission events between these hosts/habitats remain plausible. Recent studies support this hypothesis, where the genetic relatedness of human-, animal-, or plant-associated *K. var* suggests this organism has the genomic plasticity to span multiple niches within the One Health framework and, like other members of the KpSc, acts as a versatile microbial generalist.

Like the more prevalent *Klebsiella pneumoniae* (*K. pne*), *K. var* is linked to a variety of human infections, including pneumonia, bloodstream infections, urinary tract infections, and neonatal sepsis (Bagley, 1985; Holt et al., 2015; Rodríguez-Medina et al., 2019). However, the overlapping biochemical profiles of *K. var* and *K. pne* have led to the frequent misidentification of *K. var*, which may contribute to lower detection and prevalence rates of this organism (Long et al., 2017). It is estimated that 2.5% to 10% of current *K. pne* isolates are actually *K. var* (Fontana, Bonura, Lyski, & Messer, 2019). Despite its lower prevalence in human clinical infections, *K. var* has been linked to higher mortality rates in the bloodstream and neonatal

infections compared to *K. pne*. The microbial factors driving these higher mortality rates in bloodstream infections remain unknown. Moreover, urinary tract isolates exhibit significant piliation variations via the *fim* operon, which correlates with differential bacterial loads and infective outcomes in *K. var*-related UTIs. This variation in pathogenicity underscores the broad lifestyle range of *K. var*, spanning commensalism to pathogenesis.

Commensal *K. var* demonstrate symbiotic behaviour across different species, supporting biomass degradation and nitrogen fixation, which, in turn, promote fungal gardens in leaf-cutter ants (Pinto-Tomás et al., 2009) and underpin endophytic behaviour in specific crops. Recent microbiome analyses of bovine vaginal and urinary samples (Giannattasio-Ferraz et al., 2022) and faecal matter of wild green stink bugs have recovered *K. var* strains, further supporting its commensal role. Like *K. pne*, *K. var* colonises the human gut but is also linked to extraneous infections, such as bloodstream infections and UTIs. Additionally, *K. var* has been associated with diseases in non-human hosts, including collar rot of ironwood trees and bovine mastitis. This wide-ranging lifestyle, from commensalism to pathogenesis, reflects its genomic plasticity and adaptation to various niches.

Existing studies (Barrios-Camacho et al., 2019; Potter et al., 2018) indicate that *K. var*'s population structure depicts a star-like phylogeny with equidistant clades in diverse environments, all emerging from a single common ancestor. Analyses using the published *K. var* Multi-Locus Sequence Type (MLST) schema (Barrios-Camacho et al., 2019) identify "founder" ST-types (e.g., ST-types 1, 10, 60), primarily based on prevalence. Single locus variants of widely disseminated human-origin founder ST-types were found to comprise Clonal Complex 1 (Barrios-Camacho et al., 2019). While the founder ST-types and associated clonal complexes indicate a lack of restriction in both niche adaptation and geographic location, these observations are based on limited datasets, possibly overestimating the significance of founder ST-types in *K. var* population expansion.

Broadly, *K. var*'s distribution across various niches and continents supports its dissemination in a One Health context, suggesting limited barriers to gene acquisition within and between environments and species. This is evidenced by the acquisition of antibiotic resistance genes such as $bla_{\text{CTX-M}}$, bla_{OXA}, bla_{NDM}, and bla_{KPC}-like enzymes commonly associated with *K. pne* (Wyres et al., 2020). Recent work by Gorrie et al. (Gorrie et al., 2022) reveals recombination of the capsule locus between *K. pne* and *K. var*, highlighting the propensity for chromosomal genetic recombination in loci not directly under antibiotic selection pressure. Despite its low prevalence, *K. var*'s genetic proximity and gene exchange potential with *K. pne* position it as a significant emerging human pathogen.

Thus, using the largest contemporaneous dataset, we analysed the *K. var* population structure and evolution in key phenotypes related to antibiotic resistance and virulence.

3.2 *Klebsiella variicola* database

After removing duplicates and assemblies with low quality, we acquired 715 *Klebsiella variicola* genome sequences and collected strain-specific metadata, sourced from NCBI BioSample metadata or relevant research publications, including collection date, geographic origin, and source (Figure 3.1). These metadata reveal the distribution of *K. var*, spanning six continents and diverse hosts/habitats, including human, animal, plant, and environmental sources, with collection dates covering over 26 years (1994–2021). The SNP distances analyses, constructed from 723 isolates, identified 715 as *Klebsiella variicola* ssp. *variicola (K. var)* with an Average Nucleotide Identity (ANI) of 99.03%, and the remaining 11 as *Klebsiella variicola* ssp. *tropicalensis* with an ANI of 97.01%.

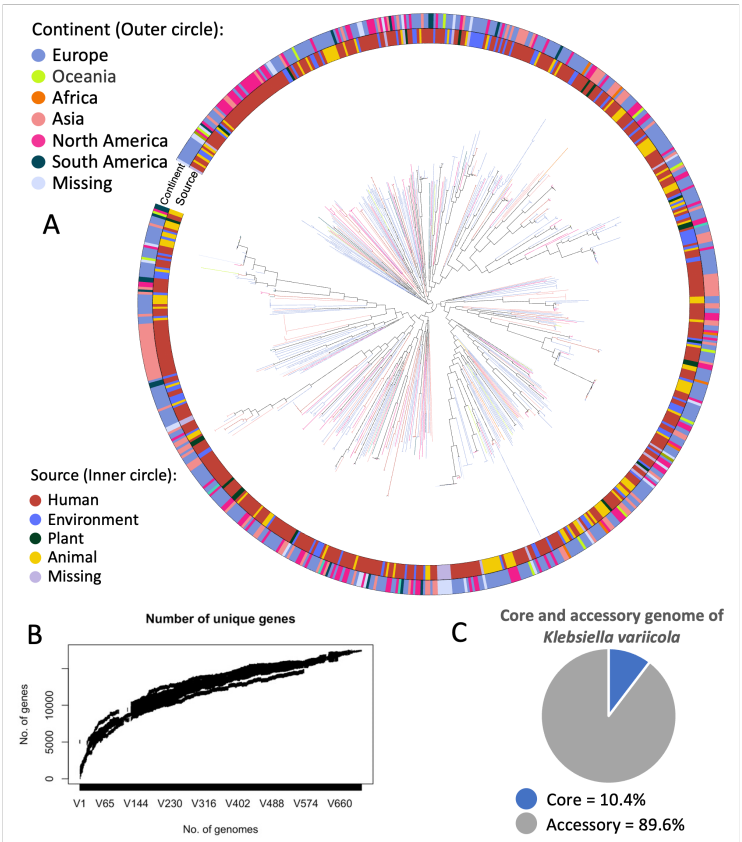


Figure 3.1: Population structure of *Klebsiella variicola*. A: Core genome phylogeny: The core genome phylogeny of 715 *Klebsiella variicola* strains. Annotations include source (inner circle) and continent (outer circle). The phylogenetic tree was constructed using ParSNP (Treangen et al., 2014), incorporating genome data and metadata from NCBI. B: The number of unique genes: A graphical representation plotting the number of uniquely identified genes within 715 *Klebsiella variicola* strains. C: Pan-genome composition: The proportion of core ($>= 95\%$ strains) and accessory genes ($< 95\%$) in 715 *Klebsiella variicola* strain's pan-genome analysis (Page et al., 2015).

We next used 94,553 SNPs identified in the core genome to reconstruct the maximum-likelihood (ML) phylogeny of *K. var* (Figure 3.1A). We presented 715 *K. var* strains as a representative subset for pan-genome analysis, as the number of unique genes (Figure 3.1B) showed convergence, indicating that this subset can sufficiently represent the genetic diversity of *K. var*. The average genome size of *K. var* is 5.71 Mb, and the total pan-genome comprises 42,231 genes, including approximately 4,373 conserved core genes (at 95% nucleotide identity) and 37,858 accessory genes (Figure **3.1**C).

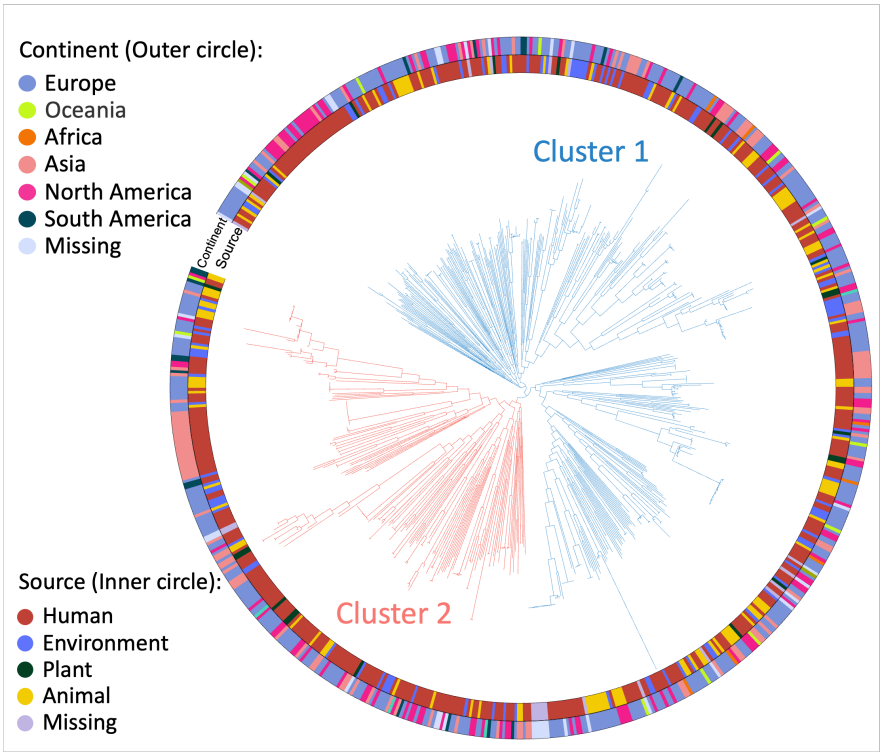


Figure 3.2: Hierarchical clustering of *Klebsiella variicola*. The hierarchical clustering was applied using RhierBAPS (Tonkin-Hill et al., 2018) and visualised on the branch colour using iTol (Letunic & Bork, 2007) The continent and source metadata are visualised on the outer and inner colour strips correspondingly.

Using hierarchical clustering, we categorised the 715 *K. var* population into two clusters. However, the average SNP distances within clusters 1 and 2 were 0.0486 and 0.0539, respectively, and the distance between the two clusters was similar to the within-cluster distances, at 0.0538. Given the similarity of these clusters, we infer that the *K. var* population should not be subdivided.

3.3 Population Structure and Evolution Dynamics

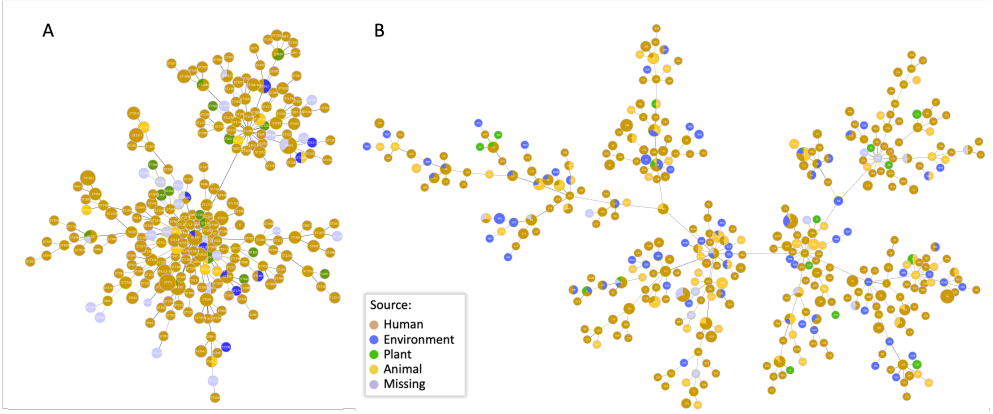


Figure 3.3: Multi-Locus Sequence Typing (MLST) of *Klebsiella variicola* A: MLST with Pasteur scheme (Diancourt et al., 2005) of 715 *Klebsiella variicola* built from Kleborate (M. Lam et al., 2021) scheme and visualised using PHYLOVIZ (Francisco et al., 2012). B: MLST with MLSTKv scheme of 715 *Klebsiella variicola* built from MLSTKv (Barrios-Camacho et al., 2019) and visualised using PHYLOVIZ (Francisco et al., 2012).

In our dataset, Multi-locus Sequence Typing (MLST) was performed using two schemes: the Pasteur scheme (Diancourt et al., 2005) and a *K. var*-specific MLST scheme (Barrios-Camacho et al., 2019). Both approaches, however, provided limited insight into the core genome phylogeny of *K. var*. The Pasteur scheme (Diancourt et al., 2005), originally designed for *K. pne*, includes the housekeeping genes *infB*, *gapA*, *pgi*, *mdh*, *phoE*, *tonB*, and *rpoB*. The concordance of *K. var*'s MLST with the core genome phylogeny (Figure A.1), indicated by Baker's Gamma (Baker, 1974), was low (0.42), particularly due to allelic variations in the *K. var infB* gene (Figure 3.3C). These variations predominantly divide the *K. var* population into two clusters, ST347 and ST2994 (founder STs), as shown in Figure 3.3A and C, thereby erroneously constraining the *K. var* Minimum Spanning Tree (MST).

The *K. var*-specific MLST scheme (Barrios-Camacho et al., 2019) encompasses *leuS*, *pgi*, *pgk*, *phoE*, *pyrG*, *rpoB*, and *fusA* genes. Applying this MLST scheme to our dataset revealed a significant shift in the distribution of founder sequence types (STs) (Figure 3.3B). In contrast to previous reports (Barrios-Camacho et al., 2019), founder STs such as ST10, ST38, ST23, and ST130 were replaced by a range of previously unassigned novel STs, including new ST21 and new ST85. Nonetheless, ST10, ST61, ST20, and ST68 remain prevalent ST-types. Importantly, over 200 strains in our study were classified into 86 novel STs, as detailed in Supplementary Table A.4. These newly identified STs currently lack official ST number assignments, suggesting that the current framework for assigning STs using the *K. var* schema MLST (*barrios2019molecular*) is limited.

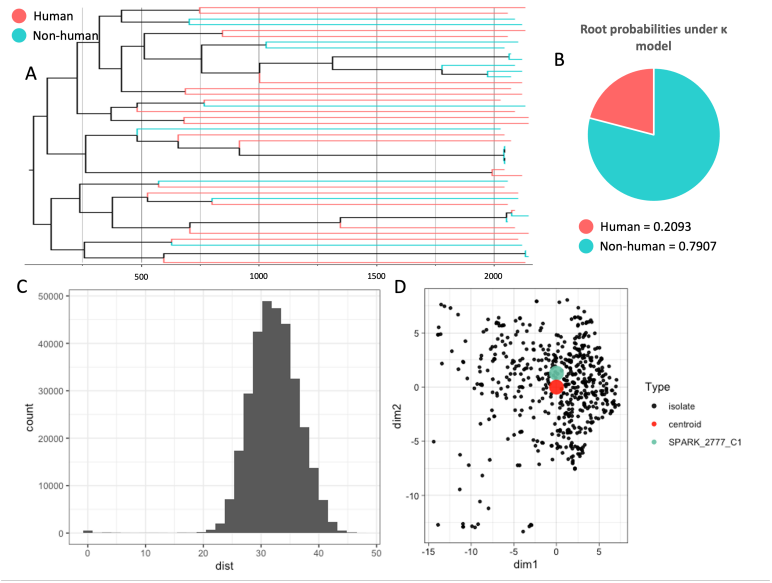


Figure 3.4: Ancestral state reconstruction and Multi-Dimensional scaling of *Klebsiella variicola*. A: Time-scaled phylogeny: A Time-scaled phylogenetic tree of a subset of *Klebsiella variicola* data that as is shown in Supplementary Table A.1. B: Ancestral state reconstruction: The inferred ancestral state (human or non-human) of the common ancestor of 715 *Klebsiella variicola*. The probabilities of these ancestral states were calculated using maximum likelihood, incorporating a model that includes the κ parameter. C: MDS distances distribution: A histogram of the distribution of Multi-Dimensional Scaling (MDS) distances. The reference point (0, 0) corresponds to the centroid of the data. D: Centroid visualization and typing: This panel visualizes the centroid point and 715 *Klebsiella variicola* isolates in two dimensions. Each dot represents an isolate, with the centroid point marked in red. The pairwise Euclidean distances of each isolate from the centroid are calculated and ranked. The 'type' *Klebsiella variicola*, defined as the strain closest to the centroid (strain SPARK_2777_C1), is highlighted in blue.

To establish the *K. var* population structure, we employed ParSNP (Treangen et al., 2014) to construct a recombination-free, SNP-based core genome phylogeny. Analysis of our 715 *K. var* genome dataset revealed that the *K. var* population forms a star-like phylogeny, potentially originating from a single common ancestor (Figure 3.1A). Unsurprisingly, this phylogeny does not exhibit distinct clustering by host/habitat or geographic location (Figure 3.1A) and suggests recent expansion into various sources, hosts, and geographic locations. To determine the timing of *K. var* expansion and transmission events, we conducted time-measured phylogenetic analyses using a strict molecular clock model in BEAST (Suchard et al., 2018), as depicted in Figure 3.4A. These analyses reveal no evidence of host- or habitat-adapted lineages, indicating that *K. var* strains have been broadly disseminated over a 1000-year timescale.

To trace the origins and dissemination of *K. var*, we utilised BayesTraits (Pagel & Meade, 2013) for ancestral state reconstruction. Root probability analysis using the kappa model suggests a non-human origin for the ancestral *K. var*, with the probability of originating from human sources (0.2093) lower than non-human (0.7907) (Figure 3.4B). Supporting this observation, we identified SPARK_2777_C1, an environmental isolate collected in Italy (Thorpe et al., 2022), as the most representative *K. var* strain (Figure 3.4C and 3.4D). Next, we calculated the rates of transition between non-human and human sources. Our data show that transition rates from non-human to human sources were more than double (0.3545) compared to human to non-human sources (0.1497), suggesting a higher probability for *K. var* to flow from non-human environments into the human host. This propensity to switch hosts is exemplified by specific STs, such as the prevalent ST61, which includes human strains (n = 3) and non-human strains (animal n = 1, plant n = 1).

Given the ease of entry into the human host, we next assessed recombination rates using Gubbins within *K. var* and between *K. var* and *K. pne*. We found that the recombination rate within *K. var* populations (1.57%) is significantly lower compared to that within *K. pne* (9.27%). However, the rate of recombination between *K. var* and *K. pne* (8.69%) is notably higher than the rate within *K. var* alone (1.57%), consistent with our hypothesis.

3.4 Adaptation of Non-Human *Kv* into Human Host

Thus far, our results suggest that the dissemination from the non-human to human host, alongside its broad distribution across multiple hosts, underscores *K. var*'s genetic plasticity and implies that genetic systems linked to host/habitat-specific success may be under selection pressure.

Accordingly, we focused on genes associated with environmental survival and host/habitat adaptation (the *nif* gene cluster) and intraspecies competition (Type VI Secretion Systems) in highly competitive environments to assess whether variations in these loci could be linked to "non-human" versus "human" *K. var* isolates.

The *nif* gene cluster is responsible for the synthesis, assembly, and regulation of nitrogenase, the enzyme responsible for converting nitrogen (N_2) to ammonia (NH_3), which plants can metabolise. The *nif* gene cluster is commonly found in diazotrophic species, particularly those forming commensal relationships with plant roots, such as *Azotobacter vinelandii* (Jacobson et al., 1989). This cluster consists of core genes, including *nifH*, encoding the nitrogenase reductase component, and *nifD* and *nifK*, encoding the α and β subunits of the nitrogenase molybdenum-iron (MoFe) protein. Additional genes involved in the nitrogen fixation process include *nifN* for FeMo-cofactor synthesis, *nifB* for the precursor biosynthesis, and *nifS* for iron-sulfur cluster assembly and insertion into nitrogenase components (Dos Santos, Fang, Mason, Setubal, & Dixon, 2012) (Fouts et al., 2008).

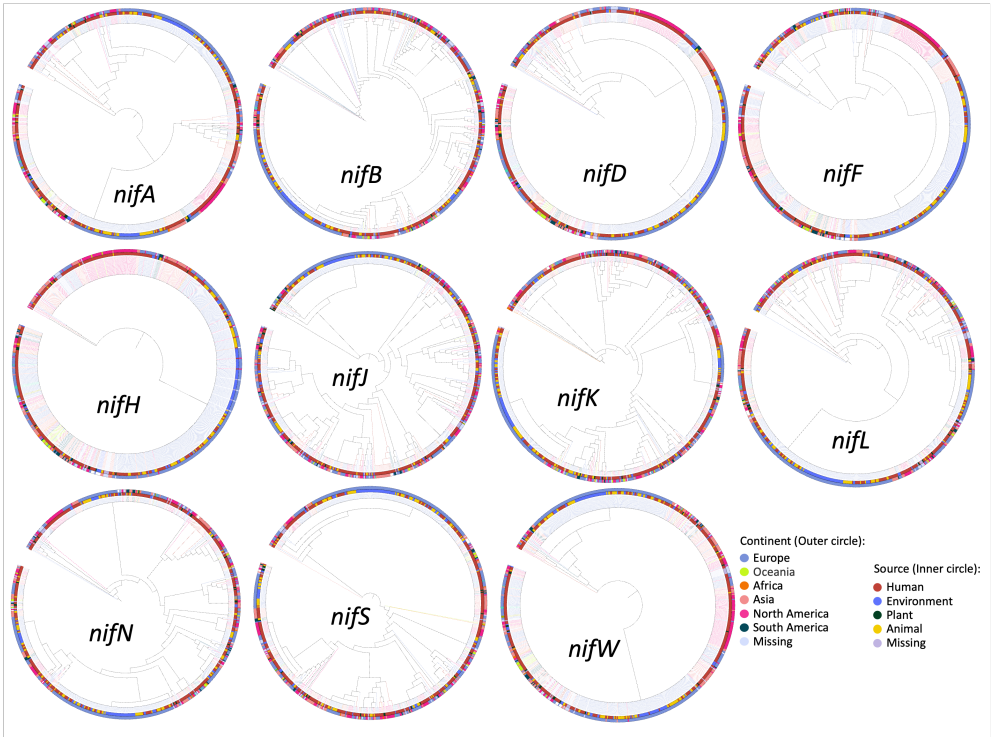


Figure 3.5: Gene phylogenies of *nif* genes in *Klebsiella variicola*. The *nif* genes were extracted from the pan-genome (Page et al., 2015), and the gene phylogenies was built using FastTree (Price et al., 2010) and visualise with iTol (Letunic & Bork, 2007). A)

The *K. var nif* cluster, including core genes *nifH*, *nifD*, *nifK*, and accessory *nif* genes *nifA*, *nifB*, *nifF*, *nifJ*, *nifL*, *nifN*, *nifS*, and *nifW*, is highly conserved. Gene phylogenies show no cladistic associations of *nif* genes with different hosts or geographic locations (Figure 3.5). The SNPs within the *nif* gene cluster do not result in deleterious mutations, as determined by PolyPhen-2 (Adzhubei et al., 2010). Together, these findings suggest that despite its association with an environmental lifestyle, the *nif* gene cluster in *K. var* is not under selective pressure related to the human context and likely has roles extending beyond agriculture or plant-associated interactions.

Next, the Type VI secretion system (T6SS) is recognised as a key factor in microbial competition and interaction. In environmental settings, T6SS can inhibit the growth of rival bacteria and eukaryotic cells (Green & Mecsas, 2016; Hernandez, Gallegos-Monterrosa, & Coulthurst, 2020). The T6SS locus includes the *hcp* gene, spike genes *vgrG* and *paar*, a contractile sheath encoded by *tssABC* genes, and a membrane component encoded by *tssJLM* (Yu, Xue, Fu, & Yang, 2021). Research indicates that the absence of *hcp* or *vgrG* genes can suppress other T6SS-related genes, thus compromising system functionality (J. Zhang et al., 2024).

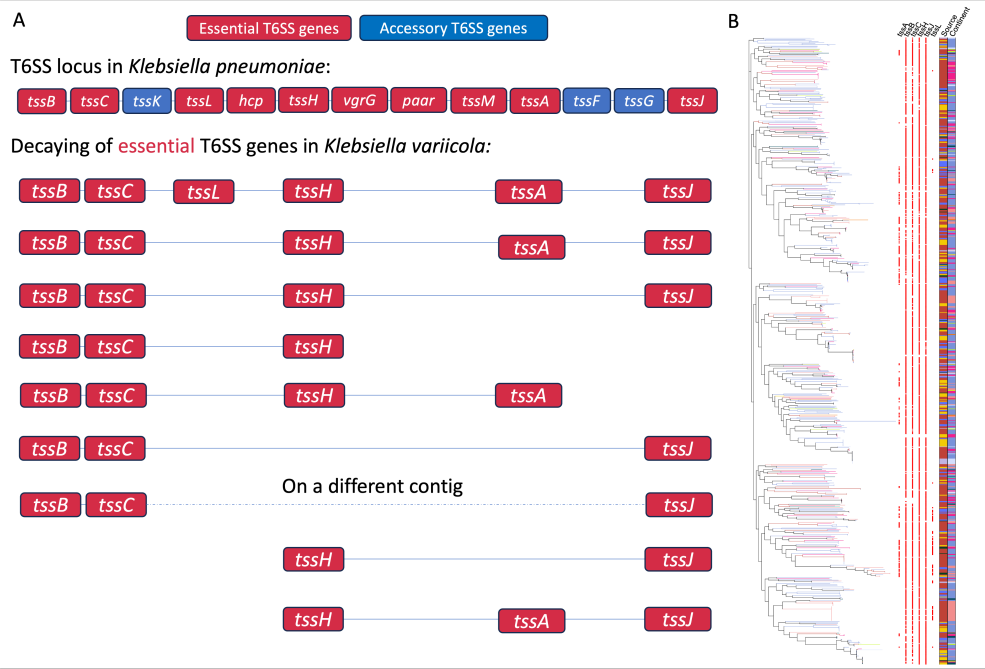


Figure 3.6: Type VI Secretion System (T6SS) in *Klebsiella variicola*. The genes in the T6SS were extracted from the pan-genome (Page et al., 2015), and the gene phylogenies was built using FastTree (Price et al., 2010) and visualise with iTol (Letunic & Bork, 2007). A: Organisation of the T6SS in *Klebsiella pneumoniae* and *Klebsiella variicola*. B: Presence of the T6SS genes in *Klebsiella variicola*.

Analysis of T6SS-associated genes shows that essential T6SS genes are not consistently present across all *K. var* strains. For example, one of the sheath-coding genes, *tssA*, is found in only 163 (22.80%) of 715 *K. var* strains. Other essential genes, such as *tssB* and *tssC*, are found in 634 and 603 strains, respectively. Among the membrane components, *tssM* is completely absent, and *tssL* is present in only 70 (9.8%) of 715 strains, while *tssJ* is present in 686 strains. Importantly, *hcp*, *vgrG*, and *paar* are absent, suggesting that *K. var* lacks the structural elements needed to produce a functional T6SS. The presence and absence patterns of essential T6SS genes are shown in Figure 3.6B. Therefore, we infer that *K. var* likely does not possess a functional T6SS system. Nevertheless, *K. var* remains the second most prevalent species in environmental and human contexts (Barrios-Camacho et al., 2019; Di, Jang, Unno, & Hur, 2017; J. Zhang et al., 2019). The T6SS locus in *K. pne* was inferred to have been acquired from *Yersinia pestis* via horizontal transfer (W. Liu et al., 2023).

Comparing the T6SS loci in *K. var* and *K. pne*, although many essential genes are absent, the remaining genes show high homology to *K. pne* with percentage identities (Pid) exceeding 95%. In contrast, comparisons to *Yersinia pestis* show lower homology (e.g., *tssB*: Pid = 28%, *tssC*: Pid = 89%). Given the evolutionary distance between *K. var* and *Yersinia pestis*, it is

likely that the T6SS found in *K. var* and other *Klebsiella* species was acquired and adapted from a shared genetic pool in common environments, potentially involving a variety of donor organisms, such as *K. pne* or *Yersinia pestis*. As shown in Figure 3.6A, the T6SS locus in *K. var* appears to be decaying compared to that in *K. pne*, with only *tssBC* consistently retained.

In *K. pne*, the T6SS locus is complete, with all essential genes present across *K. pne* strains (Storey et al., 2020). In fact, *K. pne* is the only *Klebsiella* species known to carry a full T6SS locus (Merciecca et al., 2022).

To date, our data, consistent with prior work, demonstrate that *K. var* is widely distributed across various sources and continents (Figure 3.1). This extensive distribution highlights *K. var*'s potential to colonise and infect diverse hosts, encompassing human, animal, and environmental sources. However, the specific factors in *K. var* driving virulence, drug susceptibility, or bacterial survival in different hosts and infective states remain to be systematically established.

The core genome phylogeny suggests limited differences between human- and non-human-associated *K. var* in terms of host adaptation. To validate this observation, we undertook Genome-Wide Association Analyses on the *K. var* pan-genome.

3.5 Genome-Wide Association Analyses

We have shown that there is limited distinction between human and non-human *K. var* in terms of core genome phylogeny (Figure 3.1), time-resolved phylogeny (Figure 3.4), *nif* gene cluster, and T6SS system. However, we have not investigated the accessory genome of *K. var*. To address this, we applied genome-wide association studies (GWAS) to identify genetic elements in the *K. var* pan-genome that are statistically associated with niche origin (genotype–phenotype associations). As SNP-based core genome phylogeny did not reveal cladistic links to niche adaptation, the GWAS was conducted using gene presence/absence patterns rather than SNP variation.

In our study, we employed both phylogenetic-based GWAS, which corrected for the confounding effects of clonal population structure and homologous recombination (et al., 2017), and pan-genome GWAS, which could analyse the association between gene presence/absence genotype and human/non-human phenotype (Brynildsrud et al., 2016), to examine pan-genome gene carriage differences between human (n = 441) and non-human (n = 248) *K. var* isolates. The phylogenetic-based GWAS identified 14 candidate genes, with only five passing correlation analysis and Fisher's exact test, validated through False Discovery Rate (FDR) correction (Benjamini & Hochberg, 1995) and Bonferroni correction (Abdi et al., 2007). Notably, the top

five genes identified in the pan-genome GWAS matched those from the phylogenetic-based GWAS, demonstrating concordance between these analytical methods. As shown in Table 1, the genes significantly associated with non-human *K. var* isolates include group_1575 (hypothetical protein), *cspB*, group_3155 (hypothetical protein), *chbP*, and *lpxP*.

Protein/Group	Description	Odds-ratio	Bonferroni corrected *p-value*	FDR corrected *p-value*	Present in non-human strains	Absent in human strains
group_1575	Hypothetical protein	5.221	5.00e-18	2.50e-18	150 (60.48%)	338 (76.64%)
CspB	Cold-shock domain-containing protein	5.221	5.00e-18	2.50e-18	150 (60.48%)	338 (76.64%)
ChbP	Chitin catabolism related protein	5.072	1.41e-17	3.52e-18	113 (45.56%)	287 (65.08%)
group_3155	Hypothetical protein	5.072	1.41e-17	3.52e-18	113 (45.56%)	287 (65.08%)
LpxP	Lipid A biosynthesis related protein	4.983	3.26e-17	6.53e-18	111 (44.76%)	288 (65.31%)

Table 3.1: Proteins or ortholog groups from the pan-genome that are significantly associated with the non-human sources in *Klebsiella variicola* from Genome-Wide Association Analysis (GWAS). 441 human-associated and 248 non-human-associated *Klebsiella variicola* isolates were applied in the GWAS analysis. The protein or group name and descriptions are presented in the first and second columns. The odds ratio values are reported in the third column, and Bonferroni Abdi et al. (2007) and False Discovery Rate (FDR) Benjamini and Hochberg (1995) corrected *p* values are shown in the fourth and fifth columns. The percentage of genes present in non-human strains or absent in human strains is calculated by the number of strains that carry or do not carry the gene divided by the total number of non-human (248) or human (441) associated strains.

Compared to GWAS in other enterobacteria species, where hundreds or even thousands of genes are typically identified (Sheppard et al., 2013), our findings of only five candidate genes indicate limited evidence for host/habitat adaptation in *K. var*.

Among these five candidates, CspB and LpxP are related to viability at lower temperatures. CspB is a cold-shock domain-containing protein, which plays a crucial role in bacterial survival during temperature drops (Ivancic, Jamnik, & Stopar, 2013), more common in non-human environments. CspB has the most significant *p*-value, being present in 150 (60.48%) non-human *K. var* isolates while absent in 338 (76.64%) human isolates. LpxP is crucial for modifying lipid A, a bacterial outer membrane component, under cold conditions in *E. coli* (Vorachek-Warren, Carty, Lin, Cotter, & Raetz, 2002). Specifically, LpxP incorporates a palmitoleoyl moiety into nascent lipid A, replacing the laurate chain usually added by another enzyme, LpxL. Loss of LpxP is associated with increased antibiotic sensitivity (e.g., rifampicin and vancomycin) at 12℃, suggesting that LpxP enhances the outer membrane's barrier function against harmful chemicals in lower temperatures. ChbP is involved in chitin catabolism and predominantly found in marine bacterial isolates (Ran et al., 2023), which highlights the significance of chitin metabolism for *K. var* in marine ecosystems.

Although these genes are enriched in non-human sources, they remain present in strains of human origin. We also examined sequence alignments to determine whether specific mutations or deletions were exclusive to human or non-human strains, but no changes linked to host/habitat were observed (Figure 3.7).

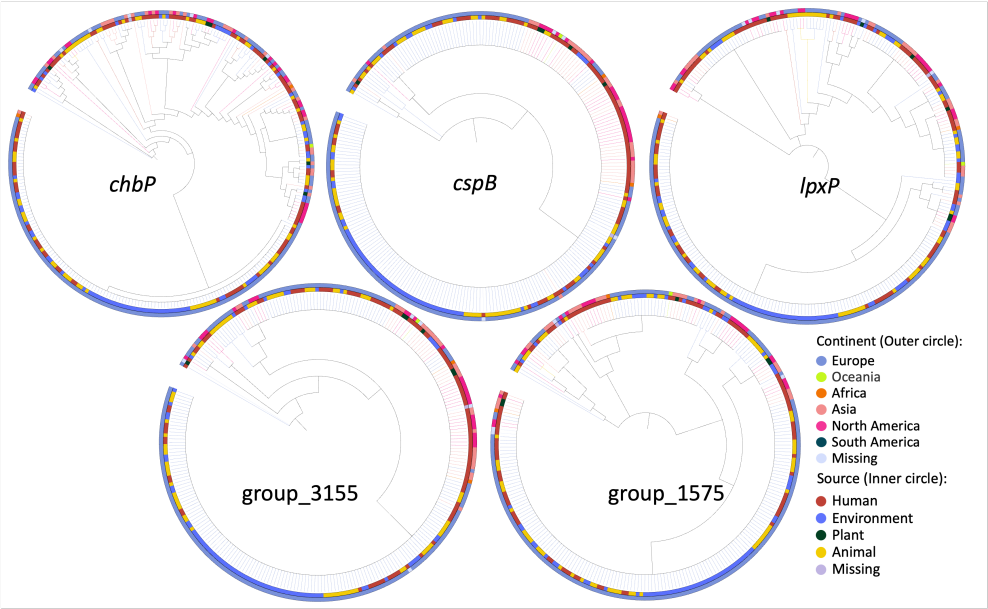


Figure 3.7: Gene phylogenies of 5 genes significantly associated with the non-human sources in *Klebsiella variicola*. The genes were extracted from the pan-genome (Page et al., 2015), and the gene phylogenies was built using FastTree (Price et al., 2010) and visualise with iTol (Letunic & Bork, 2007).

Given that our GWAS analyses suggest the enrichment of specific genes potentially facilitates adaptation to environmental contexts, we compared the pan-genome diversity between human and non-human *K. var*. We used iNext software version 3.0.0 (Chao et al., 2014) to analyse pan-genome diversity, applying interpolation and extrapolation techniques to categorise host/habitat and calculating a 95% confidence interval for both rarefaction and extrapolation curves. This approach enabled us to estimate and compare pan-genome gene diversity between human and non-human sources. As shown in Figures 3.8A and 3.8B, the overall pan-genome diversity of human-associated *K. var* strains is higher than that of non-human-associated strains, indicating that human strains are under greater selection pressure.

Similarly, diversity comparisons of AMR genes in *K. var* isolates from human and non-human environments showed that human-derived *K. var* strains exhibit a greater diversity of AMR genes compared to those from non-human sources, as depicted in Figures 3.8C and D. Additionally, comparative analysis of AMR genes across continents revealed that European *K. var* strains generally harbour fewer AMR genes, whereas Asian strains contain a higher number of AMR genes, including bla_{NDM} and *cat*, as detailed in Supplementary Table 2.

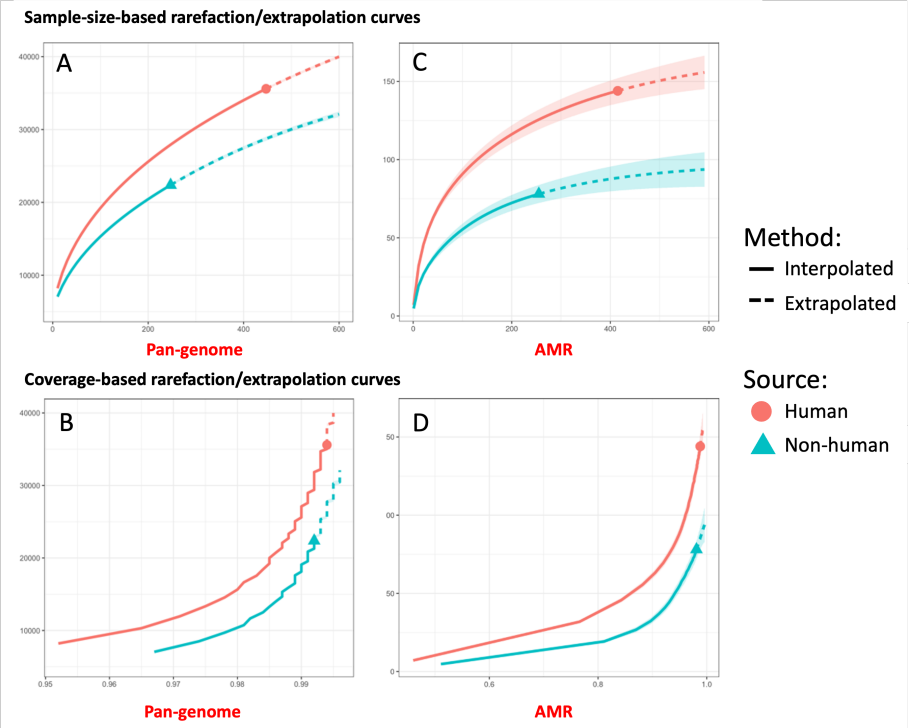


Figure 3.8: Comparative analysis of gene content diversity in human and non-human sources in 715 *Klebsiella variicola*. A: Pan-genome diversity - Sample-size-based curves: The rarefaction and extrapolation curves based on sample size for pan-genome diversity. B: Pan-genome diversity - coverage-based curves: Coverage-based rarefaction and extrapolation curves. C: AMR diversity - Sample-size-based curves: Rarefaction and extrapolation curves for antimicrobial resistance (AMR) gene diversity, calculated based on sample size. D: AMR diversity - coverage-based curves: Coverage-based rarefaction and extrapolation curves for AMR gene diversity. In all panels, solid lines represent interpolated data, while dotted lines indicate extrapolated data. The human host is depicted with red lines, contrasting with blue lines that represent a combination of non-human sources. Note the strains that do not have host/habitat annotations and the Bangladeshi isolates have been excluded from this analysis for more solid outputs.

3.6 Resistance

To underpin the antibiotic resistome content of *K. var*, we built the AMR profile for these isolates using ABRicate (Seemann, 2019) and visualised it on the core-genome phylogeny, indicating gene presence or absence (Figure 3.9). Our analyses show that the *K. var* resistome includes both intrinsically present (core genome) and extrinsically acquired (accessory genome) genes.

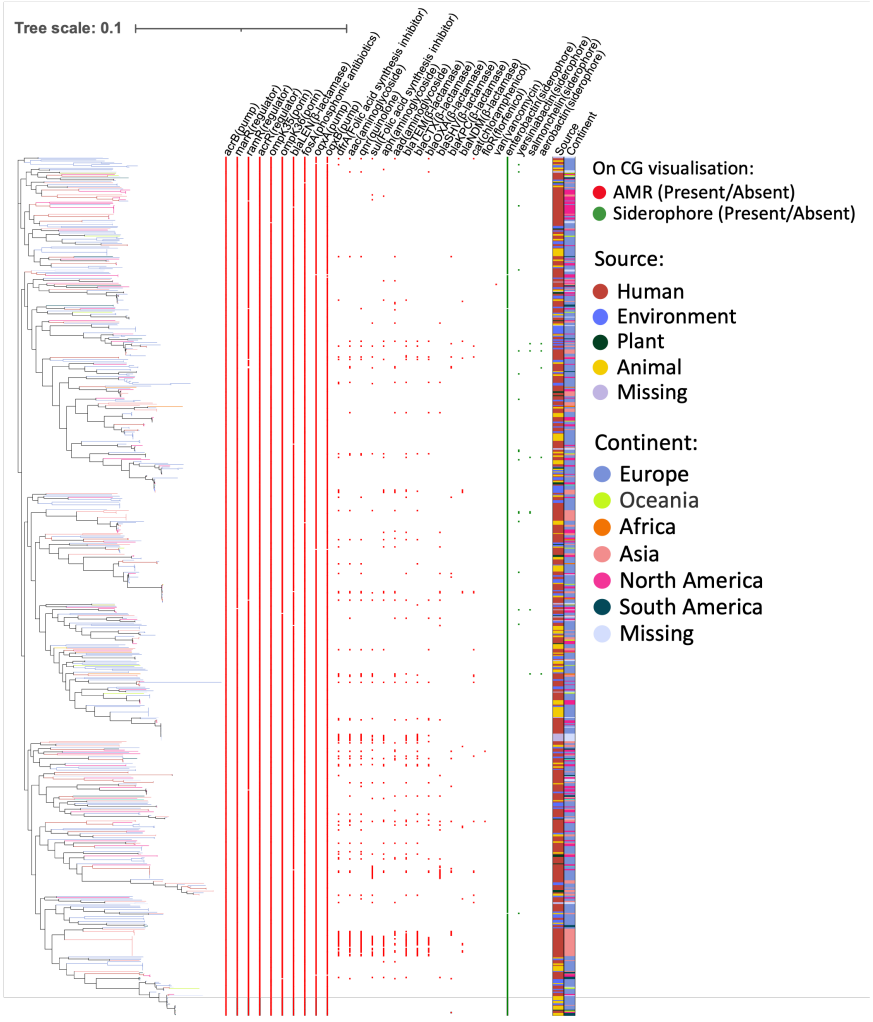


Figure 3.9: Antimicrobial resistance (AMR) and siderophore gene carriage of *Klebsiella variicola*. The resistance and siderophore profile of 715 *Klebsiella variicola* is visualised on the Core Genome (CG) phylogeny. The presence of AMR genes are coloured in red, the presence of siderophores are coloured in green. The geographic locations of the strains are annotated on the second last column, while the host/habitat of the strains are annotated on the last column and the CG tree.

We first examined the extrinsically acquired AMR genes in *K. var*. To ensure accuracy in our comparative analysis across *K. var* clusters, we excluded highly multidrug-resistant isolates from a neonatal sepsis outbreak in Bangladesh (Farzana et al., 2019), as these human-derived strains showed a significantly higher number of AMR genes (mean = 16.36) than other *K. var* strains. This exclusion aimed to avoid data skewing by outlier samples. We compared the total AMR gene content between human and non-human sources by analysing 248 *K. var* isolates from non-human sources (animal, plant, and environment) with 441 strains from the human host. Applying Welch's unequal variances *t* test to mitigate false positives, our analyses revealed that human-derived strains (mean = 7.173) carry a significantly higher

number of AMR genes (*p* value = 1.31e-12) relative to *K. var* from non-human sources (mean = 6.174). The enriched AMR gene carriage in human-associated *K. var* strains likely reflects both the diversity and abundance of the resistance gene pool and the antibiotic selection pressure imposed on *K. var* isolated from human sources.

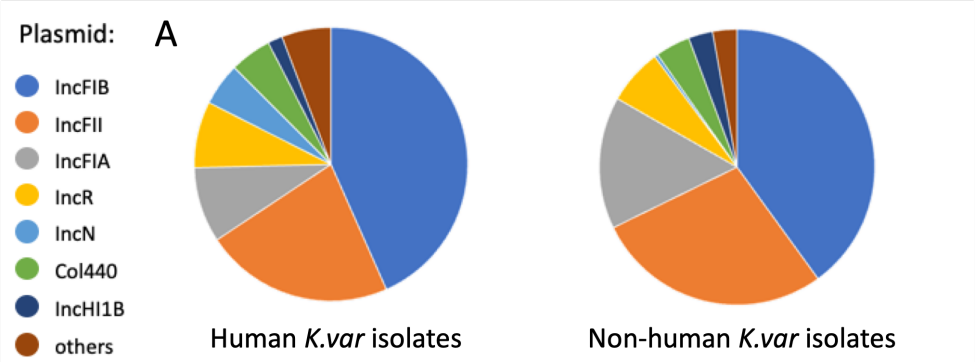
AMR gene	Class	Fisher test *p* value	Odds ratio	Genomic location
aac	Aminoglycosides	1.186e-11	39.28	plasmid
aad	Aminoglycosides	2.942e-06	7.210	both
aph	Aminoglycosides	2.559e-05	5.472	plasmid
$bla_{\text{CTX-M}}$	ESBL	5.086e-07	13.17	plasmid
bla_{KPC}	Carbapenemases	0.0002	infinity	plasmid
bla_{OXA}	Oxacillinases	2.838e-08	infinity	plasmid
bla_{SHV}	ESBL	4.247e-05	16.60	chromosome
bla_{TEM}	ESBL	1.211e-07	8.612	plasmid
cat	Chloramphenicol	0.0237	7.437	both
dfrA	Trimethoprim	4.067e-06	4.680	plasmid
qnr	Quinolone	7.541e-11	36.33	plasmid
sul	Sulfonamides	5.495e-08	14.89	plasmid

Table 3.2: Fisher test *p* value and *odds ratio* of AMR genes associated with the human host.

Next, we aimed to determine whether specific AMR genes were predominantly associated with human-derived isolates. Using Fisher's test, we analysed the extrinsic AMR genes presented in Figure 3.9. As detailed in Table 3.2, we identified 12 AMR genes significantly associated with the human host, including $bla_{\text{CTX-M}}$ (mainly $bla_{\text{CTX-M-15}}$), bla_{TEM} (mainly $bla_{\text{TEM-1}}$) (encoding class A extended-spectrum beta-lactamase), bla_{KPC} (mainly $bla_{\text{KPC-2}}$) (encoding carbapenemase), bla_{OXA} (mainly $bla_{\text{OXA-1}}$) (encoding class D beta-lactamase), and bla_{SHV} (mainly $bla_{\text{SHV-12}}$ and $bla_{\text{SHV-1}}$). Except for bla_{SHV}, which is located on the core genome, bla_{OXA}, bla_{KPC}, bla_{CTX}, and bla_{TEM} are located on the accessory genome, with *aad* and *cat* located on both core and accessory genomes.

Human *K. var* strains carry significantly more plasmids of the same types as non-human strains. The presence of IncFIB, IncFII, IncR, and IncN plasmids is associated with increased AMR gene carriage (Figure 3.10), aligning with previous observations that *K. var* strains from the human host contain more AMR genes than strains from the non-human sources.

Given the clear human host associations of extrinsically acquired AMR genes, we investigated whether variations in intrinsic genes follow a similar pattern. We constructed gene phylogenies and examined SNP variations in key intrinsic resistance genes. The chromosomally located *oqxR-oqxAB* locus, an efflux pump (Veleba et al., 2012), was present in nearly all *K. var* isolates, with only two strains (WUSM_KV_35 and 653_KVAR) lacking the *oqxA* gene. All SNPs in *oqxAB* and *oqxR* sequences were predicted to be neutral by PolyPhen-2 (Adzhubei et al., 2010), suggesting that these pumps are not overexpressed.



B

Plasmid	*t* test *p* value	Strain containing the plasmid		Strain without the plasmid	
		AMR Mean	AMR SD	AMR mean	AMR SD
IncFIB	6.224e-10	7.365	2.750	6.390	1.514
IncFII	1.126e-12	8.041	3.210	6.422	1.568
IncR	4.123e-07	8.907	3.580	6.588	1.845
IncN	2.874e-09	10.410	3.076	6.624	1.961

Figure 3.10: Plasmid Profile of *Klebsiella variicola*. The plasmid profile of each strain was built using PlasmidFinder (Carattoli et al., 2014). A) Host/habitat-based plasmid carriage: The comparison of plasmid carriage between human (n = 441) and non-human (n = 248) strains. B) Plasmids and AMR gene carriage: The plasmids associated with enriched AMR gene carriage in *Klebsiella variicola* (n = 715). In the table, the plasmid type is presented in the first column, and the *t* test *p* values are shown in the second column. The mean and Standard Deviation (SD) values of either strain containing or not containing the plasmid are shown in the following columns correspondingly.

Given that extrinsically acquired AMR genes exhibit clear associations with the human host, we wondered if variations in intrinsic chromosomally located genes which confer antibiotic resistance would follow a similar trajectory. We based the selection of these intrinsic loci on well-characterised permeability factors controlling efflux (AcrAB, OqxAB), influx (OmpK35, OmpK36) via the regulatory proteins (RamR, MarR, AcrR, OqxR, SoxR), where RamR, MarR, AcrR and OqxR function as the cognate repressors of the *ramA*, *marA*, *acrAB* and *oqxAB* genes respectively (Anes et al., 2021; De Majumdar et al., 2015; J. Li et al., 2019; Padilla et al., 2010; Rosenblum et al., 2011; Schneiders et al., 2003; Tsai et al., 2011; Veleba & Schneiders, 2012). Our analyses confirmed a trend of increased deleterious sequence variations, as predicted using PolyPhen-2 in *ramR* (p value < 0.001) amongst human *K. var* strains (human = 35, non-human = 2, missing source = 12). In contrast, *ompK35* (human = 8, missing source = 5), *ompK36* (human = 16, non-human = 12, missing source = 2), and *acrR* (human = 226, non-human = 171, missing source = 11) harboured deleterious mutations, but these were

equivalently found in both human- and non-human strains, suggestive of their broader role in nutrient uptake and cellular homeostasis. The correlation tests between the intrinsic mutations and plasmid carriage did not produce any significant associations (*p* values: *mrkD* = 0.101, *ramR* = 0.095, *acrR* = -0.303, *ompK35* = 0.192, *ompK36* = -0.071).

Interestingly, we found that 28 human *K. var* isolates in our dataset harboured the *K. pne*-specific intrinsic beta-lactamase $bla_{\text{SHV-1}}$, as shown in Figure 3.9 and Table A.2. In prior studies where *K. var* and *K. pne* strains were isolated from the same hospital and timeframe, 23% of *K. var* samples were classified as 'hybrids' due to the acquisition of the capsule locus from corresponding *K. pne* strains (Gorrie et al., 2022). The acquisition of bla_{SHV} may represent an additional horizontal gene transfer event from *K. pne* to *K. var*. bla_{SHV}-carrying *K. var* strains also contained more plasmidic ARGs (e.g., *dfrA*, *sul*, *aac*, and bla_{TEM}) (mean = 10.82) than non-bla_{SHV} strains (mean = 6.57).

To determine if bla_{SHV}-containing *K. var* strains qualify as hybrid strains, we calculated the Average Nucleotide Identity (ANI) of *K. var* strains against a *K. pne* reference genome (strain: HS11286, genome assembly: GCF_000240185.1). ANI results showed no significant difference (*p* value = 0.5868) from the *K. var* reference genome (strain: LEMB11, genome assembly: GCF_009648975.1). We calculated the pairwise distances between bla_{SHV}-positive and -negative strains and found no significant difference between them (*p* = 0.5535). Thus, while bla_{SHV} is acquired from *K. pne* by *K. var*, this event is not reflected in ANI calculations, suggesting a localised gene transfer event rather than extensive genomic evolution. In support of a localized and targeted gene transfer event, we studied the bla_{SHV} gene and the capsule locus in a bla_{SHV} containing *K. var* strain. Using the $bla_{\text{SHV-1}}$ carrying *K. var* strain SPARK_2054_C2, which was co-isolated with a $bla_{\text{SHV-1}}$ carrying *K. pne* strain (SPARK_2054_C1), we show that the ANI of SPARK_2054_C2 (99.11%) and the *K. pne* strain SPARK_2054_C1 (94.63%) are not significantly different from the ANI reported for other *K. var* and *K. pne* strains in our dataset (average ANI with *K. var*: 99.03%, average ANI with *K. pne*: 94.62%) overall. The capsule types of these strains are reported as KL11 (SPARK_2054_C1, *K. pne*) and KL123 (SPARK_2054_C2, *K. var*) respectively. In contrast, the $bla_{\text{SHV-1}}$ gene shared between these strains has 100% identity. Thus, we infer that although bla_{SHV} is obtained from *K. pne*, the *K. var* strains containing the bla_{SHV} gene did not undergo extensive genomic recombination. Taken together, our data clearly underscore the higher content of AMR gene carriage in human *K. var* isolates.

Notably, another study on a less comprehensive database by Osei et al. observed different patterns of AMR gene distribution in strain clusters (Osei Sekyere & Reta, 2021), which were not detected in our expanded *K. var* dataset, suggesting that AMR gene distribution patterns may vary based on dataset scale.

We compared AMR gene diversity between *K. var* isolates from human and non-human environments using iNext software (version 3.0.0) (Chao et al., 2014). Our results indicate that human-derived *K. var* strains exhibit greater AMR gene diversity than non-human strains (Figures 3.8C and D). Additionally, AMR gene diversity across continents reveals that European *K. var* strains generally contain fewer AMR genes, whereas Asian strains show a higher count, including bla_{NDM} and *cat* genes.

This comparative analysis underscores the impact of antibiotic use and selection pressure on AMR gene diversity between human and non-human *K. var* isolates.

3.7 Virulence

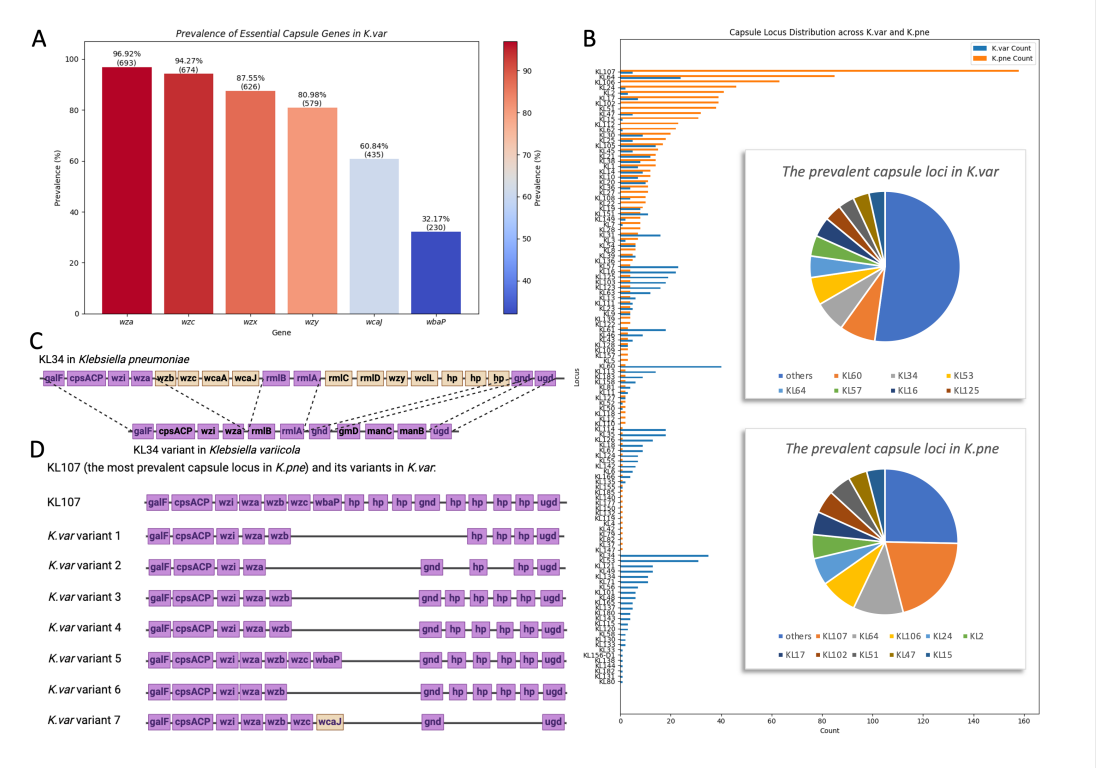


Figure 3.11: Comparative analysis of capsule loci in 715 *Klebsiella variicola*. 6A) The heatmap prevalence of essential capsule synthesis genes in *Klebsiella variicola*. 6B) The comparison of capsule locus in *Klebsiella pneumoniae* and *Klebsiella variicola*. The capsule locus is ranked by descending order in *Klebsiella pneumoniae*. The most prevalent 10 capsule loci in *Klebsiella variicola* and *Klebsiella pneumoniae* are visualised by pie charts. 6C) Novel Capsule Loci in Bangladeshi Strains: Novel capsule loci specifically found in Bangladeshi neonatal *Klebsiella variicola* strains (KL34 variant) and its comparison with KL34. 'hp' represents hypothetical proteins. 6D) A comparison of the most prevalent capsule locus in *Klebsiella pneumoniae* (KL107) and its variants in *Klebsiella variicola*.

Capsular polysaccharides (CPS) are recognised as a significant virulence factor in bacterial pathogenicity, with KL1 and KL2 as the most prevalent capsule loci in *K. pne*. These loci are almost universally present in *K. var* and *K. pne*, highlighting their fundamental role in facilitating *Klebsiella*-host interactions. Specifically, in *K. pne*, CPS plays a crucial role in evading host immune responses, enhancing bacterial survival and persistence within the host (Walker & Miller, 2020). Capsule synthesis in *K. pne* is initiated by glucose phosphotransferase WcaJ, which links glucose to the lipid carrier, or galactose phosphotransferase WbaP, which alternatively links galactose to the lipid carrier. Then, residues are flipped by flippase Wzx and polymerised by O-antigen ligase Wzy. Capsule length is regulated by polysaccharide biosynthesis tyrosine autokinase Wzc, and the mature polysaccharide chain is exported by polysaccharide export protein Wza (Rendueles, 2020).

In our study of *K. var* capsule content, we used Kaptive (Version 3.0) (Stanton et al., 2025) to construct a comprehensive profile of CPS genes and loci. We observed that 13.8% of *K. var* strains possess a novel CPS locus, compared to 5% in *K. pne*. Intriguingly, our analysis revealed 28 previously unreported novel capsule loci unique in *K. var* (Figure 3.12). Notably, the capsule locus in Bangladeshi isolates (Farzana et al., 2019), shown in Figure 3.11C, lacks key capsule synthesis genes like *wcaJ*, *wzc*, and *wzy*. In hypermucoid *K. pne* and hvKp strains, *wcaJ* loss of function correlates with a non-mucoid phenotype, capsule loss, and reduced virulence, underscoring this enzyme's importance (Pal et al., 2019; **?**). However, in Bangladeshi *K. var*, the absence of *wcaJ* does not prevent capsule formation, suggesting alternative capsule initiation pathways exist. Our analysis shows that *wcaJ* was present in just over 60% (n = 435) of *K. var* strains (Figure 3.11A). Other genes considered essential in *K. pne*, such as *wzx* and *wzy*, are absent in 12.45% and 19.02% of *K. var* strains, respectively (Figure 3.11A). Other capsule synthesis genes, such as *manCD* and *rmlABCD*, showed similar carriage in *K. var* and *K. pne*. Aside from flanking genes like *galF*, *gnd*, and *ugd*, the most prevalent capsule gene in *K. var* is *wza*, suggesting *wza*-dependent capsule synthesis in *K. var*. Based on capsule gene carriage, we infer that the combination of ten prevalent genes—*galF*, *cpsACP*, *wzi*, *wza*, *wzb*, *wzc*, *wzy*, *wzx*, *gnd*, and *ugd*—is necessary for capsule production in *K. var*.

Prevalent capsule loci in *K. pne*, such as KL1 and KL2, as reported by Choby et al. (Choby et al., 2020) and Chuang et al. (Chuang, Fang, Lai, Chang, & Wang, 2006), are rare in *K. var*. Specifically, only seven strains carry KL1, and three strains carry KL2 among 715 *K. var* strains (Figure 3.11B). In contrast, KL2 ranks as the fifth most prevalent locus in our *K. pne* database, appearing in 41 strains, with KL1 found in 14 *K. pne* strains. Figure 3.11C shows that KL107, the most common capsule locus in *K. pne* (found in 158 strains), shows notable variability, with seven variants in *K. var*.

Comparative analysis of capsule loci in *K. var* and *K. pne* reveals significant differences despite their close genetic relationship. In our dataset of 715 *K. var* and 1000 *K. pne* strains, 80 capsule loci are shared between *K. var* and *K. pne*, while 12 loci are unique to *K. pne* and 25 are unique to *K. var* (Figure 3.11B). Notably, KL107, largely present in *K. pne* (158 out of 1000 strains), is found in only five *K. var* strains, suggesting possible capsule locus acquisition from *K. pne* to *K. var*. Conversely, KL60 is more prevalent in *K. var* (40 out of 715 strains) than in *K. pne* (2 out of 1000 strains), indicating potential acquisition by *K. pne* from *K. var*. These findings suggest bi-directional capsule locus acquisition between *K. var* and *K. pne*. Using the largest available *K. var* database, we identified 64 capsule loci in *K. var* present in less than 1% of the database, implying that *K. var* may have acquired these loci from *K. pne*.

We next examined capsule regulators, as capsule levels in *K. pne* are modulated by specific regulators. Mutations in *kvrA*/*slyA*, *kvrB*/*mprA*, *kbvR*, and *rfaH* have been shown to reduce capsule production (Dorman, Feltwell, Goulding, Parkhill, & Short, 2018; Palacios et al., 2018; L. Xu et al., 2021). In contrast, mutations in *rcsB* lead to a hypermucoid phenotype (Dorman et al., 2018). Plasmid-encoded regulators, including *rmpA*, *rmpA2*, and *rmpD*, are also critical for capsule regulation, and hypervirulence in *K. pne* is largely associated with the presence of the *rmp* locus (Walker et al., 2020).

Capsule regulators *kvrA*, *kvrB*, *kbvR*, *rfaH*, and *rcsB* were highly conserved in our *K. var* database, regardless of strain origin, with sequences showing near uniform alignment with *K. pne*. The *rmp* locus was found in only six *K. var* strains (Supplementary Table A.3). Although hv*K. pne* strains commonly carry KL1 and KL2, the capsule locus is not associated with the *rmp* locus in *K. var*. Specifically, the hypervirulent Bangladeshi neonatal *K. var* strains (Farzana et al., 2019) lacked *rmpA*, *rmpA2*, and *rmpD* (J. Zhu, Wang, Chen, & Du, 2021), and did not show deleterious changes in *kvrA*, *kvrB*, or *kbvR*. Thus, we infer that capsule levels in *K. var* are likely controlled by loci other than the *rmp* locus. These findings suggest that beyond capsule locus divergence, additional capsule regulatory loci may control the *K. var* phenotype.

Our study also focused on an essential virulence factor, *mrkD*, known for its role in type III fimbriae and biofilm formation, influenced by nutrient-driven evolutionary mutations (Nucci, Rocha, & Rendueles, 2023). When examining *mrkD* variants predicted to be deleterious by PolyPhen-2 (Adzhubei et al., 2010), we identified several deleterious mutations, such as T88M and S29T. These mutations were more frequent in human-associated isolates (human = 45; non-human = 11).

Siderophores represent another crucial aspect of *K. var* virulence. They facilitate iron acquisition from the host environment, essential for bacterial growth and infection. To analyse siderophore carriage, we applied BLASTP to *K. var* pan-genome data and visualised the carriage profile of yersiniabactin, aerobactin, salmochelin, enterobactin, and colibactin (Figure 3.9). Like *K. pne*, enterobactin is highly conserved across *K. var* strains. However, yersin-

iabactin (count = 20, percentage = 2.80%), aerobactin (count = 5, percentage = 0.70%), and salmochelin (count = 7, percentage = 0.98%) are less prevalent in *K. var*. Colibactin was absent in all *K. var* isolates. Interestingly, yersiniabactin, aerobactin, and salmochelin were more common in human-derived strains, underscoring their role in host-specific adaptations. Overall, siderophore gene content in *K. var* is lower than in *K. pne*, particularly for aerobactin and salmochelin.

The significant differences in siderophore profiles between *K. var* and *K. pne* may underlie the prevalence and frequency of *K. var* infections, suggesting that the ability to cause serious systemic infections may not be associated with siderophore carriage alone.

3.8 Discussion

K. var shows remarkable plasticity and is widely distributed across a range of niches and continents. Initially originating from non-human environments, *K. var* has adeptly transitioned to infect both non-human and human hosts over a 1000-year timescale (Figure 3.4A). The host shift from non-human to human is likely facilitated by the acquisition of critical genes, such as bla_{SHV}, from its close relative, *Klebsiella pneumoniae* (*K. pne*). Indeed, *K. var* is frequently isolated from the same sample as *K. pne* (Thorpe et al., 2022).

K. pne is known for its genetic diversity due to its widespread presence in diverse environments and its ability to acquire and spread resistance genes. This diversity is reflected in a higher SNP count, as *K. pne* adapts to various hosts and conditions. The different ecological niches occupied by *K. var* and *K. pne* may lead to distinct rates of mutation and genetic drift. *K. pne* is often associated with hospital environments, where antimicrobial selection pressures are high, leading to genetic diversification. In contrast, *K. var* is commonly found in environmental and agricultural settings, likely subject to different or lower evolutionary pressures. Additionally, as *K. var* lacks the T6SS system, we infer that it might be less harmful when co-colonising with *K. pne*, facilitating recombination or other gene-exchanging events. Altogether, we further infer that *K. var* may be discarding T6SS genes to enable co-colonisation with *K. pne* and to survive by infecting humans. Beyond its role as a concerning pathogen in humans and animals, *K. var* also acts as a coloniser and endophyte within plant niches. Environmental niches further serve as important reservoirs for *K. var*.

Our AMR gene analysis identified 12 genes associated with the human niche (Table 2). *K. var* strains from human environments exhibit not only more AMR genes but also greater genetic variation within these genes (Figure 3.8). We infer that *K. var* in human environments is subject to higher selective pressure, leading to increased genetic diversity. Plasmids play a crucial role in AMR gene acquisition in *K. var*. The similarity of the plasmid pool shared between *K. var* and *K. pne* raises significant concerns. This similarity in plasmid types suggests that *K.*

var has the potential to circulate the same resistance and virulence genes, potentially evolving into a severe human pathogen like *K. pne*. Indeed, *K. var* infections have been associated with high mortality rates in cases of bloodstream infections and neonatal sepsis (Farzana et al., 2019; Maatallah et al., 2014).

The capsule locus in *K. var* exhibits notable differences from that in *K. pne*. Our analysis shows that *K. var* not only carries distinct CPS loci but also features novel capsule loci formations not observed in *K. pne*. Although lacking essential capsule synthesis genes such as *wcaJ*, *wzc*, and *wzy*, and capsule regulators like the *rmp* locus, *K. var* can still produce a capsule and display hypermucoviscosity (Farzana et al., 2019), suggesting an alternative pathway for capsule production and regulation.

When comparing the *mrkD* mutations identified by Nucci et al. (Nucci et al., 2023), we found that *mrkD* mutations in *K. var* do not correspond with their findings, suggesting that in the natural evolution of *K. var*, these mutations may not occur.

Taken together, our work provides a mechanistic basis for the emergence of *K. var* as an emerging human pathogen under antibiotic pressure with genetic and spacial proximity to *K. pne*. This, in combination with the lack of ecological niche separation of *K. var* isolates, highlights the potential for the spillover of human *K. var* into the broader environment and for *K. var* isolates within the broader environment to adapt to human infection.

Finally, we believe current bioinformatics tools, including Kleborate, Kaptive, and others, are not fully aligned with the rapidly expanding *K. var* database. Our findings highlight the importance of large-scale studies in accurately understanding the population structure and AMR gene distribution in *K. var* strains.

Figure 3.12: Novel capsule locus found in *Klebsiella variicola*. The capsule genes were identified using Kaptive (Stanton et al., 2025) and visualised using `biorender.com`.

Chapter 4

Scotland *Klebsiella variicola* case study

4.1 Introduction

Although historically misidentified as *K. pne*, *K. var* is now recognised as a distinct species with unique genetic and pathogenic traits (Fontana et al., 2019; Long et al., 2017). As discussed in the last chapter, *K. var* originates from non-human sources and has progressively disseminated into clinical settings, particularly due to its ability to cause severe infections and its potential for antimicrobial resistance. Its clinical relevance in human health is addressed by its association with bloodstream (Farzana et al., 2019) and urinary tract (Potter et al., 2018) infections, it's capacity to harbour resistance genes (chapter 3.6), and its adaptability across diverse host environments (chapters 3.4 and 3.5). Thus, understanding the genomic landscape of *K. var* is therefore for managing its impact on public health.

In the previous chapter, we analysed a comprehensive global database of *K. var*, including isolates from various continents and hosts/habitats, which provided novel insights into the global diversity and adaptation of *K. var*. Following the comprehensive worldwide multisource dataset analyses, we focused on Scotland which is associated with low rates of resistance and stricter antibiotic control measures (Pagaling et al., 2023). By concentrating on a specific geographic area with robust healthcare practices, we aim to gain deeper insights into the local epidemiology and genetic diversity of *K. var*.

This chapter addresses several key questions about the clinical *K. var* strains collected in Scotland. Firstly, we investigate the presence of MDR strains, examining their prevalence and genetic basis. Secondly, we explore whether novel capsule loci are present in these strains, given the critical role of capsular polysaccharides in virulence and the difference and variety we showed in chapter 3.7. Additionally, we identified the prevalence of the siderophores, which are important for iron acquisition and contribute to the pathogenicity of *K. var*.

Additionally, using patient data, we aim to correlate the genomic characteristics of *K. var* strains with the specific disease. Specifically, we seek to identify the diseases most commonly associated with *K. var* infections in Scotland, and determine whether there are differences in the genotype of strains linked to different diseases. These analyses will enhance our understanding of the pathogenic potential and epidemiological patterns of *K. var* in Scotland, providing a foundation for improved clinical management and infection control strategies.

4.2 Construction of the Scotland *Klebsiella variicola* database

Strain collection was undertaken with support from clinical diagnostic laboratories from January 2023 to December 2023. Throughout 2023, microbiological laboratories from Edinburgh and Dumfries & Galloway collected *K. var* samples for this study. All samples underwent MALDI-TOF mass spectrometry analyses to ensure accurate species identification as *K. var*. The strain information and corresponding host disease are shown in Table B.1.

Upon receiving the samples, we stored the samples after purification and extracted genomic DNA from each strain following the Wizard Genomic DNA Purification Kit (Promega Corporation, Madison, WI, USA) as detailed in Section 2.1. Novogene (Novogene Corporation, Beijing, China) carried out short-read whole genome sequencing of the DNA samples and genome assembly was performed using FastQC (Andrews et al., 2010), Trimmomatic (Bolger et al., 2014), and SPAdes (Bankevich et al., 2012) as described in Section 2.2. The quality of the assembled genomes was evaluated using QUAST (Gurevich et al., 2013) to ensure reliable and high-quality data for downstream analyses.

Species classification was performed using Kleborate (M. Lam et al., 2021) with the assembled genomes, which identified one *K.pne* strain (kv_scot_122) and one *K.quasi* strain (kv_scot_59) (Table B.1), despite that all samples were initially identified as *K.var* through Maldi-Tof mass spectrometry. After excluding the *K.pne* and *K.quasi* strains, a total of 108 *K.var* strains were retained for downstream analyses, including 80 strains from Edinburgh and 28 strains from Dumfries & Galloway. The majority of these *K.var* strains (89 out of 108, 82.4%) were isolated from urine samples, highlighting the significant prevalence of *K.var* in UTIs. Beyond the UTI samples, the clinical *K. var* dataset also included blood samples (count = 4, 3.70%), sputum samples (count = 4, 3.70%), gallbladder samples (count = 3, 2.78%), wound samples (count = 2, 1.85%), kidney samples (count = 2, 1.85%), an intraperitoneal tissue sample (count = 1, 0.93%), and a peritoneal fluid sample (count = 1, 0.93%).

Our comprehensive approach facilitated the creation of a robust, high-quality genomic database of *K. var* strains, which enabled detailed explorations of the genetic diversity, resistance mechanisms, and virulence factors within the clinical environment in Scotland.

4.3 Population Structure

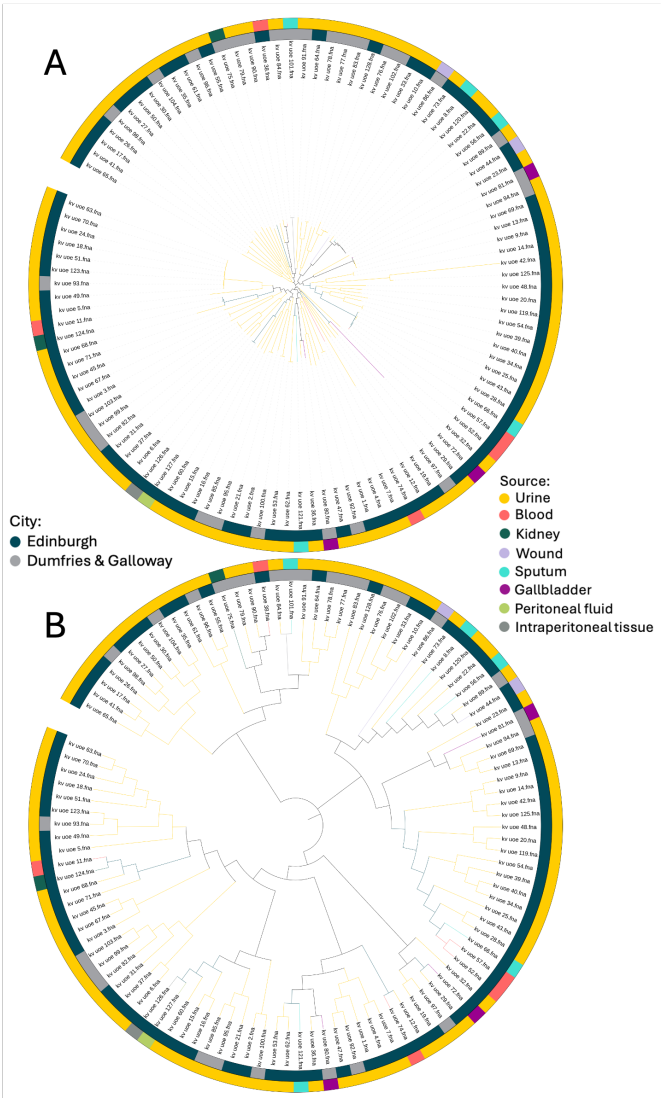


Figure 4.1: Core genome phylogeny of Scotland *Klebsiella variicola* strains. The core genome phylogeny was built using ParSNP (Treangen et al., 2014), and the metadata was visualised using iTol (Letunic & Bork, 2007). A: Recombination free unrooted core genome phylogeny of Scotland *Klebsiella variicola* clinical strains with branch length ignored. B: Recombination free unrooted core genome phylogeny of Scotland *Klebsiella variicola* clinical strains with branch length kept.

To establish the population structure of clinical *K. var* strains in Scotland, we built the recombination free unrooted core genome phylogeny using ParSNP (Treangen et al., 2014). A total of 21,655 SNPs were identified, and the average genome size of the Scottish *K. var* strains was 5.84 Mb. The Average Nucleotide Identity (ANI) of the Scottish *K. var* strains with the reference genome of *K. var* (GCF_009648975.1) is 99.01. The phylogeny of the 108 clinical strains revealed a star-like structure, with no distinct clustering based on collection location or sample type (Figure 4.1). This pattern closely resembles the core genome phylogeny of the 715 worldwide *K. var* strains described in the last chapter (Figure 3.1A), further demonstrating the extensive genetic diversity and lack of geographic or phenotypic clustering within *K. var* populations.

The ST of the Scottish strains are highly diverse, with 72 STs identified among 108 strains. The most prevalent STs are ST1563, ST250, and ST347, found in 7, 6, and 6 strains, respectively, while 61 STs are represented in only a single strain. Notably, key *K. pne* STs, such as ST11, ST15, ST258, and ST23, were not detected in the Scottish isolates.

Next, to investigate the phylogenetic relationship between the Scotland dataset and the global dataset, we combined these datasets and reconstructed the core genome phylogeny using the same methodology, as shown in Figure 4.2. As expected, the Scottish *K. var* strains did not form a distinct cluster and were randomly distributed on the phylogeny, despite being collected within a defined time frame and from a localised geographic region. This finding further underscores the extensive genetic diversity of *K. var*, highlighting its broad distribution across different geographic locations, hosts, and habitats worldwide.

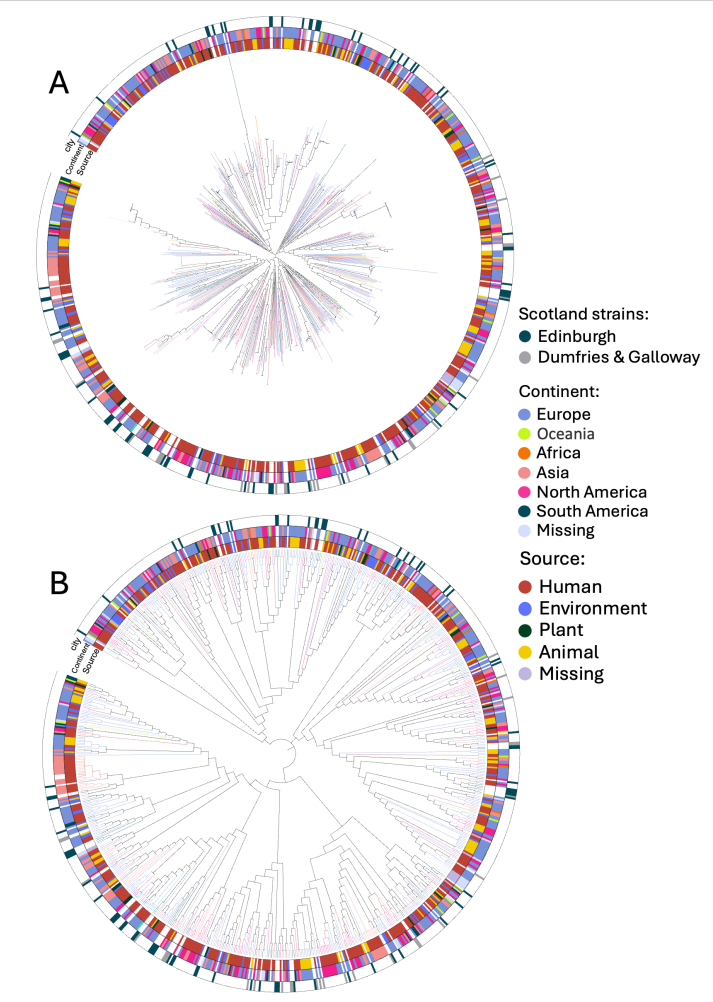


Figure 4.2: Combined core genome phylogeny of Scotland clinical strains and global dataset of *Klebsiella variicola*. The recombination-free unrooted core genome phylogeny was built using ParSNP (Treangen et al., 2014), and the metadata was visualised using iTol (Letunic & Bork, 2007). The Scottish strains were highlighted in the outer circle of the tree. A: Recombination-free unrooted core genome phylogeny of the merged dataset with branch length kept. B: Recombination-free unrooted core genome phylogeny of the merged dataset with branch length ignored.

4.4 Resistance and Virulence

Next, to investigate the pathogenicity of the Scottish *K. var* strains, we constructed their resistance profiles, including the presence or absence of AMR genes and their plasmid profiles. Additionally, we analysed their virulence profiles by examining the presence of siderophores and capsule loci. These comprehensive profiles provide insights into the genetic factors contributing to the pathogenic potential and adaptability of the Scottish *K. var* strains in the human host.

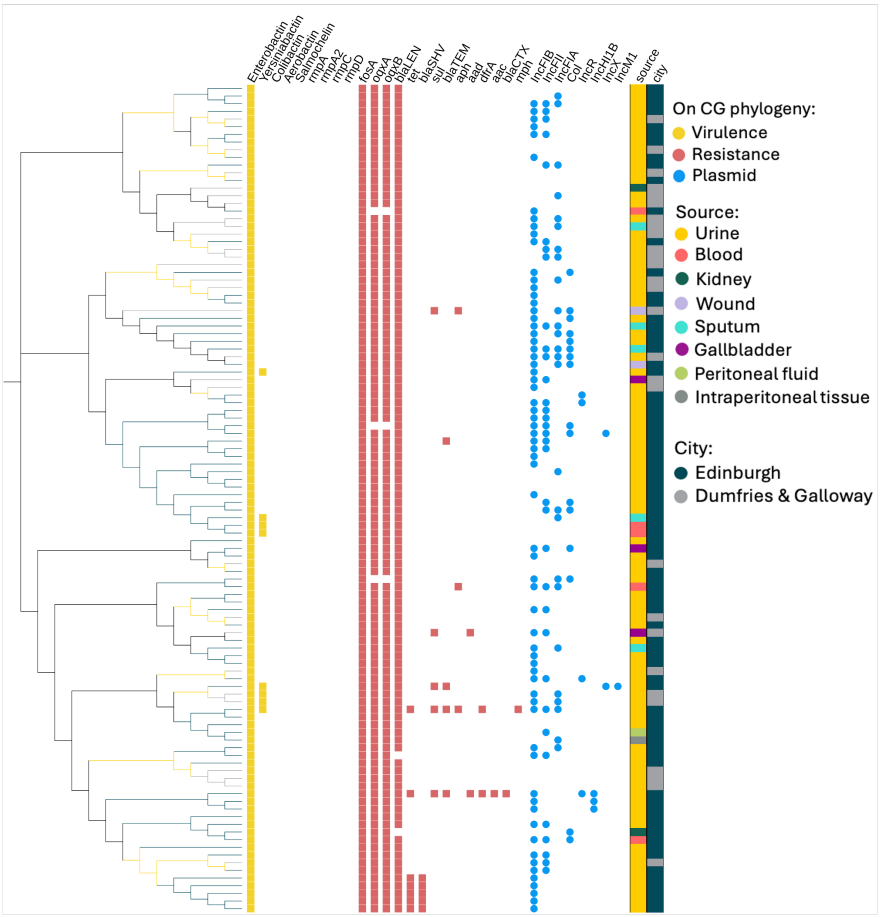


Figure 4.3: Resistance and virulence content of Scotland *Klebsiella variicola* strains. The virulence, resistance and plasmid profile of each strain was built using Kleborate (M. Lam et al., 2021), AMRFinderPlus (Feldgarden et al., 2021) and PlasmidFinder (Carattoli et al., 2014). The metadata was visualised on the core genome phylogeny using iTol (Letunic & Bork, 2007).

The average resistance gene count of Scottish *K. var* strains is 4.21, which is consistent with low levels of resistance in Scotland and is slightly lower than the global *K. var* dataset (average count = 4.78). *fosA* is carried on all the strains, which suggests fosfomycin resistance (Elliott et al., 2019). Notably, three strains in the Scottish dataset were found to have lost the *oqxAB* genes (Figure 4.3), a phenomenon also observed in the global *K. var* dataset. As observed in the previous chapter, bla_{LEN} is the intrinsic β-lactamase in *K. var*, with only two strains (1.85%) found to have lost bla_{LEN}. bla_{TEM} was found in 4 strains (percentage = 3.70%), and only one ESBL-producing strain was identified in the Scottish dataset, carrying $bla_{\text{CTX-M}}$. No strains carrying bla_{KPC} or bla_{OXA} were found in the Scottish *K. var* dataset (Figure 4.3), suggesting a less significant role in human infection. Deleterious mutations in RamR were found in four strains: kv_uoe_128 (A17D), kv_uoe_5 (G25D), kv_uoe_20 (A37S), and kv_uoe_51 (T119P). No deleterious mutations in AcrR were detected. Deleterious mutations in OmpK35 were

found in two strains: kv_uoe_12 and kv_uoe_125 (E257K). Deleterious mutations in OmpK36 were found in two strains: kv_uoe_84 (S89P) and kv_uoe_83 (S109F). The rate of deleterious mutations in these genes are lower than that in the global dataset of *K. var* as is shown in Section 3.7.

Interestingly, the bla_{SHV} gene, likely acquired through HGT from *K. pne* as discussed in the last chapter, was detected in five strains (4.63%) in the Scottish dataset (Figure 4.3). These strains were collected from UTIs in Edinburgh (Figure 4.3). The rate of horizontal bla_{SHV} transfer is slightly higher than that observed in the global *K. var* dataset, where 28 out of 715 strains (3.92%) carried bla_{SHV} (Figure 4.3).

Based on the metadata information provided by the hospital, none of the five bla_{SHV}-containing strains were co-isolated with other strains. Notably, all five bla_{SHV}-carrying strains carried the KL123 capsule locus, same as the other bla_{SHV}-carrying strain SPARK_2777_C1 which was co-isolated with *K. pne* as described in Section 3.6. KL123 was found in 4 out of 1000 *K. pne* strains as reported in the last chapter (Figure 3.11). This suggests that KL123 may be a capsule locus acquired from *K. pne* through recombination, along with the bla_{SHV} gene.

As shown in Figure 4.3, the virulence, resistance, and plasmid profiles of all Scottish *K. var* strains are visualised on the core genome phylogeny. We observed that eight strains (7.41%) carried the yersiniabactin siderophore, a rate slightly higher than that observed in the global *K. var* dataset, where 20 strains (2.80%) harboured yersiniabactin. In contrast, no strains in the Scottish *K. var* dataset was found to carry colibactin, aerobactin, salmochelin, or the *rmp* locus. In the Scottish *K. var* dataset, we did not observe the convergence of ESBL and yersiniabactin. However, there is one yersiniabactin-carrying strain that also carried *tet, sul*, bla_{TEM}, *aph, dfrA*, and *mph*, which was isolated from UTI.

The plasmid profile of the Scottish *K. var* strains (Figure 4.3) closely mirrors that of the global dataset and *K. pne* (Figure 3.10), with the IncFIB plasmid (count = 66, percentage = 61.1%) being the most prevalent plasmid type, followed by IncFII (count = 34, percentage = 31.5%) and IncFIA (count = 26, percentage = 24.1%). The count of the plasmid is not associated with the count of resistance genes (p value = 0.127).

The capsule profile of the Scottish *K. var* strains is shown in Figure 4.4. Compared to the global *K. var* dataset (Figure 3.11), the most prevalent capsule locus in the Scottish dataset is KL53 (count = 13, percentage = 10.2%), replacing KL60, which was dominant in the global dataset (count = 40, percentage = 56.0%). In the Scottish dataset, KL60 is now present in only three strains (percentage = 2.8%). We found 2 more novel capsule loci in the Scottish *K. var* strains as is shown in Figure 4.4B and C. Additionally, a significant proportion (38%) of capsule loci are found in two or fewer strains, consistent with our previous findings that the capsule locus exhibits high diversity.

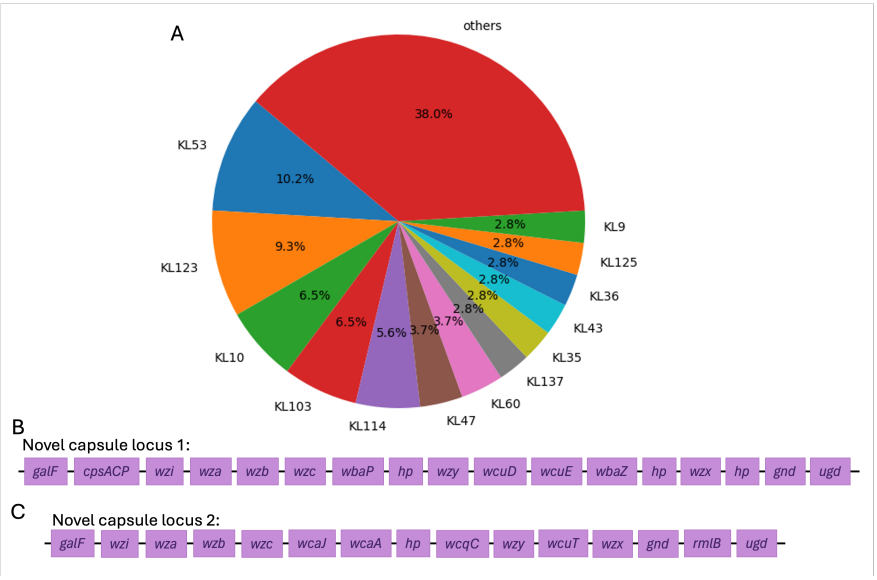


Figure 4.4: The capsule loci of Scotland clinical *Klebsiella variicola* strains. The capsule locus of each strain was built using Kaptive (Stanton et al., 2025). A: The capsule loci distribution pie-chart. B: Novel capsule locus found in strain kv_uoe_12. C: Novel capsule locus found in strain kv_uoe_99.

The *chbP*, *cspB*, and *lpxP* genes are present in 1 strain (kv_uoe_61), 4 strains (kv_uoe_45, kv_uoe_53, kv_uoe_62, kv_uoe_67), and 5 strains (kv_uoe_19, kv_uoe_45, kv_uoe_53, kv_uoe_62, kv_uoe_67), respectively.

In summary, the Scottish *K. var* strains showed a diverse resistance, virulence, plasmid, and capsule profile. The dataset showed a slightly lower average resistance gene count compared to the global dataset and higher rates of yersiniabactin carriage. The most prevalent plasmid type was IncFIB, and the capsule locus KL53 replaced KL60 as the most prevalent locus. These findings highlight the genetic diversity of the Scottish *K. var* strains and suggest the potential role of the accessory genome in their pathogenicity.

Next, to investigate whether the pathogenicity of the Scottish *K. var* strains are associated with the accessory genome rather than the core genome, we conducted pan-genome analyses to comprehensively examine the gene content of these strains.

4.5 Pan-genome and GWAS analyses

We utilised Roary (Page et al., 2015) to construct the pan-genome of the Scottish *K. var* strains and conducted GWAS to explore genetic differences between strains causing various host diseases and those collected from different Scottish cities.

The pan-genome analysis identified a total of 4426 core genes (present in $\geq$ 95% of strains) and 15,632 accessory genes (present in $<$ 95% of strains). Given that the dataset contains higher number of urine samples (count = 89) than any other sources (count = 19), we grouped the strains associated with other host diseases and compared them with those from urinary tract infections. Using Scoary (Brynildsrud et al., 2016), we analysed genes associated with host disease and collection location, with the results summarised in Table B.2 and Table B.3.

However, after applying Bonferroni and FDR corrections for *p* values, no genes were found to be significantly associated with either host disease or collection location. This suggests that there are no true genetic differences between *K. var* strains causing different diseases or those collected from different cities. These findings further highlight the plasticity of the *K. var* genome.

4.6 Discussion

The genomic analyses of the Scottish *K. var* strains provide further evidence supporting the extensive genetic diversity and plasticity of this species. The star-like population structure observed in the core genome phylogeny (Figure 4.1), with no clustering by host disease or collection location, is similar to the findings from the global *K. var* dataset (Figure 3.1).

Consistent with the findings from the global dataset where we found the European strains are the least resistant compared to the other continents, the Scottish *K. var* strains showed a lower prevalence of AMR genes and mutations, with only one ESBL-producing strain identified (Figure 4.3). This finding likely reflects the stricter antibiotic control policies in Scotland and further shows the adaptability of *K. var* as it can obtain more genes when there is antibiotic pressure, and exhibits less resistance in areas where antibiotics are better controlled. Importantly, we found 5 strains carrying bla_{SHV}, which is likely gained from *K. pne* together with the KL123 capsule locus.

The capsule locus diversity in the Scottish dataset remains high, with 2 novel capsule loci identified (Figure 4.4). Interestingly, the capsule loci found in the Scottish isolates differ significantly from those observed in the global dataset (Figure 3.11), further underscoring the variability in capsule profiles across *K. var* populations. Capsule locus KL123 was identified in all five bla_{SHV}-carrying Scottish strains, highlighting that this capsule locus was likely acquired from *K. pne* together with the bla_{SHV} gene. The plasmid profile (Figure 4.3), however, showed

remarkable similarity to the global dataset (Figure 3.10), with the IncFIB plasmid being the most prevalent type, and yersiniabactin being the only virulence factor consistently identified. The similarity in the plasmid type suggests potential gene transfer events with other *K. var* and *K. pne* isolates.

GWAS aimed at exploring differences in genetic composition by host disease or collection location did not identify any significant associations (Table B.3 and Table B.2. This suggests that there are no clear genetic distinctions between *K. var* strains causing different diseases or collected from different cities, further highlighting the plasticity and lack of disease specificity in the *K. var* genome. Interestingly, the genes identified in the previous chapter as significantly lost in human strains (Table 3.7) were rarely found in the Scottish dataset, further supporting the finding that these genes are undergoing deletion in human-associated strains.

In summary, these findings illustrate that while *K. var* displays high genetic diversity and adaptability, its population structure and genomic features remain consistent within Scotland. The gene transfer potential illustrated by the bla_{SHV} gene and the similar plasmid type reinforces the importance of continued surveillance to better understand the pathogenic dynamics of *K. var*.

Chapter 5

A Large-scale genomic analysis on the epidemiology, evolution, resistance, and virulence of *Klebsiella pneumoniae* epidemic lineage ST11

5.1 Introduction

This study focuses on examining how the most successful epidemic lineages of *K. pne* behave in human infections and investigate their population structure, resistance and virulence profiles, or evolutionary trajectories.

K. pne poses a significant threat to public health due to its capacity to cause severe infections and its increasing resistance to multiple antibiotics. The emergence of Multi-Drug Resistant (MDR), Extended-Spectrum Beta-Lactamase (ESBL)-producing, and hypervirulent strains of *Klebsiella* has escalated the urgency for vigilant surveillance and control measures. Among the various lineages, ST11 (R. Wang et al., 2024), ST15 (L. Feng, Zhang, & Fan, 2023), and ST258 (L. Chen, Mathema, Pitout, DeLeo, & Kreiswirth, 2014) are particularly concerning due to their widespread dissemination and high levels of resistance. In particular, ST11 has shown rapid and extensive spread worldwide, with China as a major epi-centre (Qi et al., 2011; R. Wang et al., 2024).

The epidemiology of ST11 in China reveals critical insights into the spread of carbapenem-resistant strains. Studies have identified a mutation in the *recC* gene as a contributor to the recombination rates in the ST11 cluster in China, underscoring the genetic factors driving resistance in this region (K. Zhou et al., 2023). Furthermore, a study involving 171 carbapenem-resistant *K. pne* isolates from ICU patients in China found ST11 to be the predominant lineage (154/171), with most strains carrying bla_{KPC} (Y. Liu et al., 2022). This study showed that within

the ST11 strains they collected, although all of them carried capsular type KL64, there are two distinct clones. Clone 1 prioritises gut colonisation and asymptomatic persistence via reduced virulence and superior ethanolamine utilisation, while clone 2 demonstrates stronger biofilm formation, environmental survival, and virulence (Y. Liu et al., 2022).

In Europe, carbapenemase-producing ST11 have also been widely disseminated (David et al., 2019), adding to the global burden of antimicrobial resistance associated with this lineage. Together, these findings highlight the adaptability and pathogenicity of ST11 across diverse geographical regions, driven by its acquisition of carbapenemase genes and other genetic mutations that promote survival under antimicrobial pressure.

In this study, we utilised the currently available ST11 strains from the NCBI RefSeq database, providing a robust dataset for analysis. The study aims to investigate the distribution, characteristics, and evolutionary trajectory of ST11 strains, with a particular focus on the genetic determinants underlying resistance, virulence, and adaptation. Understanding these factors is crucial for informing public health strategies and mitigating the threat posed by ST11 and related lineages.

5.2 Population structure of ST11, ST15 and ST258

K. pne ST11 is an emerging clade, widely disseminated in China (He et al., 2022). To understand its importance in human infection in comparison to other major clades in *Klebsiella*, we included ST15 and ST258 in our study. ST11 is a single *tonB* variant of ST258 [1], which is an epidemic clone contributing to the spread of carbapenem resistance. ST258 is predominantly found in North America, especially the USA, and in Europe, including countries such as Poland, Greece, and Norway (L. Chen et al., 2014). ST11 is widely disseminated in Asia, Brazil, and Europe, including Poland, Spain, and Greece, and is the dominant clone in China (L. Feng et al., 2023). ST15 is another widely disseminated high-risk clone, responsible for multiple hospital outbreaks (L. Feng et al., 2023).

The datasets of *K. pne* ST11 (count = 2194), ST15 (count = 848), and ST258 (count = 1193) *K. pne* strains were collected from 1997 to 2022 across six continents. Notably, the Asian isolates in ST11 and ST15 are largely from Southeast China. Most strains were isolated from humans (human strains per group: ST11 = 2033, percentage = 92.7%, ST15 = 794, percentage = 93.6%, ST258 = 1170, percentage = 98.1%), indicating strong human association of these lineages. Due to computational limitations of the computing cluster [2], we excluded ST11 strains lacking metadata (e.g., collection date, host, host disease, etc.). Consequently, our database included 1,157 ST11 strains, 1,156 of which were human strains, as shown in Figure 5.1.

1. https://bigsdb.pasteur.fr/
2. https://digitalresearchservices.ed.ac.uk/resources/eddie

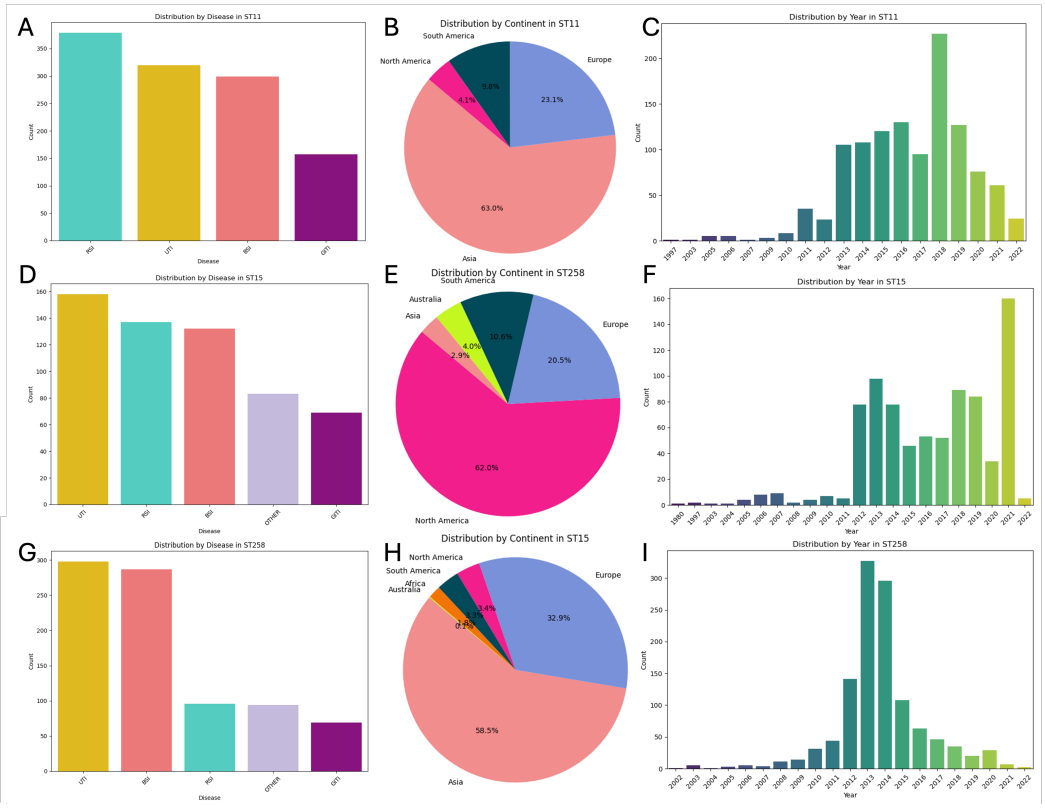


Figure 5.1: Metadata Plot of *Klebsiella pneumoniae* database. The metadata of each strain was obtained from the BioSample page on NCBI. **A, D, G**: Host disease distribution in ST11, 15 and 258 databases. The patient disease metadata was categorised into Respiratory System Infection (RSI), Urinary Tract Infection (UTI), Blood Stream Infection (BSI), Gastro-Intestinal Tract Infection (GITI) and others. **B, E, H**: Collection location distribution in ST11, 15 and 258 databases. The geographic location was categorized into continents. **C, F, I**: Collection time distribution in ST11, 15 and 258 databases.

ST11, ST15, and ST258 primarily cause Respiratory System Infection (RSI), Urinary Tract Infection (UTI), Blood Stream Infection (BSI), and Gastrointestinal Tract Infection (GITI) (Figure 5.1 A, D, G). The majority of ST11 and ST15 strains were collected from Asia, while most ST258 strains originated from North America (Figure 5.1 B, E, H).

ST11 and ST15 strains were predominantly collected after 2013 and 2012, respectively, with a marked increase in their emergence after 2018. Although the number of strains appears to have decreased in recent years, this is likely due to the lag between sequencing and publication, and the emergence of Covid. In contrast, ST258 strains emerged primarily between 2013 and 2014 but have been in decline since then, with less than 50 strains identified per year since 2017 (Figure 5.1 C, F, I). Thus, ST11 and ST15 represent more recent threats to human health compared to ST258, which experienced a significant outbreak in the past but is now coming under control.

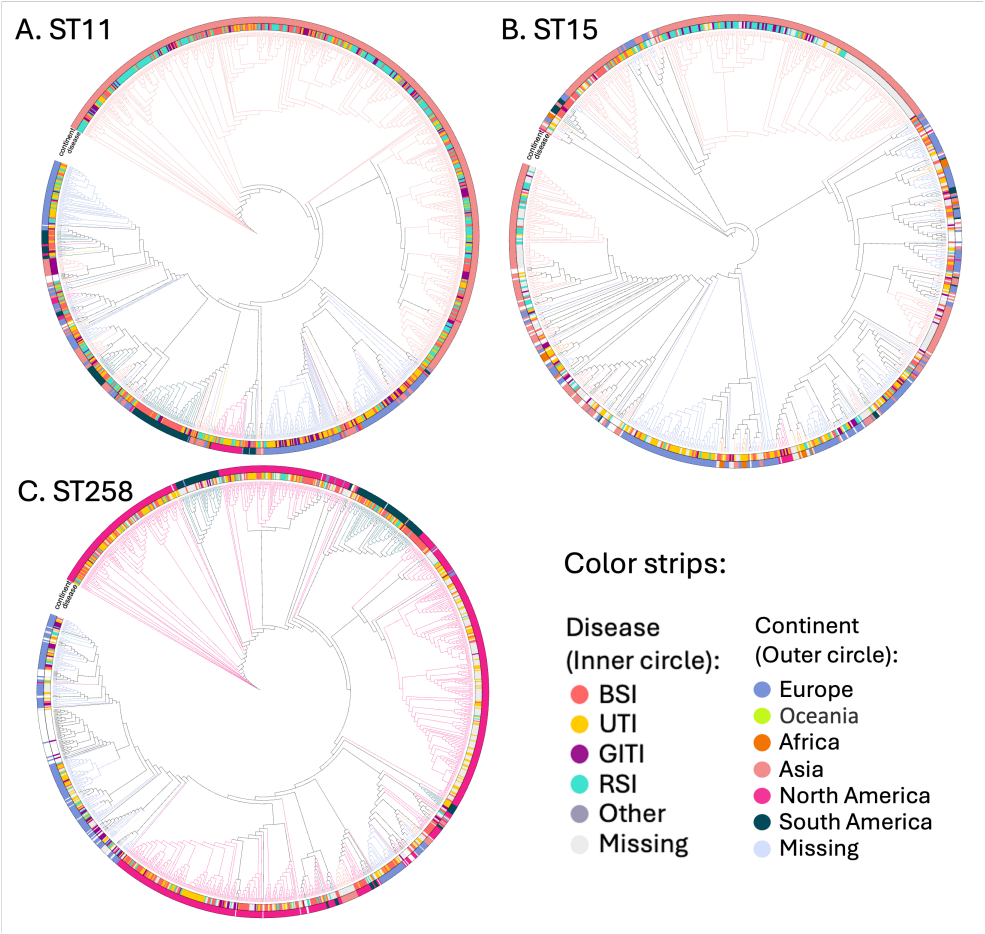


Figure 5.2: Recombination Free Unrooted Core Genome Phylogeny of ST11, 15 and 258. ParSNP (Treangen et al., 2014) was applied to remove recombination and built the core genome phylogenies for each ST. The host disease (inner circle) and location (outer circle and beach colour) information were visualised on the core genome phylogeny through iTol (Letunic & Bork, 2007). **A**: The core genome phylogeny of ST11. **B**: The core genome phylogeny of ST15. **C**: The core genome phylogeny of ST258.

Among all the strains in the ST11 dataset, more than half of the strains (63.0%) were collected from Asia (Figure 5.1), and the majority of the Asian strains (n = 621, percentage in Asian strains = 85.2%) clustered together, as shown in Figure 5.2A. Notably, strains from other continents formed a separate cluster from the Asian cluster (Figure 5.2A). The large Asian cluster in ST11 can be further divided into two major clades (Figure 5.2A.). In contrast, in the ST15 dataset, Asian strains are more sparsely distributed across the CG phylogeny, as shown in Figure 5.2B. Although there is an Asian strain cluster in the upper right of the CG phylogeny, this cluster in ST15 consists of fewer isolates and occupies a smaller section of the

CG phylogeny compared to the large Asian cluster in ST11. Figure 5.2C shows that, unlike ST11 and ST15, the CG phylogeny of ST258 includes far more isolates from North America, with Asian strains being rare. This indicates that ST11 is more closely associated with Asia, ST258 with North America, while ST15 shows a more widespread global distribution.

In terms of disease metadata (inner circles in Figure 5.2), there is no distinct clade associated with a specific host disease in any of the CG phylogenies. The disease metadata appears to be randomly distributed across all CG phylogenies, with multiple diseases occurring within the same clade, even among closely related strains.

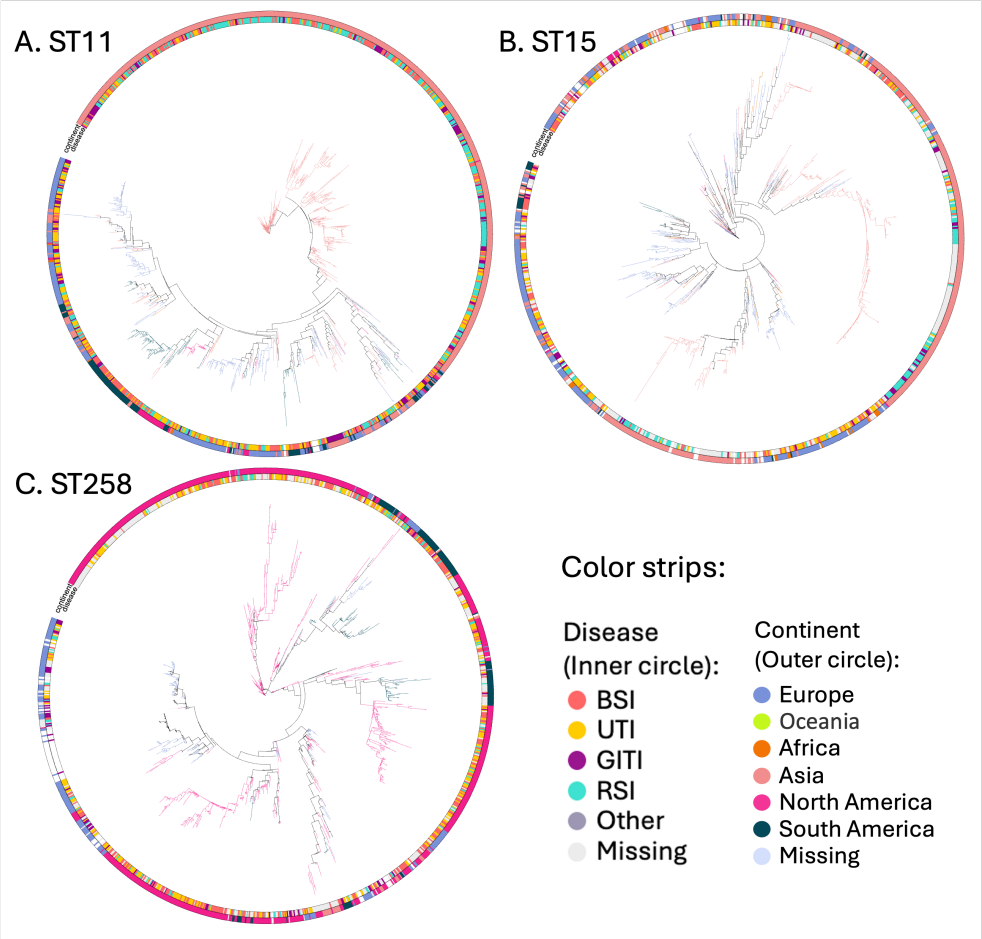


Figure 5.3: Unrooted Neighbour-joining presence and absence based accessory genome phylogeny of ST11, 15 and 258. The accessory genome of ST11, ST15 and ST258 were identified by Roary and further used to build the neighbour-joining accessory genome phylogeny (Page et al., 2015). The host disease and continent of collection strips are visualised on the phylogenies accordingly. **A:** Accessory genome phylogeny of ST11. **B:** Accessory genome phylogeny of ST15. **C:** Accessory genome phylogeny of ST258.

To better understand the accessory genome composition of ST11 in addition to the core genome phylogeny, we constructed the unrooted accessory genome phylogeny for ST11, ST15, and ST258 (Figure 5.3). As shown in Figure 5.3A, the Asian cluster remains distinct from the non-Asian cluster, which includes strains collected from various continents. For ST15 (Figure 5.3B), the smaller Asian cluster also displays a different accessory genome carriage compared to other strains. In both ST11 and ST15, the core and accessory genome phylogenies suggest the emergence of Asian clusters as distinct evolutionary lineages (Figure 5.3 and Figure 5.2). The Asian cluster in ST11 is divided into two sub-clades in both the core and accessory genome phylogenies, indicating two divergent evolutionary paths within a closely related cluster. Lastly, in ST258 (Figure 5.3C), the Asian isolates did not form a distinct cluster, suggesting they represent individual and sporadic transmission events. Once again, strains associated with the same host disease did not cluster together in any of the accessory genome phylogenies, suggesting that the genotype is not linked to the disease phenotype and may reflect just random or unrelated events.

To further investigate the Asian isolates, which are largely isolated from China, we visualized the collection locations of all isolates from China in ST11, ST15, and ST258 on a map (Figure 5.4). Both ST11 and ST15 are prevalent and widely distributed across China, especially the Southeast area. This may reflect a bias in our dataset, as Southeast China has a higher population density, and hospitals in Southwest China are better funded and equipped, allowing for more sequencing possibilities. Additionally, individuals from inland regions of China are more likely to seek medical care at these better-resourced hospitals. However, relative to ST15, ST11 has disseminated more extensively into inland regions, including the Southwest and Northeast areas of China. In contrast, ST258 is less prevalent, being found in only four locations in China, suggesting its inferior ability to compete with the other STs.

In this section, we explored the population structure of *K. pne* ST11, ST15, and ST258 across various geographic regions. ST11, the dominant clade in China, exhibited a distinct Asian cluster in both core and accessory genome phylogenies, suggesting the emergence of divergent lineages within this population. ST11 strains were more prevalent in inland regions of China compared to ST15, which, while also widely distributed, exhibited a smaller and more sporadically clustered Asian presence. ST258, on the other hand, was primarily associated with North America and showed limited dissemination in China.

The distinct evolutionary lineages in the Asian cluster of ST11, combined with its more extensive dissemination, suggest that this lineage may play a significant role in the global epidemiology of *Klebsiella*.

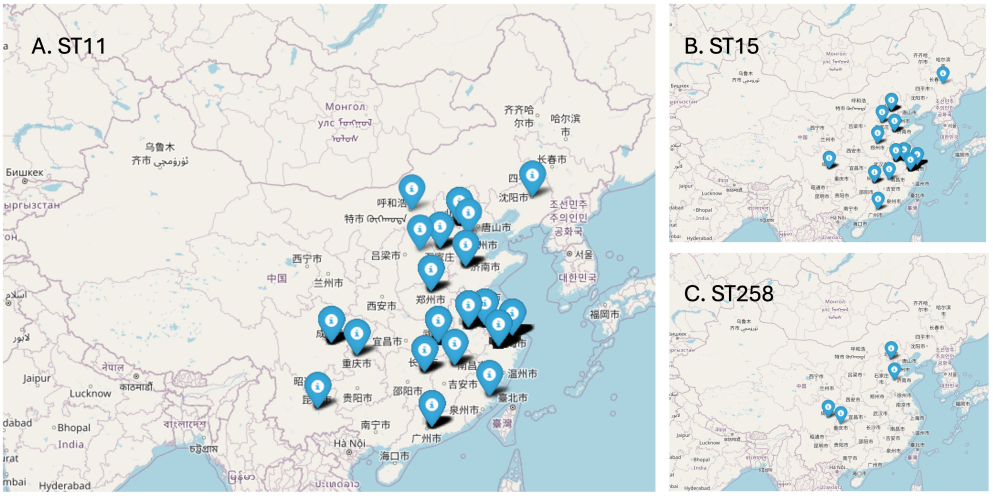


Figure 5.4: Distribution of isolates collected from China in ST11, 15 and 258. The locations of strains were visualised on the map using folium (python visualization, n.d.). **A:** Distribution of ST11 isolates in China. **B:** Distribution of ST15 isolates in China. **C:** Distribution of ST258 isolates in China.

5.3 Resistance and virulence score of ST11, ST15 and ST258

To understand the significance of ST11, ST15 and ST258 *K. pne* in human infections, we first built the resistance and virulence scores of these datasets using Kleborate (M. Lam et al., 2021). The resistance score was categorized into four levels which reflected the ESBL, carbapenem and colistin resistance, with higher scores indicating greater resistance (Table 5.1). For the virulence score, there were six levels related to yersiniabactin, colibactin and aerobactin as shown in Table 5.2.

Resistance score	ESBL	Carbapenem	Colistin
0	-	-	+/-
1	+	-	+/-
2	+/-	+	-
3	+/-	+	+

Table 5.1: Descriptions of Different Resistance Scores Provided by Kleborate manual (M. Lam et al., 2021).

Kleborate assigns resistance scores ranging from 0 to 3, where 0 represents the least resistant (no ESBLs or carbapenemases) and 3 indicates the highest resistance level (presence of carbapenemase with colistin resistance), as shown in Table 5.1 (M. Lam et al., 2021). Similarly, it assigns virulence scores from 0 to 5, with 0 indicating the absence of yersiniabactin, colibactin, and aerobactin, and 5 representing the presence of all three, as shown in Table 5.2 (M. Lam et al., 2021).

Virulence Score	Yersiniabactin	Colibactin	Aerobactin
0	-	-	-
1	+	-	-
2	+/-	+	-
3	-	-	+
4	+	-	+
5	+	+	+

Table 5.2: Descriptions of Different Virulence Scores Provided by Kleborate manual (M. Lam et al., 2021).

The resistance and virulence scores of ST11, ST15, and ST258 across different continents are visualised in Figures 5.5A and 5.5C, with comparisons between different sequence types shown in Figures 5.5B and 5.5D.

Resistance and Virulence Scores of ST11: ST11 isolates predominantly exhibit a resistance score of 2, indicating resistance to ESBLs and/or carbapenems. This distribution is consistent across continents, with no significant difference observed between strains from Europe and Asia (p value = 0.0719), although South American strains show the lowest resistance scores than other continents (p values < 0.01). Regarding virulence, ST11 shows higher virulence scores compared to ST15 (p value = 0.0414) and ST258 (p value < 0.001). The ST11 population splits into two major groups: one with a virulence score of 4 (positive for aerobactin and yersiniabactin) and another with a score of 1 (positive only for yersiniabactin). Notably, most Asian ST11 strains belong to the higher virulence score group, showing a convergence of resistance and virulence traits.

Resistance and Virulence Scores of ST15: ST15 exhibits the lowest resistance scores among the three lineages (p values < 0.001). Its resistance profile indicates less extensive exposure to selective pressures compared to ST11 and ST258. For virulence, ST15 also splits into two groups, similar to ST11: one with a virulence score of 4 (aerobactin and yersiniabactin positive) and another with a score of 1 (yersiniabactin positive only). While its virulence scores are generally lower than those of ST11 (p value = 0.0414), they are significantly higher than ST258 (p value < 0.001), reflecting an intermediate pathogenic profile.

Resistance and Virulence Scores of ST258: ST258 shows the highest resistance scores among the three lineages (p values < 0.001), with a significant proportion of isolates carrying carbapenemase genes. However, its virulence scores are the lowest (p values < 0.001), with three distinct groups observed: 0 (negative for yersiniabactin, colibactin, and aerobactin), 1 (yersiniabactin positive only), and 2 (colibactin and yersiniabactin positive). This distribution highlights ST258's emphasis on resistance mechanisms over virulence traits, making it primarily a multidrug-resistant lineage rather than a highly virulent one.

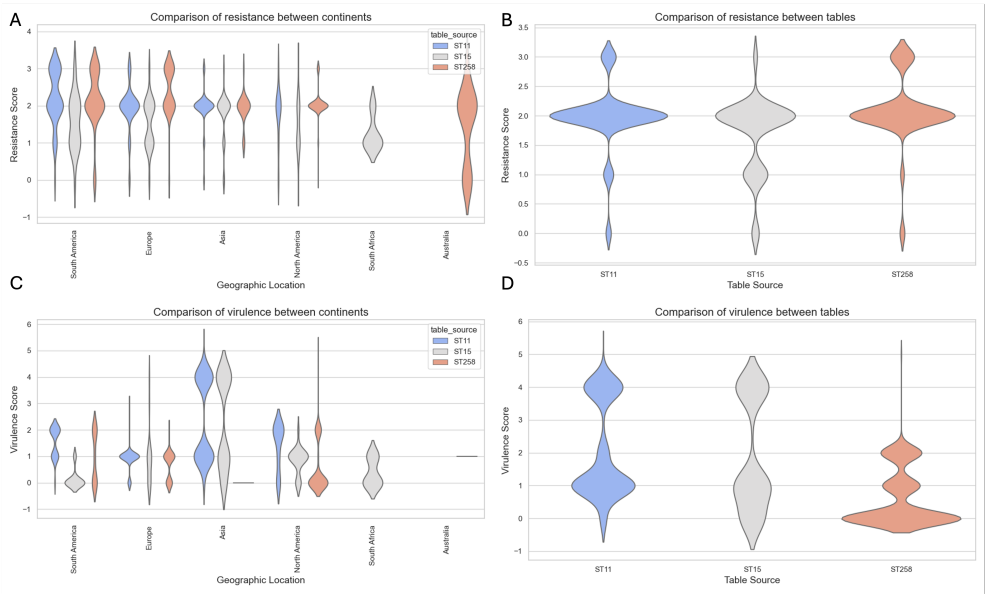


Figure 5.5: Resistance and virulence score distribution of *Klebsiella* ST11, ST15 and ST258. The resistance and virulence score of each strain were identified using Kleborate (M. Lam et al., 2021). **A**: Comparison of resistance score between ST11, ST15 and ST258 strains isolated from different continents. **B**: Comparison of resistance score between different STs. **C**: Comparison of virulence score between ST11, ST15 and ST258 strains isolated from different continents. **D**: Comparison of virulence score between different STs.

Comparison of ST11, ST15, and ST258: While ST11, ST15, and ST258 exhibit similar overall resistance score distributions, ST258 stands out with the highest resistance scores, followed by ST11, and ST15 with the lowest. In contrast, virulence scores differ more distinctly. ST11 demonstrates the highest virulence scores, particularly in Asian isolates, with a significant convergence of resistance and virulence traits. ST15, though less virulent than ST11, maintains an intermediate profile with some isolates in the higher virulence score group. ST258, however, exhibits minimal virulence, focusing on resistance instead. These differences reflect varying evolutionary trajectories and selective pressures among the three lineages.

5.4 Resistance and virulence profile of ST11

Next, we built the resistance and virulence gene carriage profiles of the ST11 dataset. For ST11, as shown in Figure 5.6, we visualized the virulence (green) and resistance profiles (red for present, blue for mutated) of ST11 strains on the core genome tree. The host disease and niche metadata are annotated using the same colours as Figure 5.1. The detailed count and percentage of virulence and resistance genes in Figure 5.1 are shown in Table 5.3 and 5.4.

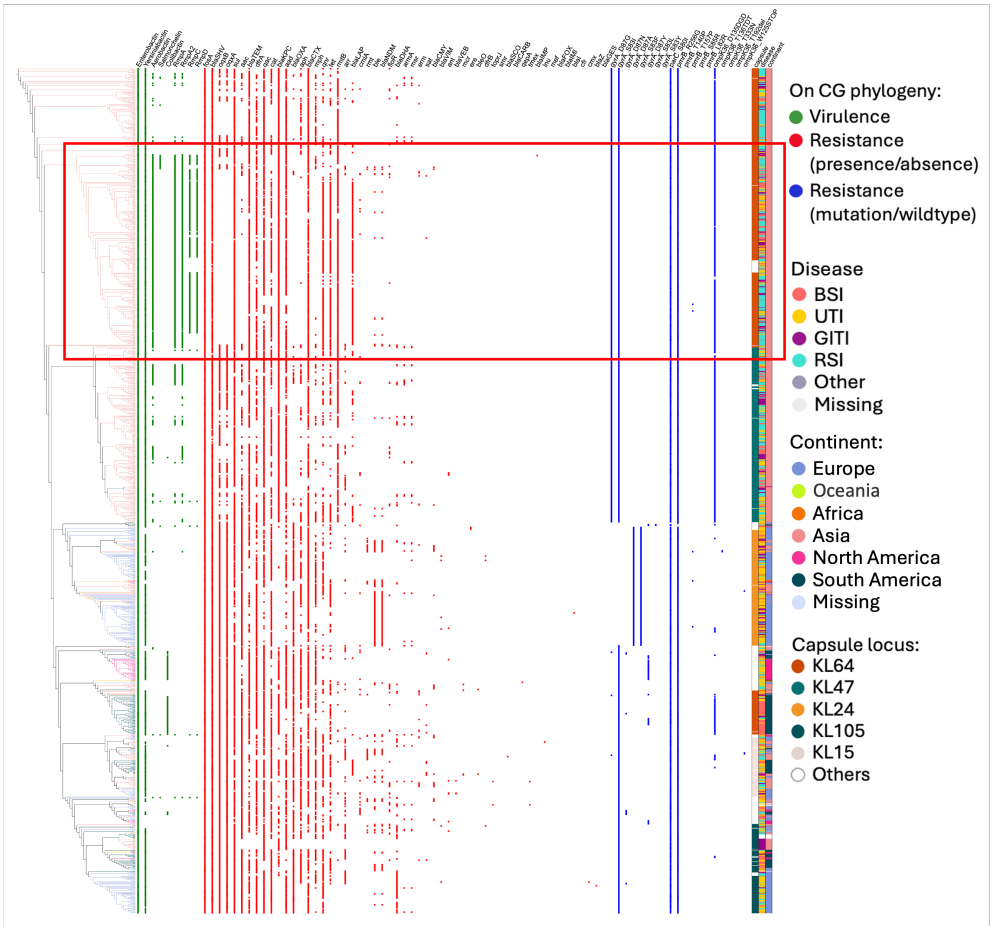


Figure 5.6: The resistance and virulence gene profile of *Klebsiella* ST11. Resistance and virulence gene carriage of ST11 strains were built using AMRfinderplus (Feldgarden et al., 2021) or Kleborate (M. Lam et al., 2021), and mapped on the core genome phylogeny using iTol (Letunic & Bork, 2007), with capsule profile, patient disease, and collection location annotated. The branch colour of the tree corresponds to the continent metadata. The Clonal Expansion (CE) cluster is highlighted in the red box.

We observed that almost all ST11 strains are enterobactin positive (count = 1151, percentage = 99.48%), consistent with previous reports (Russo et al., 2018; K. Zhou et al., 2023). Yersiniabactin is the second most prevalent siderophore in ST11 (count = 1051, percentage = 90.84%) (Table 5.3). In *Klebsiella*, enterobactin and yersiniabactin play crucial roles in iron acquisition, essential for bacterial growth and virulence, particularly in iron-limited environments such as the human host (Piperaki, Syrogiannopoulos, Tzouvelekis, & Daikos, 2017). Enterobactin, a high-affinity siderophore, sequesters iron from the host, while yersiniabactin is associated with increased virulence and the ability to evade host immune responses. Both siderophores contribute to the pathogenicity of *Klebsiella*, enabling it to establish infections, including in the urinary tract and bloodstream (Lawlor, O'connor, & Miller, 2007).

Table 5.3: Virulence factors and Corresponding Carriage in *Klebsiella pneumoniae* ST11, ST15, and ST258.

Virulence	ST11 count	ST11 percentage	ST15 count	ST15 percentage	ST258 count	ST258 percentage
enterobactin	1151	99.48	818	96.46	1193	100.00
yersiniabactin	1051	90.84	524	61.79	593	49.71
aerobactin	339	29.30	243	28.66	3	0.25
salmonchelin	23	1.99	0	0.00	1	0.08
colibactin	94	8.12	1	0.12	370	31.01
rmpA	310	26.79	40	4.72	0	0.00
rmpA2	329	28.44	221	26.06	0	0.00
rmpC	201	17.37	5	0.59	0	0.00
rmpD	202	17.46	5	0.59	0	0.00

Aerobactin is predominantly found in the Asian cluster (count = 339, percentage = 29.30%), while colibactin is enriched in a North American and a South American cluster (count = 94, percentage = 8.12%) (Table 5.3). Salmonchelin is rarely detected in ST11 (count = 23, percentage = 1.99%).

The same Asian cluster carries the entire *rmp* locus, whereas outside this cluster, the *rmp* locus is either nearly absent or incomplete (*rmpA*: count = 310, percentage = 26.80%; *rmpA2*: count = 329, percentage = 28.44%; *rmpC*: count = 201, percentage = 17.37%; *rmpD*: count = 202, percentage = 17.46%) (Table 5.3).

Resistance genes, including *fosA*, bla_{SHV}, and *sul*, are present in more than 80% of ST11 strains (Table 5.4), suggesting resistance to phosphonic antibiotics, β-lactams, and antifolates. Mutations in *gyrA* and *parC*are nearly universal in ST11 strains (Table 5.4). The resistance and virulence profiles of Asian and non-Asian ST11 clusters differ significantly, particularly in the carriage of aerobactin, the *rmp* locus, bla_{LAP}, absence of *oqxAB*, and the mutation of *ompK36*.

Table 5.4: Resistance Genes or Mutations and Corresponding Carriage in *Klebsiella pneumoniae* ST11, ST15 and ST258.

Resistance	ST11 count	ST11 percentage	ST15 count	ST15 percentage	ST258 count	ST258 percentage
fosA	1154	99.74	820	96.70	1193	100.00
bla_{SHV}	1100	95.07	812	95.75	1175	98.49
oqxB	726	62.75	812	95.75	1172	98.24
oqxA	705	60.93	813	95.87	1172	98.24

Table 5.4: Resistance Genes or Mutations and Corresponding Carriage in *Klebsiella pneumoniae* ST11, ST15 and ST258.

Resistance	ST11 count	ST11 percentage	ST15 count	ST15 percentage	ST258 count	ST258 percentage
sul	986	85.22	641	75.59	1051	88.10
aac	570	49.27	720	84.91	1130	94.72
bla_{TEM}	773	66.81	607	71.58	1010	84.66
dfrA	696	60.16	645	76.06	956	80.13
qac	872	75.37	371	43.75	943	79.04
cat	739	63.87	577	68.04	734	61.53
bla_{KPC}	766	66.21	164	19.34	1147	96.14
aad	757	65.43	277	32.67	1006	84.33
bla_{OXA}	445	38.46	583	68.75	885	74.18
aph	385	33.28	561	66.16	935	78.37
bla_{CTX}	884	76.40	670	79.01	103	8.63
mph	359	31.03	291	34.32	750	62.87
qnr	513	44.34	459	54.13	45	3.77
tet	476	41.14	341	40.21	159	13.33
rmtB	508	43.91	21	2.48	76	6.37
arr	139	12.01	268	31.60	23	1.93
bla_{LAP}	271	23.42	20	2.36	1	0.08
cmlA	38	3.28	21	2.48	223	18.69
rmt	45	3.89	202	23.82	5	0.42
ble	158	13.66	71	8.37	3	0.25
bla_{NDM}	153	13.22	66	7.78	2	0.17
floR	61	5.27	53	6.25	86	7.21
bla_{DHA}	109	9.42	51	6.01	0	0.00
armA	43	3.72	109	12.85	0	0.00
msr	41	3.54	102	12.03	4	0.34
erm	24	2.07	11	1.30	61	5.11
sat	13	1.12	1	0.12	57	4.78
bla_{CMY}	35	3.03	16	1.89	4	0.34
bla_{VIM}	6	0.52	39	4.60	0	0.00
ant	10	0.86	4	0.47	29	2.43
bla_{VEB}	3	0.26	22	2.59	8	0.67
mcr	3	0.26	16	1.89	1	0.08
ere	3	0.26	13	1.53	0	0.00
bleO	2	0.17	10	1.18	0	0.00
dfrB	3	0.26	8	0.94	1	0.08

Table 5.4: Resistance Genes or Mutations and Corresponding Carriage in *Klebsiella pneumoniae* ST11, ST15 and ST258.

Resistance	ST11 count	ST11 percentage	ST15 count	ST15 percentage	ST258 count	ST258 percentage
toprJ	2	0.17	4	0.47	0	0.00
smr	0	0.00	8	0.94	0	0.00
bla_{SCO}	1	0.09	6	0.71	1	0.08
bla_{CARB}	0	0.00	0	0.00	8	0.67
qepA	1	0.09	7	0.83	0	0.00
tmex	2	0.17	4	0.47	0	0.00
bla_{IMP}	1	0.09	5	0.59	0	0.00
lnu	1	0.09	2	0.24	0	0.00
mef	0	0.00	3	0.35	0	0.00
bla_{FOX}	0	0.00	0	0.00	3	0.25
bla_{IMI}	0	0.00	1	0.12	0	0.00
blaI	1	0.09	0	0.00	0	0.00
cfr	0	0.00	0	0.00	1	0.08
cmx	1	0.09	0	0.00	0	0.00
blaZ	1	0.09	0	0.00	0	0.00
bla_{GES}	0	0.00	1	0.12	0	0.00
*gyrA*_D87G	629	54.36	0	0.00	4	0.34
*gyrA*_S83I	985	85.13	16	1.89	1193	100.00
*gyrA*_D87N	10	0.86	0	0.00	174	14.59
*gyrA*_D87A	160	13.83	783	92.33	0	0.00
*gyrA*_S83F	160	13.83	778	91.75	0	0.00
*gyrA*_D87Y	47	4.06	0	0.00	38	3.19
*gyrA*_S83L	2	0.17	0	0.00	0	0.00
*gyrA*_S83Y	0	0.00	7	0.83	0	0.00
*parC*_S80I	1149	99.31	800	94.34	1193	100.00
*pmrB*_R256G	1148	99.22	0	0.00	1193	100.00
*pmrB*_T140P	0	0.00	1	0.12	1	0.08
*pmrB*_T157P	5	0.43	3	0.35	22	1.84
*pmrB*_S85R	0	0.00	1	0.12	1	0.08
*pmrB*_L82R	0	0.00	0	0.00	1	0.08
*ompK36*_D135DGD	671	57.99	65	7.67	349	29.25
*ompK36*_T136TDT	2	0.17	0	0.00	7	0.59
*ompK36*_T333N	0	0.00	0	0.00	1	0.08
*ompK36*_T192del	0	0.00	0	0.00	1	0.08

Table 5.4: Resistance Genes or Mutations and Corresponding Carriage in *Klebsiella pneumoniae* ST11, ST15 and ST258.

Resistance	ST11 count	ST11 percentage	ST15 count	ST15 percentage	ST258 count	ST258 percentage
*ompK36*_W125STOP	3	0.26	3	0.35	0	0.00

In *Klebsiella*, *gyrA* and *parC* mutations have been shown to be related to quinolone resistance by altering DNA gyrase and topoisomerase IV, the enzymes targeted by quinolones (Rezaei et al., 2024). In *K. pne*, strains carrying the combination of GyrA S83L, D87G, and ParC S80I mutations exhibited resistance to fluoroquinolones, including ciprofloxacin and levofloxacin (Mirshekar et al., 2024). Another study involving 42 *K. pne* strains isolated in Iraq revealed that the GyrA S83L mutation was observed in all 27 (100%) quinolone-resistant isolates, making it the primary mutation (Kareem et al., 2021). The GyrA D87G mutation was present in 24 out of 27 (88.9%) isolates. Additional mutations in the *parC* gene, such as S80I, were found in 21 out of 27 (77.8%) isolates, suggesting that the *gyrA* mutation precedes the *parC* mutation (Kareem et al., 2021). Another combination of GyrA S83I, D87G, and ParC S80I mutations was previously observed in ST11 *K. pne* isolated in China, where these mutated strains also exhibited fluoroquinolone resistance (Zeng et al., 2020). A previous study on the ST15 *K. pne* lineage showed that a different combination of GyrA mutations, including S83F and D87A, could lead to ciprofloxacin resistance (Álvarez et al., 2022).

As shown in Figure 5.6 and Table 5.4, the corresponding mutations, including ParC S80I (count = 1149, percentage = 99.31%) and mutations in GyrA (D87G: count = 629, percentage = 54.36%; S83I: count = 985, percentage = 95.13%; D87A and S83F: count = 160, percentage = 13.83%), are present in nearly all ST11 strains.

The Asian cluster in ST11 showed a different combination of GyrA D87G and S83I mutations. We infer that strains in this Asian cluster (GyrA D87G and S83I, and ParC S80I) are fluoroquinolone-resistant, as they carry the same mutations reported previously (Zeng et al., 2020). The European cluster showed a combination of GyrA S83F and D87A mutations. We infer that these strains are ciprofloxacin-resistant, given the same GyrA mutations as those in ST15 strains (Álvarez et al., 2022).

The remaining strains carry only GyrA S83I, while a North American cluster also carries the D87Y mutation. The combination of GyrA D87Y, S83L, and ParC S80I has been reported in *K. pne* ST42 strains (Takei et al., 2022). We infer that these strains may not be resistant to quinolones, as the combination of D87Y and S83L, or S83I alone in GyrA has not been reported.

Additionally, the PmrB protein is a component of the PmrA/PmrB two-component regulatory system, which is related to the expression of genes involved in lipopolysaccharide modification in the bacterial outer membrane (Mirshekar et al., 2024). Mutations in PmrB, such as T157P, have been reported to lead to constitutive activation of the PmrA protein, resulting in upregulation of the *pmrCAB* and *pmrHFIJKLM* operons (Jayol et al., 2014). This upregulation promotes the addition of positively charged groups, such as phosphoethanolamine, to LPS, reducing colistin's binding affinity and conferring colistin resistance. In the ST11 dataset, the T157P mutation in PmrB is present in 5 strains (percentage = 0.43%) (Figure 5.6), which shows that the majority of ST11 strains have not gained resistance to colistin.

Interestingly, in the ST11 dataset, the R256G mutation in PmrB is present in nearly all strains (count = 1148, percentage = 99.22%) (Figure 5.6). However, studies have suggested that the R256G substitution in PmrB has been observed in both polymyxin-resistant and polymyxin-susceptible isolates, implying it may not be directly implicated in polymyxin resistance (Sampaio et al., 2021). Based on these findings, while we cannot definitively recognise R256G as a mutation directly linked to colistin resistance, we acknowledge its correlation and potential relevance in resistance mechanisms.

Furthermore, the D135DGD insertion in OmpK36 was observed in the entire Asian cluster and in one of the South American clusters. The GD insertion is located in the extracellular loop 3 region, causing a constriction of the porin, which restricts the entry of carbapenems (David et al., 2022). A previous study showed that the D135DGD insertion is present in 3,588 (24.1%) out of 14,888 *K. pne* strains, and is specifically enriched in the ST258/512 clone, which is widely disseminated in North America and Europe (David et al., 2022). Here, we found that ST11, a close relative of ST258, exhibits the same insertion in 58.60% of the strains (615 strains from the Asian cluster, and 56 strains from the non-Asian cluster out of 1157 ST11 strains, percentage = 57.99%), indicating a cladistic relationship with the Asian cluster (Figure 5.6).

Notably, we identified a Clonal Expansion (CE) cluster within the Asian cluster of ST11, highlighted in the red box in Figure 5.6. A previous study on ST11 reported similar findings, demonstrating that the KL47 capsular type lineage was replaced by the KL64 lineage, which included a hypervirulent sublineage characterised by enhanced resistance (K. Zhou et al., 2023). Although the dataset used in this study was available on NCBI, it did not overlap with ours, as our dataset was exclusively downloaded from NCBI RefSeq, which does not include these strains. The strains in this CE cluster carry bla_{LAP} and have lost the *oqxAB* locus. The mutations, including GyrA D87G and S83I, ParC S80, suggest resistance to fluoroquinolones and probably colistin. Additionally, the D135DGD insertion in OmpK36 suggests resistance to carbapenems. This CE cluster also carries the *rmp* locus and aerobactin, which are associated with the hypervirulent phenotype in *K. pne*. Therefore, the CE cluster in ST11 demonstrates the convergence of resistance and virulence, posing a significant public health threat.

The majority of strains in the CE cluster possess the KL64 capsule locus, while the rest of the Asian strains carry KL47 (Figure 5.6). This CE cluster corresponds to the virulent cluster in Asia, as seen in Figure 5.5C. The previous report indicated that the subclone in their study (which corresponds to the CE cluster in this study) of ST11 is carbapenem-resistant, carries the KL64 capsule locus, and contains a higher load of mobile genetic elements (K. Zhou et al., 2023).

The CE cluster in ST11 has also spread to inland regions of China, including the Southwest, as shown in Figure 5.7, highlighting the urgent need for enhanced surveillance, particularly for strains demonstrating a convergence of resistance and virulence. Notably, the CE cluster appears to be present in all locations where other Chinese ST11 strains have been identified, which indicates that the CE cluster originated from the non-CE clusters.

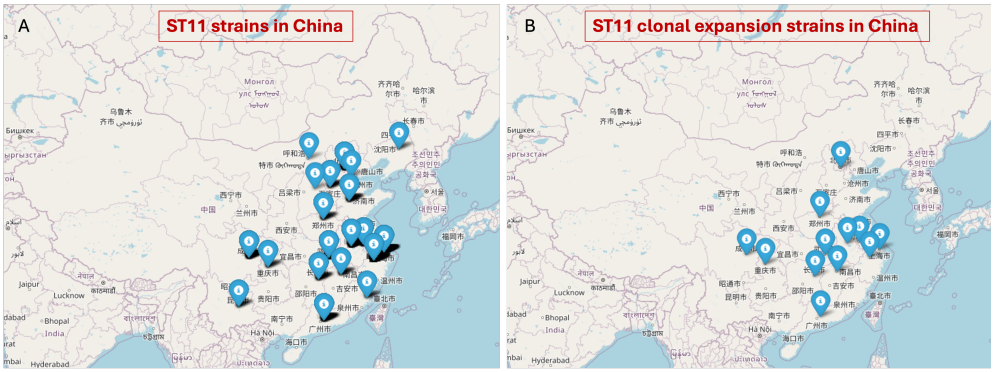


Figure 5.7: Distribution comparison of all and clonal expansion *Klebsiella pneumoniae* ST11. The distribution of ST11 was visualised using folium package in Python (python visualization, n.d.). **A**: Distribution of all ST11 strains collected in China. **B**: Distribution of only clonal expansion cluster in ST11.

In this section, we compared the resistance and virulence profiles of ST11, which demonstrated a convergence of resistance and virulence, particularly within the CE cluster found in the Asian strains. ST11 strains were highly resistant to multiple antibiotics, including fluoroquinolones, phosphonic antibiotics, and carbapenems due to key mutations such as GyrA D87G and S83I, ParC S80I, and the D135DGD insertion in OmpK36. Additionally, ST11 exhibited significant virulence, with the widespread presence of siderophores such as enterobactin, yersiniabactin, and aerobactin, along with the *rmp* locus, which are associated with hypervirulence.

5.5 Resistance and virulence profile of ST15 and ST258

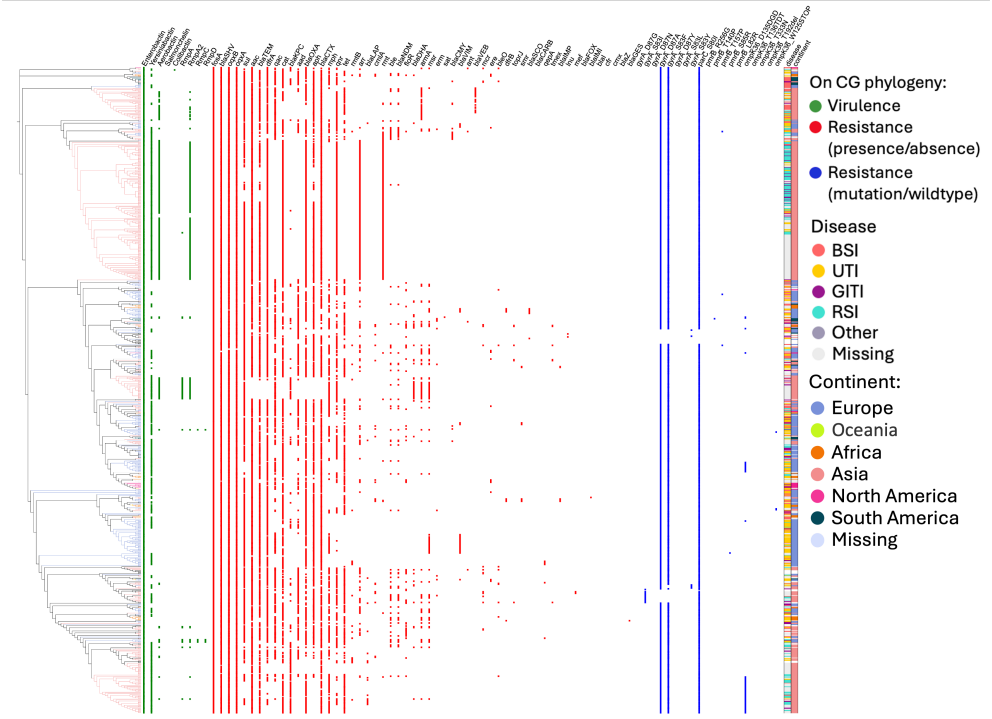


Figure 5.8: The resistance and virulence gene profile of *Klebsiella* ST15. Resistance and virulence gene carriage of ST15 strains were built using AMRfinderplus (Feldgarden et al., 2021) or Kleborate (M. Lam et al., 2021) and mapped on the core genome phylogeny using iTol (Letunic & Bork, 2007), with patient disease, collection location and capsule profile annotated. The branch colour of the tree corresponds to the continent metadata.

To contrast with ST11, we also built the resistance and virulence profiles of ST15 and ST258 as shown in Figure 5.8 and Figure 5.9. ST15 is more widely distributed globally, while strains from Asia and Europe form scattered clusters (Figure 5.8). Enterobactin is present in nearly all strains (count = 818, percentage = 96.46%), while yersiniabactin is found in 524 strains (percentage = 61.79%). Colibactin is present in only 1 strain (percentage = 0.12%), and Salmonchelin is absent in ST15 (count = 0) (Table 5.4). We observed an Asian cluster in ST15 that carries aerobactin, *rmpA2*, *cat*, *arr*, and *rmt*. The presence of the *rmpA2* gene alone is unlikely to result in the hypermucoviscosity phenotype unless accompanied by *rmpD* (Walker et al., 2020; X. Yang et al., 2023). Resistance genes such as *fosA*, bla_{SHV}, *oqxAB*, and *aac* are present in more than 80% of ST15 strains (Table 5.4). Mutations commonly observed in ST15 include ParC S80I (count = 800, percentage = 94.24%), GyrA D87A (count = 783, percentage = 92.33%), and GyrA S83F (count = 778, percentage = 91.75%), which

could lead to ciprofloxacin resistance as previously reported in ST15 (Álvarez et al., 2022) (Figure 5.12). PmrB T157P mutation only appeared in only 3 ST15 strains, suggesting limited colistin resistance. The OmpK36 D135DGD insertion was found in only 65 strains (percentage = 7.67%).

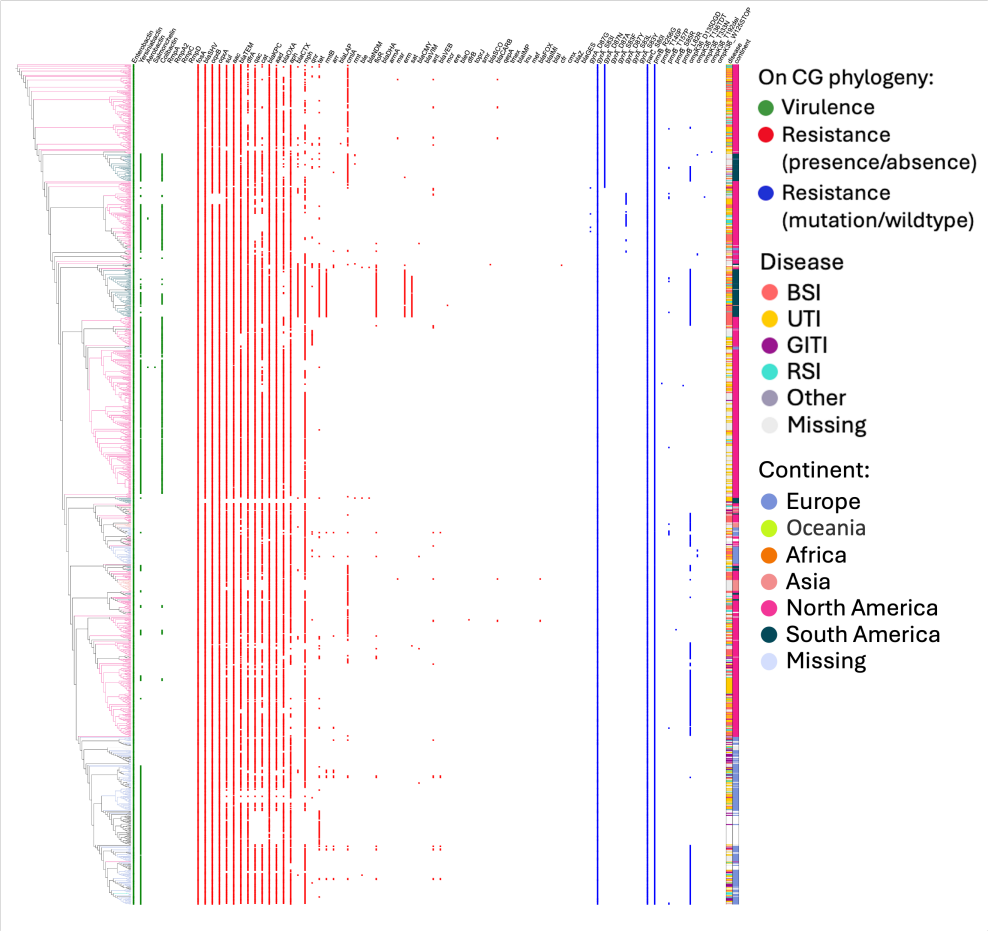


Figure 5.9: The resistance and virulence gene profile of *Klebsiella* ST258. Resistance and virulence gene carriage of ST258 strains were built using AMRfinderplus (Feldgarden et al., 2021) or Kleborate (M. Lam et al., 2021), and mapped on the core genome phylogeny using iTol (Letunic & Bork, 2007), with patient disease, collection location and capsule profile annotated. The branch colour of the tree corresponds to the continent metadata.

ST258 is primarily associated with North America and not with any other continent (Figure 5.9). Although all ST258 strains carry enterobactin (count = 1193, percentage = 100.00%), yersiniabactin is less prevalent than ST11 and ST15 (count = 593, percentage = 49.17%) (Table 5.4). Colibactin is present in one South American cluster and two North American clusters (count = 370, percentage = 31.01%), and salmonchelin was found in only 1 ST258 strain (Table 5.4). The *rmp* gene cluster was absent in our ST258 dataset (count = 0), indicating that ST258 is unlikely to exhibit a hypermucoviscous phenotype. In terms of resistance, *fosA*, bla_{SHV}, *oqxAB*, *sul*, *aac*, bla_{TEM}, *dfrA*, bla_{KPC}, and *aad* are present in more than 80%

of ST258 strains (Table 5.4). Mutations such as GyrA S83I, ParC S80I, and PmrB T157G are present in 22 ST258 strains (percentage = 1.84%), while GyrA D87G was detected in only 4 strains (percentage = 0.34%), suggesting limited resistance to fluoroquinolones (Zeng et al., 2020). The OmpK36 D135DGD insertion was found in 29.25% of the strains (count = 349) (Table 5.4).

In contrast to ST11, ST15 and ST258 displayed less resistance and virulence. Although ST15 strains were globally distributed and showed notable fluoroquinolone resistance due to mutations in GyrA and ParC, they lacked the significant virulence factors such as the *rmp* locus. ST258, primarily associated with North America, exhibited extensive resistance to carbapenems and other antibiotics but lacked key virulence factors, particularly the *rmp* locus and the siderophores, making it less likely to be hypervirulent.

Because ST11 is more resistant (especially to fluoroquinolones, phosphonic antibiotics, and carbapenems) and more virulent (due to siderophores and the *rmp* locus) than ST15 and ST258, we extended our analysis to include more porins and pumps that linked to membrane permeability to explore more about the resistance mechanism of ST11.

5.6 Membrane permeability proteins

To further investigate the resistance mechanisms in the ST11 lineage, we extended our study to include resistance genes associated with pumps and porins related to membrane permeability.

We based the selection of these intrinsic loci on well-characterised permeability factors that control efflux (AcrAB, OqxAB), influx (OmpK35, OmpK36), and the regulation of permeability through various regulatory genes (*ramR, marR, acrR, oqxR*). The AcrAB-TolC efflux pump plays a key role in excluding toxic substances, including antibiotics, from the bacterial cell. AcrR is the repressor of the AcrAB efflux pump, which is capable of pumping a broad range of antibiotics, including chloramphenicol, tetracyclines, β-lactams, and fluoroquinolones (Padilla et al., 2010). OqxR is the repressor of the OqxAB efflux pump, responsible for the exclusion of antimicrobials such as olaquindox, ciprofloxacin, and tetracycline (J. Li et al., 2019). OmpK35 and OmpK36 are critical for the permeability of the bacterial outer membrane to antibiotics, where the loss or reduced expression of these porins is associated with increased antibiotic resistance (Tsai et al., 2011). RamR is a repressor protein that inhibits the expression of RamA (Rosenblum et al., 2011). Mutations in *ramR* can lead to the overexpression of RamA, which in turn increases the expression of the AcrAB-TolC efflux pump, resulting in higher levels of antibiotic resistance (Rosenblum et al., 2011; Schneiders et al., 2003). Additionally, RamA influences the expression of outer membrane porins such as OmpK35 and OmpK36, contributing further to antibiotic resistance (De Majumdar et al., 2015).

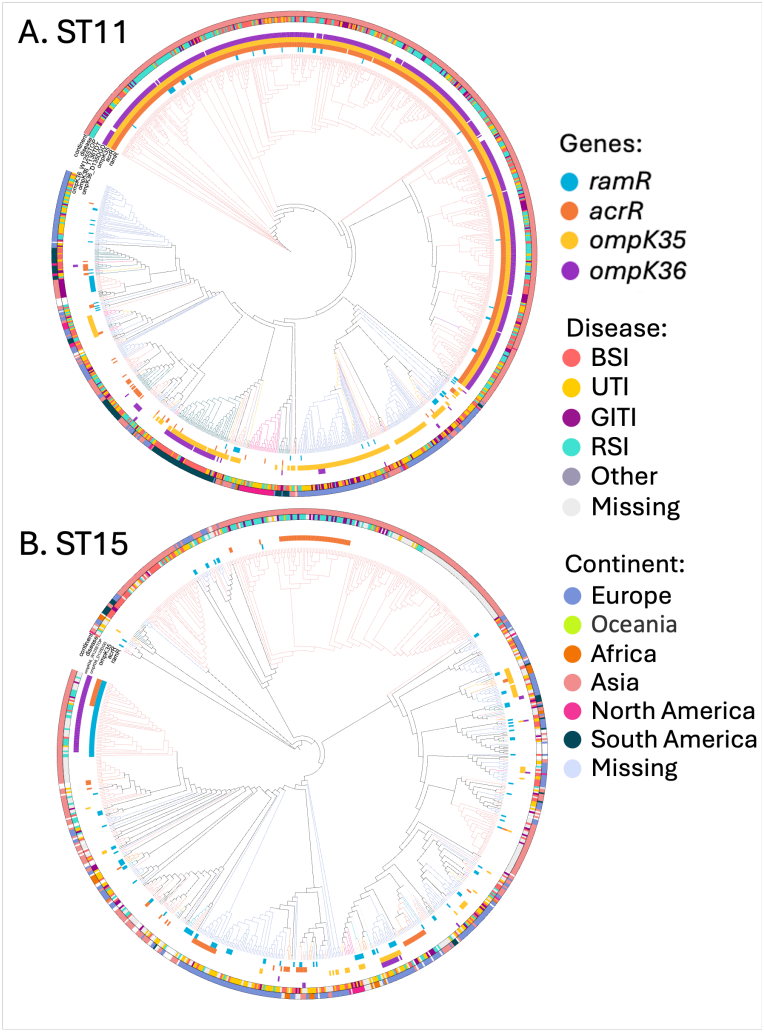


Figure 5.10: Deleterious mutations of membrane permeability genes in ST11 and ST15. The deleterious mutations in *ramR*, *acrR*, *ompK35*, *ompK36* and *oqxR* were tested against PolyPhen-2 (Adzhubei et al., 2010) and visualised on the core genome phylogenies. **A:** Deleterious mutations in ST11. **B:** Deleterious mutations in ST15.

As shown in Figure 5.10, mutations in *ramR* are almost randomly distributed in ST11 (Asia = 68, non-Asia = 26) . In contrast, mutations in *acrR* (Asia = 629, non-Asia = 47 and *ompK36* (D136DGD: Asia = 608, non-Asia = 63 ; T136TDT: Asia = 2; W125STOP: Asia = 1, non-Asia = 2) are more enriched in the Asian cluster. However, mutations in *oqxR* are more prevalent in Europe compared to other continents (Europe = 141, Asia = 56, North America = 4, South America = 2).

In comparison to ST11, mutations in *ramR* and *ompK36* in ST15 are not as widespread and are mainly enriched in an Asian cluster (*ramR*: Asia = 78, non-Asia = 62 ; *ompK36*: Asia = 48, non-Asia = 17). Interestingly, *acrR* mutations appear to cluster together and are more enriched in Asian and European strains (Asia = 78, Europe = 39, Africa = 5).

In this section, we examined the role of efflux pumps and porins in antibiotic resistance within the ST11 lineage, focusing on key permeability factors such as OmpK35 and OmpK36, as well as regulatory genes like *ramR*, *acrR*, and *oqxR*. We found that mutations in *acrR*, *ompK35* and *ompK36* were more prevalent in the Asian cluster of ST11, while mutations in *oqxR* were more enriched in European isolates. These mutations contribute to the enhanced resistance seen in ST11, particularly in Asia. In contrast to ST11, mutations in *ramR* and *ompK36* were less widespread in ST15.

Given ST11's higher resistance to antibiotics and its increased virulence, as discussed in the last section, we shifted our research focus to ST11. To better understand the evolutionary dynamics of ST11, we now turn to the core genome and hierarchical clustering of this lineage to investigate how it has evolved and adapted across different geographic regions.

5.7 Core genome and hierarchical clustering of ST11

As discussed in the previous sections, the ST11 lineage is clearly more resistant and virulent (Figure 5.6). Interestingly, we demonstrated a clear distinction between Asian and non-Asian ST11 isolates (Figure 5.2, 5.3, 5.5). To verify the evolutionary distance between ST11 strains isolated from different regions of the world, we calculated the SNP distances, as shown in Table 5.5. The average SNP distance within Asian isolates (0.01259) is the smallest among all continents, and the SNP distances between Asia and other continents are generally larger (as highlighted in Table 5.5). This suggests that the Asian isolates are evolutionarily distinct from those on other continents, which aligns with our previous finding that the Asian cluster showed different resistance and virulence content from other ST11 clusters.

Continents	Asia	Europe	Africa	North America	South America
Asia	0.01259	-	-	-	-
Europe	0.03305	0.02361	-	-	-
Africa	0.03038	0.02496	0.02214	-	-
North America	0.03389	0.03555	0.03415	0.02777	-
South America	0.03144	0.03350	0.03057	0.03237	0.02431

Table 5.5: SNP distances between ST11 strains isolated from different continents.

To compare, we also calculated the SNP distances in ST15, as shown in Table 5.6. The SNP distance within the Asian isolates is similar to, and even slightly larger than, the distances between Asia and other continents (Table 5.6). Regarding SNP distances within each continent, the Asian cluster exhibits the highest variability among all continents. In contrast to the closely related Asian cluster in ST11, this suggests that the Asian cluster in ST15 is even more variable than strains from other continents.

Continents	Asia	Europe	Africa	North America	South America
Asia	0.01088	-	-	-	-
Europe	0.00853	0.00490	-	-	-
Africa	0.00902	0.00589	0.00623	-	-
North America	0.00737	0.00347	0.00452	0.00113	-
South America	0.00934	0.00561	0.00668	0.00399	0.00562

Table 5.6: SNP distances between ST15 strains isolated from different continents.

We next investigated whether ST11 strains could be separated into different continent-associated clusters, given that the Asian cluster exhibited greater SNP distances from other continents. To explore this, we applied hierarchical clustering using pairwise SNP distances to subdivide the ST11 population (Tonkin-Hill et al., 2018). As shown in Figure 5.11A and Figure 5.11B, the Asian cluster in ST11 was primarily assigned to hierarchical groups 14 (Asian strains = 378) and 15 (Asian strains = 245, North American strain = 1). Hierarchical group 6 mainly consists of European strains (European strains = 56, Asian strain = 1), while group 2 predominantly contains South American strains (South American strains = 55, European strains = 4, Asian strains = 4, African strain = 1). Hierarchical group 18 includes only North American strains (28). Although ST11 strains appear to be undergoing evolution not only in Asia but also in Europe and South America, the number of strains in these regions (Europe = 56, South America = 55) is significantly lower compared to Asia (593). This difference highlights the predominant dissemination and evolutionary events occurring in Asia. We further demonstrated that the Asian clusters, including groups 14 and 15, are genomically distant and distinct from other ST11 isolates (Figure 5.11C), which may be a result of adaptation to different environments (e.g., antibiotic usage, clinical therapeutic approaches, etc.).

In addition to these geographically distinct groups, we observed hierarchical groups distributed across multiple continents, suggesting the worldwide dissemination of closely related ST11 strains. Group 7 is found predominantly in Europe and Asia (European strains = 21, Asian strains = 45, North American strains = 3, African strains = 3), group 22 is present on several continents (South American strains = 21, European strains = 21, Asian strains = 22, North American strains = 4), and group 21 is widely distributed across South America, Europe, and Asia (South American strains = 18, European strains = 13, Asian strains = 5) (Figure 5.11A).

The hierarchical clustering results suggest that even within a single sequence type of *Klebsiella*, further evolution and adaptation to different environments can occur. Thus, it suggests that ST11 has evolved distinctly in Asian and non-Asian environments, with Asian isolates likely exposed to higher levels of antibiotics. Within the Asian cluster, the isolates are evolving along two different trajectories (groups 14 and 15). This observation aligns with previous reports, which indicate that ST11 strains isolated from China can be separated into the KL47 and KL64 clusters (Y. Liu et al., 2022; K. Zhou et al., 2023).

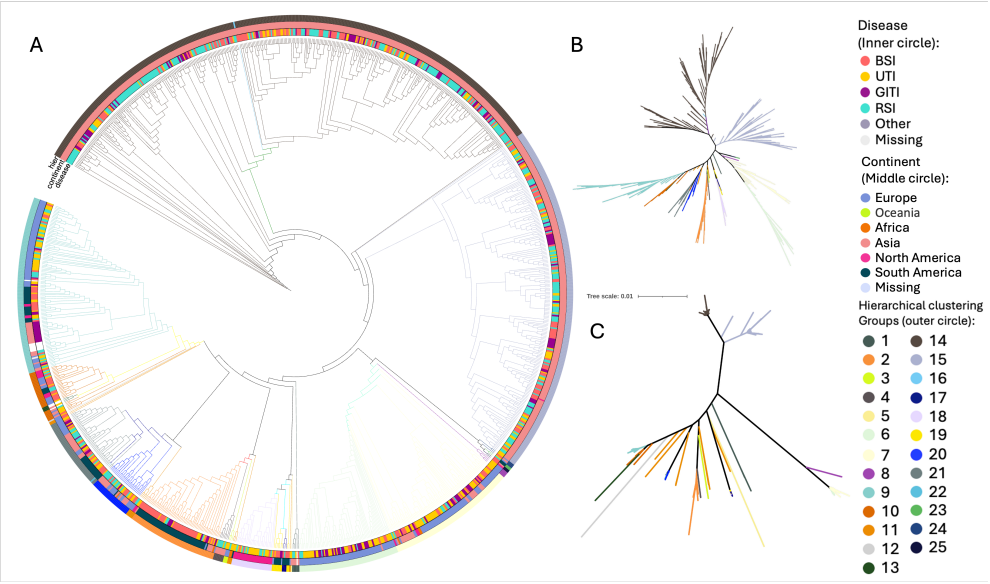


Figure 5.11: Hierarchical Clustering in ST11 *Klebsiella pneumoniae*. **A**: Hierarchical clustering groups (outer circle) visualized on the unrooted core genome phylogeny of ST11, with host disease (inner) and continent (middle) metadata. **B**: Branch length ignored unrooted core genome phylogeny of ST11 with branches coloured according to the hierarchical clustering groups, where the position of clusters in A matches with that in B. **C**: Branch length kept unrooted core genome phylogeny of ST11 with branches coloured according to the hierarchical clustering groups.

In this section, we applied hierarchical clustering to the ST11 population based on pairwise SNP distances, revealing distinct continent-associated clusters. The Asian cluster, which is genomically distant from other regions, was primarily divided into two groups, suggesting divergent evolutionary trajectories within Asia. These findings suggest that ST11 has adapted to distinct environments, with the Asian isolates likely exposed to higher levels of antibiotics, driving their evolution. Additionally, we identified groups that are distributed across multiple continents, indicating the global dissemination of closely related ST11 strains.

These insights into the population structure and adaptation of ST11 set the stage for the next section, where we will study the evolutionary dynamics and transmission patterns within and between these clusters.

5.8 Evolution and transmission

In previous sections, we demonstrated that the Asian isolates in ST11 formed a distinct and unique cluster (Table 5.5), exhibiting different antibiotic resistance and virulence profiles compared to other strains (Figure 5.6). To further investigate the evolutionary trajectory of ST11, we applied BactDating (Didelot, Croucher, Bentley, Harris, & Wilson, 2018b) to estimate the origin of ST11's common ancestor and trace its transmission (Figure 5.12). Due to computational resource limitations, we conducted time-resolved phylogenetic analysis on a subset of the ST11 data, randomly sampling 200 strains from Asia and 200 strains from other continents. We first performed core genome phylogenetic analysis, followed by time-resolved phylogenetic analysis (Figure 5.12).

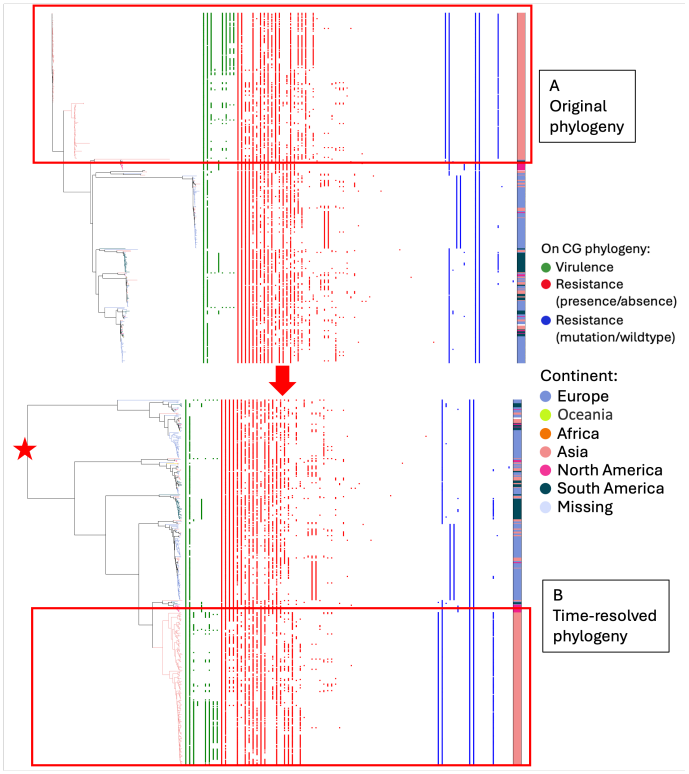


Figure 5.12: Time-resolved Phylogeny of ST11. 400 random ST11 strains are sampled to build a core genome phylogeny using ParSNP (Treangen et al., 2014), followed by BactDating (Didelot et al., 2018b) to correct and predict the root of the phylogeny. A: Original core genome phylogeny. B: Corrected core genome phylogeny. The root position is presented as a star.

As shown in Figure 5.12A, even using a subset of the ST11 population, the population structure mirrors the results from Figure 5.6, where the Asian strains cluster together, and the CE cluster could be found in the Asian cluster. This consistency confirms the reliability of using a subset of the data for analysis.

However, after correcting the phylogeny using time metadata (Figure 5.12B), the Asian and non-Asian clusters rearranged, suggesting a new temporal relationship between these clusters. The predicted root of the corrected phylogeny is marked with a red star, representing the most likely common ancestor of the ST11 population. This common ancestor maps to group 9 in the hierarchical clustering (Figure 5.11), which contains strains from Europe (count = 71), South America (count = 21), Asia (count = 21), and North America (count = 5). Therefore, we infer that the common ancestor of ST11 originated outside of Asia, probably in Europe, supporting the hypothesis that the ST11 lineage evolved in other regions before disseminating to Asia.

After correcting the phylogeny with time annotation, the CE cluster appears as the last step of ST11's evolution, positioned at the bottom of Figure 5.12B. This suggests that the CE cluster formed after the transmission of ST11 strains into Asia.

In this section, we investigated the evolutionary history and transmission patterns of ST11, revealing a two-step evolutionary process. Evolutionary step one: the common ancestor of ST11 likely originated outside of Asia and was transmitted into the region. Evolutionary step two: after its introduction to Asia, the CE cluster formed, where resistance and virulence converged. Time-resolved phylogenetic analysis supported this hypothesis, showing that the CE cluster formed later in the evolutionary timeline. This detailed insight into the temporal and geographical evolution of ST11 highlights the complexity of its transmission and adaptation.

To further understand the genetic basis behind this two-step evolution, we next applied Genome-Wide Association Studies (GWAS) to identify the specific genetic changes that supported the evolution of resistance and virulence in the CE cluster and broader ST11 lineage.

5.9 Genome Wide Association Studies

To study the genetic basis of ST11's evolution, we applied gene presence/absence-based Genome-Wide Association Studies (GWAS) (Brynildsrud et al., 2016). We compared differences between the Asian cluster (count = 621) and the non-Asian cluster (count = 528) to capture evolutionary step 1, and between the CE cluster (count = 223) and its close relative cluster (count = 157) to examine the evolutionary step 2 (Figure 5.13).

For GWAS analyses, we utilised pan-genome data generated by Roary (Page et al., 2015) for association tests. Both Bonferroni (Abdi et al., 2007) and False Discovery Rate (FDR) (Benjamini & Hochberg, 1995) corrections were applied to the *p* values, setting a significance threshold of 0.05. Genes uniquely present in one of the comparison groups were selected for further analysis.

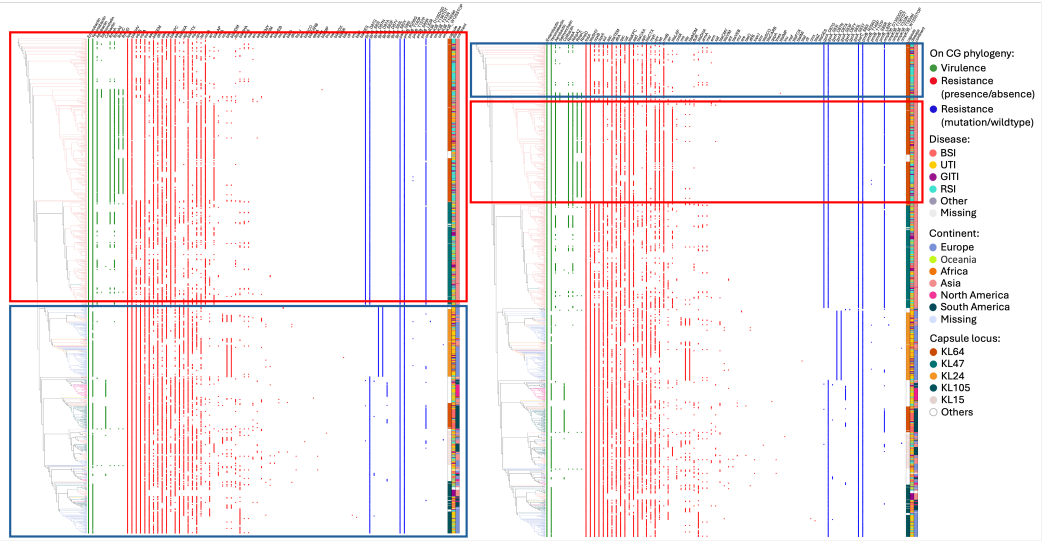


Figure 5.13: ST11 Genome-Wide Association Studies (GWAS) schematic figure. **A**: Comparison between Asian (red) cluster and non-Asian (blue) cluster (evolutionary step 1). **B**: Comparison between clonal expansion (red) and its close relative (blue) cluster (evolutionary step 2).

Table 5.7: Proteins encoded by genes significantly associated with Asian or non-Asia clusters. The GWAS analysis was applied using Scoary (Brynildsrud et al., 2016). The table presents the rank of significant genes, their gain or loss status, genomic location, odds ratio, and corrected *p* values.

Rank	Protein encoded	Gain/ Loss	Genomic location	Present non-Asia	Presnet Asia	Absent non-Asia	Absent Asia	Odds ratio	Bonferroni corrected *p* value	FDR corrected *p* value
1	CrrB	Loss	Chromosome	522	2	6	619	26926.5	1.00e-319	1.00e-319
2	CrrA	Loss	Chromosome	521	2	7	619	23035.6	4.81e-318	2.41e-318
3	HADIA family hydrolase	Loss	Chromosome	519	3	9	618	11879.3	4.47e-312	1.49e-312
4	DUF6392 family protein	Gain	Chromosome	0	604	528	17	<0.00001	3.51E-307	8.77E-308
5	hypothetical protein	Gain	Chromosome	11	615	517	6	0.00021	1.78E-302	3.57E-303
6	SUMF1/EgtB/PvdO family nonheme iron enzyme	Loss	Chromosome	508	0	20	621	inf	2.03E-301	3.39E-302
7	Partial of DUF6392 family protein	Gain	Chromosome	0	599	528	22	<0.00001	5.49E-300	7.84E-301
8	hypothetical protein	Loss	Chromosome	506	0	22	621	inf	1.82E-298	2.27E-299
9	hypothetical protein	Gain	Chromosome	0	597	528	24	<0.00001	3.02E-297	3.36E-298
10	DUF6392 family protein	Loss	Chromosome	503	0	25	621	inf	3.53E-294	3.53E-295
11	hypothetical protein	Loss	Chromosome	502	0	26	621	inf	8.79E-293	7.99E-294
12	Hok/Gef family protein	Gain	Plasmid	4	599	524	22	0.00028	2.54E-290	2.12E-291
13	PIN domain-containing protein	Gain	Plasmid	2	593	526	28	0.00018	9.27E-287	7.13E-288
14	SUMF1/EgtB/PvdO family nonheme iron enzyme	Loss	Chromosome	497	0	31	621	inf	5.00E-286	3.57E-287

Table 5.7: Proteins encoded by genes significantly associated with Asian or non-Asia clusters. The GWAS analysis was applied using Scoary (Brynildsrud et al., 2016). The table presents the rank of significant genes, their gain or loss status, genomic location, odds ratio, and corrected *p* values.

Rank	Protein encoded	Gain/ Loss	Genomic location	Present non-Asia	Presnet Asia	Absent non-Asia	Absent Asia	Odds ratio	Bonferroni corrected *p* value	FDR corrected *p* value
15	ABC transporter six transmembrane domain-containing protein	Loss	Chromosome	510	11	18	610	1571.21	2.37E-282	1.58E-283
16	hypothetical protein	Gain	Chromosome	9	600	519	21	0.00061	4.61E-282	2.88E-283
17	replication regulatory protein RepA	Gain	Plasmid	3	587	525	34	0.00033	5.20E-277	3.06E-278
18	hypothetical protein	Gain	Chromosome	16	600	512	21	0.00109	1.97E-270	1.09E-271
19	DnaT-like ssDNAbinding domain-containing protein	Gain	Chromosome	31	612	497	9	0.00092	1.80E-267	9.45E-269
20	replication protein P	Gain	Chromosome	32	612	496	9	0.00095	3.55E-266	1.78E-267
21	hypothetical protein	Gain	Chromosome	31	611	497	10	0.00102	8.67E-266	4.13E-267
22	fertility inhibition protein FinO	Gain	Plasmid	14	593	514	28	0.00129	2.55E-264	1.16E-265
23	incFII family plasmid replication initiator RepA	Gain	Plasmid	21	600	507	21	0.00145	5.99E-263	2.61E-264
24	IS1182like element ISCfr1 family transposase	Gain	Chromosome	46	618	482	3	0.00046	2.24E-260	9.32E-262
25	hypothetical protein	Gain	Chromosome	19	593	509	28	0.00176	1.18E-256	4.73E-258
26	hypothetical protein	Gain	Chromosome	11	579	517	42	0.00154	1.24E-252	4.79E-254
27	hypothetical protein	Gain	Chromosome	1	560	527	61	0.00021	2.14E-252	7.93E-254
28	headtail adaptor protein	Gain	Chromosome	12	580	516	41	0.00164	7.23E-252	2.58E-253
29	phage terminase small subunit P27 family	Gain	Chromosome	11	578	517	43	0.00158	1.55E-251	5.33E-253
30	hypothetical protein	Gain	Chromosome	3	563	525	58	0.00059	8.51E-251	2.84E-252
31	CPBP family intramembrane metalloprotease	Gain	Plasmid	30	600	498	21	0.00211	1.18E-250	3.81E-252
32	hypothetical protein	Gain	Chromosome	11	577	517	44	0.00162	1.95E-250	5.74E-252
33	terminase large subunit	Gain	Chromosome	11	577	517	44	0.00162	1.95E-250	5.74E-252
34	phage tail tape measure protein	Gain	Chromosome	11	577	517	44	0.00162	1.95E-250	5.74E-252
35	phage tail protein	Gain	Chromosome	11	574	517	47	0.00174	3.43E-247	9.54E-249
36	phage tail protein	Gain	Plasmid	11	574	517	47	0.00174	3.43E-247	9.54E-249
37	phage tail protein	Gain	Both	14	578	514	43	0.00203	1.16E-246	3.14E-248
38	head-tail connector protein	Gain	Chromosome	12	575	516	46	0.00186	2.01E-246	5.29E-248
39	hypothetical protein	Gain	Chromosome	2	556	526	65	0.00044	3.40E-246	8.73E-248
40	phage portal protein	Gain	Chromosome	12	574	516	47	0.00190	1.51E-245	3.77E-247
41	toxin YdaT family protein	Gain	Chromosome	2	555	526	66	0.00045	3.13E-245	7.64E-247
42	toxin YdaT family protein	Gain	Chromosome	11	572	517	49	0.00182	4.40E-245	1.05E-246
43	S49 family peptidase	Gain	Chromosome	12	573	516	48	0.00195	1.83E-244	4.25E-246
44	hypothetical protein	Gain	Chromosome	17	580	511	41	0.00235	2.90E-244	6.59E-246
45	tail fiber domain-containing protein	Gain	Chromosome	12	572	516	49	0.00199	2.94E-243	6.54E-245

Table 5.7: Proteins encoded by genes significantly associated with Asian or non-Asia clusters. The GWAS analysis was applied using Scoary (Brynildsrud et al., 2016). The table presents the rank of significant genes, their gain or loss status, genomic location, odds ratio, and corrected *p* values.

Rank	Protein encoded	Gain/ Loss	Genomic location	Present non-Asia	Presnet Asia	Absent non-Asia	Absent Asia	Odds ratio	Bonferroni corrected *p* value	FDR corrected *p* value
46	tail assembly protein	Gain	Both	1	549	527	72	0.00025	4.89E-242	1.06E-243
47	hypothetical protein	Gain	Chromosome	1	546	527	75	0.00026	4.05E-239	8.63E-241
48	DUF1983 domain-containing protein	Gain	Chromosome	20	578	508	43	0.00293	1.14E-237	2.38E-239
49	hypothetical protein	Gain	Chromosome	10	562	518	59	0.00203	1.27E-236	2.59E-238
50	hypothetical protein	Loss	Chromosome	451	0	77	621	inf	4.08E-235	8.17E-237

As shown in Table 5.7, we listed the top 50 proteins encoded by genes from the pan-genome analysis significantly associated with either the Asian or non-Asian clusters, based on gene presence/absence GWAS analysis (Brynildsrud et al., 2016), with corrected *p* values less than 0.05. The genomic location of these genes was predicted by the method outlined in Section 2.9.2.

The *crrAB* genes are significantly lost in the Asian isolates compared with non-Asian isolates (Table 5.7 ranks 1 and 2). Previous studies (Stojowska-Swędrzyńska, Łupkowska, Kuczyńska-Wiśnik, & Laskowska, 2021) have shown that the CrrAB two-component system interacts with PmrAB to enhance lipid A modifications, contributing to polymyxin resistance. CrrB phosphorylates CrrA, which upregulates *pmrA* (Stojowska-Swędrzyńska et al., 2021). In *K. pne*, mutations or loss of *crrAB* have been linked to decreased polymyxin resistance (Stojowska-Swędrzyńska et al., 2022). Thus, the loss of *crrAB* in Asian ST11 strains likely reduces their ability to modify lipid A, making these strains colistin-sensitive.

The analysis also revealed that the RepA protein is significantly enriched in Asian strains (Table 5.7 rank 17), which is a key plasmid replication gene found in many plasmids (Amina Nazir et al., 2020). This suggests that the cluster of Asian strains enriched for RepA likely harboured plasmids, which could potentially facilitate the horizontal transfer of antimicrobial resistance or other adaptive genes.

DnaT (Table 5.7 rank 19) plays a critical role in replication restart by linking PriC with DnaB helicase through ssDNA binding, facilitating helicase recruitment to stalled replication forks (C.-C. Huang & Huang, 2016). This function would enhance the survival of DnaT-positive Asian strains in environments where DNA damage frequently occurs.

The enrichment of the *ydaT* gene (Table 5.7 ranks 41 and 42), which encodes a putative toxin in the YdaT superfamily, suggests a role in plasmid maintenance and may support the evolution of antibiotic resistance in Asian ST11 strains through a toxin-antitoxin module (Pranavathiyani, Prava, Rajeev, & Pan, 2020).

From the GWAS analysis comparing Asian and non-Asian ST11 strains (evolutionary step 1), we identified additional genes, including *crrAB*, *repA*, *dnaT*, and *ydaT*, which may promote survival rate and adaptation of Asian strains beyond the mechanisms identified in Figure 5.6. These genetic changes likely facilitated the transition of ST11 from non-Asian to Asian environments, where selective pressures such as antibiotic usage could have driven the evolution of these traits. In the next GWAS analysis, we move on to investigate the genetic factors involved in the formation of the CE cluster (evolutionary step 2).

Table 5.8: Proteins encoded by genes significantly associated with the Clonal Expansion (CE) or its Close Relative (CR) cluster.

Rank	Protein encoded	Gain/ Loss	Genomic location	Present Presnet CR	Present Presnet CE	Absent Absent CR	Absent Absent CE	Odds ratio	Bonferroni corrected *p* value	FDR corrected *p* value
1	relaxase/mobilization nuclease domain-containing protein	Gain	Plasmid	4	201	153	22	0.00286	7.03E-72	1.00E-72
2	hypothetical protein	Gain	Plasmid	4	201	153	22	0.00286	7.03E-72	1.00E-72
3	plasmid transfer ATPase TraJ	Gain	Plasmid	4	201	153	22	0.00286	7.03E-72	1.00E-72
4	DotD/TraH family lipoprotein	Gain	Plasmid	4	201	153	22	0.00286	7.03E-72	1.00E-72
5	type IV secretion system DotC family protein	Gain	Plasmid	4	201	153	22	0.00286	7.03E-72	1.00E-72
6	conjugal transfer protein	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
7	phospholipase D family protein	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
8	conjugal transfer protein	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
9	plasmid IncI1-type surface exclusion protein ExcA	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
10	hypothetical protein	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
11	conjugal transfer protein TraY	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
12	hypothetical protein	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
13	conjugal transfer protein TraO	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
14	conjugal transfer protein	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
15	lytic transglycosylase domain-containing protein	Gain	Plasmid	4	200	153	23	0.00301	5.28E-71	1.65E-72
16	conjugal transfer protein TraM	Gain	Plasmid	4	199	153	24	0.00315	3.82E-70	1.06E-71
17	conjugal transfer protein TraN	Gain	Plasmid	4	198	153	25	0.0033	2.66E-69	6.83E-71
18	hypothetical protein	Gain	Plasmid	18	202	139	21	0.01346	2.13E-55	2.37E-57
19	plasmid segregation protein ParM	Gain	Plasmid	18	202	139	21	0.01346	2.13E-55	2.37E-57
20	hypothetical protein	Gain	Plasmid	19	201	138	22	0.01507	1.43E-53	1.24E-55
21	ferrous iron transporter B	Gain	Chromosome	40	219	117	4	0.006244	2.93E-53	2.40E-55
22	cysteine–tRNA ligase	Gain	Chromosome	40	219	117	4	0.006244	2.93E-53	2.40E-55
23	NAD(P)-binding domain-containing protein	Gain	Chromosome	40	219	117	4	0.006244	2.93E-53	2.40E-55

Table 5.8: Proteins encoded by genes significantly associated with the Clonal Expansion (CE) or its Close Relative (CR) cluster.

Rank	Protein encoded	Gain/ Loss	Genomic location	Present Presnet CR	Present Presnet CE	Absent Absent CR	Absent Absent CE	Odds ratio	Bonferroni corrected *p* value	FDR corrected *p* value
24	zinc metallochaperone GTPase ZigA	Gain	Chromosome	41	219	116	4	0.006456	1.79E-52	1.40E-54
25	ABC transporter substrate-binding protein	Gain	Chromosome	40	208	117	15	0.024655	1.15E-41	3.94E-44
26	iron ABC transporter permease	Gain	Chromosome	40	208	117	15	0.024655	1.15E-41	3.94E-44
27	RepB family plasmid replication initiator protein	Gain	Plasmid	85	219	72	4	0.021563	2.26E-24	5.75E-27
28	Protein SopB	Gain	Plasmid	85	219	72	4	0.021563	2.26E-24	5.75E-27
29	site-specific integrase	Gain	Plasmid	88	219	69	4	0.023294	8.36E-23	2.08E-25
30	plasmid-partitioning protein SopA	Gain	Plasmid	88	219	69	4	0.023294	8.36E-23	2.08E-25
31	CusA/CzcA family heavy metal efflux RND transporter	Gain	Plasmid	84	214	73	9	0.048393	2.89E-20	6.72E-23
32	hypothetical protein	Gain	Plasmid	84	214	73	9	0.048393	2.89E-20	6.72E-23
33	cation efflux system protein CusF	Gain	Plasmid	84	214	73	9	0.048393	2.89E-20	6.72E-23
34	DUF2933 domain-containing protein	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
35	copper-translocating P-type ATPase	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
36	copper resistance membrane spanning protein PcoS	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
37	copper response regulator transcription factor PcoR	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
38	copper resistance inner membrane protein PcoD	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
39	copper resistance system metallochaperone PcoC	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
40	Cu(+)/Ag(+) efflux RND transporter periplasmic adaptor subunit SilB	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
41	copper resistance protein B	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
42	Cu(+)/Ag(+) sensor histidine kinase	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
43	response regulator in two-component regulatory system with CusS	Gain	Plasmid	85	214	72	9	0.04965	9.06E-20	2.05E-22
44	hypothetical protein	Gain	Plasmid	85	213	72	10	0.055425	8.36E-19	1.88E-21
45	hypothetical protein	Loss	Plasmid	49	1	108	222	100.722	1.89E-16	3.99E-19
46	hypothetical protein	Gain	Plasmid	147	123	10	100	11.9512	6.47E-14	1.28E-16
47	hypothetical protein	Loss	Plasmid	43	1	114	222	83.7368	1.19E-13	2.35E-16
48	type IV conjugative transfer system lipoprotein TraV	Loss	Plasmid	43	1	114	222	83.7368	1.19E-13	2.35E-16
49	hypothetical protein	Gain	Chromosome	105	216	52	7	0.065438	7.96E-12	1.53E-14
50	hypothetical protein	Loss	Plasmid	34	0	123	223	inf	5.43E-11	1.02E-13

In Table 5.8, we listed the top 50 genes significantly associated with the CE or CR cluster, with corrected *p* values below 0.05. These genes play critical roles in enhancing the adaptability, pathogenicity, and survival of ST11 strains within the CE cluster.

The relaxase/mobilisation nuclease domain-containing protein, enriched in the CE cluster (Table 5.8 rank 1), facilitates horizontal gene transfer through plasmid conjugation, and could enable the transfer of plasmid DNA between cells (Añorga, Urriza, Ramos, & Murillo, 2022). This mechanism promotes the dissemination of resistance and virulence genes, and could directly strengthen the CE cluster's adaptability and ability to accumulate advantageous traits. Similarly, the enrichment of the *tra* gene cluster (Table 5.8 rank 3, 4, 11, 13, 16, 17) further supports this by encoding the conjugative apparatus essential for efficient plasmid mobilisation (J. Zhang et al., 2023). Together, these genes enhance the horizontal transfer of genetic elements, driving the acquisition of resistance and virulence factors in the CE cluster.

Additionally, the *sop* and *rep* genes (Table 5.8 rank 27, 28, 30), which are responsible for plasmid partitioning and replication, could ensure the stable inheritance and maintenance of plasmids within bacterial populations (X. Wang et al., 2021). This stability allows the CE cluster to preserve and propagate plasmids carrying critical resistance and virulence genes, which could reinforce the CE cluster's capacity to thrive under selective pressures.

The CE cluster also exhibits enrichment of iron ABC transporter genes (Table 5.8 rank 21-26), which play a pivotal role in siderophore-mediated iron acquisition (Köster, 2001). This system allows bacteria to thrive in iron-limited environments (such as the human host) by internalising siderophore-iron complexes. The combination of enriched siderophore genes and efficient ABC transporters significantly enhances bacterial fitness, virulence, and immune evasion, and could provide the CE cluster a significant advantage in host environments.

Moreover, the enrichment of the *cus* and *pco* gene systems (Table 5.8 rank 31, 33, 36-39, 41, 43) in the CE cluster equips these strains with enhanced copper resistance (Zahid, Zulfiqar, & Shakoori, 2012). By detoxifying copper through efflux pumps and sequestration mechanisms, these genes protect bacterial cells from the antimicrobial effects of copper, which is a defence used by the host immune system (Monteith & Skaar, 2021). This adaptation further strengthens the CE cluster's ability to persist in hostile host environments.

In summary, the enrichment of genes associated with horizontal gene transfer, plasmid stability, iron acquisition, and copper resistance uniquely positions the CE cluster as a highly adaptive and pathogenic sublineage. These genetic traits could lead to the CE cluster's enhanced survival, dissemination, and virulence under both environmental and host-specific selective pressures.

Table 5.9: Comparison of core and accessory genes per genome between the Asian/non-Asian and clonal expansion/its close relative clusters.

Cluster	No. of core genes	*t* statistics	*p* value	No. of accessory genes	*t* statistics	*p* value
Asian cluster	4460.51	1.986	0.047	720.00	9.656	<0.001
Non-Asian cluster	4456.02			631.90		
Clonal expansion cluster	4463.48	1.644	0.101	801.39	10.414	<0.001
Close relative cluster	4459.39			688.13		

As shown in Table 5.9, the Asian cluster contains significantly more accessory genes than the non-Asian cluster, and the CE cluster has significantly more accessory genes than the CR cluster. However, the number of core genes remains similar between these clusters. This suggests that the key driver of evolution at each step is the acquisition of accessory genes, which matches with the findings from Tables 5.7 and 5.8.

Table 5.10: Plasmid count *t* test between Asian/non-Asian and clonal expansion/its close relative clusters. The plasmid profile of each strain was built using PlasmidFinder (Carattoli et al., 2014).

Cluster	No. of plasmids	*t* statistics	*p* value
Asian cluster	3.869	-1.189	0.235
Non-Asian cluster	3.966		
Clonal expansion cluster	4.040	0.259	0.795
Close relative cluster	4.013		

However, when comparing the plasmid content between the Asian and non-Asian clusters or the CE and CR clusters, we did not observe a higher plasmid count in the Asian or CE cluster (Table 5.10). Thus, we infer that, although a similar number of plasmids are present, the gene content on these plasmids differs between clusters. This hypothesis could be explored in greater detail with long-read sequencing data, allowing for the analysis of genes on individual plasmids.

In summary, our GWAS analysis uncovered the genetic basis of the two-step evolutionary process in ST11. In the first evolutionary step, we identified genes that enhance antimicrobial resistance, plasmid stability, and adaptation to stress in Asian ST11 strains, which collectively contribute to increased resilience against antibiotics and environmental pressures. In the second evolutionary step, characterising the CE cluster, we observed significant enrichment of genes involved in horizontal gene transfer, plasmid maintenance, iron acquisition systems,

and copper resistance mechanisms. These findings highlight the adaptive strategies facilitating ST11's transition and establishment in diverse environments, where selective pressures like antibiotic usage and host immune responses are predicted to drive these evolutionary adaptations.

5.10 Discussion

In this chapter, we aimed to study the population structure, resistance, virulence, and evolutionary trajectory of the prevalent *K. pne* lineage ST11, and its comparison to ST15 and ST258. Through detailed genomic analyses and GWAS, we showed distinct evolutionary patterns, geographical adaptations, and critical genetic determinants contributing to the adaptation of ST11 to diverse environments.

Our investigation confirms the significant role of ST11 in global antimicrobial resistance, especially within Asia. The Asian cluster in ST11 shows a unique combination of high resistance and virulence profiles, characterised by a convergence of fluoroquinolone, phosphonic antibiotic, and carbapenem resistance with key virulence factors, including enterobactin, yersiniabactin, aerobactin, and the *rmp* genes. This Asian cluster, particularly the CE sub-cluster, represents an enhanced public health threat due to its dual properties of multidrug resistance and hypervirulence.

The evolutionary path of ST11 appears to have occurred in two major steps. Initially, the common ancestor of ST11 likely arose outside of Asia, as suggested by its close relationship with strains from Europe. Following its dissemination into Asia, KL47 ST11 rapidly evolved under regional selective pressures, most likely driven by high antibiotic use, to form the KL64 CE cluster within the Asian clade. This two-step evolution of ST11 reflects its remarkable adaptability to different environments and supports the hypothesis that antibiotic pressure acts as a significant evolutionary driver in the Asian strains. Previous analyses of KL64 ST11 strains from China and Brazil demonstrated that the KL64 Chinese strains are derived from KL47 strains (J. Wang et al., 2023). Our analysis, utilising a more comprehensive ST11 dataset, confirmed this as part of evolutionary step 2 and, through GWAS, identified key genes associated with plasmid stability, iron acquisition, and copper resistance driving this process.

Our hierarchical clustering analyses further highlight the unique genetic adaptations within ST11 in different geographic contexts. The clustering of Asian isolates into two major groups (groups 14 and 15) underscores the genetic divergence within the Asian clade, likely resulting from adaptation to region-specific pressures. The substantial SNP distances between Asian and non-Asian isolates further confirmed this, marking the Asian clade as a distinct lineage with unique resistance and virulence profiles.

Comparative analyses of ST15 and ST258 epidemic lineages indicate a less pronounced convergence of resistance and virulence traits than ST11. ST15, while globally disseminated and showing ciprofloxacin resistance, lacks the hypervirulent profile observed in ST11. ST258, primarily linked to North America, shows extensive carbapenem resistance but is less virulent, with fewer virulence factors such as the *rmp* locus and siderophores. These findings suggest that ST11, particularly within Asia, formed a more complex challenge due to its combination of high virulence and resistance.

Through GWAS, we identified key genes and mutations underlying the two-step evolutionary process of ST11. The initial transition from non-Asian to Asian environments involved the acquisition of genes associated with resistance and plasmid stability, such as *crrAB, repA, dnaT*, and *ydaT*. The later formation of the CE cluster in Asia involved further genetic changes that likely enhanced horizontal gene transfer and resistance mechanisms, particularly through the enrichment of the *tra* gene cluster, *sop* genes, and iron acquisition genes. This genetic adaptation in the CE cluster reflects a shift towards an even more resistant and virulent phenotype, enhancing ST11's survival and persistence in high-antibiotic and iron-limited environments like those found within the human host.

Notably, our analysis of plasmid content and core/accessory gene count suggests that plasmid gene content, rather than number, drives the phenotypic differences observed. This insight highlights the potential utility of long-read sequencing in future studies to explore plasmid-based genetic variations at a finer resolution.

In summary, the evolutionary dynamics of ST11 showed how *K. pne* adapts to local environments through genomic enrichment in resistance and virulence. These findings underscore the need for targeted surveillance of high-risk ST11 lineages, particularly the Asian CE cluster, which combines resistance and virulence factors in a way that presents heightened challenges for infection control and treatment. Enhanced surveillance and control methods, especially in regions with high antibiotic usage, will be essential to address the public health risks posed by ST11 and other emerging lineages.

Chapter 6

Discussion

6.1 General Discussion

This thesis underscores the significant role that genomic research into *Klebsiella* plays in understanding its complex dynamics within the One Health framework. By exploring the population structure, evolution, epidemiology, resistance, and virulence profiles of *K. var* (Chapter 3 and 4) and *K. pne* lineages (Chapter 5), particularly ST11, this work provides new insights into how these bacteria adapt across diverse hosts/habitats, from environmental/animal/plant sources to human infections. This work highlights the genomic plasticity of *K. var*, supporting its capacity for adaptation to the human host (Chapter 3 and 4), while also revealing the convergence of virulence and resistance factors in human-associated *K. pne*, particularly in epidemic lineages like ST11 (Chapter 5).

The findings of this research have significant public health implications and underscore the urgent need for enhanced surveillance and infection control measures targeting *Klebsiella* species. Furthermore, the detailed genomic insights provided by this study pave the way for future interventions, including the development of targeted treatments and strategies to mitigate the transmission of high-risk clones.

In a broader context, this thesis contributes to the growing understanding of how bacterial populations evolve and disseminate across different continents and environments, linking animal, human, and environmental health in line with the principles of One Health. These findings are critical for informing both clinical strategies and public health policies aimed at competing the global threat posed by antimicrobial-resistant pathogens like *Klebsiella*.

The following sections summarise the key differences between *K. var* and *K. pne*, emphasising the need to develop or update *K. pne*-based genomic analysis tools and software to accommodate comprehensive *K. var* datasets. Additionally, we highlight the significance of ST11 *K. pne* as a major human pathogen, marked by clonal expansion and a convergence of resistance and virulence. Finally, this section presents future perspectives for advancing research in this field.

6.2 The versatility of *K. var*

This work provides the most comprehensive genomic analysis of *K. var* to date, combining global and local datasets to highlight its extensive genetic diversity and adaptability. These comprehensive datasets could provide valuable insights into the population structure, resistance mechanisms, and virulence factors within a controlled geographic region.

The star-like population structure we observed (Figure 4.1, Figure 3.1) demonstrates a lack of clustering by host disease, geographic location, or sample source. Although we have *K. var* strains collected within a localised geographic area and a defined timeframe, these strains were still randomly distributed within the global phylogeny (Figure 4.2). This underscores the genomic plasticity and adaptability of *K. var*, which is a pattern distinct from the geographically restricted lineages often seen in *K. pne*. The versatility of *K. var* is further highlighted by its transmission from non-human to human sources. Originating from non-human environments with nitrogen-fixing capabilities, *K. var* has gradually adapted to human hosts by acquiring resistance genes and losing cold-shock-associated genes. Notably, SNP distance-based hierarchical clustering failed to differentiate human and non-human strains, emphasizing the ease with which *K. var* can transition between different hosts and habitats.

The lower prevalence of resistance genes in *K. var* compared to *K. pne* reflects differing antibiotic stewardship practices across hosts and habitats, as *K. pne* is predominantly associated with humans, while *K. var* is found in both human and non-human sources. The higher prevalence of intrinsic mutations, ESBL-producing genes, and critical resistance determinants such as bla_{KPC} and bla_{OXA} in human-associated *K. var* strains underscores the increased antibiotic pressure in clinical settings.

Furthermore, the presence of bla_{SHV} in *K. var* strains suggests ongoing HGT with *K. pne*. The co-occurrence of bla_{SHV} with the KL123 capsule locus in *K. var* isolates provides additional evidence for capsule locus acquisition during these recombination events. However, mutations such as those in *gyrA*, *parC*, and *pmrB*, which are prominent in *K. pne* ST11, were not observed in *K. var*.

Plasmid was found to be closely related to resistance genes HGT. The general plasmid profile of *K. var* strains closely resembles that in *K. pne*, with IncFIB as the most prevalent plasmid type. This similarity supports the hypothesis of frequent gene transfer between *K. var* and other KpSC members, including *K. pne*. Despite differences in AMR gene content, the consistent plasmid profile highlights the potential for ongoing HGT and the shared evolutionary pressures shaping *K. var* populations globally.

Capsule locus diversity remains a hallmark of *K. var*, which is significantly different from that in *K. pne* (Figure 3.11). Except for the Bangladesh cluster, we cannot find a specific cluster that contains the same capsule locus. In the global dataset, the most prevalent capsule locus (KL60) is only present in 5.60% (count = 40) of the strains. If we merge the global dataset with

the Scottish dataset, the most prevalent capsule loci, both KL60 and KL53, are only present in 5.35% (count = 44) of the strains. This diversity suggests that the capsule locus might be a recombination hot spot in *K. var*. Such variability further underscores the limited utility of current *Klebsiella* vaccines, which target capsule types rare in *K. var*. Additionally, the lack of key capsule synthesis genes such as *wcaJ*, *wzc*, and *wzy* is suggesting alternative regulatory mechanisms for capsule production in *K. var*.

Interestingly, all the siderophores are less prevalent in *K. var* than that in *K. pne*, where enterobactin is the most prevalent and yersiniabactin is the second most prevalent. Despite this, convergence between siderophore carriage and resistance genes was rarely observed in *K. var*. However, other virulence factors in *K. var*, specifically in the capsule locus, might converge with the resistance genes, as is shown in the Bangladeshi strains.

In the Scottish *K. var* dataset, GWAS failed to identify significant genetic distinctions between strains by disease type or geographic location, further illustrating the non-specialised nature of the *K. var* genome when it is adapted in the human host.

Notably, our findings reveal that current genomic analysis software/tools are predominantly designed for *K. pne* and perform poorly when applied to *K. var*. For instance, the *K. pne*-specific MLST scheme (M. Lam et al., 2021) is unsuitable for *K. var*, as the *infB* gene erroneously divides the *K. var* population into two sub-clusters. However, the *K. var* based MLSTKv scheme (Barrios-Camacho et al., 2019) proved inadequate for our comprehensive *K. var* dataset, with nearly one-third of the strains unable to be assigned a known sequence type. This limitation significantly restricts the application of molecular typing in *K. var* and underscores the urgent need for an updated, *K. var*-specific MLST scheme.

Similarly, Kaptive (Stanton et al., 2025), a tool for capsule locus detection based on *K. pne* data, failed to identify the capsule locus in over 13% of the *K. var* dataset. This is likely due to the greater diversity of capsule loci in *K. var* and the non-essentiality of certain capsule synthesis genes that are critical in *K. pne*. Overall, we suggest that genomic analysis software requires updates to incorporate *K. var* genotypes. Given its significance as an emerging human pathogen, improving the accuracy of these tools for *K. var* is essential.

Taken together, this work emphasises the genomic adaptability and plasticity of *K. var* across local and global scales, and showed the difference between *K. var* and *K. pne*. The findings reinforce the importance of continued genomic surveillance to monitor its emergence as a human pathogen and its interactions with *K. pne*. The distinct genetic features of *K. var*, including its resistance mechanisms, capsule diversity, and ability to persist across diverse environments, underscore its significance within the One Health framework.

6.3 The Importance of *K. pne* being a human pathogen

K. pne represents a critical public health challenge due to its diverse capacity to cause severe infections and its increasing resistance to antimicrobials. The emergence of MDR, ESBL-producing, and hypervirulent strains highlights the pressing need for improved surveillance and control measures. This chapter integrates the findings from this thesis with a broader understanding of *K. pne*, focusing on prominent lineages such as ST11, to show their global epidemiology, resistance, and virulence.

Among *K. pne* lineages, ST11 has shown massive worldwide dissemination, especially in China, where it dominates as the leading carbapenem-resistant lineage (Figure 5.6). Hierarchical clustering analyses reveal distinct continent-associated clusters within ST11, with the Asian cluster showing unique evolutionary adaptations (Figure 5.11). SNP distance calculations confirmed the evolutionary divergence of Asian ST11 strains, with a smaller genetic distance within the Asian cluster than between Asian and non-Asian isolates (Table 5.5). This highlights how regional selective pressures, such as high antibiotic usage, have shaped the genomic trajectory of ST11 in Asia.

Comparative analyses with ST15 and ST258 further revealed significant differences in global distribution and adaptability. While ST15 is globally distributed, its resistance and virulence profiles are less significant than those of ST11 (Figure 5.8). In contrast, ST258, predominantly associated with North America, exhibits extensive carbapenem resistance but lacks hypervirulent traits such as the *rmp* locus and aerobactin production (Figure 5.9). These findings underscore the exceptional adaptability and pathogenic potential of ST11 compared to the other lineages.

Importantly, the population structure of *K. pne* lineages ST11, ST15, and ST258 exhibited strong cladistic associations with continents and further formed distinct clusters linked to clonal expansions. For ST11, the population structure was also associated with the capsule locus, as strains within the same cluster were often found to share the same capsule locus. In contrast, the population structure of *K. var* showed no clear association with specific continents or capsule loci, highlighting a key difference in the evolutionary and ecological dynamics between these two species.

This work demonstrates a remarkable convergence of resistance and virulence in ST11, particularly within the Asian clade. Asian ST11 strains exhibit widespread resistance to carbapenems and fluoroquinolones, driven by key mutations such as *gyrA* D87G and S83I and D135DGD insertions in *ompK36*. Additionally, these strains possess high virulence scores due to the presence of siderophores such as enterobactin, yersiniabactin, and aerobactin, as well as the *rmp* locus. This convergence positions the Asian clade, particularly the CE sub-cluster, as a significant public health threat.

Compared to *K. var*, the three *K. pne* lineages studied—ST11, ST15, and ST258—appear to be generally more resistant and virulent. Notably, we did not observe the presence of a single bla_{LEN} gene in our *K. pne* datasets, suggesting that *K. pne* exhibits a greater propensity for genome content exchange within its own lineages. In contrast, *K. var* seems to function more effectively as a recipient than a donor, acquiring genes such as bla_{SHV} from *K. pne*.

The evolutionary trajectory of ST11 reveals a two-step process: (1) dissemination of a common ancestor from non-Asian regions into Asia and (2) subsequent evolution of the Asian clade under local selective pressures. GWAS and permeability gene analyses identified key genetic changes associated with these steps, including intrinsic mutations in *gyrA*, *parC*, *ompK36*, and *acrR*, as well as the loss of the *crrAB* locus in the first step. In the CE cluster, further adaptations included the acquisition of additional resistance genes such as bla_{KPC}, bla_{TEM}, $bla_{\text{CTX-M}}$ and bla_{LAP}, the loss of the *oqxAB* gene cluster, and the acquisition of siderophore genes and the *rmp* locus. These genetic changes have enhanced ST11's survival and pathogenicity, allowing it to thrive in environments with high antibiotic pressure and limited iron availability.

In comparison, the evolutionary trajectory of *K. var* appears simpler, involving primarily resistance gene acquisition and cold-shock gene loss. This suggests that *K. pne* ST11 is subject to greater evolutionary and selective pressures than *K. var*, driving more complex adaptive changes and facilitating convergence under these pressures.

The findings from this work emphasise the need for targeted interventions to control the spread of high-risk *K. pne* lineages, particularly ST11. Enhanced surveillance using genomic tools can identify emerging resistance and virulence patterns, facilitating early intervention. Additionally, strategies to reduce antibiotic overuse, particularly in regions such as Asia, are critical to slowing the evolution of such high-risk clones.

6.4 The Future of This Field

This section outlines possible future directions of this study and explores perspectives on the broader development of this field.

6.4.1 Building Unbiased Datasets

In biological research, biased datasets often present a significant challenge, preventing meaningful analyses unless additional samples are acquired or the dataset is reduced to eliminate bias. However, acquiring additional samples at the analysis stage is often impractical, as most samples are collected within fixed locations and time periods. As a result, many studies resort to excluding samples to construct a less biased database, which inherently limits the scope of statistical analyses and impedes progress.

Moreover, most researchers analyse only the data or sequencing results they have collected, which can lead to misinterpretations unless these datasets are integrated into a larger, more comprehensive framework.

In the future, the development of unbiased and high-quality datasets for bacterial genomics is crucial. Similar to the "Common Objects in Context dataset" in computer vision or the "Common Crawl" dataset for large language models, such resources could establish a foundation for robust downstream analyses. The field of bacterial genomics currently lacks such comprehensive and unbiased datasets, highlighting the urgent need for community-driven initiatives to create and share high-quality, unbiased resources.

6.4.2 Dry and Wet Lab

The same genotype does not always result in the same phenotype, and conversely, identical phenotypes can emerge from distinct genotypes. For instance, deleterious mutations or the deletion of a gene may not necessarily disrupt the function of a biosynthetic process, while identical phenotypes within the same species may arise from varying gene contents. A notable example is the absence of *wcaJ*, the capsule synthesis initiation gene, in Bangladeshi *K. var* strains; despite this, these strains are still capable of capsule production. Similarly, carbapenem resistance can be mediated by harbouring β-lactamases or through mutations in *OmpK36*, while colistin resistance may arise from the acquisition of *mcr* genes or mutations in *gyrA* and *parC*. These examples underscore the complex, non-linear relationship between genotype and phenotype, highlighting the necessity of integrating both genomic and phenotypic analyses to fully understand bacterial adaptation and resistance mechanisms.

More importantly, while studying *K. var* and *K. pne*, we realised that bioinformatics analyses can yield significantly different outcomes depending on the models, parameters, or input data used, emphasising the importance of careful methodology. To ensure reproducibility, it is critical to open source scripts and commands used in the analysis, allowing others to replicate the results. Furthermore, purely dry-lab studies must incorporate large-scale datasets to minimise bias and prevent misinterpretation.

It is also essential to recognise the gap between genomic analysis and biological reality. Large-scale patterns observed in genomic analyses may not always be reflected in laboratory experiments, and in some cases, a single robust observation in the lab can overturn genomic predictions. Therefore, integrating bioinformatics with wet lab experiments is crucial, as biological validation remains a fundamental component of meaningful scientific discovery.

6.4.3 Antibiotic Overuse and Sampling

Antibiotic overuse is a significant global issue. Even last-line antibiotics, such as carbapenems and colistin, are becoming less- or even in-effective against numerous bacterial isolates, including not only *Klebsiella* but also *Escherichia* and *Salmonella*. Developing new treatment strategies or antibiotics is not a sustainable solution, as bacteria can rapidly evolve to mutate or acquire resistance genes within just a few years. The fact that makes this issue more complex is the lag between sample collection and data publication, which often takes 2–3 years. As a result, the current state of bacterial populations is often inaccessible, even though bacteria can evolve significantly in a short period.

An additional concern lies in the difference between resource-rich and resource-poor regions. Areas with the highest levels of antibiotic overuse are often under-represented in sampling and sequencing efforts. This creates a critical knowledge gap, as the most relevant and up-to-date data on bacterial transmission and evolution may remain unexamined. To address this, future research must prioritise sampling and studying middle- and low-income regions, which can be largely affected by antibiotic resistance. Educational initiatives should also emphasise the importance of appropriate and scientific antibiotic use to control further overuse and release the global burdern of resistance.

6.4.4 Short and Long Read Sequencing

Short-read sequencing is cost-effective, high-throughput, and provides high-quality, accurate reads. However, it has limitations when dealing with repetitive regions, structural variants, and complex genome elements. Repetitive sequences are challenging to resolve because they appear identical or highly similar, leading to errors in mapping or gaps in genome assembly. Structural variants, such as DNA segments translocated to different genomic locations, may become undetected with short-read sequencing. Additionally, plasmid sequences are often fragmented across multiple contigs, making it difficult to determine whether specific genes are carried on the same plasmid or different plasmids (Amarasinghe et al., 2020).

Long-read sequencing, while less cost-effective and less accurate, can overcome these limitations. Future studies should focus more on the analysis of plasmid and chromosomal genetic content using long-read sequencing technologies to capture finer-scale variations and resolve these complexities (Amarasinghe et al., 2020).

Appendix A

Supplementary Data for Chapter 3

Table A.1: *Klebsiella variicola* strains used in the time-resolved phylogenetic analyses.

Genome Accession	Disease	Source	Country	Continent	Year
GCF_014235385.1	urine	human	America	North America	2017
GCF_015286965.1	fecal	human	China	Asia	2019
GCF_013309835.1	choanal swab	human	Brazil	South America	2012
GCF_016635235.1	blood	human	Australia	Australia	2015
GCF_000941635.2	Expectoration	human	Mexico	NorthAmerica	2011
GCF_900509745.1	Puncture Fluids	human	Germany	Europe	2013
GCF_001187805.1	/	animal	Canada	North America	2012
GCF_003386995.1	blood	human	Bangladesh	Asia	2016
GCF_012274365.1	blood	human	Brazil	South America	2015
GCF_012102105.1	/	animal	Brazil	South America	2019
GCF_002853275.1	unknown	human	China	Asia	2017
GCF_015290365.1	/	animal	China	Asia	2018
GCF_012102985.1	endodontic infection	human	Brazil	South America	2019
SPARK_1060_C1	unknown	human	Italy	Europe	2018
GCF_900406995.1	wound	human	Tanzania	Africa	2014
GCF_015354525.1	/	plant	Brazil	South America	2013
GCF_013305245.1	/	environment	China	Asia	2016
GCF_015694045.1	urine	human	Australia	Australia	2017
GCF_012393465.1	unknown	human	America	North America	2013
SPARK_972_C1	unknown	human	Italy	Europe	2017
GCF_004311265.1	unknown	human	Japan	Asia	2013
SPARK_1269_C2	/	animal	Italy	Europe	2017
GCF_006385475.1	/	plant	America	North America	2017
GCF_008364395.1	/	plant	Germany	Europe	2015
GCF_902704665.1	unknown	human	Italy	Europe	2016
GCF_903971005.1	/	plant	India	Asia	2017
GCF_002174085.1	unknown	human	China	Asia	2012
GCF_002156765.1	urine	human	America	North America	2014
GCF_001187825.1	/	animal	Canada	North America	2012
GCF_001549975.1	/	animal	China	Asia	2011
GCF_001989495.1	/	environment	South Korea	Asia	2014
GCF_014595775.1	unknown	human	America	North America	2018
GCF_002179565.1	blood	human	Netherlands	Europe	2012

Table A.1: *Klebsiella variicola* strains used in the time-resolved phylogenetic analyses.

Genome Accession	Disease	Source	Country	Continent	Year
GCF_015290565.1	/	animal	China	Asia	2019
GCF_001674935.1	/	animal	Canada	North America	2012
GCF_900175105.1	/	environment	United Kingdom	Europe	2014
GCF_003115905.1	/	animal	Canada	North America	2012
GCF_011075365.1	/	animal	China	Asia	2016
GCF_002153075.1	unknown	human	America	North America	2015
GCF_900174115.1	/	environment	United Kingdom	Europe	2015
GCF_013309855.1	/	animal	Brazil	South America	2012
GCF_008375025.1	unknown	human	France	Europe	2018
SPARK_1012_C1	/	animal	Italy	Europe	2018
GCF_001006575.1	/	plant	Malaysia	Asia	2013
GCF_014595765.1	unknown	human	America	North America	2019

Table A.2: Metadata information of *Klebsiella variicola* genomes that contain bla_{SHV}.

Accession	Pasteur ST	AMR count	ANI with *K. var* ref	ANI with *K. pne* ref	Capsule	Source	Continent
GCF_009497715.1	ST1562	15	99.17	94.63	KL114	human	Asia
GCF_002851635.1	0 (unknown)	12	99.09	94.64	KL58	human	South America
GCF_015718035.1	ST1563	7	99.13	94.66	None	human	North. America
GCF_000826565.1	ST919	9	99.1	94.6	None	human	North America
GCF_002264565.1	ST3942	17	99.03	94.73	KL13	human	Asia
GCF_002851415.1	ST533	8	98.84	94.63	KL64	human	South America
GCF_002886665.1	ST454	11	99.17	94.69	KL67	human	North America
GCF_003285185.1	ST1562	13	99.14	94.7	KL114	human	Asia
GCF_008123395.1	ST347	11	99.05	94.61	KL25	human	South America
GCF_900084795.1	ST3707	14	99.06	94.59	KL126	human	Europe
GCF_900085165.1	ST3699	12	99.13	94.57	KL46	human	Europe
GCF_900085175.1	ST1683	13	99.15	94.66	KL71	human	Europe
GCF_900085185.1	ST3716	9	99.04	94.61	KL56	human	Europe
GCF_900085825.1	ST925	12	99.15	94.66	KL38	human	Europe
GCF_900086175.1	ST1562	15	99.11	94.56	KL114	human	Europe
GCF_900086385.1	ST2273	11	99.12	94.57	KL9	human	Europe
GCF_900509625.1	ST1563	8	99.08	94.69	KL123	human	Europe
GCF_900509745.1	ST1469	9	99.2	94.5	KL56	human	Europe
GCF_900509815.1	ST360	11	99.2	94.61	KL31	human	Europe
GCF_900559395.1	ST368-1LV	6	99.13	94.65	KL121	human	Europe
R008921.fna	ST1563	15	99.07	94.67	KL123	human	Europe
R011353.fna	ST1563	15	99.09	94.71	KL123	human	Europe
C014465.fna	ST2627	8	99.09	94.63	KL123	human	Europe
SPARK_2545_C1.fna	ST4175	10	99.16	94.68	KL149	animal	Europe
SPARK_1744_C1.fna	ST1563	8	99.1	94.66	KL123	human	Europe
SPARK_2054_C2.fna	ST1563	8	99.11	94.63	KL123	human	Europe
SPARK_26_C1.fna	ST1563	8	99.09	94.43	KL123	human	Europe
SPARK_39_C1.fna	ST1563	8	99.1	94.45	KL123	human	Europe

Table A.3: Metadata of *Klebsiella variicola* strains carrying *rmp* locus. The *rmp* genes were identified using Kleborate (M. Lam et al., 2021).

Genome Accession	Source	Continent	Plasmid	AMR	Siderophore	Capsule
GCF_004014895.1	human	Asia (China)	IncX5_2__MF062700, IncFIB(pNDM-Mar)_1__JN420336, IncFIB(K)(pCAV1099-114)_1__CP011596, IncFII(pKP91)_1__CP000966	*oqxB7, oqxA6*, $bla_{\text{LEN-34}}$, *fosA*, $bla_{\text{KPC-2}}$;	salmonchelin, aerobactin	KL16
GCF_900636185.1	missing	missing	IncFIB(K)_1__JN233704, IncFII(K)_1__CP000648	*oqxB16, oqxA6*, $bla_{\text{LEN-2}}$, *fosA*	yersiniabactin, salmonchelin	None
GCF_004312545.1	human	Asia (Japan)	none	$bla_{\text{LEN-11}}$, *oqxA6*,	yersiniabactin, salmonchelin	KL2
GCF_004311405.1	human	Asia (Japan)	IncHI1B(pNDM-MAR)_1__JN420336, repB_KLEB_VIR_AP006726	$bla_{\text{LEN-11}}$, *oqxB15, oqxA6, fosA*	yersiniabactin, salmonchelin	KL2
GCF_002852915.4	human	Asia (China)	IncHI1B(pNDM-MAR)_1__JN420336, repB_KLEB_VIR_AP006726	*oqxB7, oqxA6*, *blaLEN-34*, *fosA*	yersiniabactin, salmonchelin, aerobactin	KL16
RIVM_C036606	human	Europe (Netherland)	IncFIA(HI1)_1__AF250878, IncFIA_1__AP001918, FIA(pBK30683)_1__KF954760, IncFII(K)_1__CP000648	*oqxA6, oqxB15*, *fosA*, $bla_{\text{LEN-34}}$	yersiniabactin, salmonchelin	KL101

Table A.4: Novel STs identified in *Klebsiella variicola* using MLSTKv (Barrios-Camacho et al., 2019).

ST type	Genome Accession	*leuS*	*pgi*	*pgk*	*phoE*	*pyrG*	*rpoB*	*fusA*
new1	GCF_013305245.1	1	19	10	12	1	1	4
new10	GCF_015286725.1	1	6	12	37	1	1	4
new10	GCF_015288045.1	1	6	12	37	1	1	4
new10	GCF_015288785.1	1	6	12	37	1	1	4
new11	GCF_900622625.1	46	39	5	43	1	13	4
new12	GCF_012273075.1	1	3	12	12	1	1	2
new13	GCF_902166725.1	23	17	5	1	34	1	2
new13	GCF_902162515.1	23	17	5	1	34	1	2
new14	GCF_012393465.1	2	24	6	50	1	6	2
new16	GCF_901420855.1	7	1	12	1	56	1	2
new17	GCF_900452335.1	1	10	2	25	1	13	2
new18	GCF_009825415.1	10	27	12	44	11	4	2
new19	GCF_004311265.1	11	27	12	22	1	7	2
new20	GCF_900774435.1	9	4	6	25	1	1	2
new21	GCF_900084135.1	1	4	3	1	11	1	2
new21	GCF_900085165.1	1	4	3	1	11	1	2
new21	GCF_002740845.1	1	4	3	1	11	1	2
new21	GCF_002740885.1	1	4	3	1	11	1	2
new22	GCF_900774485.1	1	27	6	1	1	1	2

Table A.4: Novel STs identified in *Klebsiella variicola* using MLSTKv (Barrios-Camacho et al., 2019).

ST type	Genome Accession	*leuS*	*pgi*	*pgk*	*phoE*	*pyrG*	*rpoB*	*fusA*
new23	GCF_004313485.1	11	28	1	2	23	1	2
new24	GCF_900173315.1	1	10	14	42	22	13	6
new24	GCF_900174175.1	1	10	14	42	22	13	6
new24	GCF_900173285.1	1	10	14	42	22	13	6
new24	GCF_900174165.1	1	10	14	42	22	13	6
new25	GCF_013103315.1	16	1	5	41	1	1	6
new26	GCF_011075275.1	25	1	12	8	1	1	2
new26	GCF_011075365.1	25	1	12	8	1	1	2
new27	GCF_900774135.1	9	4	6	1	1	1	2
new28	GCF_016544565.1	9	4	6	1	1	1	12
new29	GCF_014235385.1	9	18	12	2	3	1	2
new3	GCF_015284185.1	1	19	1	1	35	1	1
new3	GCF_015288445.1	1	19	1	1	35	1	1
new3	GCF_015286965.1	1	19	1	1	35	1	1
new3	GCF_015289325.1	1	19	1	1	35	1	1
new3	GCF_015287425.1	1	19	1	1	35	1	1
new3	GCF_015285645.1	1	19	1	1	35	1	1
new31	GCF_900494265.1	2	24	6	2	34	1	2
new32	GCF_900559295.1	11	4	6	8	2	2	2
new33	GCF_004312985.1	9	9	13	1	1	1	2
new34	GCF_004311545.1	1	3	1	2	5	13	6
new35	GCF_004313105.1	1	19	12	22	3	1	2
new36	GCF_004312565.1	2	27	3	55	2	1	4
new37	GCF_900171985.1	9	6	6	35	22	1	2
new38	GCF_004313525.1	11	19	5	44	35	3	4
new39	GCF_014283115.1	11	27	2	8	45	4	2
new4	GCF_015287805.1	11	1	6	24	2	1	4
new40	GCF_902702985.1	1	19	5	32	9	1	4
new41	GCF_902166225.1	23	17	5	1	34	1	2
new42	GCF_004312425.1	22	4	6	38	1	1	5
new43	GCF_900174355.1	11	1	3	8	15	1	4
new43	GCF_900173385.1	11	1	3	8	15	1	4
new43	GCF_900173565.1	11	1	3	8	15	1	4
new43	GCF_900174345.1	11	1	3	8	15	1	4
new43	GCF_900174375.1	11	1	3	8	15	1	4
new44	GCF_001549975.1	2	10	1	29	1	1	4

Table A.4: Novel STs identified in *Klebsiella variicola* using MLSTKv (Barrios-Camacho et al., 2019).

ST type	Genome Accession	*leuS*	*pgi*	*pgk*	*phoE*	*pyrG*	*rpoB*	*fusA*
new45	GCF_902704665.1	11	17	6	51	34	13	2
new46	GCF_900978805.1	14	22	21	21	51	17	7
new47	GCF_004801715.1	11	27	5	1	11	1	6
new49	GCF_004312405.1	2	39	6	9	55	1	2
new50	GCF_902508975.1	18	4	5	1	1	1	2
new51	GCF_014842805.1	2	5	10	3	1	1	2
new52	GCF_014286295.1	1	6	5	2	22	1	6
new54	GCF_016635665.1	1	1	4	2	2	2	2
new55	GCF_900978675.1	30	25	1	39	40	17	4
new56	GCF_900978645.1	30	25	1	39	51	17	4
new57	GCF_014595765.1	1	20	1	1	11	16	6
new58	GCF_900084775.1	1	27	2	24	5	1	4
new59	GCF_003196115.1	6	39	16	11	11	1	6
new6	GCF_013103345.1	16	1	5	41	1	-	7
new60	GCF_900978625.1	20	25	20	49	51	17	7
new61	GCF_900978755.1	30	6	21	21	40	17	4
new62	GCF_016635235.1	10	9	6	1	45	1	6
new63	GCF_900085015.1	1	4	6	20	1	1	11
new64	GCF_016634785.1	46	28	12	11	11	1	4
new65	GCF_014902675.1	10	3	12	8	3	5	2
new66	GCF_900085185.1	23	28	5	2	15	7	2
new67	GCF_900084765.1	11	36	2	8	3	1	2
new68	GCF_902509095.1	1	5	7	2	2	1	2
new69	GCF_014285035.1	9	12	6	32	1	1	2
new70	GCF_900576775.1	1	6	6	48	32	1	2
new71	GCF_014595735.1	51	39	6	33	39	13	4
new71	GCF_014595745.1	51	39	6	33	39	13	4
new72	GCF_903971005.1	16	1	5	1	5	1	6
new72	GCF_903970905.1	16	1	5	1	5	1	6
new73	GCF_900084795.1	23	17	2	1	34	1	2
new74	GCF_004312005.1	1	9	1	22	37	11	10
new76	GCF_900084995.1	9	4	3	8	1	1	6
new77	GCF_016635725.1	1	6	6	8	1	7	6
new78	GCF_900978435.1	30	25	21	21	40	17	4
new8	GCF_015290365.1	23	19	5	1	52	1	2
new8	GCF_015290565.1	23	19	5	1	52	1	2

Table A.4: Novel STs identified in *Klebsiella variicola* using MLSTKv (Barrios-Camacho et al., 2019).

ST type	Genome Accession	*leuS*	*pgi*	*pgk*	*phoE*	*pyrG*	*rpoB*	*fusA*
new80	GCF_002851635.1	47	4	5	52	1	13	4
new81	GCF_015679375.1	11	5	5	3	5	1	4
new82	GCF_011394875.1	44	25	20	21	51	17	7
new83	GCF_001068855.1	18	9	1	22	4	7	6
new84	GCF_003196055.1	13	19	1	38	2	1	6
new85	GCF_016138015.1	1	10	1	1	1	1	2
new86	GCF_015670265.1	46	18	10	38	1	1	2

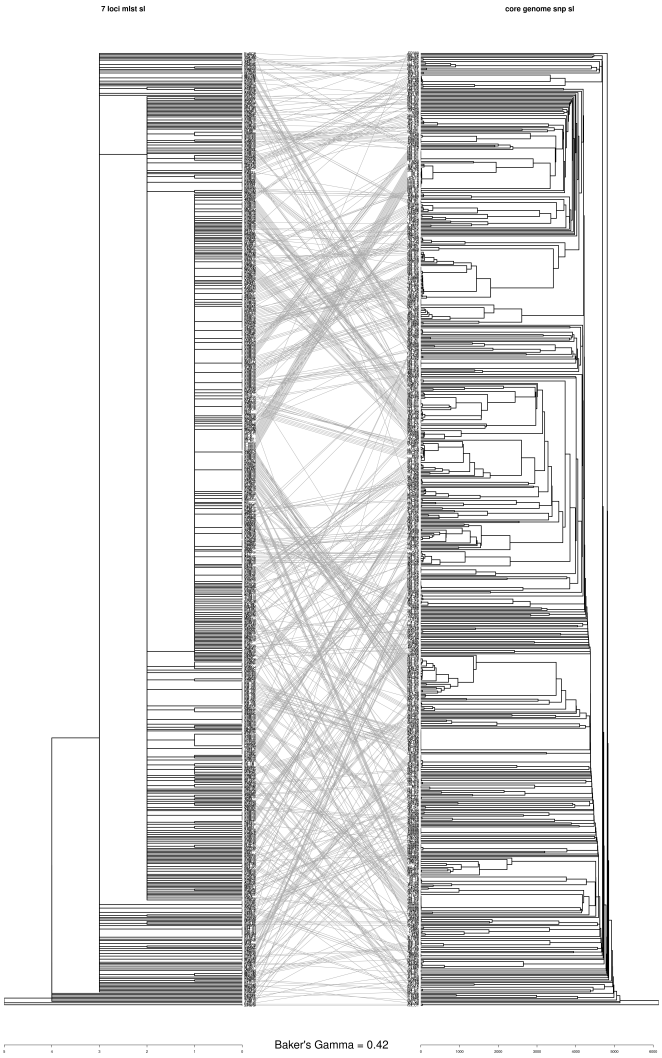


Figure A.1: Comparison of Multi-Locus Sequence Typing scheme and the core genome phylogeny in *Klebsiella variicola*

Appendix B

Supplementary Data for Chapter 4

Table B.1: Metadata of Scotland *Klebsiella variicola* dataset

Count	Strain	Location	Source	Species
1	kv_uoe_1	Edinburgh	urine	*K. var*
2	kv_uoe_2	Edinburgh	urine	*K. var*
3	kv_uoe_3	Edinburgh	urine	*K. var*
4	kv_uoe_4	Edinburgh	urine	*K. var*
5	kv_uoe_5	Edinburgh	urine	*K. var*
6	kv_uoe_6	Edinburgh	urine	*K. var*
7	kv_uoe_7	Edinburgh	urine	*K. var*
8	kv_uoe_8	Edinburgh	sputum	*K. var*
9	kv_uoe_9	Edinburgh	urine	*K. var*
10	kv_uoe_10	Edinburgh	urine	*K. var*
11	kv_uoe_11	Edinburgh	blood	*K. var*
12	kv_uoe_12	Edinburgh	urine	*K. var*
13	kv_uoe_13	Edinburgh	urine	*K. var*
14	kv_uoe_14	Edinburgh	urine	*K. var*
15	kv_uoe_15	Edinburgh	urine	*K. var*
16	kv_uoe_16	Edinburgh	urine	*K. var*
17	kv_uoe_17	Edinburgh	urine	*K. var*
18	kv_uoe_18	Edinburgh	urine	*K. var*
19	kv_uoe_19	Edinburgh	urine	*K. var*
20	kv_uoe_20	Edinburgh	urine	*K. var*
21	kv_uoe_21	Edinburgh	urine	*K. var*
22	kv_uoe_22	Edinburgh	urine	*K. var*
23	kv_uoe_23	Edinburgh	urine	*K. var*
24	kv_uoe_24	Edinburgh	urine	*K. var*
25	kv_uoe_25	Edinburgh	urine	*K. var*
26	kv_uoe_26	Edinburgh	urine	*K. var*

Count	Strain	Location	Source	Species
27	kv_uoe_27	Edinburgh	urine	*K. var*
28	kv_uoe_28	Edinburgh	urine	*K. var*
29	kv_uoe_29	Edinburgh	urine	*K. var*
30	kv_uoe_30	Edinburgh	urine	*K. var*
31	kv_uoe_31	Edinburgh	urine	*K. var*
32	kv_uoe_32	Edinburgh	urine	*K. var*
33	kv_uoe_33	Edinburgh	urine	*K. var*
34	kv_uoe_34	Edinburgh	urine	*K. var*
35	kv_uoe_35	Edinburgh	urine	*K. var*
36	kv_uoe_36	Edinburgh	urine	*K. var*
37	kv_uoe_37	Edinburgh	urine	*K. var*
38	kv_uoe_38	Edinburgh	blood	*K. var*
39	kv_uoe_39	Edinburgh	urine	*K. var*
40	kv_uoe_40	Edinburgh	urine	*K. var*
41	kv_uoe_41	Edinburgh	urine	*K. var*
42	kv_uoe_42	Edinburgh	urine	*K. var*
43	kv_uoe_43	Edinburgh	urine	*K. var*
44	kv_uoe_44	Edinburgh	wound	*K. var*
45	kv_uoe_45	Edinburgh	urine	*K. var*
46	kv_uoe_47	Edinburgh	urine	*K. var*
47	kv_uoe_48	Edinburgh	urine	*K. var*
48	kv_uoe_49	Edinburgh	urine	*K. var*
49	kv_uoe_50	Edinburgh	urine	*K. var*
50	kv_uoe_51	Edinburgh	urine	*K. var*
51	kv_uoe_52	Edinburgh	blood	*K. var*
52	kv_uoe_53	Edinburgh	urine	*K. var*
53	kv_uoe_54	Edinburgh	urine	*K. var*
54	kv_uoe_55	Edinburgh	urine	*K. var*
55	kv_uoe_56	Edinburgh	sputum	*K. var*
56	kv_uoe_57	Edinburgh	blood	*K. var*
57	kv_uoe_59	Edinburgh	urine	*K. quasi*
58	kv_uoe_60	Edinburgh	urine	*K. var*
59	kv_uoe_61	Edinburgh	urine	*K. var*
60	kv_uoe_62	Edinburgh	urine	*K. var*
61	kv_uoe_63	Edinburgh	urine	*K. var*
62	kv_uoe_64	Edinburgh	urine	*K. var*
63	kv_uoe_65	Edinburgh	urine	*K. var*
64	kv_uoe_66	Edinburgh	sputum	*K. var*

Count	Strain	Location	Source	Species
65	kv_uoe_67	Edinburgh	urine	*K. var*
66	kv_uoe_68	Edinburgh	urine	*K. var*
67	kv_uoe_69	Edinburgh	urine	*K. var*
68	kv_uoe_70	Edinburgh	urine	*K. var*
69	kv_uoe_71	Edinburgh	urine	*K. var*
70	kv_uoe_72	Edinburgh	gallbladder	*K. var*
71	kv_uoe_73	Edinburgh	urine	*K. var*
72	kv_uoe_74	Edinburgh	blood	*K. var*
73	kv_uoe_75	Dumfries & Galloway	kidney	*K. var*
74	kv_uoe_76	Dumfries & Galloway	urine	*K. var*
75	kv_uoe_77	Dumfries & Galloway	urine	*K. var*
76	kv_uoe_78	Dumfries & Galloway	urine	*K. var*
77	kv_uoe_79	Dumfries & Galloway	urine	*K. var*
78	kv_uoe_80	Dumfries & Galloway	gallbladder	*K. var*
79	kv_uoe_81	Dumfries & Galloway	gallbladder	*K. var*
80	kv_uoe_82	Dumfries & Galloway	urine	*K. var*
81	kv_uoe_83	Dumfries & Galloway	urine	*K. var*
82	kv_uoe_84	Dumfries & Galloway	urine	*K. var*
83	kv_uoe_85	Dumfries & Galloway	urine	*K. var*
84	kv_uoe_86	Dumfries & Galloway	wound	*K. var*
85	kv_uoe_89	Dumfries & Galloway	urine	*K. var*
86	kv_uoe_90	Dumfries & Galloway	urine	*K. var*
87	kv_uoe_91	Dumfries & Galloway	urine	*K. var*
88	kv_uoe_92	Dumfries & Galloway	urine	*K. var*
89	kv_uoe_93	Dumfries & Galloway	urine	*K. var*
90	kv_uoe_94	Dumfries & Galloway	urine	*K. var*
91	kv_uoe_95	Dumfries & Galloway	urine	*K. var*
92	kv_uoe_96	Dumfries & Galloway	urine	*K. var*
93	kv_uoe_97	Dumfries & Galloway	urine	*K. var*
94	kv_uoe_98	Dumfries & Galloway	urine	*K. var*
95	kv_uoe_99	Dumfries & Galloway	urine	*K. var*
96	kv_uoe_100	Dumfries & Galloway	urine	*K. var*
97	kv_uoe_101	Dumfries & Galloway	sputum	*K. var*
98	kv_uoe_102	Dumfries & Galloway	urine	*K. var*
99	kv_uoe_103	Dumfries & Galloway	urine	*K. var*
100	kv_uoe_104	Dumfries & Galloway	urine	*K. var*
101	kv_uoe_119	Edinburgh	urine	*K. var*
102	kv_uoe_120	Edinburgh	urine	*K. var*

Count	Strain	Location	Source	Species
103	kv_uoe_121	Edinburgh	sputum	*K. var*
104	kv_uoe_122	Edinburgh	urine	*K. pne*
105	kv_uoe_123	Edinburgh	urine	*K. var*
106	kv_uoe_124	Edinburgh	kidney	*K. var*
107	kv_uoe_125	Edinburgh	urine	*K. var*
108	kv_uoe_126	Edinburgh	intraperitoneal tissue	*K. var*
109	kv_uoe_127	Edinburgh	peritoneal fluid	*K. var*
110	kv_uoe_128	Edinburgh	urine	*K. var*

Table B.2: Genome-Wide Association Analyses (GWAS) of Scottish *K. var* isolates on the difference of Urinary Tract Infection (UTI) and other host diseases. The pan-genome was built using Roary (Page et al., 2015), and the GWAS analyses was applied using Scoary (Brynildsrud et al., 2016).

Rank	Protein encoded	Present Others	Present UTI	Absent Others	Absent UTI	Odds ratio	Bonferroni corrected *p* value	FDR corrected *p* value
1	IS110 family transposase ISKpn43	7	3	12	86	16.72	1	0.789
2	Antitoxin CcdA	10	11	9	78	7.879	1	0.789
3	Toxin CcdB	10	12	9	77	7.130	1	0.789
4	hypothetical protein	4	0	15	89	inf	1	0.789
5	hypothetical protein	4	0	15	89	inf	1	0.789
6	hypothetical protein	6	3	13	86	13.23	1	0.789
7	IS110 family transposase IS5075	11	17	8	72	5.824	1	0.789
8	Cyclic di-GMP phosphodiesterase PdeH	1	39	18	50	0.071	1	0.789
9	hypothetical protein	9	11	10	78	6.382	1	0.789
10	hypothetical protein	7	6	12	83	8.069	1	0.789
11	hypothetical protein	17	45	2	44	8.311	1	0.789
12	hypothetical protein	17	46	2	43	7.946	1	0.789
13	hypothetical protein	2	43	17	46	0.126	1	0.789
14	Hemolysin transporter protein ShlB	2	43	17	46	0.126	1	0.789
15	hypothetical protein	1	37	18	52	0.078	1	0.789
16	hypothetical protein	19	61	0	28	inf	1	0.789
17	hypothetical protein	4	1	15	88	23.47	1	0.789
18	hypothetical protein	4	1	15	88	23.47	1	0.789
19	hypothetical protein	4	1	15	88	23.47	1	0.789
20	Na(+)/H(+) antiporter NhaA	4	1	15	88	23.47	1	0.789

Table B.3: Genome-Wide Association Analyses (GWAS) of Scottish *K. var* isolates on the difference of Edinburgh and Dumphries & Galloway (border) isolates. The pan-genome was built using Roary (Page et al., 2015), and the GWAS analyses was applied using Scoary (Brynildsrud et al., 2016).

Rank	Protein encoded	Present Border	Present Edinburgh	Absent Border	Absent Edinburgh	Odds ratio	Bonferroni corrected *p* value	FDR corrected *p* value
1	hypothetical protein	0	34	28	46	0	0.060	0.060
2	HTH-type transcriptional regulator IscR	27	46	1	34	19.96	0.861	0.350
3	hypothetical protein	26	44	2	36	10.64	1	0.350
4	hypothetical protein	6	0	22	80	inf	1	0.350
5	hypothetical protein	24	37	4	43	6.973	1	0.350
6	N-carbamoyl-L-amino acid hydrolase	4	43	24	37	0.143	1	0.350
7	hypothetical protein	4	43	24	37	0.143	1	0.350
8	hypothetical protein	4	43	24	37	0.143	1	0.350
9	Serine transporter	4	43	24	37	0.143	1	0.350
10	Leucine-responsive regulatory protein	4	43	24	37	0.143	1	0.350
11	Diaminopropionate ammonia-lyase	4	43	24	37	0.143	1	0.350
12	hypothetical protein	26	45	2	35	10.11	1	0.350
13	hypothetical protein	1	30	27	50	0.062	1	0.350
14	putative diguanylate cyclase DgcC	27	51	1	29	15.36	1	0.350
15	hypothetical protein	19	23	9	57	5.232	1	0.350
16	Anaerobic nitric oxide reductase transcription regulator NorR	11	7	17	73	6.748	1	0.350
17	Staphyloferrin B transporter	11	7	17	73	6.748	1	0.350
18	Glutamyl-tRNA(Gln) amidotransferase subunit A	11	7	17	73	6.748	1	0.350
19	hypothetical protein	11	7	17	73	6.748	1	0.350
20	Isochorismatase family protein YecD	11	7	17	73	6.748	1	0.350

Appendix C

Supplementary Data for Chapter 5

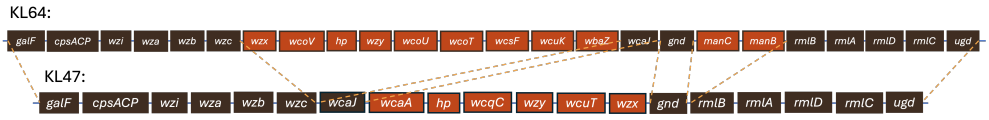


Figure C.1: Comparison of capsule locus in *Klebsiella pneumoniae* ST11

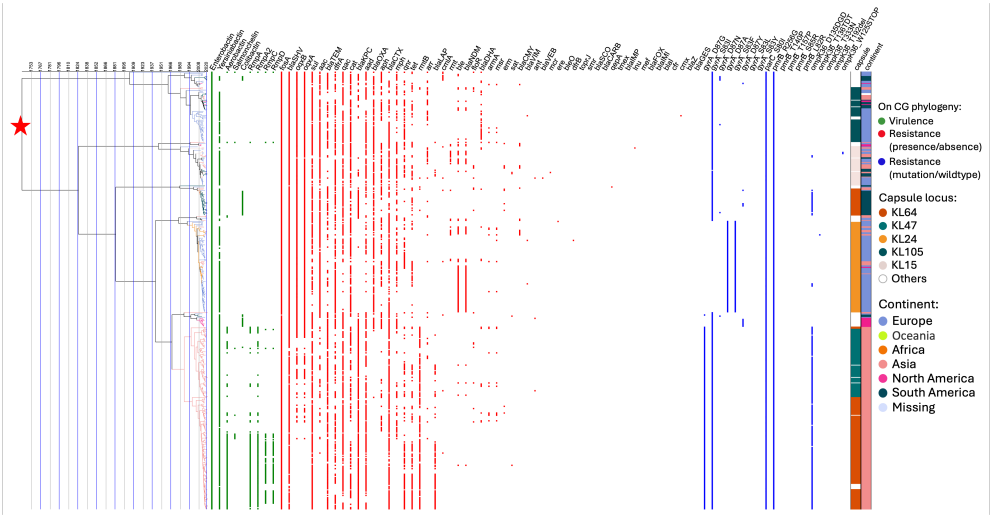


Figure C.2: Time-resolved phylogeny of *Klebsiella pneumoniae* ST11

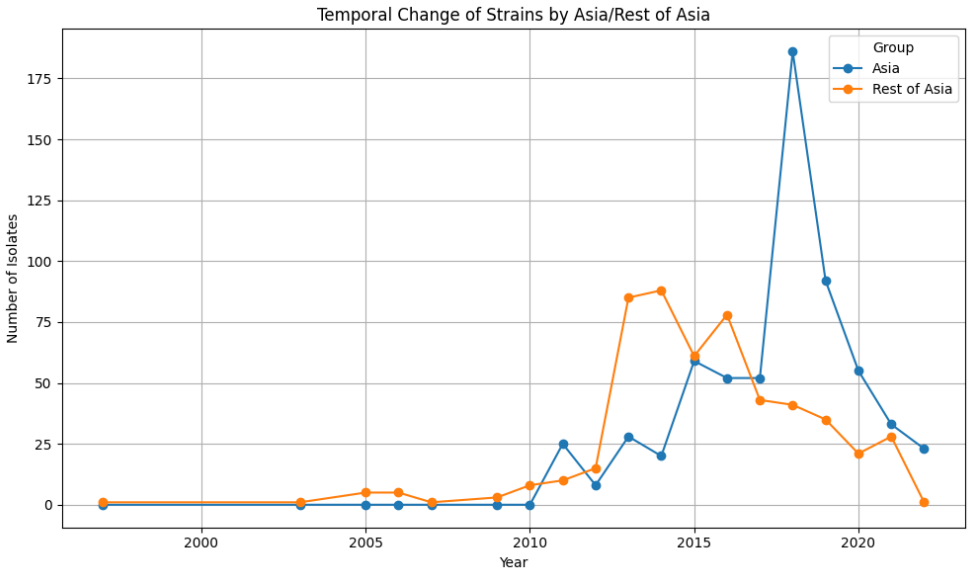


Figure C.3: Temporal change plot of Asian cluster and rest of Asian cluster in *Klebsiella pneumoniae* ST11

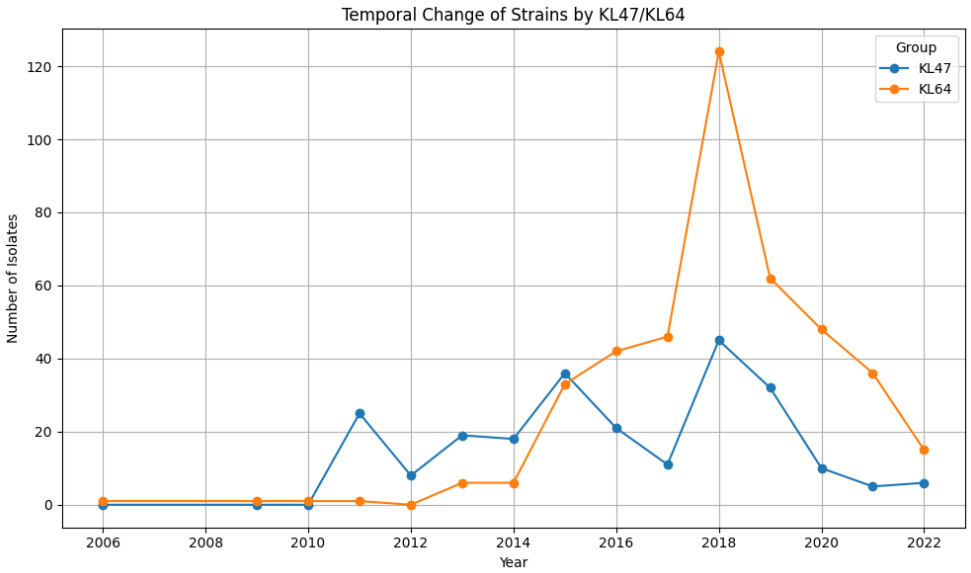


Figure C.4: Temporal change plot of KL47 and KL64 clusters in *Klebsiella pneumoniae* ST11

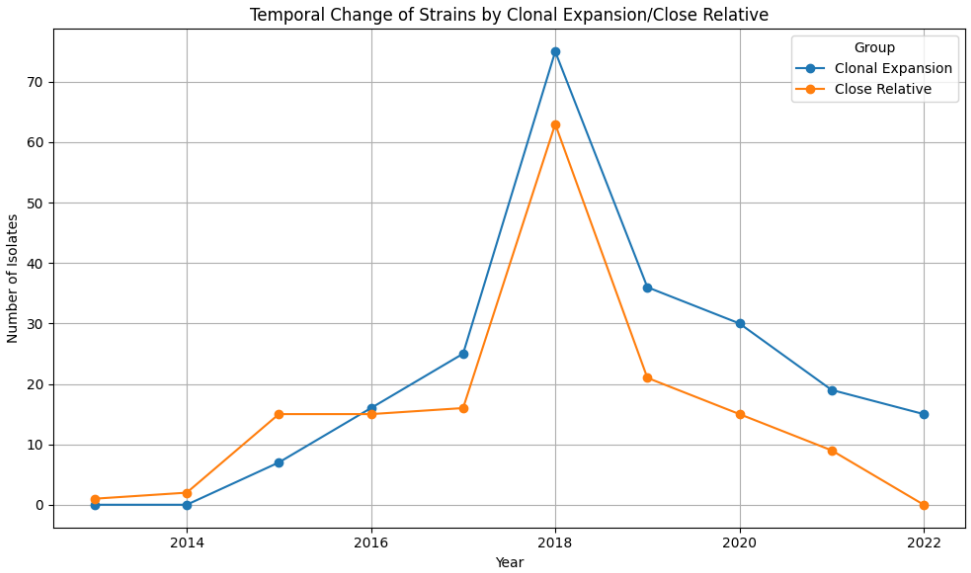


Figure C.5: Temporal change plot of the clonal expansion and its close relative clusters in *Klebsiella pneumoniae* ST11

Bibliography

Abdi, H., et al. (2007). Bonferroni and šidák corrections for multiple comparisons. *Encyclopedia of Measurement and Statistics*, *3*(01), 2007.

Abozahra, R., Gaballah, A., & Abdelhamid, S. M. (2023). Prevalence of the colistin resistance gene MCR-1 in colistin-resistant *Klebsiella pneumoniae* in Egypt. *AIMS microbiology*, *9*(2), 177.

Acman, M., Wang, R., van Dorp, L., Shaw, L. P., Wang, Q., Luhmann, N., ... others (2022). Role of mobile genetic elements in the global dissemination of the carbapenem resistance gene bla_{NDM}. *Nature communications*, *13*(1), 1131.

Addington, E., Sandalli, S., & Roe, A. J. (2024). Current understandings of colibactin regulation. *Microbiology*, *170*(2), 001427.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., ... Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–249.

Aghamohammad, S., Khazani Asforooshani, M., Malek Mohammadi, Y., Sholeh, M., & Badmasti, F. (2023). Decoding the genetic structure of conjugative plasmids in international clones of klebsiella pneumoniae: A deep dive into bla_{KPC}, bla_{NDM}, $bla_{\text{OXA-48}}$, and bla_{GES} genes. *Plos one*, *18*(11), e0292288.

Akenten, C. W., Khan, N. A., Mbwana, J., Krumkamp, R., Fosu, D., Paintsil, E. K., ... others (2023). Carriage of ESBL-producing *Klebsiella pneumoniae* and *Escherichia coli* among children in rural Ghana: a cross-sectional study. *Antimicrobial Resistance & Infection Control*, *12*(1), 60.

Akinola, O. T., & Dahunsi, S. O. (2024). Whole genome sequencing reveals antibiotic resistance pattern and virulence factors in *Klebsiella quasipneumoniae* subsp. *Similipneumoniae* from hospital wastewater in south-west, Nigeria. *Microbial Pathogenesis*, *197*, 107040.

Al-Busaidi, B., Al-Muzahmi, M., Al-Shabibi, Z., Rizvi, M., Al-Rashdi, A., Al-Jardani, A., ... Al-Jabri, Z. (2024). Hypervirulent capsular serotypes K1 and K2 *Klebsiella pneumoniae* strains demonstrate resistance to serum bactericidal activity and Galleria mellonella lethality. *International Journal of Molecular Sciences*, *25*(3), 1944.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Álvarez, V. E., Massó, M. G., González, G. D., Gambino, A. S., Knecht, C. A., Mc Cormick, B. P., ... others (2022). Emergence of colistin resistance in *Klebsiella pneumoniae* ST15 disseminating *bla*KPC-2 in a novel genetic platform. *Journal of Global Antimicrobial Resistance*, *29*, 537–539.

Al-Zahrani, I. A., & Alasiri, B. A. (2018). The emergence of carbapenem-resistant *Klebsiella pneumoniae* isolates producing OXA-48 and NDM in the Southern (Asir) province, Saudi Arabia. *Saudi medical journal*, *39*(1), 23.

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome biology*, *21*(1), 30.

Amina Nazir, Y. Z., et al. (2020). Structural genomics of *repA*, *repB1*-carrying IncFIB family pA1705-*qnrS*, P911021-*tetA*, and P1642-*tetA*, multidrug-resistant plasmids from *Klebsiella pneumoniae*. *Infection and Drug Resistance*, *13*, 1889–1903. Retrieved from `https://www.tandfonline.com/doi/abs/10.2147/IDR.S228704` doi: 10.2147/IDR.S228704

Andrews, S., et al. (2010). FastQC: a quality control tool for high throughput sequence data. *GitHub Repository*.

Anes, J., Dever, K., Eshwar, A., Nguyen, S., Cao, Y., Sivasankaran, S. K., ... others (2021). Analysis of the oxidative stress regulon identifies *soxS* as a genetic target for resistance reversal in multidrug-resistant *Klebsiella pneumoniae*. *MBio*, *12*(3), 10–1128.

Añorga, M., Urriza, M., Ramos, C., & Murillo, J. (2022). Multiple relaxases contribute to the horizontal transfer of the virulence plasmids from the tumorigenic bacterium *Pseudomonas syringae pv. savastanoi* NCPPB 3335. *Frontiers in Microbiology*, *13*, 1076710.

Arabaghian, H., Salloum, T., Alousi, S., Panossian, B., Araj, G. F., & Tokajian, S. (2019). Molecular characterization of carbapenem resistant klebsiella pneumoniae and klebsiella quasipneumoniae isolated from lebanon. *Scientific Reports*, *9*(1), 531.

Arcari, G., & Carattoli, A. (2023). Global spread and evolutionary convergence of multidrug-resistant and hypervirulent *Klebsiella pneumoniae* high-risk clones. *Pathogens and Global Health*, *117*(4), 328–341.

Arena, F., Henrici De Angelis, L., Pieralli, F., Di Pilato, V., Giani, T., Torricelli, F., ... Rossolini, G. M. (2015). Draft genome sequence of the first hypermucoviscous *Klebsiella quasipneumoniae subsp. quasipneumoniae* isolate from a bloodstream infection. *Genome Announcements*, *3*(5), 10–1128.

Assessment, R. (2021). Emergence of hypervirulent *Klebsiella pneumoniae* ST23 carrying carbapenemase genes in EU/EEA countries. *European Centre for Disease Prevention and Control [Internet]*, *17*.

Azam, M., Gaind, R., Yadav, G., Sharma, A., Upmanyu, K., Jain, M., & Singh, R. (2021). Colistin resistance among multiple sequence types of *Klebsiella pneumoniae* is associated with diverse resistance mechanisms: A report from India. *Frontiers in microbiology*, *12*, 609840.

Bachman, M. A., Oyler, J. E., Burns, S. H., Caza, M., Lépine, F., Dozois, C. M., & Weiser, J. N. (2011). *Klebsiella pneumoniae* yersiniabactin promotes respiratory tract infection through evasion of lipocalin 2. *Infection and immunity*, *79*(8), 3309–3316.

Bagley, S. T. (1985). Habitat association of *Klebsiella* species. *Infection Control & Hospital Epidemiology*, *6*(2), 52–58.

Bailey, D. C., Alexander, E., Rice, M. R., Drake, E. J., Mydy, L. S., Aldrich, C. C., & Gulick, A. M. (2018). Structural and functional delineation of aerobactin biosynthesis in hypervirulent *Klebsiella pneumoniae*. *Journal of Biological Chemistry*, *293*(20), 7841–7852.

Bain, W., Ahn, B., Peñaloza, H. F., McElheny, C. L., Tolman, N., van der Geest, R., ... others (2024). In vivo evolution of a *Klebsiella pneumoniae* capsule defect with *wcaJ* mutation promotes complement-mediated opsonophagocytosis during recurrent infection. *The Journal of Infectious Diseases*, jiae003.

Baker, F. B. (1974). Stability of two hierarchical grouping techniques case I: sensitivity to data errors. *Journal of the American Statistical Association*, *69*(346), 440–445.

Bandeira, M., Almeida Carvalho, P., Duarte, A., & Jordao, L. (2014). Exploring dangerous connections between *Klebsiella pneumoniae* biofilms and healthcare-associated infections. *Pathogens*, *3*(3), 720–731.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... others (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477.

Barrios-Camacho, H., Aguilar-Vera, A., Beltran-Rojel, M., Aguilar-Vera, E., Duran-Bedolla, J., Rodriguez-Medina, N., ... Garza-Ramos, U. (2019). Molecular epidemiology of *Klebsiella variicola* obtained from different sources. *Scientific Reports*, *9*(1), 1–10.

Bej, A. K., Mahbubani, M. H., & Atlas, R. M. (1991). Amplification of nucleic acids by polymerase chain reaction (PCR) and other methods and their applications. *Critical reviews in biochemistry and molecular biology*, *26*(3-4), 301–334.

Bellifa, S., Hassaine, H., Balestrino, D., Charbonnel, N., M'hamedi, I., Terki, I. K., ... Forestier, C. (2013). Evaluation of biofilm formation of *Klebsiella pneumoniae* isolated from medical devices at the university hospital of Tlemcen, Algeria. *Afr J Microbiol Res*, *7*(49), 5558–64.

Bengoechea, J. A., & Sa Pessoa, J. (2019). *Klebsiella pneumoniae* infection biology: living to counteract host defences. *FEMS microbiology reviews*, *43*(2), 123–144.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289–300.

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

Boonstra, M. B., Spijkerman, D. C., Voor, A. F., van der Laan, R. J., Bode, L. G., van Vianen, W., ... others (2020). An outbreak of ST307 extended-spectrum beta-lactamase (ESBL)–producing *Klebsiella pneumoniae* in a rehabilitation center: An unusual source and route of transmission. *Infection Control & Hospital Epidemiology*, *41*(1), 31–36.

Boralli, C. M. d. S., Paganini, J. A., Meneses, R. S., Mata, C. P. S. M. d., Leite, E. M. M., Schürch, A. C., ... Camargo, I. L. B. C. (2023). Characterization of $bla_{\text{KPC-2}}$ and $bla_{\text{NDM-1}}$ plasmids of a *K. pneumoniae* ST11 outbreak clone. *Antibiotics*, *12*(5), 926.

Boskova, V., & Stadler, T. (2020). PIQMEE: Bayesian phylodynamic method for analysis of large data sets with duplicate sequences. *Molecular Biology and Evolution*, *37*(10), 3061–3075.

Breilh, D., Texier-Maugein, J., Allaouchiche, B., Saux, M.-C., & Boselli, E. (2013). Carbapenems. *Journal of Chemotherapy*, *25*(1), 1–17.

Brynildsrud, O., Bohlin, J., Scheffer, L., & Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome biology*, *17*(1), 1–9.

Burmeister, A. R. (2015). Horizontal gene transfer. *Evolution, medicine, and public health*, *2015*(1), 193–194.

Bush, K., & Bradford, P. A. (2016). β-lactams and β-lactamase inhibitors: an overview. *Cold Spring Harbor perspectives in medicine*, *6*(8), a025247.

Calland, J. K., Haukka, K., Kpordze, S. W., Brusah, A., Corbella, M., Merla, C., ... others (2023). Population structure and antimicrobial resistance among *Klebsiella* isolates sampled from human, animal, and environmental sources in ghana: a cross-sectional genomic One Health study. *The Lancet Microbe*, *4*(11), e943–e952.

Carattoli, A., Zankari, E., Garcìa-Fernandez, A., Larsen, M. V., Lund, O., Villa, L., ... Hasman, H. (2014). PlasmidFinder and pMLST: in silico detection and typing of plasmids. *Antimicrobial Agents and Chemotherapy*, AAC–02412.

Carvalho, I., Chenouf, N. S., Carvalho, J. A., Castro, A. P., Silva, V., Capita, R., ... others (2021). Multidrug-resistant *Klebsiella pneumoniae* harboring extended spectrum β-lactamase encoding genes isolated from human septicemias. *PLoS One*, *16*(5), e0250525.

Castanheira, M., Simner, P. J., & Bradford, P. A. (2021). Extended-spectrum β-lactamases: an update on their characteristics, epidemiology and detection. *JAC-antimicrobial resistance*, *3*(3), dlab092.

Chao, A., Gotelli, N. J., Hsieh, T. C., Sande, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, *84*, 45–67.

Chen, H. D., & Groisman, E. A. (2013). The biology of the PmrA/PmrB two-component system: the major regulator of lipopolysaccharide modifications. *Annual review of microbiology*, *67*(1), 83–112.

Chen, L., Mathema, B., Pitout, J. D., DeLeo, F. R., & Kreiswirth, B. N. (2014). Epidemic *Klebsiella pneumoniae* ST258 is a hybrid strain. *MBio*, *5*(3), e01355–14.

Chen, M., Li, Y., Li, S., Tang, L., Zheng, J., & An, Q. (2016). Genomic identification of nitrogen-fixing *Klebsiella variicola*, *K. pneumoniae* and *K. quasipneumoniae*. *Journal of Basic Microbiology*, *56*(1), 78–84.

Chen, Y.-T., Chang, H.-Y., Lai, Y.-C., Pan, C.-C., Tsai, S.-F., & Peng, H.-L. (2004). Sequencing and analysis of the large virulence plasmid pLVPK of *Klebsiella pneumoniae* CG43. *Gene*, *337*, 189–198.

Choby, J., Howard-Anderson, J., & Weiss, D. (2020). Hypervirulent *Klebsiella pneumoniae*–clinical and molecular perspectives. *Journal of Internal Medicine*, *287*(3), 283–300.

Chuang, Y.-P., Fang, C.-T., Lai, S.-Y., Chang, S.-C., & Wang, J.-T. (2006). Genetic determinants of capsular serotype K1 of *Klebsiella pneumoniae* causing primary pyogenic liver abscess. *The Journal of Infectious Diseases*, *193*(5), 645–654.

Clemente, A. M., Castronovo, G., Antonelli, A., D'Andrea, M. M., Tanturli, M., Perissi, E., ... others (2017). Differential Th17 response induced by the two clades of the pandemic ST258 *Klebsiella pneumoniae* clonal lineages producing KPC-type carbapenemase. *PLoS One*, *12*(6), e0178847.

Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., ... Harris, S. R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic acids research*, *43*(3), e15–e15.

Dangor, Z., Benson, N., Berkley, J. A., Bielicki, J., Bijsma, M. W., Broad, J., ... others (2024). Vaccine value profile for *Klebsiella pneumoniae*. *Vaccine*.

Daoud, L., Al-Marzooq, F., Moubareck, C. A., Ghazawi, A., & Collyns, T. (2022). Elucidating the effect of iron acquisition systems in *Klebsiella pneumoniae* on susceptibility to the novel siderophore-cephalosporin cefiderocol. *PLoS One*, *17*(12), e0277946.

David, S., Reuter, S., Harris, S. R., Glasner, C., Feltwell, T., Argimon, S., ... others (2019). Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. *Nature microbiology*, *4*(11), 1919–1929.

David, S., Wong, J. L., Sanchez-Garrido, J., Kwong, H.-S., Low, W. W., Morecchiato, F., ... others (2022). Widespread emergence of OmpK36 loop 3 insertions among multidrug-resistant clones of *Klebsiella pneumoniae*. *PLoS Pathogens*, *18*(7), e1010334.

De Majumdar, S., Yu, J., Fookes, M., McAteer, S. P., Llobet, E., Finn, S., ... others (2015). Elucidation of the RamA regulon in *Klebsiella pneumoniae* reveals a role in LPS regulation. *PLOS Pathogens*, *11*(1), e1004627.

de Oliveira Alves, W., Scavuzzi, A. M. L., Beltrão, E. M. B., de Oliveira, É. M., dos Santos Vasconcelos, C. R., Rezende, A. M., & de Souza Lopes, A. C. (2022). Occurrence of *bla*$_{\text{NDM-7}}$ and association with *bla*$_{\text{KPC-2}}$, *bla*$_{\text{CTX-M15}}$, *aac*, *aph*, *mph (a)*, *catB3* and virulence genes in a clinical isolate of *Klebsiella pneumoniae* with different plasmids in Brazil. *Archives of Microbiology*, *204*(8), 459.

Di, D. Y., Jang, J., Unno, T., & Hur, H.-G. (2017). Emergence of *Klebsiella variicola* positive for NDM-9, a variant of New Delhi metallo-β-lactamase, in an urban river in South Korea. *Journal of Antimicrobial Chemotherapy*, *72*(4), 1063–1067.

Diancourt, L., Passet, V., Verhoef, J., Grimont, P. A., & Brisse, S. (2005). Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *Journal of clinical microbiology*, *43*(8), 4178–4182.

Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R., & Wilson, D. J. (2018a). Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Research*, *46*(22), e134–e134.

Didelot, X., Croucher, N. J., Bentley, S. D., Harris, S. R., & Wilson, D. J. (2018b). Bayesian inference of ancestral dates on bacterial phylogenetic trees. *Nucleic Acids Research*, *46*, e134. doi: 10.1093/nar/gky783

Domingues, S., da Silva, G. J., & Nielsen, K. M. (2012). Integrons: vehicles and pathways for horizontal dissemination in bacteria. *Mobile genetic elements*, *2*(5), 211–223.

Dong, N., Lin, D., Zhang, R., Chan, E. W.-C., & Chen, S. (2018). Carriage of *bla*$_{\text{KPC-2}}$ by a virulence plasmid in hypervirulent *Klebsiella pneumoniae*. *Journal of Antimicrobial Chemotherapy*, *73*(12), 3317–3321.

Dong, N., Yang, X., Zhang, R., Chan, E. W.-C., & Chen, S. (2018). Tracking microevolution events among ST11 carbapenemase-producing hypervirulent *Klebsiella pneumoniae* outbreak strains. *Emerging microbes & infections*, *7*(1), 1–8.

Dorado-Morales, P., Vilanova, C., P. Garay, C., Martí, J. M., & Porcar, M. (2015). Unveiling bacterial interactions through multidimensional scaling and dynamics modeling. *Scientific Reports*, *5*(1), 18396.

Dorman, M. J., Feltwell, T., Goulding, D. A., Parkhill, J., & Short, F. L. (2018). The capsule regulatory network of *Klebsiella pneumoniae* defined by density-TraDISort. *MBio*, *9*(6), 10–1128.

Dos Santos, P. C., Fang, Z., Mason, S. W., Setubal, J. C., & Dixon, R. (2012). Distribution of nitrogen fixation and nitrogenase-like sequences amongst microbial genomes. *BMC genomics*, *13*, 1–12.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, *7*(1), 1–8.

Du, F.-l., Huang, Q.-s., Wei, D.-d., Mei, Y.-f., Long, D., Liao, W.-j., ... Zhang, W. (2021). Prevalence of carbapenem-resistant *Klebsiella pneumoniae* co-harboring *bla*$_{\text{KPC}}$-carrying plasmid and pLVPK-like virulence plasmid in bloodstream infections. *Frontiers in Cellular and Infection Microbiology*, *10*, 556654.

Duan, Q., Wang, Q., Sun, S., Cui, Q., Ding, Q., Wang, R., & Wang, H. (2022). ST11 carbapenem-resistant *Klebsiella pneumoniae* clone harboring *bla*$_{\text{NDM}}$ replaced a *bla*$_{\text{KPC}}$ clone in a tertiary hospital in China. *Antibiotics*, *11*(10), 1373.

Duran-Bedolla, J., Garza-Ramos, U., Rodríguez-Medina, N., Aguilar Vera, A., & Barrios-Camacho, H. (2021). Exploring the environmental traits and applications of *Klebsiella variicola*. *Brazilian Journal of Microbiology*, 1–13.

Eisler, P. (2012). Deadly superbugs invade US health care facilities. *USA Today*, *29*.

El Fertas-Aissani, R., Messai, Y., Alouache, S., & Bakour, R. (2013). Virulence profiles and antibiotic susceptibility patterns of *Klebsiella pneumoniae* strains isolated from different clinical specimens. *Pathologie Biologie*, *61*(5), 209–216.

Elliott, Z. S., Barry, K. E., Cox, H. L., Stoesser, N., Carroll, J., Vegesana, K., ... others (2019). The role of *fosA* in challenges with fosfomycin susceptibility testing of multispecies *Klebsiella pneumoniae* carbapenemase-producing clinical isolates. *Journal of clinical microbiology*, *57*(10), 10–1128.

Ernst, C. M., Braxton, J. R., Rodriguez-Osorio, C. A., Zagieboylo, A. P., Li, L., Pironti, A., ... others (2020). Adaptive evolution of virulence and persistence in carbapenem-resistant *Klebsiella pneumoniae*. *Nature medicine*, *26*(5), 705–711.

et al., A. C. (2017). Bayesian analysis of genetic association across tree-structured routine healthcare data in the UK Biobank. *Nature Genetics*. doi: NatureGenetics

Farooq, N., Bain, W., Ahn, B., Van Der Geest, R., Peñaloza, H., Stolz, D., ... Lee, J. (2020). Wcaj mutation in a carbapenemase-producing *Klebsiella pneumoniae* clinical isolate is associated with capsule alterations and increased serum resistance. In *D34. infections and the immune response* (pp. A6513–A6513). American Thoracic Society.

Farzana, R., Jones, L. S., Rahman, M. A., Andrey, D. O., Sands, K., Portal, E., ... Walsh, T. R. (2019). Outbreak of hypervirulent multidrug-resistant *Klebsiella variicola* causing high mortality in neonates in Bangladesh. *Clinical Infectious Diseases*, *68*(7), 1225–1227.

Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J. G., Haendiges, J., Haft, D. H., ... others (2021). AMRFinderPlus and the reference gene catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Scientific Reports*, *11*(1), 12728.

Feldman, M. F., Mayer Bridwell, A. E., Scott, N. E., Vinogradov, E., McKee, S. R., Chavez, S. M., ... Harding, C. M. (2019). A promising bioconjugate vaccine against hypervirulent *Klebsiella pneumoniae*. *Proceedings of the National Academy of Sciences*, *116*(37), 18655–18663.

Feng, J., Yin, Z., Zhao, Q., Zhao, Y., Zhang, D., Jiang, X., ... others (2017). Genomic characterization of novel IncFII-type multidrug resistant plasmids p0716-KPC and p12181-KPC from *Klebsiella pneumoniae*. *Scientific Reports*, *7*(1), 5830.

Feng, L., Zhang, M., & Fan, Z. (2023). Population genomic analysis of clinical ST15 *Klebsiella pneumoniae* strains in China. *Frontiers in Microbiology*, *14*, 1272173.

Fevre, C., Passet, V., Weill, F.-X., Grimont, P. A., & Brisse, S. (2005). Variants of the *Klebsiella pneumoniae* OKP chromosomal beta-lactamase are divided into two main groups, OKP-A and OKP-B. *Antimicrobial agents and chemotherapy*, *49*(12), 5149–5152.

Fleeman, R. M., Macias, L. A., Brodbelt, J. S., & Davies, B. W. (2020). Defining principles that influence antimicrobial peptide activity against capsulated *Klebsiella pneumoniae*. *Proceedings of the National Academy of Sciences*, *117*(44), 27620–27626.

Fontana, L., Bonura, E., Lyski, Z., & Messer, W. (2019). The brief case: *Klebsiella variicola*—identifying the misidentified. *Am Soc Microbiol*.

Fouts, D. E., Tyler, H. L., DeBoy, R. T., Daugherty, S., Ren, Q., Badger, J. H., ... others (2008). Complete genome sequence of the N2-fixing broad host range endophyte *Klebsiella pneumoniae* 342 and virulence predictions verified in mice. *PLoS genetics*, *4*(7), e1000141.

Francisco, A. P., Vaz, C., Monteiro, P. T., Melo-Cristino, J., Ramirez, M., & Carriço, J. A. (2012). PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics*, *13*(1), 1–10.

Fu, L., Huang, M., Zhang, X., Yang, X., Liu, Y., Zhang, L., ... Zhou, Y. (2018). Frequency of virulence factors in high biofilm formation blakpc-2 producing *Klebsiella pneumoniae* strains from hospitals. *Microbial pathogenesis*, *116*, 168–172.

Galani, I., Karaiskos, I., Karantani, I., Papoutsaki, V., Maraki, S., Papaioannou, V., ... others (2018). Epidemiology and resistance phenotypes of carbapenemase-producing *Klebsiella pneumoniae* in Greece, 2014 to 2016. *Eurosurveillance*, *23*(31), 1700775.

Garza-Ramos, U., Silva-Sánchez, J., Martínez-Romero, E., Tinoco, P., Pina-Gonzales, M., Barrios, H., ... Tellez-Sosa, J. (2015). Development of a multiplex-PCR probe system for the proper identification of *Klebsiella variicola*. *BMC Microbiology*, *15*(1), 1–14.

Giannattasio-Ferraz, S., Ene, A., Johnson, G., Maskeri, L., Oliveira, A. P., Banerjee, S., ... Putonti, C. (2022). Multidrug-resistant *Klebsiella variicola* isolated in the urine of healthy bovine heifers, a potential risk as an emerging human pathogen. *Applied and Environmental Microbiology*, *88*(9), e00044–22.

Gomez-Simmonds, A., & Uhlemann, A.-C. (2017). Clinical implications of genomic adaptation and evolution of carbapenem-resistant *Klebsiella pneumoniae*. *The Journal of infectious diseases*, *215*(suppl_1), S18–S27.

Gondal, A. J., Saleem, S., Jahan, S., Choudhry, N., & Yasmin, N. (2020). Novel carbapenem-resistant *Klebsiella pneumoniae* ST147 coharboring $bla_{\text{NDM-1}}$, $bla_{\text{OXA-48}}$ and extended-spectrum β-lactamases from Pakistan. *Infection and Drug Resistance*, 2105–2115.

Gorrie, C. L., Mirčeta, M., Wick, R. R., Judd, L. M., Lam, M., Gomi, R., ... others (2022). Genomic dissection of *Klebsiella pneumoniae* infections in hospital patients reveals insights into an opportunistic pathogen. *Nature Communications*, *13*(1), 1–17.

Granström, M., Wretlind, B., Markman, B., & Cryz, S. (1988). Enzyme-linked immunosorbent assay to evaluate the immunogenicity of a polyvalent *Klebsiella* capsular polysaccharide vaccine in humans. *Journal of clinical microbiology*, *26*(11), 2257–2261.

Green, E. R., & Mecsas, J. (2016). Bacterial secretion systems: an overview. *Virulence mechanisms of bacterial pathogens*, 213–239.

Gu, D., Dong, N., Zheng, Z., Lin, D., Huang, M., Wang, L., ... others (2018). A fatal outbreak of ST11 carbapenem-resistant hypervirulent *Klebsiella pneumoniae* in a Chinese hospital: a molecular epidemiological study. *The Lancet infectious diseases*, *18*(1), 37–46.

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

Han, R., Niu, M., Liu, S., Mao, J., Yu, Y., & Du, Y. (2022). The effect of siderophore virulence genes *entB* and *ybtS* on the virulence of carbapenem-resistant *Klebsiella pneumoniae*. *Microbial Pathogenesis*, *171*, 105746.

Hatrongjit, R., Wongsurawat, T., Jenjaroenpun, P., Chopjitt, P., Boueroy, P., Akeda, Y., ... Kerdsin, A. (2024). Genomic analysis of carbapenem-and colistin-resistant *Klebsiella pneumoniae* complex harbouring *mcr-8* and *mcr-9* from individuals in Thailand. *Scientific Reports*, *14*(1), 16836.

Haudiquet, M., Le Bris, J., Nucci, A., Bonnin, R. A., Domingo-Calap, P., Rocha, E. P., & Rendueles, O. (2024). Capsules and their traits shape phage susceptibility and plasmid conjugation efficiency. *Nature communications*, *15*(1), 2032.

He, Z., Xu, W., Zhao, H., Li, W., Dai, Y., Lu, H., ... Sun, B. (2022). Epidemiological characteristics an outbreak of st11 multidrug-resistant and hypervirulent *Klebsiella pneumoniae* in Anhui, China. *Frontiers in Microbiology*, *13*, 996753.

Heng, H., Yang, X., Ye, L., Tang, Y., Guo, Z., Li, J., ... Chen, S. (2024). Global genomic profiling of *Klebsiella pneumoniae*: a spatio-temporal population structure analysis. *International Journal of Antimicrobial Agents*, *63*(2), 107055.

Hernandez, R. E., Gallegos-Monterrosa, R., & Coulthurst, S. J. (2020). Type VI secretion system effector proteins: effective weapons for bacterial competitiveness. *Cellular microbiology*, *22*(9), e13241.

Hillenkamp, F., Karas, M., Beavis, R. C., & Chait, B. T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Analytical chemistry*, *63*(24), 1193A–1203A.

Holt, K. E., Wertheim, H., Zadoks, R. N., Baker, S., Whitehouse, C. A., Dance, D., ... others (2015). Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proceedings of the National Academy of Sciences*, *112*(27), E3574–E3581.

Hsieh, T., Ma, K., & Chao, A. (2016). iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, *7*(12), 1451–1456.

Hu, F., Pan, Y., Li, H., Han, R., Liu, X., Ma, R., ... others (2024). Carbapenem-resistant *Klebsiella pneumoniae* capsular types, antibiotic resistance and virulence factors in China: a longitudinal, multi-centre study. *Nature Microbiology*, *9*(3), 814–829.

Hu, Y., Tang, R., Jin, S., Guan, J., Meng, X., Dan, Z., ... Lu, J. (2024). Molecular characterization of ST15 carbapenem-resistant *Klebsiella pneumoniae* isolated in a single patient. *Journal of Global Antimicrobial Resistance*.

Huang, C.-C., & Huang, C.-Y. (2016). DnaT is a pric-binding protein. *Biochemical and Biophysical Research Communications*, *477*(4), 988-992. doi: 10.1016/j.bbrc.2016.07.018

Huang, L., Fu, L., Hu, X., Liang, X., Gong, G., Xie, C., ... Zhou, Y. (2021). Co-occurrence of *Klebsiella variicola* and *Klebsiella pneumoniae* both carrying bla_{KPC} from a respiratory Intensive Care Unit patient. *Infection and Drug Resistance*, 4503–4510.

Huang, Y.-H., Chou, S.-H., Liang, S.-W., Ni, C.-E., Lin, Y.-T., Huang, Y.-W., & Yang, T.-C. (2018). Emergence of an XDR and carbapenemase-producing hypervirulent *Klebsiella pneumoniae* strain in Taiwan. *Journal of Antimicrobial Chemotherapy*, *73*(8), 2039–2046.

Huang, Y.-J., Liao, H.-W., Wu, C.-C., & Peng, H.-L. (2009). MrkF is a component of type 3 fimbriae in *Klebsiella pneumoniae*. *Research in Microbiology*, *160*(1), 71–79.

Hutchings, M. I., Truman, A. W., & Wilkinson, B. (2019). Antibiotics: past, present and future. *Current opinion in microbiology*, *51*, 72–80.

Ichikawa, M., Nakamoto, N., Kredo-Russo, S., Weinstock, E., Weiner, I. N., Khabra, E., ... others (2023). Bacteriophage therapy against pathological *Klebsiella pneumoniae* ameliorates the course of primary sclerosing cholangitis. *Nature communications*, *14*(1), 3261.

Imai, K., Ishibashi, N., Kodana, M., Tarumoto, N., Sakai, J., Kawamura, T., ... others (2019). Clinical characteristics in blood stream infections caused by *Klebsiella pneumoniae*, *Klebsiella variicola*, and *Klebsiella quasipneumoniae*: a comparative study, japan, 2014–2017. *BMC infectious diseases*, *19*, 1–10.

Ito, R., Mustapha, M. M., Tomich, A. D., Callaghan, J. D., McElheny, C. L., Mettus, R. T., ... Doi, Y. (2017). Widespread fosfomycin resistance in gram-negative bacteria attributable to the chromosomal fosa gene. *MBio*, *8*(4), 10–1128.

Ivancic, T., Jamnik, P., & Stopar, D. (2013). Cold shock CspA and CspB protein production during periodic temperature cycling in *Escherichia coli*. *BMC research notes*, *6*, 1–9.

Iwasaki, Y., Inokuchi, R., Harada, S., Aoki, K., Ishii, Y., & Shinohara, K. (2017). Bacterial meningitis caused by hypervirulent *Klebsiella pneumoniae* capsular genotype K54 with development of granuloma-like nodal enhancement in the brain during the subacute phase. *Internal Medicine*, *56*(3), 373–376.

Jacobson, M. R., Brigle, K., Bennett, L., Setterquist, R., Wilson, M., Cash, V., ... Dean, D. (1989). Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. *Journal of bacteriology*, *171*(2), 1017–1027.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 1–8.

Janssen, A. B., De Bakker, V., Aprianto, R., Trebosc, V., Kemmer, C., Pieren, M., & Veening, J.-W. (2024). *Klebsiella pneumoniae* ompr facilitates lung infection through transcriptional regulation of key virulence factors. *Microbiology Spectrum*, *12*(1), e03966–23.

Jayol, A., Poirel, L., Brink, A., Villegas, M.-V., Yilmaz, M., & Nordmann, P. (2014). Resistance to colistin associated with a single amino acid change in protein PmrB among *Klebsiella pneumoniae* isolates of worldwide origin. *Antimicrobial Agents and Chemotherapy*, *58*(8), 4762–4766.

Jiménez-Castellanos, J.-C., Wan Ahmad Kamil, W. N. I., Cheung, C. H. P., Tobin, M. S., Brown, J., Isaac, S. G., ... Avison, M. B. (2016). Comparative effects of overproducing the AraC-type transcriptional regulators MarA, SoxS, RarA and RamA on antimicrobial drug susceptibility in *Klebsiella pneumoniae*. *Journal of Antimicrobial Chemotherapy*, *71*(7), 1820–1825.

Karaman, E., Çiçek, A. Ç., Şemen, V., & Şaban Beriş, F. (2024). Characterization of resistance genes and replicon typing in Carbapenem-resistant *Klebsiella pneumoniae* strains. *Annals of Clinical Microbiology and Antimicrobials*, *23*(1), 19.

Kareem, S. M., Al-Kadmy, I. M., Kazaal, S. S., Mohammed Ali, A. N., Aziz, S. N., Makharita, R. R., ... others (2021). Detection of *gyrA* and *parC* mutations and prevalence of plasmid-mediated quinolone resistance genes in *Klebsiella pneumoniae*. *Infection and Drug Resistance*, 555–563.

Karki, D., Dhungel, B., Bhandari, S., Kunwar, A., Joshi, P. R., Shrestha, B., ... Banjara, M. R. (2021). Antibiotic resistance and detection of plasmid mediated colistin resistance *mcr-1* gene among *Escherichia coli* and *Klebsiella pneumoniae* isolated from clinical samples. *Gut pathogens*, *13*, 1–16.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, *30*(4), 772–780.

Kaye, K. S., Marchaim, D., Thamlikitkul, V., Carmeli, Y., Chiu, C.-H., Daikos, G., ... others (2022). Colistin monotherapy versus combination therapy for carbapenem-resistant organisms. *NEJM evidence*, *2*(1), EVIDoa2200131.

Kim, H. B., Wang, M., Park, C. H., Kim, E.-C., Jacoby, G. A., & Hooper, D. C. (2009). *oqxAB* encoding a multidrug efflux pump in human clinical isolates of *Enterobacteriaceae*. *Antimicrobial agents and chemotherapy*, *53*(8), 3582–3584.

Kiratisin, P., Apisarnthanarak, A., Laesripa, C., & Saifon, P. (2008). Molecular characterization and epidemiology of extended-spectrum-β-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* isolates causing health care-associated infection in thailand, where the CTX-M family is endemic. *Antimicrobial agents and chemotherapy*, *52*(8), 2818–2824.

Kochan, T. J., Nozick, S. H., Valdes, A., Mitra, S. D., Cheung, B. H., Lebrun-Corbin, M., ... others (2023). *Klebsiella pneumoniae* clinical isolates with features of both multidrug-resistance and hypervirulence have unexpectedly low virulence. *Nature communications*, *14*(1), 7962.

Kolls, J., Norton, E., & Chen, K. (2022, May 26). *Multivalent vaccines derived from Klebsiella out membrane proteins.* Google Patents. (US Patent App. 17/601,051)

Köster, W. (2001). Abc transporter-mediated uptake of iron, siderophores, heme and vitamin B12. *Research in Microbiology*, *152*(3-4), 291–301.

Ku, Y.-H., Chuang, Y.-C., Chen, C.-C., Lee, M.-F., Yang, Y.-C., Tang, H.-J., & Yu, W.-L. (2017). *Klebsiella pneumoniae* isolates from meningitis: epidemiology, virulence and antibiotic resistance. *Scientific reports*, *7*(1), 6634.

Kumar, A., Chakravorty, S., Yang, T., Russo, T. A., Newton, S. M., & Klebba, P. E. (2024). Siderophore-mediated iron acquisition by *Klebsiella pneumoniae*. *Journal of Bacteriology*, *206*(5), e00024–24.

Lai, Y.-C., Lin, A.-C., Chiang, M.-K., Dai, Y.-H., Hsu, C.-C., Lu, M.-C., ... Chen, Y.-T. (2014). Genotoxic *Klebsiella pneumoniae* in Taiwan. *PloS one*, *9*(5), e96292.

Lam, M., Wick, R. R., Watts, S. C., Cerdeira, L. T., Wyres, K. L., & Holt, K. E. (2021). A genomic surveillance framework and genotyping tool for *Klebsiella pneumoniae* and its related species complex. *Nature Communications*, *12*(1), 1–16.

Lam, M. M., Wick, R. R., Judd, L. M., Holt, K. E., & Wyres, K. L. (2022). Kaptive 2.0: updated capsule and lipopolysaccharide locus typing for the *Klebsiella pneumoniae* species complex. *Microbial Genomics*, *8*(3).

Lam, M. M., Wyres, K. L., Judd, L. M., Wick, R. R., Jenney, A., Brisse, S., & Holt, K. E. (2018). Tracking key virulence loci encoding aerobactin and salmochelin siderophore synthesis in *Klebsiella pneumoniae*. *Genome medicine*, *10*, 1–15.

Lam, M. M., Wyres, K. L., Wick, R. R., Judd, L. M., Fostervold, A., Holt, K. E., & Löhr, I. H. (2019). Convergence of virulence and mdr in a single plasmid vector in MDR *Klebsiella pneumoniae* ST15. *Journal of antimicrobial chemotherapy*, *74*(5), 1218–1222.

Lawlor, M. S., O'connor, C., & Miller, V. L. (2007). Yersiniabactin is a virulence factor for *Klebsiella pneumoniae* during pulmonary infection. *Infection and Immunity*, *75*(3), 1463–1472.

Letunic, I., & Bork, P. (2007). Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics*, *23*(1), 127–128.

Li, J., Zhang, H., Ning, J., Sajid, A., Cheng, G., Yuan, Z., & Hao, H. (2019). The nature and epidemiology of OqxAB, a multidrug efflux pump. *Antimicrobial Resistance & Infection Control*, *8*, 1–13.

Li, Y., Hu, D., Ma, X., Li, D., Tian, D., Gong, Y., & Jiang, X. (2021). Convergence of carbapenem resistance and hypervirulence leads to high mortality in patients with postoperative *Klebsiella pneumoniae* meningitis. *Journal of Global Antimicrobial Resistance*, *27*, 95–100.

Li, Y., Kumar, S., Zhang, L., Wu, H., & Wu, H. (2023). Characteristics of antibiotic resistance mechanisms and genes of *Klebsiella pneumoniae*. *Open Medicine*, *18*(1), 20230707.

Ligowska-Marzęta, M., Barbier, E., Rodrigues, C., Piveteau, P., Schrøder Hansen, D., Hartmann, A., ... Brisse, S. (2024). Triplex real-time PCR ZKIR-T assay for simultaneous detection of the *Klebsiella pneumoniae* species complex and identification of *K. pneumoniae* sensu stricto. *Microbiology spectrum*, *12*(12), e00336–24.

Liu, C., Wu, Y., Fang, Y., Sang, Z., Huang, L., Dong, N., ... Chen, G. (2022). Emergence of an ST1326 (CG258) multi-drug resistant *Klebsiella pneumoniae* co-harboring *mcr-8.2*, ESBL genes, and the resistance-nodulation-division efflux pump gene cluster *tmexCD1-toprJ1* in China. *Frontiers in Microbiology*, *13*, 800993.

Liu, H., Xiang, Y., Xiong, M., Xiao, X., Zhou, J., Tian, H., ... Li, Y. (2024). Prevalence of ST1049-KL5 carbapenem-resistant *Klebsiella pneumoniae* with a $bla_{\text{KPC-2}}$ and $bla_{\text{NDM-1}}$ co-carrying hypertransmissible IncM1 plasmid. *Communications Biology*, *7*(1), 695.

Liu, W., Li, M., Cao, S., Ishaq, H. M., Zhao, H., Yang, F., & Liu, L. (2023). The biological and regulatory role of type VI secretion system of *Klebsiella pneumoniae*. *Infection and Drug Resistance*, 6911–6922.

Liu, X., Wu, Y., Zhu, Y., Jia, P., Li, X., Jia, X., ... others (2022). Emergence of colistin-resistant hypervirulent *Klebsiella pneumoniae* (CoR-HvKp) in China. *Emerging Microbes & Infections*, *11*(1), 648–661.

Liu, Y., Zhu, S., Wei, L., Feng, Y., Cai, L., Dunn, S., ... Zong, Z. (2022). Arm race among closely-related carbapenem-resistant *Klebsiella pneumoniae* clones. *ISME communications*, *2*(1), 76.

Long, S. W., Linson, S. E., Ojeda Saavedra, M., Cantu, C., Davis, J. J., Brettin, T., & Olsen, R. J. (2017). Whole-genome sequencing of human clinical *Klebsiella pneumoniae* isolates reveals misidentification and misunderstandings of *Klebsiella pneumoniae*, *Klebsiella variicola*, and *Klebsiella quasipneumoniae*. *mSphere*, *2*(4), e00290–17.

Lu, Y., Feng, Y., McNally, A., & Zong, Z. (2018). Occurrence of colistin-resistant hypervirulent *Klebsiella variicola*. *Journal of Antimicrobial Chemotherapy*, *73*(11), 3001–3004.

Luo, Y., Wang, Y., Ye, L., & Yang, J. (2014). Molecular epidemiology and virulence factors of pyogenic liver abscess causing *Klebsiella pneumoniae* in China. *Clinical Microbiology and Infection*, *20*(11), O818–O824.

Maatallah, M., Vading, M., Kabir, M. H., Bakhrouf, A., Kalin, M., Naucler, P., ... Giske, C. G. (2014). *Klebsiella variicola* is a frequent cause of bloodstream infection in the Stockholm area, and associated with higher mortality compared to *K. pneumoniae*. *PloS one*, *9*(11), e113539.

Mahillon, J., & Chandler, M. (1998). Insertion sequences. *Microbiology and molecular biology reviews*, *62*(3), 725–774.

Malchione, M. D., Torres, L. M., Hartley, D. M., Koch, M., & Goodman, J. L. (2019). Carbapenem and colistin resistance in enterobacteriaceae in southeast asia: review and mapping of emerging and overlapping challenges. *International journal of antimicrobial agents*, *54*(4), 381–399.

March, C., Cano, V., Moranta, D., Llobet, E., Pérez-Gutiérrez, C., Tomás, J. M., ... Bengoechea, J. A. (2013). Role of bacterial surface structures on the interaction of *Klebsiella pneumoniae* with phagocytes. *PloS one*, *8*(2), e56847.

Marsh, J. W., Mustapha, M. M., Griffith, M. P., Evans, D. R., Ezeonwuka, C., Pasculle, A. W., ... others (2019). Evolution of outbreak-causing carbapenem-resistant *Klebsiella pneumoniae* ST258 at a tertiary care hospital over 8 years. *MBio*, *10*(5), 10–1128.

Martin, M. J., Stribling, W., Ong, A. C., Maybank, R., Kwak, Y. I., Rosado-Mendez, J. A., ... others (2023). A panel of diverse *Klebsiella pneumoniae* clinical isolates for research and development. *Microbial genomics*, *9*(5), 000967.

Martin, R. M., & Bachman, M. A. (2018). Colonization, infection, and the accessory genome of *Klebsiella pneumoniae*. *Frontiers in cellular and infection microbiology*, *8*, 4.

Martínez-Romero, E., Rodríguez-Medina, N., Beltrán-Rojel, M., Silva-Sánchez, J., Barrios-Camacho, H., Pérez-Rueda, E., & Garza-Ramos, U. (2018). Genome misclassification of *Klebsiella variicola* and *Klebsiella quasipneumoniae* isolated from plants, animals and humans. *salud pública de méxico*, *60*(1), 56–62.

Mathers, A. J., Crook, D., Vaughan, A., Barry, K. E., Vegesana, K., Stoesser, N., ... others (2019). *Klebsiella quasipneumoniae* provides a window into carbapenemase gene transfer, plasmid rearrangements, and patient interactions with the hospital environment. *Antimicrobial agents and chemotherapy*, *63*(6), 10–1128.

Matsuda, N., Aung, M. S., Urushibara, N., Kawaguchiya, M., Ohashi, N., Taniguchi, K., ... Kobayashi, N. (2023). Prevalence, clonal diversity, and antimicrobial resistance of hypervirulent *Klebsiella pneumoniae* and *Klebsiella variicola* clinical isolates in northern Japan. *Journal of Global Antimicrobial Resistance*, *35*, 11–18.

Medeiros, G. S., Rigatto, M. H., Falci, D. R., & Zavascki, A. P. (2019). Combination therapy with polymyxin b for carbapenemase-producing *Klebsiella pneumoniae* bloodstream infection. *International journal of antimicrobial agents*, *53*(2), 152–157.

Merciecca, T., Bornes, S., Nakusi, L., Theil, S., Rendueles, O., Forestier, C., & Miquel, S. (2022). Role of *Klebsiella pneumoniae* Type VI secretion system (T6SS) in long-term gastrointestinal colonization. *Scientific Reports*, *12*(1), 16968.

Mills, R. O., Dadzie, I., Le-Viet, T., Baker, D. J., Addy, H. P., Akwetey, S. A., ... others (2024). Genomic diversity and antimicrobial resistance in clinical *Klebsiella pneumoniae* isolates from tertiary hospitals in Southern Ghana. *Journal of Antimicrobial Chemotherapy*, dkae123.

Mirshekar, M., Zadeh, R. G., Moghadam, M. T., Shahbazi, S., & Jazi, F. M. (2024). Upregulation of *pmrA, pmrB, pmrC, phoQ, phoP*, and *arnT* genes contributing to resistance to colistin in superbug *Klebsiella pneumoniae* isolates from human clinical samples in Tehran, Iran. *New Microbes and New Infections*, *59*, 101275.

Mohd Asri, N. A., Ahmad, S., Mohamud, R., Mohd Hanafi, N., Mohd Zaidi, N. F., Irekeola, A. A., ... others (2021). Global prevalence of nosocomial multidrug-resistant *Klebsiella pneumoniae*: a systematic review and meta-analysis. *Antibiotics*, *10*(12), 1508.

Monaco, M., Giani, T., Raffone, M., Arena, F., Garcia-Fernandez, A., Pollini, S., ... others (2014). Colistin resistance superimposed to endemic carbapenem-resistant *Klebsiella pneumoniae*: a rapidly evolving problem in Italy, November 2013 to April 2014. *Eurosurveillance*, *19*(42), 20939.

Monteiro, A. d. S. S., Cordeiro, S. M., & Reis, J. N. (2024). Virulence factors in *Klebsiella pneumoniae*: A literature review. *Indian Journal of Microbiology*, 1–13.

Monteith, A. J., & Skaar, E. P. (2021). The impact of metal availability on immune function during infection. *Trends in Endocrinology & Metabolism*, *32*(11), 916–928.

Morales-León, F., Opazo-Capurro, A., Caro, C., Lincopan, N., Cardenas-Arias, A., Esposito, F., ... others (2021). Hypervirulent and hypermucoviscous extended-spectrum β-lactamase-producing *Klebsiella pneumoniae* and *Klebsiella variicola* in *Chile*. *Virulence*, *12*(1), 35–44.

Mora-Ochomogo, M., & Lohans, C. T. (2021). β-lactam antibiotic targets and resistance mechanisms: from covalent inhibitors to substrates. *RSC Medicinal Chemistry*, *12*(10), 1623–1639.

Morgado, S., Fonseca, E., & Vicente, A. C. (2022). Genomics of *Klebsiella pneumoniae* species complex reveals the circulation of high-risk multidrug-resistant pandemic clones in human, animal, and environmental sources. *Microorganisms*, *10*(11), 2281.

Mshana, S. E., Hain, T., Domann, E., Lyamuya, E. F., Chakraborty, T., & Imirzalioglu, C. (2013). Predominance of *Klebsiella pneumoniae* ST14 carrying CTX-M-15 causing neonatal sepsis in Tanzania. *BMC Infectious Diseases*, *13*, 1–8.

Murphy, C. N., Mortensen, M. S., Krogfelt, K. A., & Clegg, S. (2013). 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. *Infection and immunity*, *81*(8), 3009–3017.

Nadasy, K. A., Domiati-Saad, R., & Tribble, M. A. (2007). Invasive *Klebsiella pneumoniae* syndrome in North America. *Clinical Infectious Diseases*, *45*(3), e25–e28.

Namikawa, H., Niki, M., Niki, M., Yamada, K., Nakaie, K., Sakiyama, A., ... others (2019). Clinical and virulence factors related to the 30-day mortality of *Klebsiella pneumoniae* bacteremia at a tertiary hospital: a case–control study. *European Journal of Clinical Microbiology & Infectious Diseases*, *38*, 2291–2297.

Nikaido, H., & Takatsuka, Y. (2009). Mechanisms of rnd multidrug efflux pumps. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, *1794*(5), 769–781.

Nucci, A., Rocha, E. P., & Rendueles, O. (2023). Latent evolution of biofilm formation depends on life-history and genetic background. *BioRxiv*, 2023–04.

Octavia, S., Kalisvar, M., Venkatachalam, I., Ng, O. T., Xu, W., Sridatta, P. S. R., ... others (2019). *Klebsiella pneumoniae* and *Klebsiella quasipneumoniae* define the population structure of $bla_{\text{KPC-2}}$ *Klebsiella*: a 5 year retrospective genomic study in Singapore. *Journal of Antimicrobial Chemotherapy*, *74*(11), 3205–3210.

Osei Sekyere, J., & Reta, M. A. (2021). Global evolutionary epidemiology and resistome dynamics of citrobacter species, *Enterobacter hormaechei*, *Klebsiella variicola*, and Proteeae clones. *Environmental Microbiology*.

Padilla, E., Llobet, E., Doménech-Sánchez, A., Martínez-Martínez, L., Bengoechea, J. A., & Albertí, S. (2010). *Klebsiella pneumoniae* AcrAB efflux pump contributes to antimicrobial resistance and virulence. *Antimicrobial Agents and Chemotherapy*, *54*(1), 177–183.

Pagaling, E., Hough, R., Avery, L., Robinson, L., Freitag, T., Coull, M., ... others (2023). Antibiotic resistance patterns in soils across the Scottish landscape. *Communications Earth & Environment*, *4*(1), 403.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., ... Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693.

Pagel, M., & Meade, A. (2013). BayesTraits v. 2.0. *Reading: University of Reading*.

Pajand, O., Rahimi, H., Badmasti, F., Gholami, F., Alipour, T., Darabi, N., ... Leekitcharoenphon, P. (2023). Various arrangements of mobile genetic elements among CC147 subpopulations of *Klebsiella pneumoniae* harboring $bla_{\text{NDM-1}}$: a comparative genomic analysis of carbapenem resistant strains. *Journal of Biomedical Science*, *30*(1), 73.

Pal, S., Verma, J., Mallick, S., Rastogi, S. K., Kumar, A., & Ghosh, A. S. (2019). Absence of the glycosyltransferase WcaJ in *Klebsiella pneumoniae* ATCC13883 affects biofilm formation, increases polymyxin resistance and reduces murine macrophage activation. *Microbiology*, *165*(8), 891–904.

Palacios, M., Miner, T. A., Frederick, D. R., Sepulveda, V. E., Quinn, J. D., Walker, K. A., & Miller, V. L. (2018). Identification of two regulators of virulence that are conserved in *Klebsiella pneumoniae* classical and hypervirulent strains. *MBio*, *9*(4), 10–1128.

Pankok, F., Taudien, S., Dekker, D., Thye, T., Oppong, K., Wiafe Akenten, C., ... others (2022). Epidemiology of plasmids in *Escherichia coli* and *Klebsiella pneumoniae* with acquired extended spectrum beta-lactamase genes isolated from chronic wounds in Ghana. *Antibiotics*, *11*(5), 689.

Pedersen, T., Tellevik, M. G., Kommedal, Ø., Lindemann, P. C., Moyo, S. J., Janice, J., ... Langeland, N. (2020). Horizontal plasmid transfer among *Klebsiella pneumoniae* isolates is the key factor for dissemination of extended-spectrum β-lactamases among children in Tanzania. *Msphere*, *5*(4), 10–1128.

Pei, N., Sun, W., He, J., Li, Y., Chen, X., Liang, T., ... Li, J. (2022). Genome-wide association study of *Klebsiella pneumoniae* identifies variations linked to carbapenems resistance. *Frontiers in Microbiology*, *13*, 997769.

Petrosillo, N., Taglietti, F., & Granata, G. (2019). Treatment options for colistin resistant *Klebsiella pneumoniae*: present and future. *Journal of clinical medicine*, *8*(7), 934.

Pinto-Tomás, A. A., Anderson, M. A., Suen, G., Stevenson, D. M., Chu, F. S., Cleland, W. W., ... Currie, C. R. (2009). Symbiotic nitrogen fixation in the fungus gardens of leaf-cutter ants. *Science*, *326*(5956), 1120–1123.

Piperaki, E.-T., Syrogiannopoulos, G. A., Tzouvelekis, L. S., & Daikos, G. L. (2017). *Klebsiella pneumoniae*: virulence, biofilm and antimicrobial resistance. *The Pediatric infectious disease journal*, *36*(10), 1002–1005.

Pons, M. J., Marí-Almirall, M., Ymaña, B., Moya-Salazar, J., Muñoz, L., Sauñe, S., ... Roca, I. (2020). Spread of ST348 *Klebsiella pneumoniae* producing NDM-1 in a Peruvian hospital. *Microorganisms*, *8*(9), 1392.

Potter, R. F., Lainhart, W., Twentyman, J., Wallace, M. A., Wang, B., Burnham, C.-A. D., ... Dantas, G. (2018). Population structure, antibiotic resistance, and uropathogenicity of *Klebsiella variicola*. *MBio*, *9*(6), e02481–18.

Pranavathiyani, G., Prava, J., Rajeev, A. C., & Pan, A. (2020). Novel target exploration from hypothetical proteins of *Klebsiella pneumoniae* mgh 78578 reveals a protein involved in host-pathogen interaction. *Frontiers in Cellular and Infection Microbiology*, *10*, 109. doi: 10.3389/fcimb.2020.00109

Pray, L. A. (2008). Transposons: The jumping genes. *Nature education*, *1*(1), 204.

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS One*, *5*(3), e9490.

python visualization. (n.d.). *Folium.* Retrieved from `https://python-visualization.github.io/folium/`

Qi, Y., Wei, Z., Ji, S., Du, X., Shen, P., & Yu, Y. (2011). ST11, the dominant clone of KPC-producing *Klebsiella pneumoniae* in China. *Journal of Antimicrobial Chemotherapy*, *66*(2), 307–312.

Qin, X., Wu, S., Hao, M., Zhu, J., Ding, B., Yang, Y., ... Hu, F. (2020). The colonization of carbapenem-resistant *Klebsiella pneumoniae*: epidemiology, resistance mechanisms, and risk factors in patients admitted to intensive care units in China. *The Journal of infectious diseases*, *221*(Supplement_2), S206–S214.

Rambaut, A. (2023). FigTree. *GitHub repository*. `http://tree.bio.ed.ac.uk/software/figtree/`.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology*, *67*(5), 901–904.

Ran, L., Wang, X., He, X., Guo, R., Wu, Y., Zhang, P., & Zhang, X.-H. (2023). Genomic analysis and chitinase characterization of *Vibrio harveyi* WXL538: insight into its adaptation to the marine environment. *Frontiers in Microbiology*, *14*.

Remya, P., Shanthi, M., & Sekar, U. (2018). Occurrence and characterization of hyperviscous K1 and K2 serotype in *Klebsiella pneumoniae*. *Journal of laboratory physicians*, *10*(03), 283–288.

Remya, P., Shanthi, M., & Sekar, U. (2019). Characterisation of virulence genes associated with pathogenicity in *Klebsiella pneumoniae*. *Indian journal of medical microbiology*, *37*(2), 210–218.

Rendueles, O. (2020). Deciphering the role of the capsule of *Klebsiella pneumoniae* during pathogenesis: A cautionary tale. *Molecular microbiology*, *113*(5), 883–888.

Rezaei, S., Tajbakhsh, S., Naeimi, B., & Yousefi, F. (2024). Investigation of *gyrA* and *parC* mutations and the prevalence of plasmid-mediated quinolone resistance genes in *Klebsiella pneumoniae* clinical isolates. *BMC Microbiology*, *24*(1), 265.

Rice, P. M., Rice, P. M., Bleasby, A. J., & Ison, J. C. (2011). *EMBOSS user's guide: Practical bioinformatics with EMBOSS*. Cambridge University Press.

Riwu, K. H. P., Effendi, M. H., Rantam, F. A., Khairullah, A. R., & Widodo, A. (2022). A review: virulence factors of *Klebsiella pneumoniaw* as emerging infection on the food chain. *Veterinary world*, *15*(9), 2172.

Rocha, J., Henriques, I., Gomila, M., & Manaia, C. M. (2022). Common and distinctive genomic features of *Klebsiella pneumoniae* thriving in the natural environment or in clinical settings. *Scientific reports*, *12*(1), 10441.

Rodrigues, C., Passet, V., Rakotondrasoa, A., & Brisse, S. (2018). Identification of *Klebsiella pneumoniae*, *Klebsiella quasipneumoniae*, *Klebsiella variicola* and related phylogroups by MALDI-TOF mass spectrometry. *Frontiers in Microbiology*, *9*, 3000.

Rodríguez-Medina, N., Barrios-Camacho, H., Duran-Bedolla, J., & Garza-Ramos, U. (2019). *Klebsiella variicola*: an emerging pathogen in humans. *Emerging microbes & infections*, *8*(1), 973–988.

Rodríguez-Medina, N., Rodríguez-Santiago, J., Alvarado-Delgado, A., Sagal-Prado, A., Silva-Sánchez, J., De la Cruz, M. A., ... others (2024). Comprehensive study reveals phenotypic heterogeneity in *Klebsiella pneumoniae* species complex isolates. *Scientific Reports*, *14*(1), 5876.

Rojas, L. J., Salim, M., Cober, E., Richter, S. S., Perez, F., Salata, R. A., ... others (2017). Colistin resistance in carbapenem-resistant *Klebsiella pneumoniae*: laboratory detection and impact on mortality. *Clinical Infectious Diseases*, *64*(6), 711–718.

Rolbiecki, D., Paukszto, Ł., Krawczyk, K., Korzeniewska, E., Sawicki, J., & Harnisz, M. (2024). Genomic and metagenomic analysis reveals shared resistance genes and mobile genetic elements in *E. coli* and *Klebsiella spp.* isolated from hospital patients and hospital wastewater at intra-and inter-genus level. *International Journal of Hygiene and Environmental Health*, *261*, 114423.

Rosenblueth, M., Martínez, L., Silva, J., & Martínez-Romero, E. (2004). *Klebsiella variicola*, a novel species with clinical and plant-associated isolates. *Systematic and applied microbiology*, *27*(1), 27–35.

Rosenblum, R., Khan, E., Gonzalez, G., Hasan, R., & Schneiders, T. (2011). Genetic regulation of the *ramA* locus and its expression in clinical isolates of *Klebsiella pneumoniae*. *International Journal of Antimicrobial Agents*, *38*(1), 39–45.

Russo, T. A., & Marr, C. M. (2019). Hypervirulent *Klebsiella pneumoniae*. *Clinical microbiology reviews*, *32*(3), 10–1128.

Russo, T. A., Olson, R., Fang, C.-T., Stoesser, N., Miller, M., MacDonald, U., ... Johnson, J. R. (2018). Identification of biomarkers for differentiation of hypervirulent *Klebsiella pneumoniae* from classical *K. pneumoniae*. *Journal of clinical microbiology*, *56*(9), 10–1128.

Ryan, K. J., Ray, C. G., & Sherris, J. C. (2010). Sherris medical microbiology. *(No Title)*.

Sakai, J., & Maesaki, S. (2022). Prevalence, molecular characterization, and drug susceptibility of extended-spectrum β-lactamase-producing *Klebsiella pneumoniae, K. quasipneumoniae*, and *K. variicola* in japan. *Microbial Pathogenesis*, *169*, 105663.

Sampaio, S. C. F., Bigelli Carvalho, R. d. L., Mimica, M. J., da Silva, C. B., Mimica, L. M. J., de Lima, A. V., ... Sampaio, J. L. M. (2021). Genotyping of paired KPC-producing *Klebsiella pneumoniae* isolates with and without divergent polymyxin B susceptibility profiles. *Brazilian Journal of Microbiology*, *52*(4), 1981–1989.

Sands, K., Carvalho, M. J., Portal, E., Thomson, K., Dyer, C., Akpulu, C., ... others (2021). Characterization of antimicrobial-resistant Gram-negative bacteria that cause neonatal sepsis in seven low-and middle-income countries. *Nature microbiology*, *6*(4), 512–523.

Sauer, M. M., Jakob, R. P., Eras, J., Baday, S., Eriş, D., Navarra, G., ... Glockshuber, R. (2016). Catch-bond mechanism of the bacterial adhesin FimH. *Nature communications*, *7*(1), 10738.

Schneiders, T., Amyes, S., & Levy, S. (2003). Role of AcrR and RamA in fluoroquinolone resistance in clinical *Klebsiella pneumoniae* isolates from Singapore. *Antimicrobial Agents and Chemotherapy*, *47*(9), 2831–2837.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069.

Seemann, T. (2018). snp-dists. *GitHub repository*. `https://github.com/tseemann/snp-dists`.

Seemann, T. (2019). ABRicate. *GitHub repository*. `https://github.com/tseemann/abricate`.

Segerman, B. (2012). The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Frontiers in cellular and infection microbiology*, *2*, 116.

Semet, C., Efe, K., Akalın, H., İşçimen, R., Girgin, N. K., Özakın, C., ... Kahveci, F. (2024). Outcome of carbapenem or colistin resistant *Klebsiella pneumoniae* bacteremia in the intensive care unit. *Scientific Reports*, *14*(1), 25805.

Sheppard, S. K., Didelot, X., Meric, G., Torralbo, A., Jolley, K. A., Kelly, D. J., ... Falush, D. (2013). Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. *Proceedings of the National Academy of Sciences*, *110*(29), 11923–11927.

Shu, H.-Y., Fung, C.-P., Liu, Y.-M., Wu, K.-M., Chen, Y.-T., Li, L.-H., ... Tsai, S.-F. (2009). Genetic diversity of capsular polysaccharide biosynthesis in *Klebsiella pneumoniae* clinical isolates. *Microbiology*, *155*(12), 4170–4183.

Singhai, M., Malik, A., Shahid, M., Malik, M. A., & Goyal, R. (2012). A study on device-related infections with special reference to biofilm production and antibiotic resistance. *Journal of global infectious diseases*, *4*(4), 193–198.

Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P., & de la Cruz, F. (2010). Mobility of plasmids. *Microbiology and Molecular Biology Reviews*, *74*(3), 434–452.

Solomkin, J. S., Mazuski, J. E., Bradley, J. S., Rodvold, K. A., Goldstein, E. J., Baron, E. J., ... others (2010). Diagnosis and management of complicated intra-abdominal infection in adults and children: guidelines by the Surgical Infection Society and the Infectious Diseases Society of America. *Surgical infections*, *11*(1), 79–109.

Stahlhut, S. G., Chattopadhyay, S., Kisiela, D. I., Hvidtfeldt, K., Clegg, S., Struve, C., ... Krogfelt, K. A. (2013). Structural and population characterization of MrkD, the adhesive subunit of type 3 fimbriae. *Journal of bacteriology*, *195*(24), 5602–5613.

Stanton, T. D., Hetland, M. A., Löhr, I. H., Holt, K. E., & Wyres, K. L. (2025). Fast and accurate in silico antigen typing with Kaptive 3. *Microbial Genomics*, *11*(6), 001428.

Stoesser, N., Giess, A., Batty, E., Sheppard, A., Walker, A., Wilson, D., ... others (2014). 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. *Antimicrobial agents and chemotherapy*, *58*(12), 7347–7357.

Stojowska-Swędrzyńska, K., Łupkowska, A., Kuczyńska-Wiśnik, D., & Laskowska, E. (2021). Antibiotic heteroresistance in klebsiella pneumoniae. *International journal of molecular sciences*, *23*(1), 449.

Stojowska-Swędrzyńska, K., et al. (2022). Antibiotic heteroresistance in *Klebsiella pneumoniae*. *International Journal of Molecular Sciences*, *23*(1). Retrieved from `https://www.mdpi.com/1422-0067/23/1/449` doi: 10.3390/ijms23010449

Storey, D., McNally, A., Åstrand, M., sa Pessoa Graca Santos, J., Rodriguez-Escudero, I., Elmore, B., ... others (2020). *Klebsiella pneumoniae* type VI secretion system-mediated microbial competition is PhoPQ controlled and reactive oxygen species dependent. *PLoS pathogens*, *16*(3), e1007969.

Struve, C., Bojer, M., & Krogfelt, K. A. (2008). Characterization of *Klebsiella pneumoniae* type 1 fimbriae by detection of phase variation during colonization and infection and impact on virulence. *Infection and immunity*, *76*(9), 4055–4065.

Struve, C., Bojer, M., & Krogfelt, K. A. (2009). Identification of a conserved chromosomal region encoding klebsiella pneumoniae type 1 and type 3 fimbriae and assessment of the role of fimbriae in pathogenicity. *Infection and immunity*, *77*(11), 5016–5024.

Struve, C., & Krogfelt, K. A. (2003). Role of capsule in *Klebsiella pneumoniae* virulence: lack of correlation between in vitro and in vivo studies. *FEMS microbiology letters*, *218*(1), 149–154.

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1), vey016.

Surveillance atlas of infectious diseases. (2021). *European Centre for Disease Prevention and Control Stockholm (Sweden)*.

Takei, S., Lu, Y. J., Tohya, M., Watanabe, S., Misawa, S., Tabe, Y., ... others (2022). Spread of Carbapenem-resistant *Klebsiella pneumoniae* clinical isolates producing NDM-type metallo-β-lactamase in Myanmar. *Microbiology Spectrum*, *10*(4), e00673–22.

Tan, T. Y., Ong, M., Cheng, Y., & Ng, L. S. Y. (2019). Hypermucoviscosity, *rmpA*, and aerobactin are associated with community-acquired *Klebsiella pneumoniae* bacteremic isolates causing liver abscess in *Singapore*. *Journal of Microbiology, Immunology and Infection*, *52*(1), 30–34.

Tan, Y. H., Chen, Y., Chu, W. H., Sham, L.-T., & Gan, Y.-H. (2020). Cell envelope defects of different capsule-null mutants in K1 hypervirulent *Klebsiella pneumoniae* can affect bacterial pathogenesis. *Molecular microbiology*, *113*(5), 889–905.

Team, R. C., et al. (2013). R: A language and environment for statistical computing. *Foundation for Statistical Computing, Vienna, Austria*.

Thorpe, H. A., Booton, R., Kallonen, T., Gibbon, M. J., Couto, N., Passet, V., ... others (2022). A large-scale genomic snapshot of *Klebsiella* spp. isolates in Northern Italy reveals limited transmission between clinical and non-clinical settings. *Nature Microbiology*, *7*(12), 2054–2067.

Tian, G.-B., Doi, Y., Shen, J., Walsh, T. R., Wang, Y., Zhang, R., & Huang, X. (2017). MCR-1-producing *Klebsiella pneumoniae* outbreak in China. *The Lancet. Infectious diseases*, *17*(6), 577.

Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D., & Corander, J. (2018). RhierBAPS: an R implementation of the population clustering algorithm hierBAPS. *Wellcome Open Research*, *3*.

Törönen, P., Medlar, A., & Holm, L. (2018). PANNZER2: a rapid functional annotation web server. *Nucleic acids research*, *46*(W1), W84–W88.

Treangen, T. J., Ondov, B. D., Koren, S., & Phillippy, A. M. (2014). The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome biology*, *15*, 1–15.

Tsai, Y.-K., Fung, C.-P., Lin, J.-C., Chen, J.-H., Chang, F.-Y., Chen, T.-L., & Siu, L. K. (2011). *Klebsiella pneumoniae* outer membrane porins OmpK35 and OmpK36 play roles in both antimicrobial resistance and virulence. *Antimicrobial Agents and Chemotherapy*, *55*(4), 1485–1493.

Uzairue, L. I., Rabaan, A. A., Adewumi, F. A., Okolie, O. J., Folorunso, J. B., Bakhrebah, M. A., ... others (2022). Global prevalence of colistin resistance in *Klebsiella pneumoniae* from bloodstream infection: a systematic review and meta-analysis. *Pathogens*, *11*(10), 1092.

Veleba, M., Higgins, P. G., Gonzalez, G., Seifert, H., & Schneiders, T. (2012). Characterization of RarA, a novel AraC family multidrug resistance regulator in *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy*, *56*(8), 4450–4458.

Veleba, M., & Schneiders, T. (2012). Tigecycline resistance can occur independently of the *ramA* gene in *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy*, *56*(8), 4466–4467.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... others (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.

Vorachek-Warren, M. K., Carty, S. M., Lin, S., Cotter, R. J., & Raetz, C. R. (2002). An *Escherichia coli* mutant lacking the cold shock-induced palmitoleoyltransferase of lipid A biosynthesis: absence of unsaturated acyl chains and antibiotic hypersensitivity at 12 C. *Journal of Biological Chemistry*, *277*(16), 14186–14193.

Walker, K. A., & Miller, V. L. (2020). The intersection of capsule gene expression, hypermucoviscosity and hypervirulence in *Klebsiella pneumoniae*. *Current Opinion in Microbiology*, *54*, 95–102.

Walker, K. A., Treat, L. P., Sepúlveda, V. E., & Miller, V. L. (2020). The small protein RmpD drives hypermucoviscosity in *Klebsiella pneumoniae*. *MBio*, *11*(5), 10–1128.

Wang, J., Coipan, C., Franz, E., Kuijper, E., Wessels, E., Gally, D., ... Schneiders, T. (2024). Genome analyses of a worldwide multisource collection of klebsiella variicola reveal adaptation relevant for human infection. *bioRxiv*. Retrieved from `https://www.biorxiv.org/content/early/2024/10/28/2024.10.28.617863` doi: 10.1101/2024.10.28.617863

Wang, J., Feng, Y., & Zong, Z. (2023). The origins of st11 kl64 *Klebsiella pneumoniae*: a genome-based study. *Microbiology Spectrum*, *11*(2), e04165–22.

Wang, J., & Wang, J. (2012). Classical multidimensional scaling. *Geometric structure of high-dimensional data and dimensionality reduction*, 115–129.

Wang, R., Van Dorp, L., Shaw, L. P., Bradley, P., Wang, Q., Wang, X., ... others (2018). The global distribution and spread of the mobilized colistin resistance gene *mcr-1*. *Nature communications*, *9*(1), 1179.

Wang, R., Zhang, A., Sun, S., Yin, G., Wu, X., Ding, Q., ... others (2024). Increase in antioxidant capacity associated with the successful subclone of hypervirulent carbapenem-resistant *Klebsiella pneumoniae* ST11-KL64. *Nature Communications*, *15*(1), 67.

Wang, X., Zhao, J., Ji, F., Chang, H., Qin, J., Zhang, C., ... others (2021). Multiple-replicon resistance plasmids of *Klebsiella* mediate extensive dissemination of antimicrobial genes. *Frontiers in Microbiology*, *12*, 754931.

Wickham, H., Hester, J., & Ooms, J. (2018). xml2: Parse XML. *R package version*, *1*(0).

Wong, J. L., Romano, M., Kerry, L. E., Kwong, H.-S., Low, W.-W., Brett, S. J., ... Frankel, G. (2019). OmpK36-mediated Carbapenem resistance attenuates ST258 *Klebsiella pneumoniae in vivo*. *Nature communications*, *10*(1), 3957.

Wu, C.-C., Huang, Y.-J., Fung, C.-P., & Peng, H.-L. (2010). Regulation of the *Klebsiella pneumoniae* Kpc fimbriae by the site-specific recombinase KpcI. *Microbiology*, *156*(7), 1983–1992.

Wyres, K. L., Lam, M., & Holt, K. E. (2020). Population genomics of *Klebsiella pneumoniae*. *Nature Reviews Microbiology*, *18*(6), 344–359.

Wyres, K. L., Wick, R. R., Judd, L. M., Froumine, R., Tokolyi, A., Gorrie, C. L., ... Holt, K. E. (2019). Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. *PLoS Genetics*, *15*(4), e1008114.

Xanthopoulou, K., Carattoli, A., Wille, J., Biehl, L. M., Rohde, H., Farowski, F., ... others (2020). Antibiotic resistance and mobile genetic elements in extensively drug-resistant *Klebsiella pneumoniae* sequence type 147 recovered from Germany. *Antibiotics*, *9*(10), 675.

Xie, M., Yang, X., Xu, Q., Ye, L., Chen, K., Zheng, Z., ... others (2021). Clinical evolution of ST11 carbapenem resistant and hypervirulent *Klebsiella pneumoniae*. *Communications biology*, *4*(1), 650.

Xu, L., Wang, M., Yuan, J., Wang, H., Li, M., Zhang, F., ... Li, B. (2021). The KbvR regulator contributes to capsule production, outer membrane protein biosynthesis, antiphagocytosis, and virulence in *Klebsiella pneumoniae*. *Infection and Immunity*, *89*(5).

Xu, Q., Sheng, Z., Hao, M., Jiang, J., Ye, M., Chen, Y., ... Wang, M. (2021). Rama upregulates multidrug resistance efflux pumps acrab and oqxab in *Klebsiella pneumoniae*. *International journal of antimicrobial agents*, *57*(2), 106251.

Yan, Q., Zhou, M., Zou, M., & Liu, W.-e. (2016). Hypervirulent *Klebsiella pneumoniae* induced ventilator-associated pneumonia in mechanically ventilated patients in China. *European journal of clinical microbiology & infectious diseases*, *35*, 387–396.

Yang, J., Li, Y., Tang, N., Li, J., Zhou, J., Lu, S., ... others (2022). The human gut serves as a reservoir of hypervirulent *Klebsiella pneumoniae*. *Gut Microbes*, *14*(1), 2114739.

Yang, Q., Jia, X., Zhou, M., Zhang, H., Yang, W., Kudinha, T., & Xu, Y. (2020). Emergence of ST11-K47 and ST11-K64 hypervirulent carbapenem-resistant *Klebsiella pneumoniae* in bacterial liver abscesses from China: a molecular, biological, and epidemiological study. *Emerging microbes & infections*, *9*(1), 320–331.

Yang, X., Liu, X., Chan, E., Zhang, R., Chen, S., & Zhou, X. (2023). *Functional characterization of plasmid-borne rmpADC homologues in Klebsiella pneumoniae. microbiol spectr 11: e0308122.*

Yao, B., Xiao, X., Wang, F., Zhou, L., Zhang, X., & Zhang, J. (2015). Clinical and molecular characteristics of multi-clone carbapenem-resistant hypervirulent (hypermucoviscous) *Klebsiella pneumoniae* isolates in a tertiary hospital in Beijing, China. *International journal of infectious diseases*, *37*, 107–112.

Yoshida, T., Qin, L., Egger, L. A., & Inouye, M. (2006). Transcription regulation of *ompF* and *ompC* by a single transcription factor, ompr. *Journal of Biological Chemistry*, *281*(25), 17114–17123.

Yu, K.-W., Xue, P., Fu, Y., & Yang, L. (2021). T6SS mediated stress responses for bacterial environmental survival and host adaptation. *International Journal of Molecular Sciences*, *22*(2), 478.

Zahid, N., Zulfiqar, S., & Shakoori, A. R. (2012). Functional analysis of cus operon promoter of *Klebsiella pneumoniae* using *E. coli lacZ* assay. *Gene*, *495*(1), 81–88.

Zeng, L., Zhang, J., Li, C., Fu, Y., Zhao, Y., Wang, Y., ... Zhang, X. (2020). The determination of *gyrA* and *parC* mutations and the prevalence of plasmid-mediated quinolone resistance genes in carbapenem resistant *Klebsiella pneumonia* ST11 and ST76 strains isolated from patients in Heilongjiang province, China. *Infection, Genetics and Evolution*, *82*, 104319.

Zhang, J., Li, W., Huang, X., Li, D., Liu, X., Jiang, X., ... Li, X. (2024). Genomic and transcriptomic characteristics of type VI secretion system in *Klebsiella pneumoniae*. *bioRxiv*, 2024–01.

Zhang, J., Liang, S., Wang, X., Lu, Z., Sun, P., Zhang, H., & Sun, F. (2019). Biodegradation of atrazine by the novel *Klebsiella variicola* strain FH-1. *BioMed Research International*, *2019*.

Zhang, J., Xu, Y., Wang, M., Li, X., Liu, Z., Kuang, D., ... Qu, J. (2023). Mobilizable plasmids drive the spread of antimicrobial resistance genes and virulence genes in *Klebsiella pneumoniae*. *Genome Medicine*, *15*(1), 106.

Zhang, J., Zhou, K., Zheng, B., Zhao, L., Shen, P., Ji, J., ... Xiao, Y. (2016). High prevalence of ESBL-producing *Klebsiella pneumoniae* causing community-onset infections in China. *Frontiers in microbiology*, *7*, 1830.

Zhang, Y., Zeng, J., Liu, W., Zhao, F., Hu, Z., Zhao, C., ... others (2015). Emergence of a hypervirulent carbapenem-resistant *Klebsiella pneumoniae* isolate from clinical infections in China. *Journal of Infection*, *71*(5), 553–560.

Zhao, H., He, Z., Li, Y., & Sun, B. (2022). Epidemiology of carbapenem-resistant *Klebsiella pneumoniae* ST15 of producing KPC-2, SHV-106 and CTX-M-15 in Anhui, China. *BMC microbiology*, *22*(1), 262.

Zheng, B., Xu, H., Lv, T., Guo, L., Xiao, Y., Huang, C., ... others (2020). Stool samples of acute diarrhea inpatients as a reservoir of ST11 hypervirulent KPC-2-producing *Klebsiella pneumoniae*. *MSystems*, *5*(3), 10–1128.

Zhou, K., Xiao, T., David, S., Wang, Q., Zhou, Y., Guo, L., ... others (2020). Novel subclone of carbapenem-resistant *Klebsiella pneumoniae* sequence type 11 with enhanced virulence and transmissibility, China. *Emerging infectious diseases*, *26*(2), 289.

Zhou, K., Xue, C.-X., Xu, T., Shen, P., Wei, S., Wyres, K. L., ... others (2023). A point mutation in *recC* associated with subclonal replacement of carbapenem-resistant *Klebsiella pneumoniae* ST11 in china. *Nature Communications*, *14*(1), 2464.

Zhou, M., Lan, Y., Wang, S., Liu, Q., Jian, Z., Li, Y., ... Liu, W. (2020). Epidemiology and molecular characteristics of the type VI secretion system in *Klebsiella pneumoniae* isolated from bloodstream infections. *Journal of Clinical Laboratory Analysis*, *34*(11), e23459.

Zhu, J., Wang, T., Chen, L., & Du, H. (2021). Virulence factors in hypervirulent *Klebsiella pneumoniae*. *Frontiers in Microbiology*, *12*, 642484.

Zhu, Z., Huang, H., Xu, Y., Wang, M., Lv, J., Xu, L., ... others (2021). Emergence and genomics of OXA-232-producing *Klebsiella pneumoniae* in a hospital in Yancheng, China. *Journal of Global Antimicrobial Resistance*, *26*, 194–198.