Identification of Clinical and Bacterial Genetic Risk Factors Associated with
***Klebsiella pneumoniae* Infection**

by

Rebekah Michal Martin

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in the University of Michigan
2017

Doctoral Committee:

     Assistant Professor Michael A. Bachman, Chair
     Professor Nicholas W. Lukacs
     Professor Harry L.T. Mobley
     Professor Gabriel Núñez
     Assistant Professor Evan S. Snitkin

Rebekah Michal Martin

rmmar@umich.edu

ORCID iD: 0000-0001-9729-0283

Rebekah Michal Martin

rmmar@umich.edu

ORCID iD: 0000-0001-9729-0283

DEDICATION

To my grandfathers:

George Bernard Skidmore,

who gifted me my first microscope

&

Robert James Martin,

who believed I would achieve great things

ACKNOWLEDGEMENTS

owed to Robert Tibbetts & Linoj Samuel who continue to answer my many questions and to help me navigate the field of clinical microbiology.

I am also indebted to the many other individuals who mentored me in various ways throughout graduate school: my committee, Gary Huffnagle, Nick Lukacs, Harry Mobley, Gabriel Núñez, and Evan Snitkin who provided continual advice and assistance. I am truly honored to have each of them on my committee and impacting my project. Maria Sandkvist and Lynne Shetron-Rama for their guidance and advice on how to be an effective university instructor. Duane Newton for allowing me to take up space in the University of Michigan Clinical Microbiology Lab, and for taking the time to provide career advice and share his journey.

I have been fortunate to have a fantastic group of friends, both from graduate school and from other stages of my life: the MCP and PIBS students, who provided much needed stress relief through trivia nights, movie and game nights, and sharing in day-to-day struggles and adventures. My best friend Marie Tyrrell, who has been my sounding board and encouragement from afar.

My family have supported my passion for science since before I can remember, and have been a constant source of encouragement through each milestone and each setback. My parents have reminded me of everything I am capable of and that they support me whatever comes. My sister, Elisa, was my first hero & role model and continues to be a source of support even from across the ocean.

Without my husband and partner, Ben Dettmar, this journey undoubtedly would have never happened. I still remember the moment he told me I was capable of pursuing a graduate degree, and from that moment I never looked back. He has always encouraged me to fulfill my potential, and he has taught me to slow down and embrace life along the way. I am forever grateful for the

support, encouragement, & advice he still provides daily. I cannot imagine this life adventure with

anyone else x.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

ANOVA   Analysis of variance

AUROC   Area under the receiver operator characteristic curve

Aer   Aerobactin

AST   Antimicrobial susceptibility testing

BAP   Blood agar plate

BSI   Bloodstream infection

CA   Community-acquired

CAUTI   Catheter-associated urinary tract infection

CDC   Centers for Disease Control and Prevention

CFU   Colony forming units

cgMLST   Core genome multilocus sequence typing

CRE   Carbapenem-resistant *Enterobacteriaceae*

CR-Kp   Carbapenem-resistant *Klebsiella pneumoniae*

EMR   Electronic medical record

Ent   Enterobactin

ESBL   Extended-spectrum-β-lactamase

GWAS   Genome wide association study

HA   Hospital-acquired

HAI   Hospital-acquired infection

HAP   Hospital-acquired pneumonia

HV	Hypervirulent
hvKP	Hypervirulent *Klebsiella pneumoniae*
ICU	Intensive care unit
IDSA	Infectious Diseases Society of America
KPC	*Klebsiella pneumoniae* carbapenemase
KpI	*Klebsiella pneumoniae* group I (*K. pneumoniae*)
KpII	*Klebsiella pneumoniae* group II (*K. quasipneumoniae*)
KpIII	*Klebsiella pneumoniae* group III (*K. variicola*)
LB	Luria-Bertani
LPS	Lipopolysaccharide
LTACH	Long-term acute care hospital
MAC	MacConkey agar
MALDI-TOF	Matrix-assisted laser desorption/ionization time of flight
MCIK	MacConkey-inositol-potassium tellurite agar
MIC	Minimum inhibitory concentration
MLF	Mucoid lactose fermenter
MLST	Multilocus sequence typing
MRSA	Methicillin-resistant *Staphylococcus aureus*
NGS	Next-generation sequencing
NHSN	National Healthcare Safety Network
OD	Optical density
OR	Odds Ratio
PAL-Seq	Pathogenicity-associated locus sequencing

PCA   Principal component analysis

PLA   Pyogenic liver abscess

PNA   Pneumonia

rep-PCR   Repetitive sequence-based polymerase chain reaction

ROC   Receiver operator characteristic curve

SIBO   Small intestine bacterial overgrowth

ST   Sequence type

T2D   Type two diabetes mellitus

ter   Tellurite resistance locus

UK   Unknown

UMHS   University of Michigan Health System

UPEC   Uropathogenic *Escherichia coli*

UPGMA   Unweighted pair group method with arithmetic mean

UTI   Urinary tract infection

VAP   Ventilator-associated pneumonia

VRE   Vancomycin-resistant *Enterococci*

WGS   Whole genome sequencing

WT   Wild type

Ybt   Yersiniabactin

ABSTRACT

Klebsiella pneumoniae is a leading cause of hospital-acquired infections, including bacteremia, pneumonia, and urinary tract infections. Traditionally, this organism has been considered an opportunistic pathogen, infecting immunocompromised patients. In recent years, however, there has been development of highly antibiotic-resistant *K. pneumoniae* strains and the emergence of hypervirulent strains. This had led to complications in treating patients infected by *K. pneumoniae* and highlighted the need to understand acquisition and pathogenesis of these bacteria. Being able to identify patients at risk for infection, as well as identify potential disease-causing strains of *K. pneumoniae*, could aid in decreasing patient morbidity and mortality. This thesis dissertation is focused on the host and bacterial genetic risk factors that are associated with *K. pneumoniae* clinical infection. The central hypothesis of this work is that colonized patients are more likely to develop infection, and that specific genes in the accessory genome predict clinical infection. To test the link between colonization and infection, a cohort study of intensive care patients was performed. Among several patient risk factors identified, there was a significant and independent association between intestinal colonization and extra-intestinal infection with *K. pneumoniae*. To identify bacterial genetic risk factors for infection, a case-control study incorporating clinical modeling and a novel comparative genomics method called Pathogenicity-Associated Locus Sequencing (PAL-Seq) was performed. This approach identified several bacterial pathogenicity-associated loci (PALs). Combined with patient risk factors, these PALs were highly predictive of infection in our sample set. Together, these findings further our understanding of *K. pneumoniae* pathogenesis and have potential implications for identifying

1

patients at risk for developing infection. Identifying colonization as a significant step in progression to infection provides an opening for potential interventions to prevent infections. Furthermore, identification of bacterial genes associated with clinical infection provides potential diagnostic targets for predicting infection in colonized patients. This work supports a paradigm for *K. pneumoniae* pathogenesis where colonized patients either progress to infection or remain asymptomatically colonized based on the combined effects of patient risk factors and pathogenicity associated gene loci in the accessory genome.

CHAPTER 1

Introduction

Summary

Klebsiella pneumoniae is a leading cause of healthcare acquired infections (HAIs) in the United States. Typically, *K. pneumoniae* is considered an opportunistic pathogen, since it frequently infects critically ill and immunocompromised patients. However, risk factors for infection with *K. pneumoniae* are not fully understood. For decades, intestinal colonization has been suspected as a reservoir for strains that cause extra-intestinal HAIs such as pneumonia, urinary tract infections, and bloodstream infections. However, the factors influencing progression from colonization to infection remain unclear. In recent years, the effects of *K. pneumoniae* evolution on human infections has become apparent. Acquisition of plasmids and diversity of chromosomal content has created a diverse accessory genome resulting in both an increased resistance to antibiotics and the emergence of hypervirulent strains. In 2013, the Centers for Disease Control and Prevention identified carbapenem-resistant *Enterobacteriaceae* (CRE) as an urgent threat to public health since these bacteria are resistant to all or nearly all available antibiotics. *Klebsiella* species are responsible for approximately 80% of CRE infections. Concurrently, hypervirulent (HV) community-acquired strains of *K. pneumoniae* emerged in the Asian Pacific Rim, and have recently been isolated in several other locations, including the United States. These strains have a hypermucoviscous capsule (K1 or K2) and cause severe diseases such as pyogenic liver abscess (PLA), endophthalmitis, and meningitis in immunocompetent patients.

Finally, approximately 20% of isolates previously identified as *K. pneumoniae* are actually other species, specifically either *K. quasipneumoniae* or *K. variicola*. Though colonization appears to play a role in infection with each of these *K. pneumoniae* strain types, the dynamics of acquisition of colonizing and infecting strains as well as epidemiology of infections differs. This chapter will explore the associations between colonization and infection with opportunistic, antibiotic-resistant, and hypervirulent *K. pneumoniae* strains as well as the similarities and differences in patient risk factors. As *K. pneumoniae* infections become progressively more difficult to treat in the face of antibiotic resistance and hypervirulent strains, an increased understanding of the epidemiology and pathogenesis of these bacteria is vital.

1.1 Introduction

Klebsiella pneumoniae was first described by Carl Friedlander in 1882 as a bacterium isolated from the lungs of patients who had died from pneumonia [1]. Initially called Friedlander's bacillus, the genus was re-named *Klebsiella* after Edward Klebs. *Klebsiella* species are found ubiquitously in nature, including in plants, animals, and humans. They are the causative agent of several types of infections in humans, including respiratory tract infections, urinary tract infections, and bloodstream infections. Classically, these infections occur in hospitalized or otherwise immunocompromised patients and are routinely treated with β-lactams and other antibiotics effective against *Enterobacteriaceae*. In recent years, however, we have seen the development of antibiotic-resistant *K. pneumoniae* as well as the emergence of hypervirulent *K. pneumoniae* strains. Understanding how these strains are similar and how they differ from one another, as well as the genetic factors contributing to their epidemiology, is necessary to successfully combat these infections. Several species of *Klebsiella* are known to cause clinical infection, with *K. pneumoniae*

4

being the most prevalent, though recent advances in molecular capabilities have shown that a portion of clinical isolates identified as *K. pneumoniae* are in fact other species.

This chapter will focus on the epidemiological and genetic similarities and differences between endemic opportunistic, epidemic antibiotic-resistant, and emerging hypervirulent strains of *K. pneumoniae*, as well as the role the accessory genome plays in these differences. First, the classic epidemiology of this species will be discussed, focusing on sources of human gastrointestinal colonization, sites of infection, and risk factors for infection. Well-defined virulence factors and the accessory genome will also be discussed. Next, the development of antibiotic resistance in *K. pneumoniae* will be explored, including mechanisms of resistance, risk factors, colonization and infection, and the role of the accessory genome in resistance. The emergence of hypervirulent strains of *K. pneumoniae* will be also be discussed. This will include characterization of these strains, and epidemiology of infections caused by these bacteria. Finally, clinical misidentification of *Klebsiella* species will be discussed. We will also explore how the accessory genome plays a role in virulence, pathogenicity, and bacterial identification. This chapter will outline the current understanding of *K. pneumoniae* pathogenesis and the gaps in knowledge that require addressing. Ultimately, it will provide the information necessary to support the hypotheses discussed in this dissertation.

1.2 *Klebsiella pneumoniae*: an opportunistic, hospital-acquired infection

Klebsiella pneumoniae is a Gram-negative pathogenic bacterium. On agar media, it has a mucoid phenotype that is conferred by the polysaccharide capsule attached to the bacterial outer membrane and ferments lactose. *K. pneumoniae* is part of the *Enterobacteriaceae* family, which is comprised of other familiar pathogens such as *Escherichia coli*, *Yersinia* species, *Salmonella*

species, and *Shigella* species. *K. pneumoniae*, a leading cause of hospital-acquired infections (HAIs) in the United States, has classically been considered an opportunistic pathogen, since it has typically caused infections primarily in hospitalized or otherwise immunocompromised individuals. As the virulence and patient demographics of these bacteria and the infections they cause begin to shift, understanding how *K. pneumoniae* is transmitted and the factors responsible for pathogenicity is important in treating infected patients.

       1.2.1   *Klebsiella pneumoniae* commonly colonize human mucosal surfaces

The environment likely acts as a reservoir for acquisition of *K. pneumoniae*, either as colonization or infection. *K. pneumoniae* is frequently found in the environment in locations such as water, sewage, soil, and plant surfaces [2, 3]. Several studies have shown that *K. pneumoniae* in the environment are very similar to their clinical counterparts in biochemical patterns, virulence and pathogenicity, and bacteriocin susceptibility patterns [4-7], though capsule type representation differs between clinical/fecal and environmental sources [8]. However, environmental *K. pneumoniae* are significantly more susceptible to antibiotics than clinical *K. pneumoniae* [4], suggesting that selective pressure exists in a clinical setting. Interestingly, both environmental temperature and dewpoint are positively predictive of increased bloodstream infection caused by *K. pneumoniae*, suggesting that acquisition from the environment may vary with season and climate [9].

Once acquired, *K. pneumoniae* colonizes the of mucosal surfaces in humans, including the nasopharynx and the gastrointestinal tract. These bacteria can be found on skin, but are considered transient at this site rather than colonizing [10]. Colonization rates differ at each body site and based on whether bacteria are community- (CA) or hospital-acquired (HA). Rates of CA colonization of the nasopharynx have been reported from 3-15% [11-14], and are typically higher

in adults than children [12, 13]. Nasopharyngeal colonization has also been associated with alcohol consumption [14, 15]. Rates of colonization tend to increase upon hospitalization, as evidenced by HA nasopharyngeal colonization rates reaching as high as 19% [16]. The gastrointestinal tract is also a common site of colonization in humans. Rates of CA gastrointestinal colonization vary but can reach as high as 35% [11]. As with nasopharyngeal colonization, rates of gastrointestinal colonization increase among hospitalized patients and have been reported as high as 77% [17], though recent studies demonstrate rates around 20% [18]. Colonization rates are also shown to increase following antibiotic treatment [19]. Though colonization can occur at both body sites, gastrointestinal colonization is likely the more common and significant reservoir in terms of risk of transmission and infection. These colonization rates are based on detectable colonization by nasopharyngeal, rectal, or fecal sampling. It is possible that many more people are truly colonized, but that colonization is only detected by culture when the density exceeds a certain threshold.

There are several potential sources of transmission for patients colonized with *K. pneumoniae*, which may also act as sources of HAIs. As stated, the environment itself likely provides a route for acquisition. Another source of transmission is person-to-person contact between healthcare workers and patients, with healthcare workers' hands being a significant source [20, 21]. Contaminated surfaces and instrumentation have also been identified as sources of transmission [20].

1.2.2 Risk factors for progression from colonization to infection

Podschun noted in 1990 that the capsule serotypes of *Klebsiella* clinical isolates were more similar to fecal sample serotypes than those of environmental samples, suggesting that colonization and infection may be linked [8]. Several studies have suggested that the intestinal tract is a reservoir for hospital-acquired pathogens and have further suggested mechanisms for dissemination,

7

however most have focused on pathogens other than *K. pneumoniae* [22]. Though the progression from intestinal *K. pneumoniae* colonization to infection is not clearly understood, there are apparent risk factors that play a role in this progression. Nasopharyngeal colonization has been posited has a source of acquisition in pneumonia, but further studies need to be done to test this hypothesis [14]. Intestinal domination by Proteobacteria, which taxonomically includes *K. pneumoniae*, leads to a 5-fold increase in the risk of bacteremia in allogenic hematopoietic stem cell transplant patients [23]. This suggests that bacterial density of colonizing strains plays a role in progression to disease. Procedures such as endoscopy are a potential further source of endogenous infection [24]. Several underlying diseases have also been identified as risk factors for infection, since they weaken host defenses and therefore increase susceptibility to infection. Among diseases that are associated with both CA and HA *K. pneumoniae* infection, are cancer, diabetes mellitus, and alcoholism [25-27]. A final risk factor for progression to disease is variation in the bacterial accessory genome. Not all *K. pneumoniae* strains cause disease in animal models of infection. Similarly, not all colonizing strains go on to cause disease in humans. In fact, these bacteria are traditionally considered commensals and opportunistic pathogens [28]. Of those that do cause disease, certain genes, operons, or high pathogenicity islands have been identified in *K. pneumoniae* that are associated with virulence [29-32]. This indicates that progression to disease is not only reliant on host immunocompetence, but also on genes the bacteria possess.

1.2.3   Hospital-acquired infections

Klebsiella species have been identified as the third leading cause of hospital-acquired infections in the United States (9.9%) behind *Clostridium difficile* and *Staphylococcus aureus* [33]. *K. pneumoniae* causes serious infections such as pneumonia, wound infections, urinary tract infections (UTIs), and bloodstream infections [33]. In fact, *Klebsiella* species have been identified

as the number three cause of HA pneumonia in the United States [33]. Hospital-acquired pneumonia (HAP) is defined as a pneumonia occurring ≥48 hours after hospital admission. *Klebsiella* species are also a leading cause of ventilator-associated pneumonia (VAP) among patients in intensive care units (ICUs) [34, 35], and VAP is responsible for 83% of hospital-acquired pneumonias [36]. CA cases of *K. pneumoniae* pneumonia are frequently associated with chronic alcoholism [37]. Mortality rates in *K. pneumoniae* pneumonia have been reported as high as 50% [17].

K. pneumoniae are the second leading cause of bloodstream infections (BSI) caused by Gram-negative bacteria, behind only *E. coli* [17, 33]. Cancer is the primary underlying disease associated with hospital-acquired BSI, while liver disease and diabetes mellitus had the highest association among CA *K. pneumoniae* BSI [38]. BSI can be a primary infection with no identifiable source. However, BSI is often a secondary infection that results from dissemination into the bloodstream from a known source. Common sources of secondary BSI include the urinary tract, the gastrointestinal tract, intravenous or urinary catheters, and respiratory sites [39]. The 30-day mortality rate of HA BSI due to *K. pneumoniae* is approximately 30%, although mortality rates of up to 50% have been noted [17].

The urinary tract is the most common site of infection by *K. pneumoniae* [17]. As with other infections, UTI due to *K. pneumoniae* are associated with diabetes mellitus [40]. Catheter-associated UTIs (CAUTIs) are another infection caused by *K. pneumoniae*. It is thought that these are facilitated by the ability for form biofilms and adhere to catheters [41]. *Klebsiella* are also responsible for wound/surgical site infections. This site represents approximately 13% of all infections caused by *Klebsiella* [17, 33]. Together, *K. pneumoniae* infections at each of these body

sites represent an endemic opportunistic pathogen that demonstrates a substantial burden for healthcare.

1.2.4   The accessory genome

Within a bacterial species, there is typically a set of genes that is conserved amongst all members. This set of genes is considered the core genome, and contains genes necessary for essential life processes such as cellular replication and metabolism. In *K. pneumoniae* the core genome is estimated to be comprised of approximately 2,000 genes [30]. Genes that vary between isolates are referred to as the accessory genome. This includes chromosomally encoded genes and genes on plasmids. Since *K. pneumoniae* genomes are typically between 5,000-6,000 genes, this means that each isolate possesses an accessory genome of at least 3,000 genes. Genes in the accessory genome can aid in specific processes, such as nitrogen fixation. They can also encode specific virulence factors, as discussed below. The accessory genome also carries genes encoding various antibiotic-resistant enzymes and mechanisms. A recent study of 328 *Klebsiella* isolates identified almost 30,000 unique protein-coding sequences [30], defining the *Klebsiella* pan-genome. The authors further demonstrated that the "pangenome" is open, indicating that there are more accessory genes yet to be identified and characterized.

1.2.5   *Klebsiella pneumoniae* virulence factors

Several factors for *K. pneumoniae* have been identified that contribute to virulence and pathogenicity in disease progression. Established virulence factors in *K. pneumoniae* include capsule, lipopolysaccharide, siderophores, and pili [17]. The polysaccharide capsule is one of the most important virulence factors used by *K. pneumoniae*. It is primarily used to assist in evading the immune system during infection, either by protecting bacteria from opsonophagocytosis [42] and from serum killing [43]. The capsule is generated by the capsular polysaccharide synthesis

(*cps*) locus in *K. pneumoniae*, and is a structure that lies on the outside of the bacterial cell attached to the outer membrane. It is composed of repeating subunits of four to six sugars, as well as uronic acids [17]. Certain capsule types are associated with increased virulence, specifically capsule types K1 and K2 [44]. Based on serological testing, 77 capsular types have been identified [45]. These various capsule types are generated by different alleles of the genes in the *cps* locus, which is part of the core genome.

Lipopolysaccharide (LPS), also termed endotoxin, is a major component decorating the outer membrane of Gram-negative bacteria. Genes encoding LPS are therefore part of the core genome. LPS is widely recognized as the most powerful mediator of septic shock caused by bacteria. Host sensing of LPS via Toll-like receptor 4 (TLR4) leads to an inflammatory cascade [46]. It is this host response, rather than LPS itself, that leads to the devastating pathogenesis of sepsis and septic shock. LPS molecules are comprised of lipid A, a core domain, and the O-antigen. LPS have an overall negative charge. LPS has been shown to have multiple virulence mechanisms in Gram-negative bacteria. Certain LPS O-antigens confer resistance to complement in Gram-negative bacteria, including *K. pneumoniae* [43]. LPS is also thought to play a role in protecting bacteria from antimicrobial peptides, including polymyxin antibiotics [47, 48].

Siderophores are high affinity, low-molecular weight iron-chelating molecules secreted by various bacteria to aid in iron acquisition [49]. *K. pneumoniae* secrete multiple types of siderophores [50]. One of the common catecholates secreted by *K. pneumoniae* is enterobactin (Ent), which is encoded in the *K. pneumoniae* core genome [51, 52]. Since Ent is a common siderophore, the innate immune system has developed a way to bind Ent, preventing bacteria from acquiring iron [53]. Bacteria have therefore developed other siderophores, which are encoded in the accessory genome, to counteract this. A second catecholate siderophore is a glucosylated Ent-

derivative, Salmochelin (Sal). Sal, encoded by the accessory genome, is a mechanism to evade the innate immune system [54]. Mixed-type siderophores such as yersiniabactin (Ybt) and aerobactin (Aer) are also commonly secreted by *K. pneumoniae*. The various siderophores are associated with different sites and severity of infection, and their pathogenic functions extend beyond simply iron acquisition. Ybt was first identified in *Yersinia* species, and the genes encoding the biosynthesis, transport, and regulation of Ybt are considered part of the *K. pneumoniae* accessory genome. They are located on a transposable chromosomal fragment termed a "high pathogenicity island" after to the resulting expression of a high pathogenicity phenotype [55]. Secretion and utilization of Ybt is associated with respiratory tract infections in patients and is sufficient to promote pneumonia in a murine model [56]. World-wide, Ybt is the most common virulence factor associated with human infections [30]. Aer has also been identified as a major virulence factor in *K. pneumoniae* infections. Aer is plasmid-encoded and is also part of the accessory genome [57]. In addition to providing iron for *K. pneumoniae* replication, secretion of siderophores by *K. pneumoniae* isolates also induces bacterial dissemination and lung inflammation [58]. Siderophore-dependent dissemination requires the host transcriptional regulatory protein HIF-1α that is activated by iron chelation.

A critical step in progression to infection for bacteria is to gain a foothold at the site of infection by adhering to host surfaces. In *K. pneumoniae*, this is frequently achieved using pili (fimbriae). Pili are filamentous structures extending from the surface of bacteria. They can be as long as 10μm and between 1-11 nm in diameter [59]. There are two common types of pili found on *K. pneumoniae*: type 1 pili and type 3 pili. Type 1 pili are thought to aid virulence through their ability to adhere to human mucosal or epithelial surfaces, and thus colonize those areas and create a reservoir for infection. Type 3 pili similarly adhere to cell surfaces, but importantly have been

identified as strong promoters of biofilm formation [41]. Both type 1 and type 3 pili are considered part of the core genome [60, 61]. It is thought that both types of pili play a role in colonization of urinary catheters, leading to catheter-associated urinary tract infections [62].

1.3 Emergence of antibiotic resistance in *Klebsiella pneumoniae*

The Centers for Disease Control and Prevention estimates that more than two million people contract infections due to antibiotic resistant microorganisms each year in the United States [63]. Of those infected, approximately 23,000 die. There are multiple factors believed to contribute to the spread of antibiotic resistance, including inappropriate antibiotic use in healthcare and agriculture, and lack of new antimicrobial therapeutics. *K. pneumoniae* are one of several bacteria that have experienced a dramatic increase in antibiotic resistance in the past decades. Several mechanisms of antibiotic resistance are found in *K. pneumoniae*, with resistance to β-lactams having the greatest impact on effective treatment. Colonization with antibiotic-resistant *K. pneumoniae* has been associated with subsequent infection with antibiotic-resistant *K. pneumoniae* in hospitalized patients, although the progression from colonization to infection is incompletely understood. With antibiotic-resistant *K. pneumoniae* emerging as an urgent threat to public health, understanding the epidemiology of these infections may help treat and control infections due to these bacteria.

1.3.1 β-lactamase-producing *Klebsiella*

Resistance of bacteria to β-lactam antibiotics emerged before penicillin was widely used to treat infections. Alexander Fleming was the first to note that *E. coli* and other bacteria were not inhibited by penicillin [64], a resistance which was later attributed to an enzyme produced by these bacteria [65]. Resistance to β-lactam antibiotics is achieved through hydrolysis of the antibiotic β-

lactam ring by β-lactamases. In several Gram-negative bacteria, this resistance to β-lactams is naturally occurring since the enzyme is chromosomally encoded in the core genomes of the species. In *K. pneumoniae* SHV is consistently found in the chromosome, and corresponding ampicillin resistance is a hallmark of the species [66, 67]. In the 1960s the first plasmid-mediated β-lactamase, TEM-1, was discovered in *E. coli* [68]. *K. pneumoniae* are also known to harbor plasmid-mediated β-lactamases, such as AmpC enzymes which confer resistance to most penicillin antibiotics [69]. β-lactamase enzymes are thought to have evolved from penicillin binding proteins due to selective pressure in the environment [70-73].

1.3.2   Extended-spectrum β-lactamases

Extended spectrum β-lactamase (ESBL) -producing *Klebsiella pneumoniae* were first identified in Europe in 1983 [74] and in the United States in 1989 [75]. ESBLs are classified as β-lactamases able to hydrolyze oxyimino-cephalosporins, such as third-generation cephalosporins and aztreonam, but are inhibited by clavulanic acid [76]. Frequently plasmids encoding β-lactamases also possess resistance genes to other antibiotics as well as heavy metals, meaning bacteria harboring these plasmids are frequently multidrug-resistant [77]. A recent study identified that *K. pneumoniae* clonal group 307 (CG307) is associated with ESBL infections [78].

Several ESBL enzymes have been discovered, many of which are derivatives of TEM-1 and SHV-1 β-lactamases through point mutations that alter amino acid sequences. These mutations alter the enzymatic capabilities of ESBL-producing bacteria, allowing them to hydrolyze third-generation cephalosporins and aztreonam. TEM-3 was the first β-lactamase to display the ESBL phenotype in the late 1980s [79]. Recently a new class of plasmid-mediated ESBLs has emerged called CTX-M, which favor hydrolyzing cefotaxime. These ESBLs are found in several *Enterobacteriaceae* worldwide [80]. The OXA family of ESBLs has also recently emerged. These

enzymes confer resistance to cloxacillin and oxacillin by hydrolyzing these antibiotics, including extended spectrum oxyimino-β-lactams [81]. OXA enzymes are often derivatives of OXA-10. These are most frequently found in *Pseudomonas aeruginosa* and *E. coli* isolates in Turkey and France. Carbapenems have typically been the drug of choice to treat severe infections caused by ESBL-producing bacteria.

1.3.3   Carbapenem-resistant *Klebsiella pneumoniae* (CR-Kp)

Perhaps due to the selective pressure of treating ESBL infections with carbapenems, carbapenem resistance has emerged and *K. pneumoniae* is the most common carbapenem-resistant *Enterobacteriaceae*. Emergence of carbapenem-resistant organisms, which are also frequently multidrug-resistant, complicates treatment of infections since it leaves few therapeutic options available. In 2013 the CDC declared carbapenem-resistant *Enterobacteriaceae* (CRE) an urgent threat to public health in the United States, bringing these "superbug" organisms to the forefront of the battle against antibiotic resistance [63]. Of the approximately 9,000 infections due to CRE *Klebsiella* species are responsible for approximately 80% of infections, representing a significant healthcare burden and a field in need of further investigation.

There are multiple mechanisms responsible for carbapenem-resistance in *K. pneumoniae*, including up-regulation of efflux pumps and alteration of outer membrane porins and permeability, hyperproduction of ESBL enzymes or AmpC β-lactamases, as well as production of carbapenem-specific β-lactamases. Carbapenemases are classified as belonging to either molecular Ambler class A serine carbapenemases, class B metalloenzymes, or class D oxacillinases [81]. Frequently carbapenem-resistant bacteria both produce a carbapenemase, and also have decreased porin function. Typically, the resistance phenotype in *K. pneumoniae* requires both a porin mutation and hyperproduction of an ESBL or a carbapenemase enzyme. Carbapenem-resistant *K. pneumoniae*

are also often associated with clonal group 258 (CG258) [82, 83]. Interestingly, the siderophore Ybt is found in several CG258 antibiotic-resistant strains, adding to the difficulty of combating these strains [30].

Klebsiella pneumoniae carbapenemase (KPC) β-lactamases are class A carbapenemases and the most frequent carbapenemases found in *K. pneumoniae*. They are mainly plasmid-encoded enzymes (bla_{KPC} genes) and are also produced by several other *Enterobacteriaceae*. Among *K. pneumoniae*, these enzymes are found primarily in ST258 strains. KPCs contain serine at their active site, and act by hydrolyzing the β-lactam ring of β-lactam molecules including carbapenems [84]. The first KPC, KPC-1, was discovered in a carbapenem-resistant *K. pneumoniae* during the late 1990s in North Carolina, although it was later determined to be identical to KPC-2 [85, 86]. Since then several KPC variants have been identified worldwide.

Class B metallo-β-lactamases are characterized by a requirement for zinc at their active site. Because of this, they are frequently inhibited by metal chelators. The New Delhi metallo-β-lactamase-1 (NDM-1) is characteristic of this class. NDM-1 was discovered in a *K. pneumoniae* clinical urinary culture isolate from a Swedish patient who had recently traveled to India [87]. The bla_{NDM-1} gene is frequently plasmid-encoded, facilitating easy transfer between isolates, although it has also been carried chromosomally [88]. Acquisition of isolates that secrete NDM-1 seems to be associated with travel and hospitalization.

Class D enzymes are characterized by their ability to hydrolyze cloxacillin or oxacillin and are therefore designated as OXA enzymes [81]. As stated above, OXA enzymes are also found as ESBLs. A newer subclass of these enzymes has further been found to hydrolyze carbapenems. Most relevant here is the plasmid encoded OXA-48 (bla_{OXA-48}), which is found in *K. pneumoniae* and confers a high level of resistance to imipenem [89].

1.3.4   Colistin resistance in *Klebsiella pneumoniae*

In recent years, colistin resistance has become a major concern in *Enterobacteriaceae*. Colistin is among the polymyxin class of antibiotics. Polymyxins were used to treat infections due to Gram-negative bacteria in the 1960s and 1970s. Their use was soon discontinued, due to renal- and neurotoxicity [90]. The recent emergence of CRE has made it necessary to return to colistin as a drug of last resort to treat multidrug-resistant Gram-negative infections. Resistance to colistin in Gram-negative bacteria can be intrinsic or acquired. Colistin resistance is acquired in *K. pneumoniae*, typically through various mutations to a two-component regulatory system that modifies bacteria lipid A, which is the target of polymyxin antibiotics, decreasing the ability of polymyxins to interact [91-96]. In 2015, plasmid-mediated resistance to colistin was discovered in an *E. coli* isolate in China [97], conferred by the *mcr-1* gene. This discovery heralds a real potential for easily transmissible pan-resistance, which is the ultimate concern in antimicrobial resistance. The first reported incidence of *mcr-1* in the United States occurred in 2016 in *E. coli* [98]. In September of 2016 a pan-resistant isolate of *K. pneumoniae* was isolated [99]. Colistin resistance in this isolate was not mediated by *mcr-1*.

1.3.5   Colonization as a reservoir of antibiotic resistant K. pneumoniae

In the 1970s, intestinal colonization with antibiotic-resistant *K. pneumoniae* was first considered a reservoir for subsequent infection in hospitalized patients [100]. This study determined that patients colonized with multidrug-resistant *K. pneumoniae* after hospital admission developed infection with multidrug-resistant *K. pneumoniae* at a higher percentage within 21 days compared to those who did not become intestinal carriers. They further compared serotypes of colonization isolates and subsequent infecting isolates, finding that 45.2% of patients colonized after admission were infected with the same serotype, indicating that hospitalized

patients are often colonized with the same isolate they become infected with. There was, however, a predominant serotype circulating, confounding the significance of these findings. Since then, studies have identified colonization with antibiotic resistant *K. pneumoniae* as a risk factor for infection, but little has been done to determine a more concrete association or progression.

1.3.6   Infections caused by antibiotic resistant *Klebsiella pneumoniae*

Risk factors for colonization and infection with antibiotic-resistant *K. pneumoniae* are often considered together, so the risk factors for infection specifically are unclear [16, 100, 101]. Risk factors associated with ESBL colonization and infection include prior treatment with antibiotics, prolonged hospitalization, prolonged ICU stay, and mechanical ventilation [102-104]. Intestinal colonization with ESBL bacteria has also been associated with ESBL infection. Risk factors associated with carbapenem-resistant *K. pneumoniae* colonization and infection include, prior antibiotic treatment, renal dysfunction, older age, surgical procedures and ICU admission [105, 106]. As with endemic strains of *K. pneumoniae*, hospitalization seems to be a key factor associated with infection. Though antibiotic resistant strains can infect an array of body sites similar to endemic strains, they frequently cause urinary tract infections [100, 103]. This may represent inoculation of the urinary tract with *K. pneumoniae* from the gastrointestinal tract across the perineum.

1.4 Community-acquired hypervirulent strains

In the 1980s and 1990s, reports began to emerge from the Asian Pacific Rim detailing severe infections due to *K. pneumoniae* [107-109]. These infections were unique in that they were community-acquired (CA), a departure from the classic presentation of *K. pneumoniae* infections in hospitalized patients. Alarmingly, these hypervirulent (HV) strains have begun to emerge

worldwide, including in the United States [110-115]. Several virulence factors are unique to these isolates, which will be discussed below.

1.4.1 Hypervirulent *K. pneumoniae* possess unique virulence factors

The most striking aspect of HV *K. pneumoniae* (hvKP) isolates is their ability to cause unnervingly severe infections in otherwise healthy patients. The severity of infections, termed hypervirulence, is attributed to virulence factors encoded by the accessory genome. Common infections due to hvKP include pyogenic liver abscess (PLA); endophthalmitis, an infection inside the eye; meningitis; and bloodstream infections [116]. Symptoms of PLA vary between individuals and are frequently non-specific. Diagnosis requires radiographic imaging [117]. Approximately 3-11% of patients with PLA will go on to develop endophthalmitis [118, 119]. These isolates are further characterized by their hypermucoviscous phenotype. This phenotype has been determined to be conferred by two proteins: RmpA, which regulates capsule production [120, 121], and MagA, which is associated with the hypermucoviscous phenotype [122]. Genes encoding RmpA and MagA are highly associated with hvKP, especially in Asia, and are considered virulence factors. However recently K1 and K2 capsule type have been identified as playing a more important role in virulence than these to genes [123]. These isolates are characterized as community-acquired (CA) infections since they frequently infect non-hospitalized patients. Patient risk factors for severe infection with hvKP include being aged 55-60 years, male, and having diabetes mellitus [120, 124, 125]. Frequently, hvKP strains are K1 or K2 capsule type [120, 126-128]. They are also often sequence type 23 (ST23) a phylogroup strongly associated with the K1 capsule type [129]. Fortunately, hvKP isolates are highly susceptible to most antibiotics [116]. The siderophore Ybt has also been associated with ST23 hypervirulent strains [30], and the siderophore Aer has been distinguished as the most common siderophore secreted by hypervirulent *K. pneumoniae* [130].

1.4.2 Colonization as a reservoir

It is unknown whether gastrointestinal hvKP colonization is associated with infection, although due to the high virulence of these strains it might be expected that asymptomatic colonization would be infrequent. If colonization is a potential reservoir for infection with hvKP strains, then understanding the community carriage rates is important. With that in mind, a recent study sought to identify the fecal carriage rate of K1 *K. pneumoniae* in healthy Koreans in order to identify this potential reservoir [131]. They determined that 4.9% of individuals tested carry K1 *K. pneumoniae*. Of these, 94.7% were found to be ST23, which is strongly associated with PLA. They further determined that the K1 carriage rate was higher in those who lived in Korea compared to those of Korean descent who live outside of Korea (24.1% vs 5.6%, *P* = .024), suggesting that exposure to the Korean peninsula plays a role in K1 colonization. These findings establish that a hvKP colonization state exists, and further suggest colonization as a key step in progression to infection with K1 hvKP strains. Further studies are needed to determine if colonization with ST23 or K1 strains precedes infection.

1.5 *Klebsiella* species are distinguished by their accessory genome

Increased molecular epidemiology and sequencing capabilities have recently demonstrated that isolates frequently identified as *K. pneumoniae* can be divided into three distinct *Klebsiella* species, largely based on their accessory genome [132-136]. Phylogroup KpI represents approximately 80% of these isolates and is the species *K. pneumoniae* (*sensu stricto*). Phylogroup KpII is identified as *K. quasipneumoniae* and KpIII as *K. variicola*. As three distinct species, these groups vary in their epidemiology of colonization and infection.

Not much is known about the smallest of the three Kp phylogroups, *K. quasipneumoniae*. This species was identified as phylogroup KpII in 2001 [132], though it was not described as *quasipneumoniae* species until over a decade later [136]. Approximately 94% of *K. quasipneumoniae* are found in humans, with over 50% of human isolates being associated with intestinal carriage rather than infection (compared to 24% and 39% intestinal carriage in KpI and KpIII respectively) [30]. Infections caused by *K. quasipneumoniae* are primarily urine and respiratory infections, and a high frequency of isolates are ESBL-producers [30].

Klebsiella variicola was proposed as a new, distinct species based on both genetic and biochemical differences from *K. pneumoniae* [136, 137]. A discrete trait of *K. variicola* is the ability of this species to fix nitrogen, which explains their endophytic relationship with several plants such as maize, banana, sugarcane, and wheat. Genes responsible for nitrogen fixation, such as those in the *nif* operon, are found in the *K. variicola* accessory genome [138]. While colonization of plants is common, colonization rates in humans is unknown. Though this species is classically thought to be associated solely with environmental sources, the clinical importance of *K. variicola* is becoming more apparent. In fact, a recent study determined that patients with bloodstream infections due to *K. variicola* demonstrate a higher 30-day mortality rate (29.4%) compared to *K. pneumoniae* (13.5%) and *K. quasipneumoniae* (11.1%) species [135]. This increased mortality was not due to any known virulence factors and was significant after controlling for patient co-morbidities, suggesting that the accessory genome of this species harbors yet undiscovered and potentially clinically relevant genes.

A comparison of the antimicrobial resistance patterns of clinical isolates of the three Kp phylogroups (n=420) to ten antimicrobial agents showed that KpI has the highest resistance levels, KpII has intermediate resistance levels, and KpIII has the lowest resistance, with KpI having a

resistance rate of 2- to 3-fold higher than KpIII [133]. A more recent study of isolates from the three phylogroups collected worldwide (n=328) determined that approximately 50% of KpII isolates display an ESBL phenotype [30].

Since it is only relatively recently that these species have been differentiated, there is a paucity of information in the literature regarding colonization rates and virulence factors in *K. quasipneumoniae* and *K. variicola*. These three distinct but related species vary in their epidemiology of infection and antimicrobial resistance, likely due to the variation in genes each species possesses, indicating that exploration of the *Klebsiella* accessory genome is imperative for better understanding the nuances of how these bacteria cause disease.

1.6   Gaps in knowledge

There have been a number of studies involving virulence factors and epidemiology of *K. pneumoniae* in recent years, as well as increased application of molecular techniques to explore variations and similarities in *K. pneumoniae* genomes. However, several key questions remain underexplored. The strength of association between colonization and subsequent infection is poorly defined. Also, variables controlling the progression of colonization to infection are unclear. These variables have been explored to some extent in both antibiotic-resistant and hypervirulent strains, but not in endemic isolates. Additionally, there is frequently a conflation of risk factors for colonization and infection. Understanding them as two distinct stages with potentially varying risk factors will further aid in understanding epidemiology of infection.

Another critical gap in knowledge is the relative lack of understanding of *K. pneumoniae* pathogenesis in general. Several *K. pneumoniae* virulence factors identified to date are based on animal models or association with hypervirulent phenotypes, such as PLA. Furthermore, most

studies involving *K. pneumoniae* tend to focus on one clonal group, such as carbapenem-resistant or hypervirulent isolates). While understanding of pathogenesis for each distinct type is important, a broad understanding of how *K. pneumoniae* causes the common infections of pneumonia, bacteremia and urinary tract infections in could better aid in targeting common factors for diagnosis and treatment. Finally, it is only recently that we have begun to recognize that approximately 20% of isolates identified as *K. pneumoniae* are actually other species. These species have distinct epidemiological and resistance profiles, and may have further clinically relevant distinctions. Identification of pathogenicity-associated loci will aid in a more thorough understanding of how these bacteria cause disease.

1.7 Conclusions

Though classically considered an opportunistic, hospital-acquired infection that infects only immunocompromised hosts, *K. pneumoniae* has emerged as two distinct threats with the increased development of antibiotic resistance and emergence of hvKP strains. The obvious concern with these is a dual-risk isolate, one that is both antibiotic-resistant and hypervirulent. With treatment options already limited, this result could be devastating. Because of the threat of dual-risk isolates, increasing our understanding of the broad epidemiology and pathogenesis of *K. pneumoniae* is important for combating these infections.

1.8 Outline of thesis

The focus of this thesis dissertation is defining the risk factors for infection with *K. pneumoniae* in hospitalized patients. The central hypothesis of this work is that colonized patients are more likely to develop infection, and that genes in the accessory genome promote clinical infection.

These studies identified host and bacterial risk factors associated with *K. pneumoniae* infection, which may lead to new screening methods to identify at-risk patients and eventually new therapeutic targets for disease treatment. In Chapter II, the association between intestinal colonization and subsequent infection with *K. pneumoniae* will be explored. Clinical models are used to assess the significant and independent association between colonization and infection. Molecular tools are used to show that patients are frequently extra-intestinally infected with the same strain they are colonized with. Chapter III will delve into identification of pathogenicity-associated loci (PALs). A novel comparative genomics technique will be described, and identified PALs will be characterized *in silico*, *in vitro*, and *in vivo*. In Chapter IV, the findings of these studies will be discussed as will the potential implications for clinical laboratories and healthcare facilities. Finally, future questions and studies to verify and further characterize current findings will be discussed.

CHAPTER II

Molecular Epidemiology of Colonizing and Infecting *Klebsiella pneumoniae* Clinical Isolates

Summary

Klebsiella pneumoniae is among the most common causes of hospital-acquired infections and has emerged as an urgent threat to public health due to carbapenem antimicrobial resistance. *K. pneumoniae* commonly colonizes hospitalized patients and causes extra-intestinal infections such as urinary tract infection, bloodstream infection (septicemia), and pneumonia. If colonization is an intermediate step before infection, then detection and characterization of colonizing isolates could enable strategies to prevent or empirically treat *K. pneumoniae* infections in hospitalized patients. However, the strength of association between colonization and infection is unclear. To test the hypothesis that hospitalized patients become infected with their colonizing strain, 1765 patients were screened for rectal colonization with *K. pneumoniae* and extra-intestinal isolates from these same patients were collected over a three-month period in a cohort study design. The overall colonization prevalence was 23.0%. After adjustment for other patient factors, colonization was significantly associated with subsequent infection: 21 of 406 (5.2%) colonized patients later had extra-intestinal infection, compared to 18 of 1359 (1.3%) non-colonized patients (adjusted odds ratio [OR] 4.01 (95% confidence interval 2.08–7.73, *P* <.001). Despite high diversity of colonizing isolates, 7/7 respiratory, 4/4 urinary, and 2/5 bloodstream isolates from colonized patients matched the patients' corresponding rectal swab isolates based on *wzi* capsular typing,

multilocus sequence typing (MLST), and whole genome sequence analysis. These results suggest that *K. pneumoniae* colonization is directly associated with progression to extra-intestinal infection.

Importance

K. pneumoniae commonly infects hospitalized patients and these infections are increasingly resistant to carbapenems, antibiotics of last resort for life-threatening bacterial infections. To prevent and treat these infections, we must better understand how *K. pneumoniae* causes disease and discover new ways to predict and detect infections. This study demonstrates that colonization with *K. pneumoniae* in the intestinal tract is strongly linked to subsequent infection. This helps to identify a potential time frame and possible approach for intervention: the colonizing strain from a patient could be isolated as part of a risk assessment and antibiotic susceptibility testing could guide empiric therapy if the patient becomes acutely ill.

2.1 Introduction

Klebsiella pneumoniae is a Gram-negative bacillus and a member of the *Enterobacteriaceae* family. *Klebsiella* spp. are among the most common causes of hospital-acquired infections (HAIs) in the U. S., responsible for about 10% of all infections [33]. However, the major sources of *K. pneumoniae* that cause HAI remain unclear. Intestinal colonization [100], presence of *K. pneumoniae* in the environment, contaminated instruments [20], and healthcare workers' hands [20, 21] have all been implicated in transmission. *K. pneumoniae* gastrointestinal colonization rates in hospitalized patients are estimated to be 20-38%, based largely on studies before 1980 [17, 100, 139, 140], and a more recent study identified a 21.1% fecal carriage rate in

healthy adults in Korea [131] with a high proportion of Sequence Type (ST) 23 isolates that are associated with pyogenic liver abscess. Prior treatment with antimicrobials has been reported as a risk factor for colonization [16, 100, 101] but may be specific for antimicrobial resistant *Klebsiella*. Earlier work identified gastrointestinal colonization with *K. pneumoniae* as a reservoir for infection with *K. pneumoniae* [100], but may reflect the virulence potential of two predominant serotypes in this cohort. Regardless of transmission route, *K. pneumoniae* appears to be transmitted efficiently as evidenced by reported outbreaks [141].

New techniques in molecular strain typing offer the opportunity to measure concordance among colonizing and infecting isolates of *K. pneumoniae* in patients. Repetitive sequence-based PCR (rep-PCR) has been widely used to characterize isolates in antibiotic-resistant *K. pneumoniae* [142-144]. Multilocus sequence typing (MLST) characterizes *K. pneumoniae* based on polymorphisms of seven conserved genes (*rpoB*, *gapA*, *mdh*, *pgi*, *phoE*, *infB,* and *tonB*) [145], and is widely used as a common language for *K. pneumoniae* strain typing. Sequencing of the *wzi* gene is a rapid and inexpensive approach to differentiate *K. pneumoniae* capsular types [146]. Recent studies reported that *wzi* sequencing has similar discriminatory power to MLST [147-149], suggesting that *wzi* could be used as a rapid and inexpensive method to screen for genetic differences among strains. Ultimately, whole genome sequencing (WGS) has become the gold standard for strain typing. For *K. pneumoniae*, a core genome MLST scheme, based on WGS and 634 conserved genes, has been validated as a way to characterize strains in a systematic and reproducible manner. These new tools provide methods to both screen for strain differences and confirm strain concordance with the power of WGS [66].

The objective of this study was to test the hypothesis that intestinal colonization leads to subsequent infection with *K. pneumoniae* in hospitalized patients. To test this hypothesis, we

determined the association and strain concordance between intestinal *K. pneumoniae* colonization and subsequent extra-intestinal infections in a large cohort. The rationale for this study was that, if colonizing isolates are highly likely to cause disease, this provides a focus for pathogenesis research and a potential window for infection prevention interventions.

2.2 Results

2.2.1 Patient demographics

During a three-month period, 1800 patients were screened for *K. pneumoniae* colonization by rectal swab culture; extra-intestinal infection with *K. pneumoniae* among this group was assessed based on positive clinical cultures. After excluding 35 patients whose first rectal swabs were collected after their first positive *K. pneumoniae* isolate at an extra-intestinal site, a total of 1765 patients were included in subsequent analysis (Figure 2.1). Of 77 patients with a positive blood, respiratory or urine culture, 39 patients met case definitions of infection (11 cases of bloodstream infection [BSI], 15 of pneumonia, and 14 of UTI; one patient met case definitions for both pneumonia and UTI). The demographic characteristics of patients with and without clinical infections are shown in Table 2.1. There were no significant differences in age, sex, or self-reported race/Hispanic ethnicity. Antibiotic exposure was numerically higher in the uninfected group (26.2% vs 12.8% of infected patients), but the difference did not reach statistical significance ($P = .067$). Neurologic disorders and fluid and electrolyte disorders were significantly more frequent in infected patients than in non-infected patients. Baseline albumin levels were significantly lower in the infected group ($P = .009$) and length of stay was significantly longer (14.9 vs 11.6 days, $P = .01$).
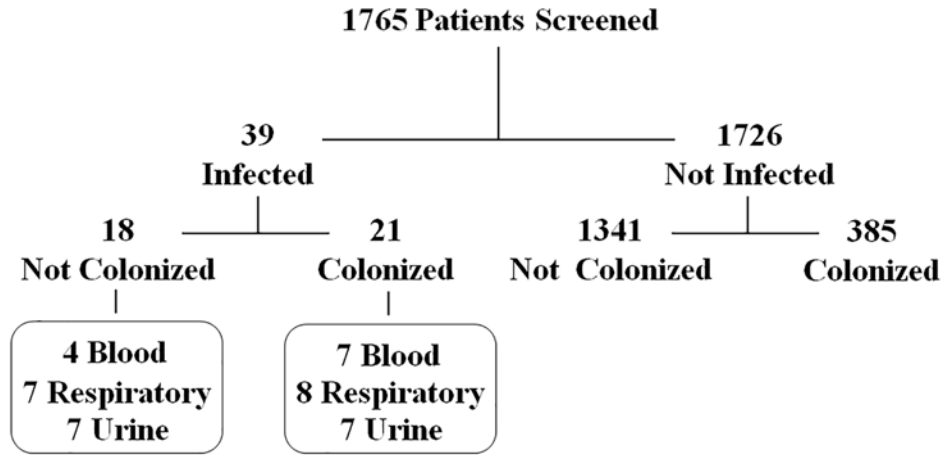
Figure 2.1: Study Population.

Adult patients in the University of Michigan Health System intensive care units (ICUs) and adult hematology/oncology patients were screened for colonization and extra-intestinal infection with *K. pneumoniae* between July and October of 2014 (n=1765), divided into "Infected" and "Not Infected" groups, and further divided into "Colonized" and "Not Colonized". Colonization is defined as a *K. pneumoniae* positive rectal swab. Number of infections by body site are shown in boxes; one colonized patient met case definitions for both pneumonia and UTI.

Table 2.1: Demographic characteristics of patients with and without infection.

Variable	Infection (n = 39)	No Infection (n = 1726)	P value[a]
	No (%) or mean ± standard deviation		
Female	19 (48.7)	835 (48.3)	>.99
White race	31 (79.5)	1438 (83.3)	.488
Hispanic	0 (0)	33 (1.9)	.984
Prior admit (28 days)	27 (69.2)	947 (54.8)	.079
Length of stay (days)	14.9 ±14.9	11.6 ±18.9	.01
Neurologic disorder[b]	6 (15.4)	109 (6.3)	.029
Fluid & electrolyte disorder[b]	20 (51.3)	569 (33)	.019
Renal disease[b]	5 (12.8)	286 (16.6)	.533
Liver disease[b]	5 (12.8)	112 (6.5)	.123
Alcohol abuse[b]	1 (2.6)	94 (5.4)	.442
Solid organ tumor[b]	13 (33.3)	403 (23.3)	.151
Diabetes mellitus, uncomplicated[b]	8 (20.5)	242 (14)	.218
K. pneumoniae colonization	21 (53.8)	385 (22.3)	<.001
Central line before colonization	27 (69.2)	963 (55.8)	.099
Antibiotics[c]	5 (12.8)	453 (26.2)	.067
Aminoglycoside	0 (0)	178 (10.3)	.984
Fluoroquinolone	0 (0)	0 (0)	N/A
Macrolide	2 (5.1)	151 (8.7)	.433
Cephalosporin	1 (2.6)	139 (8.1)	.237
Carbapenem	0 (0)	0 (0)	N/A
Clindamycin	2 (5.1)	97 (5.6)	.895
Age (years)	62.7 ±12.8	58.2 ±16.1	.078
Hemoglobin, baseline (g/dl)	10.4 ±2.4	11.1 ±2.5	.062
Platelets, baseline ($\times 10^3/\mu l$)	167.9 ±90.9	207.3 ±115.5	.012
Albumin, baseline (g/dl)	3.3 ±0.6	3.5 ±0.6	.009
Body mass index (kg/m^2)	27 ±6.7	29.5 ±9.5	.128

[a] *P* values were obtained using the Student's t-test for continuous variables and the chi-squared or Fisher's exact test for categorical variables
[b] As defined in the Elixhauser score [150]
[c] all classes combined, with receipt before colonization; selected individual classes also listed

2.2.2 Association of colonization with K. pneumoniae and infection

Of the 1765 patients analyzed, 406 (23%) were identified as colonized (Table 2.2). Of those colonized, 5.2% (n=21) later developed infection with *K. pneumoniae* at an extra-intestinal site, compared to only 1.3% (n=18) of non-colonized patients (unadjusted odds ratio [OR] 4.06; 95% confidence interval [CI] 2.14–7.7; *P* <.0001). In terms of specific sites, colonization was significantly associated with BSI (OR = 5.94, 95% CI 1.73–20.41, *P* =.005), pneumonia (OR = 3.88, 95% CI 1.40–10.77, *P* =. 01) and UTI (OR = 3.39, 95% CI 1.18–9.72, *P* =.024) (Table 2.3). For 20 of 21 colonized patients who became infected, colonization was detected on their initial rectal swab; one patient became positive on their second rectal swab six days later.

Table 2.2: Prior colonization with K. pneumoniae vs. subsequent infection.

	Colonized n (%)	Not colonized n (%)	Total	Odds ratio for infection (95% CI)	P value[a]
Infection	21 (5.2)	18 (1.3)	39		
				4.06 (2.14–7.7)	<.001
No infection	385 (94.8)	1341(98.7)	1726		
Total	406 (23)	1359 (77)	1765		

[a] Fisher's exact test

Table 2.3: Association with prior colonization for each site of infection.

	Colonization Frequency (%)			
Site of Infection	Infected	Not Infected	Odds ratio (95% CI)	P value[a]
Blood	7/11 (64)	399/1754 (23)	5.94 (1.73–20.41)	.005
Respiratory	8/15 (53)	398/1750 (23)	3.88 (1.40–10.77)	.01
Urine	7/14 (50)	399/1751 (23)	3.39 (1.18–9.72)	.024

[a] P values were obtained using Fisher's Exact test

In the final multivariable model, colonization with *K. pneumoniae* had the highest association with infection (all sites) after adjustment for other potential confounders (OR 4.01; 95% CI 2.08–7.73; P <.001; Table 2.4). In addition, fluid and electrolyte disorder, neurologic disorder, and previous hospital admissions within the past 28 days were independently associated with infection. Low baseline platelet levels approached but did not reach significance (P =.058); however, this variable was retained as it significantly improved the performance of the model (P =.046 for the likelihood ratio test) without significantly altering the other variables' estimates. The AUROC demonstrated acceptable performance of the model (AUROC =0.78, 95% CI 0.718–0.842; Figure 2.2) and the Hosmer-Lemeshow test did not indicate poor model fit (P =0.135).

Table 2.4: Multiple logistic regression model of risk factors for infection.

Variable	Odds Ratio	95% CI	*P* value
Colonized	4.01	2.08–7.73	<.001
Fluid & Electrolyte Disorder[a]	2.37	1.22–4.59	.011
Neurologic Disorder[a]	3.31	1.28–8.54	.013
Prior Admit (28 Days)	2.16	1.04–4.48	.038
Baseline Platelet count/100 units ($\times10^3$/µL)[b]	0.73	0.53–1.01	.058

[a] As defined by the Elixhauser score [150]
[b] For every 100-unit increase in baseline platelet count, the odds of infection was 0.73-fold lower
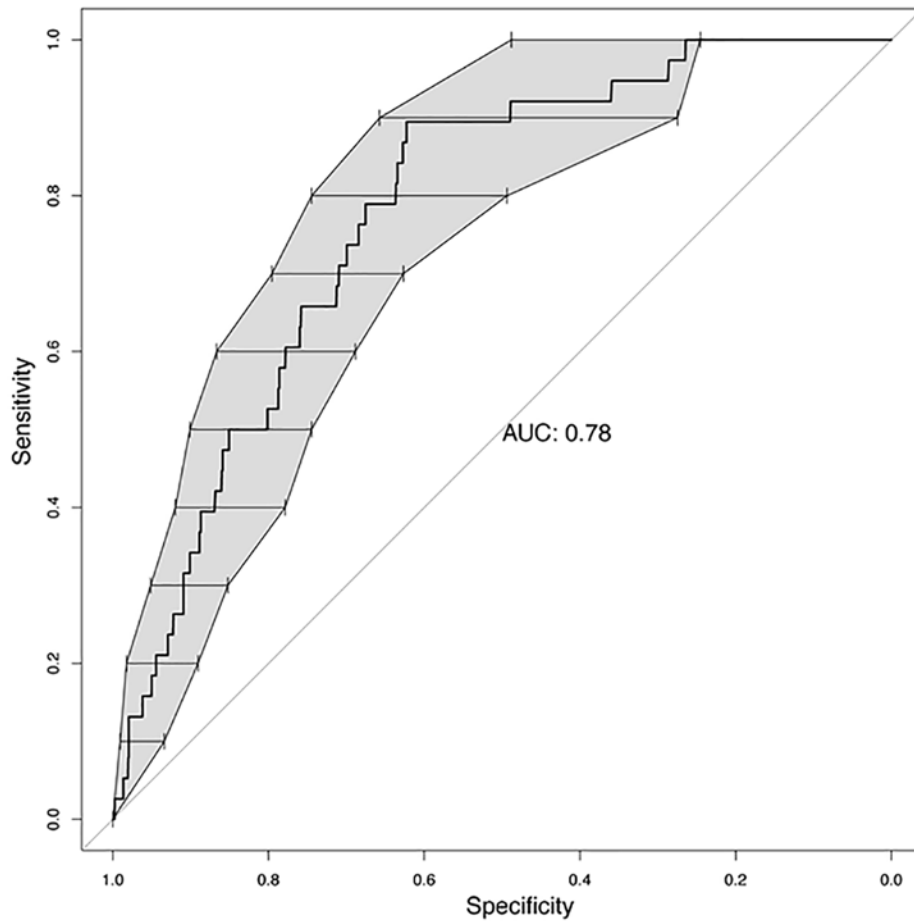
Figure 2.2: Receiver Operator Characteristic curve for multivariable model of risk factors for clinical infection.

Multiple logistic regression on *K. pneumoniae* infection was used to generate a predictive model using five patient variables (Table 2.4). Bars and shaded area of ROC curve represent bootstrapped 95% confidence intervals (10,000 replicates) for specificity at each level of sensitivity (AUC=0.78, 95% CI 0.72–0.84).

2.2.3 Concordance of colonizing and infecting isolate pairs

based on molecular strain typing

To determine if patients become infected with strains they are previously colonized with, we first screened for genetic differences using *wzi* gene sequencing. Preliminary results from 17 patients' isolates indicated that *wzi* sequencing had similar discriminatory power to MLST, with both distinguishing 16 sequence types among 20 isolates (Figure 2.3). In order to assess the diversity of strains patients were colonized with, we determined the *wzi* types of colonizing isolates from 40 patients. Sixteen of these colonized patients had subsequent positive clinical cultures and met case definitions for BSI, pneumonia, or UTI; twenty-four patients did not. A total of 110 rectal swab isolates were tested; up to three isolates obtained from each patient. From these 40 patients, only 8 patients had two *wzi* types detected and no patients had three types within a sample. Despite the homogeneity within individual patients, 43 different *wzi* types were identified between these 40 patients, suggesting high genetic diversity of colonizing *K. pneumoniae* in this patient population (Figure 2.4).
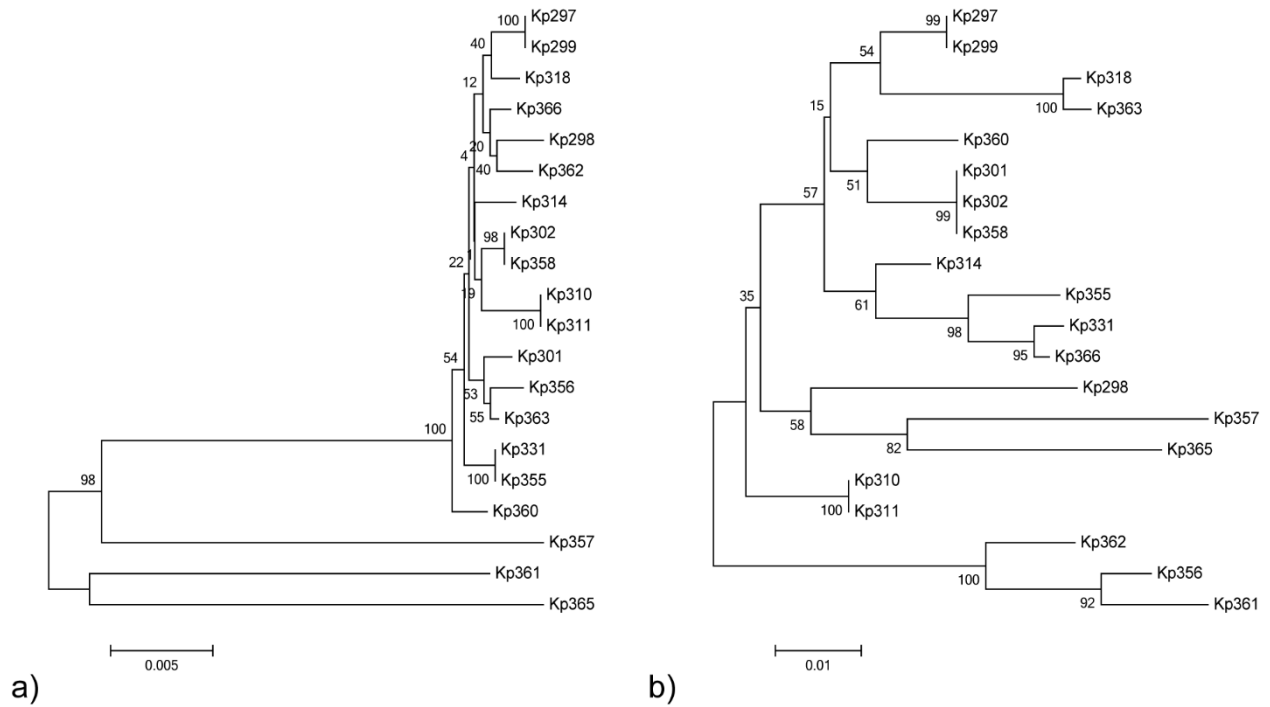
Figure 2.3: wzi sequencing method has similar discriminatory power to MLST.

Phylogenetic trees based on MLST results (a) and *wzi* sequencing results (b) are shown, each distinguishing 16 sequence types among 20 *K. pneumoniae* isolates. The scale bar represents the amount of genetic change. The numbers next to each node are the percentage of iterations that recover the same node.
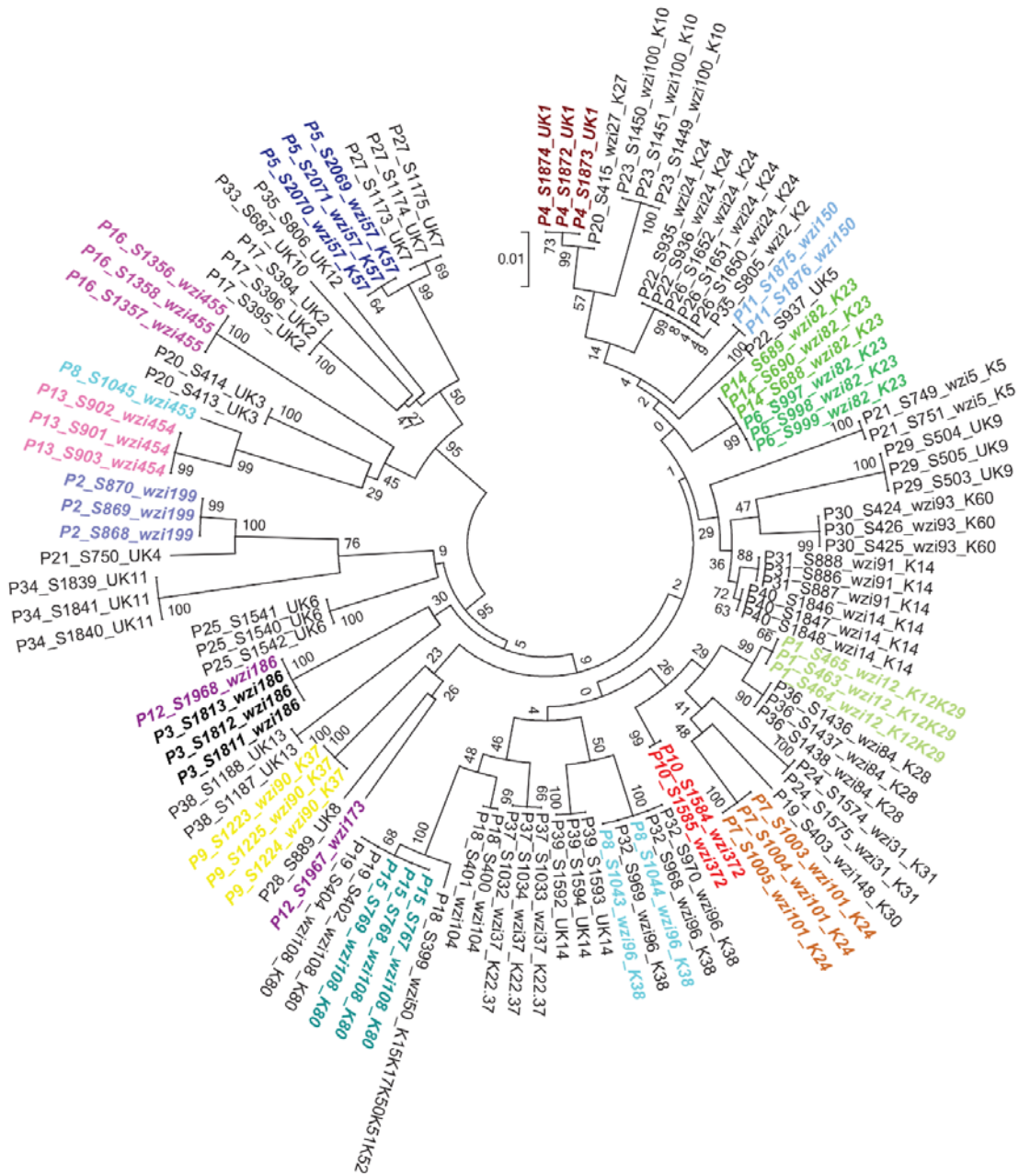
Figure 2.4: Phylogenetic tree for wzi sequence of patient rectal swab isolates.

Unique patients (P) are numbered (P1-P40). Rectal swab (S) isolate is indicated after patient number and immediately before isolate number (ex. S463 is stool isolate number 463). Isolate *wzi* type is indicated, and novel alleles are designated as unknown (UK). 110 rectal swab isolates from 40 unique patients were tested for strain types using *wzi* gene sequencing. 43 different *wzi* types of strains were identified. Rectal swab isolates for patients with *K. pneumoniae* colonization prior to infection were all included in the analysis (P1-P16, colored font). The scale bar represents the amount of genetic change; 0.01 equals 1 change per 100 nucleotide sites. The numbers next to each node are the percentage of iterations that recover the same node.

Of 21 colonized patients who developed infection, 16 sets of colonizing and infecting isolates were available for analysis. Two out of five patients with BSI (40%) had concordant pairs based on *wzi* sequencing of blood and rectal swab isolates. Respiratory and rectal swab isolates from patients with pneumonia (n = 7) demonstrated perfect concordance (7/7; Figure 2.5). Although two patients with pneumonia were each colonized with 2 different *wzi* types, one was concordant with each patient's respiratory isolate (stool isolate 1043 matched respiratory isolates 733 and 734, and stool isolate 1967 matched respiratory isolate 2005; Figure 2.5). Analysis of urine and rectal swab isolates from patients with UTI (n = 4) also demonstrated perfect concordance (4/4).

Despite high concordance of colonizing and infecting isolate pairs by *wzi* sequencing, using a single gene typing method may not be sufficient to determine true isolate concordance. To confirm that isolate pairs are the same strain, we performed whole genome sequencing (WGS) and determined isolate sequence type (ST) by both 7-gene multilocus sequence typing (MLST) and 634-gene core genome multilocus sequence typing (cgMLST). We analyzed 13 preliminarily concordant pairs and one un-matched pair as a discordant control (pair 463/1946). MLST analysis showed perfect agreement with *wzi* sequence typing results in identifying 13 concordant pairs (Figure 2.6). Two novel STs were identified, ST2359 and ST2360. For patients where one of two colonizing isolates matched the infecting isolate, MLST distinguished between the colonizing isolates and indicated that only one was concordant with the infecting isolate.

a)

b)

c)

41
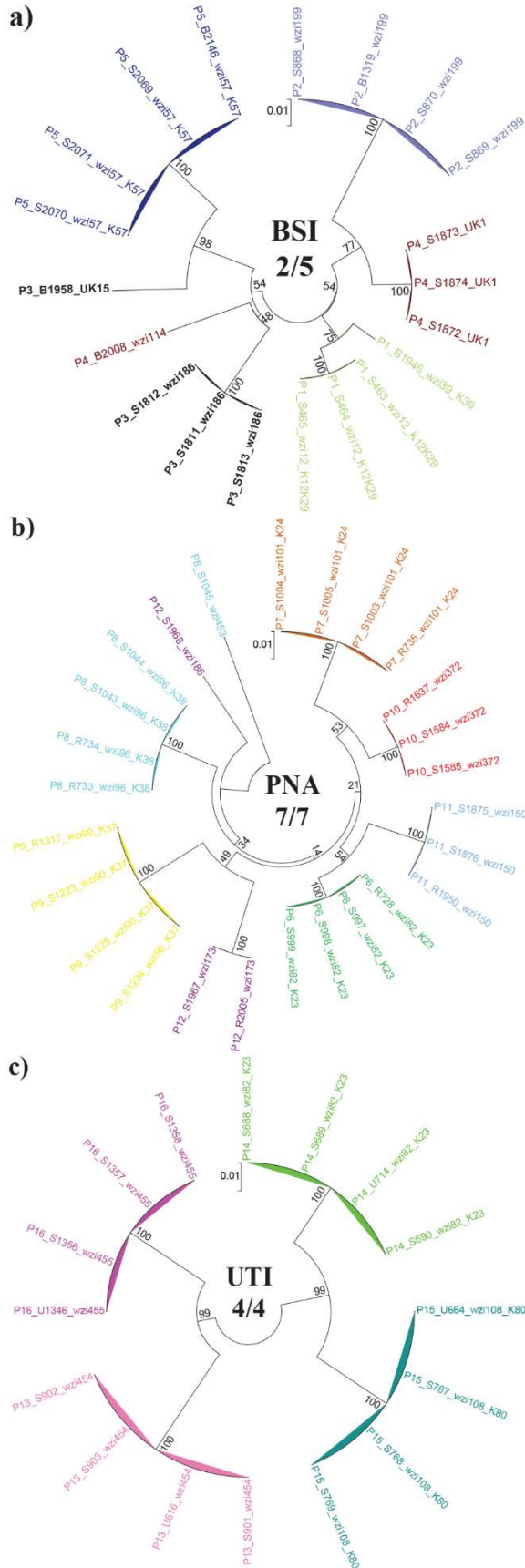
Figure 2.5: Phylogenetic trees of rectal swab and infecting isolates based on wzi gene sequencing.

Phylogenetic trees, built using the neighbor joining method, for infecting and colonizing isolates from patients with (a) bloodstream infection (BSI), (b) pneumonia (PNA) and (c) urinary tract infection (UTI) are shown along with the fraction of patients with a concordant colonizing-infecting isolate pair. Unique patients are indicated by different colors and labeled as patient number_ isolate number_wzi allele (unless a novel allele) K-type (if known), where the isolate number prefixes indicate rectal swab (S), blood (B), respiratory (R) or urine (U). The scale bar represents the amount of genetic change. The numbers next to each node are the percentage of iterations that recover the same node.

| Isolate	Patient	Source	ST	wzi	Associated K-type
688	14	stool	280	82	K23
714	14	urine	280	82	K23
1584	10	stool	1322	372	unknown
1637	10	respiratory tract	1322	372	unknown
1319	2	blood	1691	199	unknown
868	2	stool	1691	199	unknown
1223	9	stool	643	90	K37
1317	9	respiratory tract	643	90	K37
1043	8	stool	34	96	K38
733	8	respiratory tract	34	96	K38
734	8	respiratory tract	34	96	K38
664	15	urine	37	108	K80
767	15	stool	37	108	K80
1003	7	stool	45	101	K24
735	7	respiratory tract	45	101	K24
1967	12	stool	433	173	unknown
2005	12	respiratory tract	433	173	unknown
728	6	respiratory tract	17	82	K23
997	6	stool	17	82	K23
1968	12	stool	20	186	unknown
463	1	stool	2359	12	K12K29
1875	11	stool	152	150	unknown
1950	11	respiratory tract	152	150	unknown
1946	1	blood	327	39	K39
616	13	urine	1562	454	unknown
901	13	stool	1562	454	unknown
2069	5	stool	208	57	K57
2146	5	blood	208	57	K57
1346	16	urine	2360	455	unknown
1356	16	stool	2360	455	unknown
1045	8	stool	935	453	unknown

Figure 2.6: Core genome similarity between infecting and colonizing strains within patients.

Patients who had concordant colonizing and infecting isolates based on *wzi* sequencing were further analyzed by WGS and core genome MLST and represented by a UPGMA dendrogram along with isolate number, MLST type (ST), *wzi* type, and body site of culture (source). Each color represents an individual patient. Isolates 463 and 1946 (Patient 1) were discordant by *wzi* and included as a control.

The cgMLST method provides a more discriminatory approach to defining concordance since it is based on allelic similarities of 634 *K. pneumoniae* genes [66]. All concordant pairs, based on *wzi* and MLST analyses, also clustered together based on cgMLST (Figure 2.6). The only pairs that did not group together were our discordant control, and the discordant colonizing isolates (1045 and 1968) from patients with another colonizing strain that matched the pneumonia isolates. These isolates were also discordant by *wzi* and MLST analyses. To measure the strength of cgMLST concordance of colonizing-infecting pairs within patients compared to between patients, a Minimum Spanning Tree (Figure 2.7) was also generated based on the cgMLST data. For example, there are two allelic differences between stool isolate 1223 and pneumonia isolate 1317 in the same patient while there are 189 allelic differences between these isolates and their closest neighbors from a different patient (pair 1319/868). Overall, there was an average of 2 allelic mismatches between concordant pairs (range: 0 – 7), and 449 allelic mismatches between patients (range: 189 - 629). Taken together the *wzi*, MLST and cgMLST data indicate that 100% of urinary and pneumonia isolates tested corresponded to the previously colonizing strain of *K. pneumoniae*.
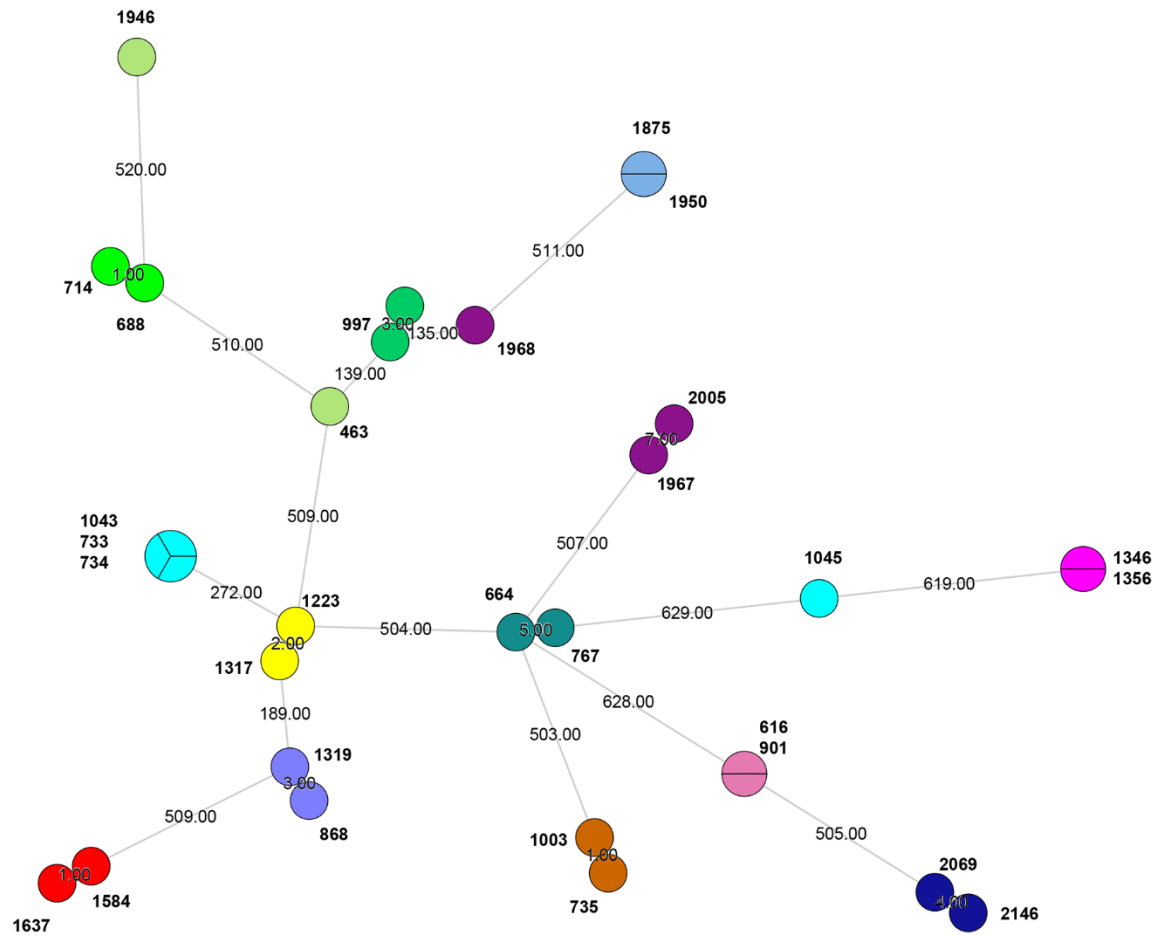
Figure 2.7: Concordant colonizing-infecting isolate pairs show high core genome allelic similarity.

Minimum Spanning Tree based on the core genome MLST profiles, onto which the number of allelic differences between isolates is indicated along the links. Note that the numbers are not additive and that the tree should not be interpreted as depicting phylogenetic relationships. Isolate names are in bold. Each color represents an individual patient. Isolates 463 and 1946 (Patient 1) were discordant by *wzi* and included as a control. cgMLST profiles of isolates within single circle were totally identical.

2.2.4 Categorical agreement of antimicrobial susceptibility

of colonizing and infecting isolate pairs

To measure antibiotic susceptibility agreement between concordant and discordant colonizing-infecting isolate pairs, we tested seventeen antimicrobials active against Gram-negative bacteria and measured categorical agreement of susceptible (S), intermediate (I), or resistant (R) based on MIC breakpoints (Table 2.5) [151]. Categorical agreement was greater than 90% for all antimicrobials tested. However, this isolate collection had a low prevalence of antibiotic resistance. In 12 of 16 patients, both isolates were susceptible to all antimicrobials tested, including the 3 patients with discordant pairs based on sequence type. Two patients with concordant strain types had initially discordant susceptibility results. Only one discrepancy was reproducible by broth microdilution (trimethoprim/sulfamethoxazole in 1967/2005). The remaining 11 patients with concordant isolate pairs by sequence type had identical susceptibility patterns including one ESBL *K. pneumoniae* (767/664).

Table 2.5: Categorical agreement of colonizing-infecting isolate pairs in case patients.

	Categorical result of isolate pair by infection site, patient number, and *wzi* Match (Y/N)[a]																Categorical agreement (%)
	BSI					Pneumonia							UTI				
Antibiotic	1	2	3	4	5	6	7[b]	8	9	10	11	12[c]	13	14	15	16	
	N	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	
Amp/Sul	S	I	S	S	S	S	R	S	S	S	S	R/S	S	S	R	S	94
Pip/Tazo	S	S	S	S	S	S	I/R	S	S	S	S	S	S	S	R	S	94
Cefazolin	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	100
Ceftazidime	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	100
Ceftriaxone	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	100
Cefepime	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	100
Aztreonam	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	100
Ertapenem	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	100
Imipenem	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	100
Meropenem	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	100
Amikacin	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	100
Gentamicin	S	S	S	S	S	S	S	S	S	S	S	I/S	S	S	S	S	94
Tobramycin	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	100
Ciprofloxacin	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	100
Levofloxacin	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	100
Tigecycline	S	R	S	S	S	S	S/R	S	S	S	S	S	S	S	S	S	94
TMP/Sulfa	S	S	S	S	S	S	S	S	S	S	S	R/S	S	S	R	S	94

[a] Combined categorical results of Susceptible (S), Intermediate (I), or Resistant (R) for each pair are shown; discrepant results confirmed by broth microdilution are shown as the result for colonizing/infecting isolate and highlighted in red; discrepant results that could not be confirmed are in yellow; concordant non-susceptible results are in green.

[b] Patient 7 isolates were S/S for both Pip/Tazo and Tigecycline by broth microdilution

[c] Patient 12 isolates were I/S for Ampicillin/Sulbactam and S/S for Gentamicin by broth microdilution

2.3 Discussion

2.3.1 Summary of findings

The objective of this study was to examine the association between *K. pneumoniae* rectal colonization and subsequent extra-intestinal *K. pneumoniae* infection. Based on data from 1765 intensive care and hematology/oncology patients, we found that approximately 1 in 4 patients were rectal carriers of *K. pneumoniae*. We also observed a significant association between rectal *K. pneumoniae* colonization and subsequent infection, even after adjusting for patient variables. Furthermore, there was high concordance among colonizing isolates and subsequent infecting isolates, particularly for pneumonia and UTI, as measured by *wzi*, MLST, and cgMLST analyses. Taken together, these results implicate colonization as a critical step in the pathogenesis of hospital-acquired infections. These results also identify a possible window for intervention to decolonize patients or characterize their colonizing strain in order to predict risk of disease and inform empiric therapy if infection develops.

2.3.2 Strengths and Limitations

Our study has several strengths. Firstly, whereas previous studies focused on drug-resistant *K. pneumoniae* or strains involved in outbreaks [152-154], we tested all isolates during a three-month collection period across multiple wards and units in the hospital. This approach provided comprehensive information on *K. pneumoniae* colonization in the hospital setting and minimized potential selection bias. The large sample size (n=1765) provided sufficient power to examine the relationship between colonization and patients that met strict case definitions of infection. Secondly, we used *wzi* gene sequencing to rapidly screen for genetic differences between *K. pneumoniae* isolates [146]. We then confirmed concordant pairs using WGS-based MLST and cgMLST. In a hospital laboratory setting, *wzi* could be used to screen for a potential outbreak strain

as a triage step before more costly WGS [146]. Lastly, molecular strain typing indicates high *K. pneumoniae* strain diversity in our study population. Molecular epidemiology studies show clonal spread of carbapenem-resistant *K. pneumoniae* in the United States [154, 155]. If a dominant clone existed in our population, it would obscure the true association between colonization and subsequent infection. In our diverse setting, the high concordance between colonizing and infecting strains suggests a robust pathogenic mechanism in which patients become infected by their colonizing strain.

This study also has some limitations. First, 35 patients with a positive clinical culture had unknown colonization status prior to the culture date and, thus, were excluded. By excluding this subset of data, we potentially lost cases of infection, and cannot predict in which direction it biases the results. Although we collected three rectal swab isolates per patient, most extra-intestinal *K. pneumoniae* provided by the clinical lab represented one isolate per site. It is possible that multiple strains may be present at an extra-intestinal site but only one isolate was sampled. For *wzi* sequencing, rectal swab isolates from fewer than 10% of colonized patients were tested. Given 40 unique patients with 43 unique strains, almost every patient was colonized with a different strain. This high level of diversity is unlikely to be maintained in the larger sample set. A limitation of the susceptibility data was that the majority of isolates had no detectable acquired resistance. With low diversity of resistance phenotypes, we were unable to rigorously test the agreement of susceptibility testing between colonizing isolates and subsequent infecting isolates in the same patient. Future studies should determine if high categorical agreement holds in a larger, more resistant colonized-infected patient population.

2.3.3 Conclusions

We conclude, based on three distinct methods, that there is high concordance between colonizing and infecting isolates, particularly for pneumonia and UTI. The discordance in bloodstream infections could be due to exogenous sources of *K. pneumoniae*, such as insertion of an intravenous catheter or a healthcare worker's hands. The perfect concordance for UTI is consistent with the paradigm for *E. coli* UTI, where fecal colonizing strains contaminate the urogenital tract [156]. However, the perfect concordance between rectal isolates and pneumonia isolates was striking. This may indicate simultaneous colonization of the respiratory tract at the time of intestinal acquisition of the strain. This strong concordance suggests that infection prevention approaches or guidance of empiric therapy based on detection and characterization of colonizing *K. pneumoniae* isolates is feasible.

In addition to prior colonization, prior admission, low baseline platelets, and comorbidities of neurologic and fluid and electrolyte disorders were highly predictive of *K. pneumoniae* infection in a multivariable model. The association between healthcare exposure and subsequent infection is plausible, even after adjustment for colonization, since it likely indicates overall poor health status, itself a risk factor for infection. This is likely also true of the other comorbidities included in the final model. The components of this model include information readily available at admission as part of routine testing and chart review. If validated in an independent cohort, rectal screening paired with these variables could rapidly predict risk of *K. pneumoniae* infection.

The finding that patients often become infected with their colonizing strain has strong implications for both infection control and patient care interventions. A recent study in a long-term acute care hospital (LTACH) determined that interventions based on screening for carbapenemase-producing *K. pneumoniae* (KPC) decreased both the colonization rate of patients as well as the

rate of clinical infections [157]. Moreover, characterization of colonizing strains could inform treatment decisions. Understanding the pathogenic mechanisms of progression from *K. pneumoniae* colonization to disease could enable novel diagnostics and therapeutics to prevent and rapidly treat these common nosocomial infections.

2.4 Experimental procedures

Patient population and setting. The study was conducted at the University of Michigan Health System (UMHS), a tertiary care hospital with more than 1,000 beds, in Ann Arbor, Michigan. Approval for this study was granted by the Institutional Review Board of the University of Michigan Medical School. During a three-month period from July 30 to October 31, 2014, rectal swabs from 1800 adult ($\geq$18 years) patients from the intensive care unit (ICU) or hematology/oncology wards were screened for *K. pneumoniae*. Concurrently, extra-intestinal *K. pneumoniae* isolates were obtained from the clinical microbiology lab. 1765 patients had either a rectal swab performed prior to a positive clinical culture or a rectal swab and no positive clinical culture, and were included for analysis of the association between colonization and subsequent infection. Patient demographic characteristics and clinical information was obtained through the electronic medical record (EMR).

Bacterial identification and growth conditions. Rectal swabs were collected during the course of clinical care (upon unit admission, weekly, and at discharge) and were transported and stored in the ESwab Transport System (BD, Franklin Lakes, NJ) at room temperature. Within 24 hours of receipt, 1µl of inoculated ESwab media was plated to MacConkey agar (Remel, Lenexa, KS) streaked for quantification and incubated for 18–24 hours at 35ºC. To ensure collection of the

dominant clone in each sample, three mucoid lactose fermenting (MLF) colonies were isolated as potential *K. pneumoniae* and subcultured onto blood agar plates (BAP) (Remel, Lenexa, KS) [158]. If fewer than three MLF colonies were present in a particular sample, all were subcultured. Bacterial identification was performed using matrix-assisted laser desorption/ionization time of flight (MALDI-TOF). Isolates were stored at −80ºC in Luria-Bertani (LB) Broth containing 20% glycerol, and were grown on either BAP or LB plates at 30ºC overnight unless otherwise indicated.

Definitions. In patients without *K. pneumoniae* infection, *K. pneumoniae* colonization was defined as a positive rectal swab culture for *K. pneumoniae* at any point during the hospital admission. For patients with *K. pneumoniae* infection, colonization was only considered positive if detected prior to the date of the documented infection. If a patient was identified as colonized on more than one date of collection, we used the first positive sample before infection for *wzi* sequencing, MLST, cgMLST, and antimicrobial susceptibility testing. Patient EMRs were reviewed for any positive culture for *K. pneumoniae* within 90 days of rectal swab culture. Bloodstream infection was defined as any positive blood culture for *K. pneumoniae*. Pneumonia was defined based on a positive *K. pneumoniae* respiratory culture and other Infectious Diseases Society of America (IDSA) diagnostic criteria [159]. Patients with positive urine cultures were identified as cases based on the CDC National Healthcare Safety Network (NHSN) urinary tract infection (UTI) case definitions [160]. Patients not meeting a case definition of infection within 90 days of a rectal swab culture were considered uninfected. Comorbid disease definitions were extracted from the EMR based on ICD-9 and DRG codes as specified in the Elixhauser index [150]. To ensure that we were capturing antibiotic exposure as a risk factor and not an outcome, this variable was constructed to ensure that exposure was present prior to colonization/infection.

Exposure to an antibiotic was defined as true for patients without *K. pneumoniae* infection or colonization if administered at any point during the admission. For those patients with colonization, the antibiotic exposure variable was true if antibiotics were started at least 48 hours prior to the detected colonization. For those patients with infection but no preceding colonization, antibiotic exposure was positive if started at least 48 hours prior to documented infection.

wzi gene sequencing. DNA preparation and polymerase chain reaction (PCR) amplification were performed as described by Brisse *et al.* [146] with the following modifications: PCR products were diluted 1:20 in sterile water and sequencing primers (wzi_for2 and wzi_rev) were diluted to 1pM/µL prior to submission for Sanger sequencing. Forward and reverse chromatograms were assembled using Lasergene SeqMan (DNASTAR, Madison, Wisconsin). Complete alignment was done using ClustalX 2.1 [161]. Initial phylogenetic trees were constructed using MEGA 6 [162] based on the neighbor-joining method (500 bootstrap replicates) and Jukes-Cantor distance. The *wzi* library obtained from Brisse *et al.* [146] was used as a reference in the current analysis.

Whole genome sequencing and assembly. Bacterial genomic DNA (gDNA) was purified using the UltraClean® Microbial DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, California). Purified gDNA was sent to the University of Michigan DNA Sequencing Core, where it was sheared (320 bp) and prepared as a multiplex library with unique barcodes for each sample. Whole genome sequencing was performed using the HiSeq 4000 Sequencing System (Illumina, San Diego, California). Reads were preprocessed for each sample by trimming bases at both ends if the quality score is below 10, using Trimmomatic (v0.32) [163], removing read duplicates using FastUniq [164], and performing error-correction using SOAPec (v2.01) [165]. Preprocessed reads

were assembled using VelvetOptimiser (v2.2.5). In this process, the reads were assembled by Velvet [166] using stepwise K-mer size at step=2, from 51 to 149 (for PE samples), or from 25 to 51 (for SE samples). The assembly with the largest N50 value was used for subsequent analysis.

Multilocus sequence typing (MLST) and core genome MLST. The gene sequences corresponding to the international MLST scheme of Institut Pasteur [145] were extracted from the genomic assemblies using the BIGSdb platform [167] using the BLASTN algorithm and their corresponding allelic number was defined by comparison with the reference nomenclature database (http://bigsdb.pasteur.fr/klebsiella). Core genome MLST (cgMLST) was performed in the same way using the strict core genome MLST scheme defined by Bialek-Davenet et al [66]. Novel alleles and MLST profiles were submitted to the nomenclature database. MLST profiles were compared using the categorical mismatch method using BioNumerics version 6.6 (Applied-Maths, Sint-Martens Latem, Belgium). Uncalled alleles were not considered as mismatches in pairwise profile comparisons.

Antimicrobial susceptibility testing. Antimicrobial susceptibility testing was performed on the VITEK 2 (bioMérieux, Marcy-l'Étoile, France) automated system, using AST-GN82 cards loaded per the manufacturer's instructions. Isolates were grown on BAP at 37°C overnight. Colonizing and infecting isolates from the same patient were tested in the same batch. Susceptibility testing was performed on one rectal isolate for each patient. ESBL phenotypes were determined with the use of the VITEK 2 Advanced Expert System (AES). Pairs with discrepant results for any antibiotic were tested by Sensititre broth microdilution (Trek Diagnostics Systems, Oakwood Village, OH).

Statistical analysis. Initial tests included examination of variables for out-of-range values, measures of central tendency/spread for continuous variables, and frequencies for categorical variables. These initial analyses assisted in constructing variables including transformations (for example, length of stay was log transformed prior to analysis given the non-normal distribution). Initial bivariable analyses were conducted with the Student's t-test for continuous variables and the chi-squared or Fisher's exact test for categorical variables. Based on these initial analyses, variables with a *P* <.2 on bivariable tests were eligible for inclusion in the final multiple logistic regression model. This final model of *K. pneumoniae* infection was constructed via backwards elimination using a likelihood ratio test for variable retention, with a cutoff α of 0.05. Interactions between variables in the final model were tested and included if significant. Additional model regression assessments included the Hosmer-Lemeshow test for goodness of fit and calculation of the area under the receiver-operator characteristic (AUROC) curve. For interpretation of the results, a *P* value of 0.05 was considered statistically significant for all analyses. The analyses were performed using SAS 9.3 (SAS Institute, Cary, North Carolina) and R 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria).

Data availability. Whole genome sequencing files have been deposited in the NCBI Sequence Read Archive (PRJNA341404) under accession numbers SAMN05722982, SAMN05722983, SAMN05722984, SAMN05722985, SAMN05722986, SAMN05722987, SAMN05722988, SAMN05722989, SAMN05722990, SAMN05722991, SAMN05722992, SAMN05722993, SAMN05722994, SAMN05722995, SAMN05722996, SAMN05722997, SAMN05722998, SAMN05722999, SAMN05723000, SAMN05723001, SAMN05723002,

SAMN05723003, SAMN05723004, SAMN05723005, SAMN05723006, SAMN05723007, SAMN05723008, SAMN05723009, SAMN05723010, SAMN05723011, SAMN05723012. Sequences of *wzi* alleles used for strain comparison are included in supplemental materials. All other data is available upon request.

2.5 Notes

RMM, JC, KR, PNM, and MAB conceived and designed the study. RMM acquired patient isolates and susceptibility data. JC acquired *wzi* sequence data. JC and KR acquired patient demographic data. WW assembled the sequenced genomes. SB and VP acquired core genome MLST data. RMM, JC, SB, VP, KR, and LZ analyzed the data. RMM and JC drafted the manuscript. SB, VP, WW, LZ, PNM, KR and MAB critically revised the manuscript. MAB supervised the study.

CHAPTER III

Identification of Pathogenicity-Associated Loci in

Klebsiella pneumoniae Clinical Isolates

Summary

Klebsiella pneumoniae, a leading cause of healthcare-associated infections (HAI) in the U.S., commonly colonizes hospitalized patients. A subset of colonized patients goes on to develop extra-intestinal infections such as urinary tract infections, bacteremia, and pneumonia. Though several virulence genes have been identified to date, the bacterial factors that determine whether an isolate causes disease or remains a colonizer are poorly understood. We hypothesize that genes in the bacterial accessory genome, which are present in a subset of strains, promote extra-intestinal infection in colonized patients. To identify host factors and bacterial genes associated with *K. pneumoniae* infection, we performed a case-control study comparing infected patients meeting standard case definitions (n=38) and asymptomatic colonized patients (n=76), matched by age, sex, and sample collection date. A conditional logit model identified patient factors significantly associated with infection (area under the receiver operator characteristic curve [AUROC] = 0.88). We then measured the frequency of gene loci among colonizing and disease-causing strains by read mapping from whole genome sequencing to a *K. pneumoniae* pan-genome. Using this Pathogenicity-Associated Locus Sequencing (PAL-Seq) method, we identified five gene loci associated with infection after adjustment for patient factors (P = .014–.03), including sugar utilization and tellurite resistance loci. Incorporation of three of these gene loci improved the

57

performance of our model (AUROC = 0.95; P = .011). A permease mutant in the sugar utilization locus was unable to use psicose for growth. Our results suggest that genes in the accessory genome of *K. pneumoniae* are associated with HAIs and could provide a basis for novel predictive diagnostic approaches to identify colonized patients at risk of subsequent *K. pneumoniae* infection.

3.1 Introduction

Klebsiella pneumoniae is a Gram-negative bacterial pathogen and a leading cause of hospital-acquired infections (HAIs) in the United States [33]. In addition, hypervirulent community-acquired (CA) strains have emerged that infect seemingly healthy individuals. These strains are particularly worrisome since they cause severe diseases such as pyogenic liver abscess (PLA), meningitis, and endophthalmitis [107-109]. Originating in the Asian Pacific Rim, these strains are now being reported globally including in the United States [110, 111, 113]. Similarly, antibiotic resistance has become an emerging problem with *K. pneumoniae.* In fact, in 2013 the Centers for Disease Control and Prevention (CDC) designated carbapenem-resistant *Enterobacteriaceae* (CRE), of which *Klebsiella* species are the most prevalent, as an urgent threat [63]. Further complicating the treatment of these infections is that approximately 20% of isolates identified as *K. pneumoniae* are actually not *pneumoniae* species (Kp group I), but are either *K. quasipneumoniae* (Kp group II) or *K. variicola* (Kp group III) [132-135]. About 50% of *K. quasipneumoniae* are extended-spectrum beta-lactamase (ESBL) producers [30], and *K. variicola* has been associated with higher mortality in bloodstream infections [135]. Therefore, correct identification of Kp groups may impact patient care. Together, *K. pneumoniae* and closely related species represent a significant healthcare burden.

K. pneumoniae commonly colonizes the nasopharynx and gastrointestinal tract of hospitalized patients [17]. We recently demonstrated that intestinal colonization with *K. pneumoniae* is significantly and independently associated with infection with *K. pneumoniae*, and that patients are frequently infected with the same strain they are colonized with [168]. This association was independently confirmed [18] and suggests colonization as a reservoir for infection, and thus as an important step in the pathogenesis of *K. pneumoniae* infection. Infections caused by these bacteria include pneumonia, septicemia, urinary tract infections, and wound infections. In an effort to understand the broad pathogenesis determinants for *K. pneumoniae*, several virulence genes have been identified to date including evasion of complement, capsule, adhesins, and use of siderophores for iron acquisition [17]. However, a subset of colonized patients do not proceed to disease, instead remaining asymptomatically colonized. The bacterial factors that determine whether an isolate causes disease or remains a colonizer are poorly understood.

With the increasing capabilities of whole genomes sequencing (WGS) and comparative genomics we are better able to effectively identify virulence factors in bacteria, including *K. pneumoniae* [30]. These techniques can delineate the bacterial accessory genome, which varies between isolates, from the conserved core genome and have identified genes that aid in virulence in uropathogenic *Escherichia coli* (UPEC) [169] and *K. pneumoniae* [30]. This opens the door for genome wide association studies (GWAS) to compare strains that do or do not cause disease. GWAS has been performed successfully with human genomes to identify genetic variants significantly associated with a particular disease phenotype. Recently, bacterial GWAS has identified associations between bacterial polymorphisms or novel loci and specific bacterial phenotypes [170-173].

In this study, we combined bacterial GWAS and clinical modeling to identify patient and bacterial factors associated with *K. pneumoniae* infection compared to asymptomatic colonization. We hypothesized that genes in the accessory genome of *K. pneumoniae* promote extra-intestinal infection in colonized patients. To test this hypothesis, we performed a case-control study with 1:2 matching and analyzed the gene frequency differences between bacterial isolates using a novel comparative genomics technique we termed Pathogenicity-Associated Locus Sequencing (PAL-Seq). We then assessed whether candidate virulence genes were independent predictors of infection, whether they improved prediction of infection when incorporated into a clinical model, and whether they had a distinct phenotype in a murine model of pneumonia.

3.2 Results

3.2.1 Study design and patient demographics

During a three-month period from July 30, 2014 to October 31, 2014, over 2000 patients were screened for *K. pneumoniae* colonization by rectal swab culture. Simultaneously, patients were screened for extra-intestinal infection with *K. pneumoniae* based on positive clinical cultures from the blood or respiratory tract. We identified 38 patients meeting case definitions for extra-intestinal infection, either bacteremia or pneumonia [159]. One patient had two isolates separated in time that met the case definition. Since we did not design the study for repeated measures and did not want to bias towards this particular patient's characteristics, only the first invasive infection from this patient was counted for the case-control analysis. However, all isolates were included for sequencing. Each case patient (n=38) was matched to two asymptomatically colonized controls (n=76) based on age range (± 10 years), gender, and sample collection date (± 3 weeks).

3.2.2 Patient variables are associated with invasive disease

To determine if there were any patient factors that were associated with clinical infection, we performed a bivariable analysis to identify significant differences in clinical characteristics between cases (n=36) and controls (n=72) (Table 3.1). Some patients were removed from this analysis due to missing clinical data.

In bivariable analysis comparing cases to controls, only white race showed a significant difference (66.7% vs 84.7; P = .033), with an inverse association with infection. Though other variables did not show a significant association, some trends toward association were seen. To further assess this, we built a multivariable using backwards elimination on all covariates where the unadjusted analysis resulted in a P <.2 (white race, baseline white blood cell count, minimum glucose, baseline body mass index, central venous catheter, fluid & electrolyte disorders, peripheral vascular disease, other neurological disorders, and liver disease). The final model included six patient variables that independently associated with *K. pneumoniae* infection in this cohort (Table 3.2), based on an Area under the Receiver Operator Characteristic Curve (AUROC) of 0.88 (Figure 3.1). Interactions among these variables in the final model were tested and none was significant. Fluid & electrolyte disorders and minimum serum glucose were positively associated with infection, whereas baseline BMI, white race, and peripheral vascular disease were inversely association with infection. Presence of a central venous catheter was not an independent predictor of infection (P =.075 when adjusted for covariates in the model).

Table 3.1: Clinical characteristics of cases and controls.

Variable n (%) or mean ±SD	CASE Bloodstream Infection or Pneumonia (n=36)	CONTROL Asymptomatic Rectal Colonization (N=72)	*P* Value (logit)
Age, yrs.	61.5 ±18.5	61.6 ±17	.857
Female	18 (50)	36 (50)	>.99
White Race	24 (66.7)	61 (84.7)	.033
Admitted from Nursing Home	1 (2.8)	1 (1.4)	.624
White Blood Cell Count Baseline	10 ±6.2	16.5 ±28.4	.138
Hemoglobin Baseline	10.3 ±2.2	10.7 ±2.3	.354
Platelet Baseline	177.8 ±90.8	185.1 ±112.5	.798
Creatinine Baseline	1.2 ±1.6	1.4 ±1.7	.599
Albumin Baseline	3.4 ±0.57	3.3 ±0.64	.615
Glucose Baseline	137.9 ±59.3	131.2 ±55.7	.572
Glucose Minimum	93.3 ±21.5	86.9 ±21.4	.181
Glucose Maximum	209.8 ±101.2	193.1 ±85.3	.337
Body Mass Index Baseline	25.2 ±8	30.5 ±8.5	.134
Central Venous Catheter	13 (36.1)	36 (50)	.17
Fluid & Electrolyte Disorders	21 (58.3)	32 (44.4)	.193
Peripheral Vascular Disease	1 (2.8)	8 (11.1)	.191
Other Neurological Disorders	4 (11.1)	1 (1.4)	.063
Pulmonary Disorders	5 (13.9)	8 (11.1)	.684
Diabetes without Chronic Complications	8 (22.2)	11 (15.3)	.506
Renal Failure	5 (13.9)	13 (18.1)	.572
Liver Disease	1 (2.8)	8 (11.1)	.166
Lymphoma	2 (5.6)	4 (5.6)	>.99
Metastatic Cancer	3 (8.3)	3 (4.2)	.355
Solid Tumor Without Metastasis	9 (25)	18 (25)	>.99
Alcohol Abuse	0 (0)	5 (6.9)	>.99
Anemia Deficiency	10 (27.8)	16 (22.2)	.538

Table 3.2: Multivariable model for clinical infection with K. pneumoniae.

Patient Variable	Odds Ratio	95% CI	*P* Value
Fluid & Electrolyte Disorders	5.1	1.29-20.17	.02
Minimum Serum Glucose (mg/dL)	1.04	1.01-1.08	.009
Body Mass Index Baseline (kg/m^2)	0.86	0.77-0.97	.011
White Race	0.13	0.03-0.59	.008
Peripheral Vascular Disease	0.03	0.001-0.67	.027
Central Venous Catheter	0.29	0.08-1.13	.075

A.

Receiver Operator Characteristic Curve

AUC: 0.882

B.

Receiver Operator Characteristic Curve

AUC: 0.951
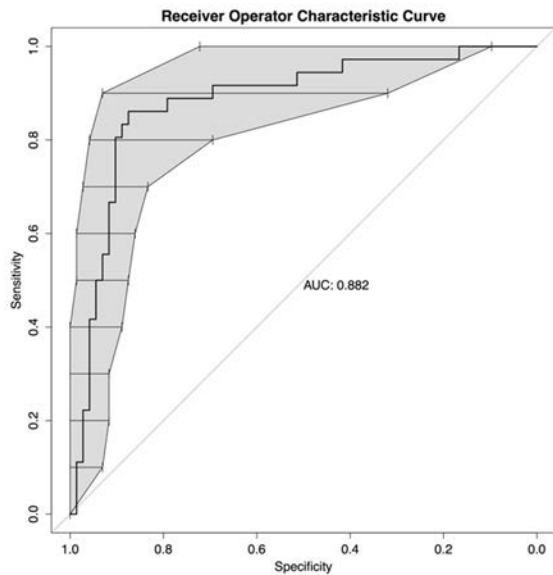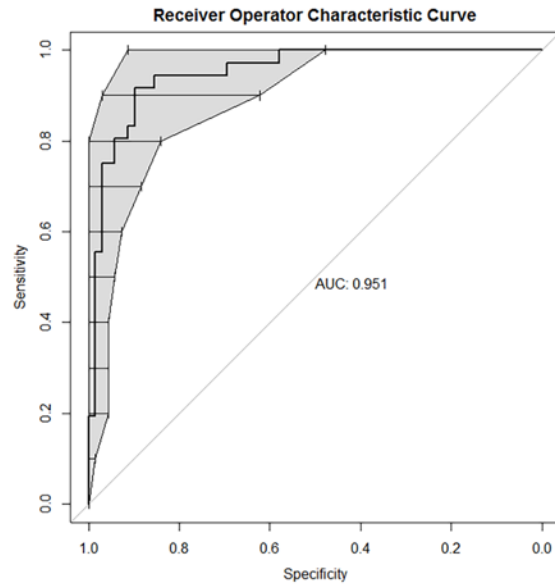
C.

Gene Locus Tags	Bacteria gene loci that improve the multivariable model
KPK_RS27195, KP1_RS26720	Tellurium resistance protein
KPNJ1_01715	Hypothetical open reading frame
KPN_01782	Putative Deoxygluconate dehydrogenase

Figure 3.1: Receiver operator characteristic curve for a multivariable model of factors predictive for patient infection.

Conditional logistic regression of patient factors was used to generate a predictive model (AUC=0.88) (A). Six clinical variables associated with invasive infection in this sample set. Conditional logistic regression identified bacterial genes that are associated with disease and improved AUC to 0.95 when added to the clinical model ($P = .011$) (B). Central line was removed in the final model (B). Bars and shaded area of ROC curve represent bootstrapped 95% confidence intervals for specificity at each level of sensitivity. (C) Lists the bacterial gene loci that were included in analysis for (B).

3.2.3 Pathogenicity-Associated Locus Sequencing

We next wanted to identify bacterial genes that are associated with infection in our cohort. To do this, we first created a reference "pan-genome" comprised of the entire genomes of five *K. pneumoniae* strains that are representative of pathogenic isolates of varying virulence potential and whose genome sequences are publicly available [138, 174, 175]. MGH 78578 is a hospital-acquired strain of *K. pneumoniae* that was isolated from a patient with pneumonia. NTUH K2044 is a hypervirulent strain that caused a liver abscess and meningitis [176]. NJST258_1 is a *Klebsiella pneumoniae* carbapenemase- (KPC) producing ST258 strain from a urinary tract infection. KP342 is an isolate of *Klebsiella variicola*, a closely related species to *K. pneumoniae* that is often misidentified in clinical laboratories and causes the same spectrum of disease [135]. KPPR1 is a genetically tractable strain frequently used in mouse models of infection [174]. Approximately 3900 orthologous genes are common to all five strains and are considered part of the core genome for this project (Figure 3.2). The accessory genomes range from approximately 1,100 genes to just under 1,900 genes, and represent an average of 26% of each genome.

A. Construct Pan-Genome from Reference Sequences
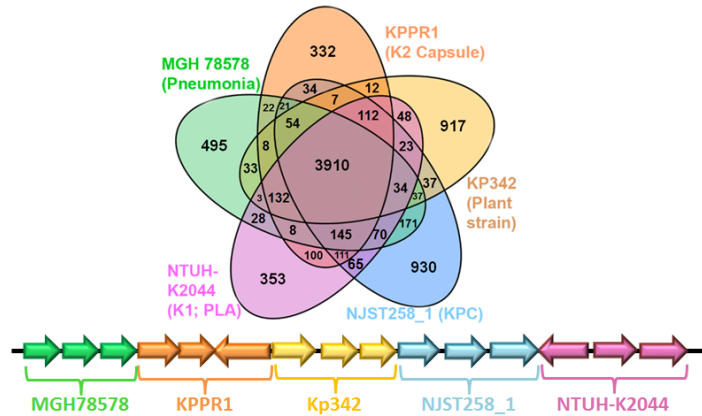
B. Map Illumina Sequencing Reads from Each Isolate to Pan-Genome
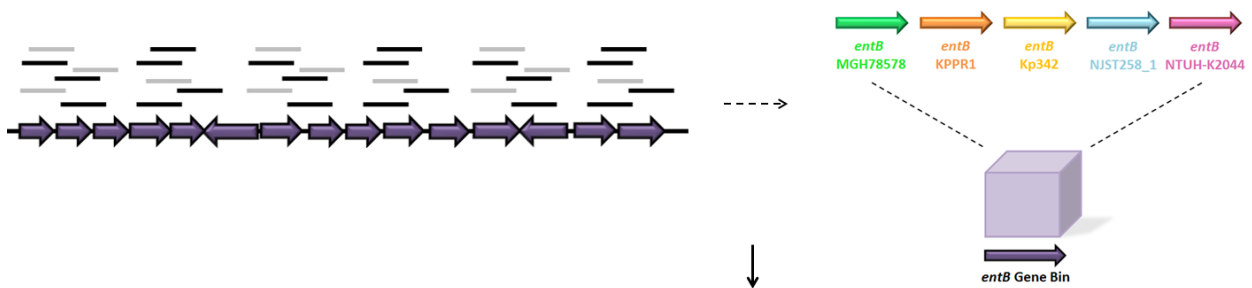
C. Calculate Gene Frequencies
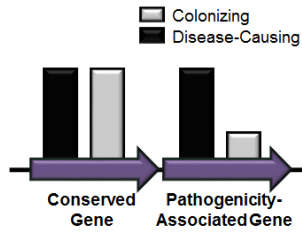
Figure 3.2: Pathogenicity-Associated Locus Sequencing (PAL-Seq).

Genes from five reference genomes are strung together to create a "pan-genome" (A). Orthologous pan-genome genes are collapsed into gene loci. Isolates are then sequenced using the Illumina HiSeq platform and raw reads are mapped to the pan-genome (B), allowing us to determine presence or absence of each gene in each genome. Gene frequencies are then measured and compared between colonizing and disease-causing groups (C).

To assess the differences in gene frequencies between the case and control isolates, we first sequenced all 114 clinical isolates using the Illumina HiSeq 4000 sequencing system. We also included KPPR1 that is part of the pan-genome as a positive control. KPPR1 had a mapping ratio of 98.9, meaning that 98.9% of reads mapped to the pan-genome. Reads from each sequenced sample were mapped to the pan-genome. Samples Kp723, Kp499, and Kp891 had the lowest mapping ratio, suggesting these strains are significantly different from the strains included in the pan-genome and may be *K. quasipneumoniae* (Table 3.3)**.** Sixteen samples have most of their reads mapped to Kp342, indicating they are *K. variicola*. The remaining 95 samples have most of their alignments equally distributed in KPPR1, MGH 78578, NJST258_1, and NTUH K2044. To determine the lineage of these outliers, multilocus sequence typing (MLST) was performed and the isolates were integrated with a reference set of MLST sequences from each species (Figure 3.3). As predicted, this collection of strains comprised 95 *K. pneumoniae* (34 infecting and 61 colonizing), 16 *K. variicola* (3 infecting vs 13 colonizing), and 3 *K. quasipneumoniae* (1 infecting vs 2 colonizing) (Figure 3.3). To assess the gene content of each strain's accessory genome, normalized counts were summed for each gene and dichotomized as present or absent based on k-means analysis. As validation of the k-mean analysis, PAL-seq successfully detected 5,016 of the 5,102 known KPPR1 genes and was negative for 3,157 of the 3,195 genes in the pan-genome that KPPR1 does not possess (sensitivity=98.3%; specificity=98.8%). Hierarchal clustering of present and absent genes in the accessory genomes (Figure 3.4), as well as PCA analysis of the accessory genomes (Figure 3.5), also distinguished three groups based on species.

Table 3.3: Gene Alignment Counts.

sampleID	RawReads	RawAlignment	MappingRatio	DeduppedAlignment
Kp1711	9,468,448	8,939,825	94.4	7,353,935
Kp1820	8,214,982	7,664,883	93.3	6,479,836
Kp1946	8,897,647	7,895,803	88.7	6,575,618
Kp1606	9,289,831	8,087,874	87.1	5,704,805
Kp1509	8,026,765	7,537,188	93.9	6,409,620
Kp1775	8,826,711	8,335,972	94.4	6,957,720
Kp1368	7,028,486	6,605,087	94.0	5,714,784
Kp1165	9,114,084	8,087,953	88.7	5,750,299
Kp1318	6,318,081	5,847,038	92.5	5,115,797
Kp783	9,230,499	8,233,480	89.2	6,861,919
Kp1432	9,886,288	9,236,843	93.4	7,583,138
Kp723	12,078,640	7,931,408	65.7	5,857,358
Kp744	6,957,767	6,274,024	90.2	5,412,266
Kp698	7,697,670	6,779,168	88.1	5,766,215
Kp701	8,077,407	7,615,668	94.3	6,431,518
Kp2106	8,353,463	7,761,927	92.9	6,572,819
Kp2103	7,061,495	6,473,661	91.7	5,608,186
Kp2145	9,315,864	8,392,280	90.1	7,001,760
Kp1232	9,820,821	8,769,413	89.3	6,044,456
Kp1082	9,561,848	8,585,173	89.8	7,096,503
Kp1319	8,334,307	7,466,941	89.6	6,333,331
Kp1282	7,426,550	6,699,545	90.2	5,733,972
Kp499	6,382,563	4,259,207	66.7	3,562,771
Kp590	8,486,905	7,099,731	83.7	5,987,364
Kp2168	7,913,000	7,004,476	88.5	5,162,107
Kp2089	7,638,285	6,964,015	91.2	5,980,822
Kp2009	6,731,351	5,879,313	87.3	5,150,915
Kp1000	10,333,805	9,400,445	91.0	7,692,162
Kp780	7,911,974	7,364,138	93.1	6,262,558
Kp860	8,012,430	7,106,140	88.7	6,067,355
Kp1026	8,423,996	7,829,745	92.9	6,595,412
Kp697	12,378,730	10,710,936	86.5	8,553,427
Kp725	9,713,420	8,689,158	89.5	7,135,762
Kp1866	10,653,588	9,790,959	91.9	7,956,931
Kp1912	15,527,184	14,708,442	94.7	10,961,189
Kp2026	12,525,518	11,674,427	93.2	9,194,453
Kp965	9,833,607	8,282,762	84.2	5,767,532
Kp469	9,354,662	8,677,018	92.8	7,191,779
Kp630	9,116,944	8,259,824	90.6	6,917,235
Kp1824	11,564,081	10,058,761	87.0	6,654,198
Kp1921	12,348,059	11,255,586	91.2	7,131,677
Kp1961	11,662,486	10,724,572	92.0	8,583,968
Kp1854	9,538,185	8,765,343	91.9	6,005,661
Kp1799	11,147,178	10,255,517	92.0	8,267,527
Kp1958	8,634,489	7,444,692	86.2	6,272,496
Kp2116	9,003,916	8,328,226	92.5	6,977,710
Kp2086	6,951,020	6,330,017	91.1	4,762,022
Kp2008	7,244,761	6,513,473	89.9	5,619,195
Kp2083	10,162,284	9,417,781	92.7	6,376,153
Kp1997	9,302,939	7,605,697	81.8	6,384,992
Kp2156	7,742,132	7,064,438	91.2	6,058,680
Kp1089	7,444,841	6,782,131	91.1	5,003,198
Kp1434	8,039,489	7,438,825	92.5	6,338,843
Kp953	7,865,538	7,213,038	91.7	6,176,450
Kp1207	6,709,081	6,162,683	91.9	5,380,261
Kp989	9,250,160	8,795,498	95.1	7,312,478
Kp731	10,720,200	9,924,512	92.6	8,046,672

Kp1439	12,167,515	10,038,175	82.5	6,705,138
Kp848	9,504,346	8,889,096	93.5	7,374,063
Kp709	9,699,150	8,470,317	87.3	7,082,219
Kp1038	8,614,308	7,970,351	92.5	6,715,938
Kp1440	8,192,162	7,261,228	88.6	5,304,237
Kp730	8,311,417	7,258,733	87.3	5,314,549
Kp1238	8,493,556	8,039,788	94.7	6,776,650
Kp1237	9,314,389	8,556,854	91.9	7,135,985
Kp1386	11,366,864	10,632,403	93.5	8,501,535
Kp2062	7,203,999	6,329,697	87.9	5,426,971
Kp2111	6,940,438	6,684,704	96.3	5,758,127
Kp2146	6,578,918	5,916,982	89.9	4,538,308
Kp1814	7,565,700	6,757,758	89.3	5,807,549
Kp1909	7,898,390	7,211,988	91.3	6,168,495
Kp2005	10,022,618	9,075,582	90.6	7,489,978
Kp1647	11,892,308	11,172,237	93.9	8,900,876
Kp1497	10,876,264	9,519,784	87.5	7,775,944
Kp1492	9,276,561	8,250,657	88.9	6,881,425
Kp2125	9,432,660	8,708,446	92.3	7,247,564
Kp1926	8,648,506	7,866,597	91.0	6,665,529
Kp2007	9,417,815	8,683,877	92.2	7,233,178
Kp891	6,888,902	4,337,811	63.0	3,609,402
Kp457	7,942,791	7,305,172	92.0	6,244,709
Kp625	10,485,747	9,129,757	87.1	6,178,637
Kp1295	10,949,655	9,844,050	89.9	8,013,739
Kp1375	11,916,865	10,930,636	91.7	8,749,556
Kp1317	9,858,444	9,146,086	92.8	7,549,166
Kp1062	8,776,386	8,149,372	92.9	6,844,573
Kp1108	9,942,622	9,014,579	90.7	7,437,317
Kp733	11,441,306	10,342,648	90.4	8,279,396
Kp1827	10,078,787	9,391,960	93.2	7,718,818
Kp1890	11,794,451	11,002,791	93.3	8,792,778
Kp1950	9,688,289	8,859,520	91.4	7,336,019
Kp1693	10,706,572	9,539,971	89.1	7,797,937
Kp1626	10,475,495	9,496,803	90.7	7,762,786
Kp1637	7,988,432	7,410,706	92.8	6,325,392
Kp1732	7,595,527	7,098,601	93.5	6,095,923
Kp1885	7,904,396	7,291,015	92.2	6,237,680
Kp1641	9,165,629	8,632,673	94.2	7,225,030
Kp466	11,098,112	9,989,334	90.0	8,139,382
Kp511	10,995,503	10,523,853	95.7	8,470,744
Kp609	9,754,236	8,994,808	92.2	7,489,096
Kp1411	9,583,972	8,750,620	91.3	7,303,802
Kp706	8,566,646	7,765,039	90.6	6,557,319
Kp671	6,980,172	6,091,123	87.3	5,331,808
Kp832	7,701,342	6,991,405	90.8	5,988,026
Kp1014	12,963,872	11,631,583	89.7	9,211,528
Kp728	9,846,351	8,435,009	85.7	7,050,510
Kp1009	12,420,897	11,478,150	92.4	9,112,689
Kp1035	9,038,808	8,046,603	89.0	6,805,513
Kp735	8,713,242	8,006,383	91.9	6,778,394
Kp1669	7,109,915	6,455,356	90.8	5,577,015
Kp807	7,114,015	6,460,925	90.8	5,594,098
Kp710	7,399,837	6,841,442	92.5	5,879,519
Kp1559	10,264,235	9,135,528	89.0	7,542,299
Kp1601	10,714,906	9,795,412	91.4	8,016,780
Kp1469	9,336,291	8,551,719	91.6	7,127,147
KPPR1	9,118,862	9,022,393	98.9	7,409,588

Tree scale: 0.01

70
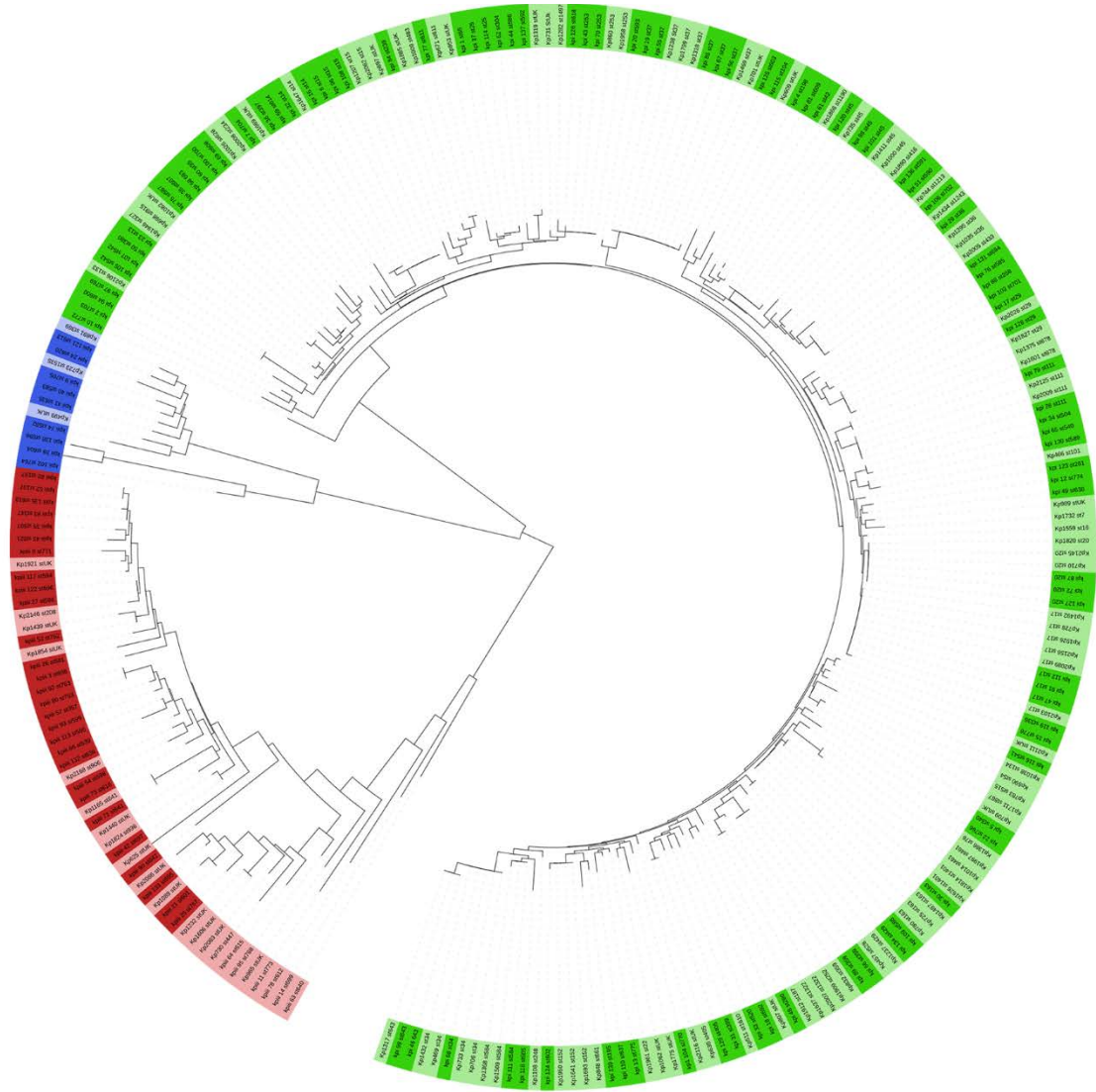
Figure 3.3: Phylogenetic tree of Kp phylogroups.

Clinical isolates analyzed using PAL-Seq were compared to reference isolates from each three identified Kp groups based on MLST sequence. The resulting phylogenetic tree shows three distinct branches, which distinguish Kp groups: KpI – *K. pneumoniae* (green); KpII – *K. quasipneumoniae* (blue); and KpIII – *K. variicola* (red). Reference isolates are shaded darker, which clinical PAL-Seq isolates are shaded lighter.
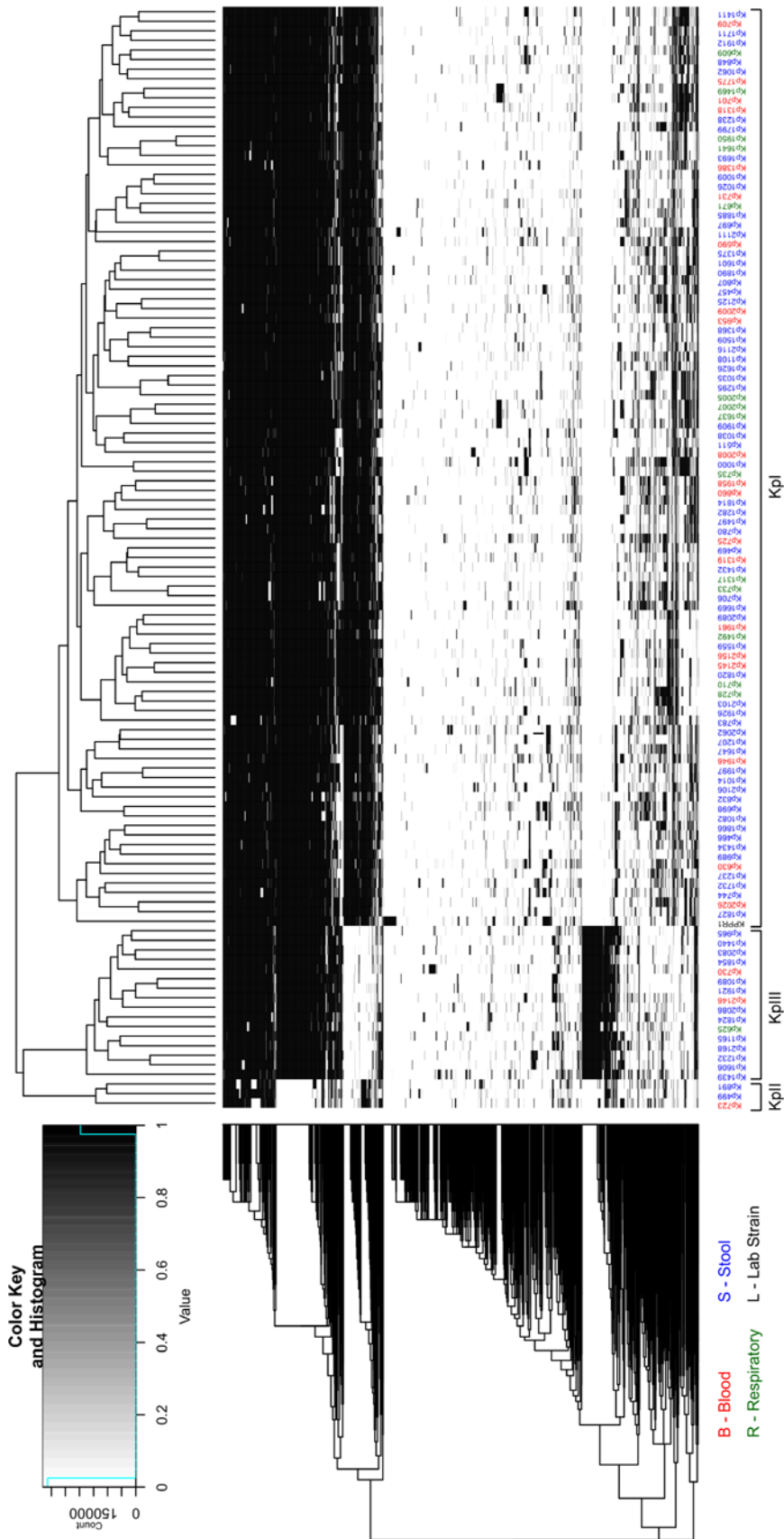
Figure 3.4: Hierarchical clustering of accessory genes.

Based on the binary classification data, 3169 bins were classified as present, and 703 bins were classified as absent in all samples. After removal of these uniform bins, the remaining bins underwent hierarchical clustering of their classification, shown as a heatmap in which white and black represent absent and present respectively. The column dendrogram can be viewed as a measurement of similarity between samples. We hypothesize that this clustering would be consistent, in some extent, with the biological characteristics of the samples. Kp phylogroups are indicated for each cluster.

Figure 3.5: Principal Component Analysis of normalized count sum data.

Instead of binary classification, original normalized count sum data was used in PCA to see if samples can be separated in clusters. The data was first processed by in-row quantile normalization and PCA was performed between columns. Principal component 1 versus 2 was plotted. The clustering of samples is consistent with what the alignment distribution reflects.

3.2.4 Bacterial genes are significantly associated with disease

To determine if there were significant differences in the frequency of any genes between case isolates and control isolates, a conditional logistic regression was performed using the binary classification. This analysis was limited to genes with frequencies of 5-95% within our sample set. After ranking by *P* value, certain genes clustered together based on their frequencies and sequential location in the pan-genome. Analysis of gene annotation and their locations in reference genomes indicated that these genes were likely located within operons. To facilitate further analysis, genes from the ten most significant operons (Table 3.4) were collapsed into one representative gene, with five loci as potential virulence factors and five as potential protective factors.

Table 3.4: Genes present in significantly different frequencies between cases and controls.

Reference Genome Gene Locus Tags	Annotation	Frequency, No (%)		Q value
		Case Bloodstream Infection or Pneumonia (N=38)	**Control** Asymptomatic Rectal Colonization (N=76)	
KPK_RS16860, VK055_1478, KP1_RS09340	Iron ABC transporter permease	4 (10.5)	29 (38.1)	**0.004369**
KPNJ1_RS13400, KPN_01704, KP1_RS12850	DeoR family protein/alkaline phosphatase	32 (84.2)	43 (56.6)	**0.005060**
KPK_RS16855, VK055_1477, KP1_RS09345	Amino acid ABC transporter substrate-binding protein	4 (10.5)	28 (36.8)	**0.005286**
KPK_RS16810, VK055_1467, KP1_RS09395	PTS fructose transporter subunit IIA	5 (13.1)	30 (39.4)	**0.006209**
KPK_RS16815, VK055_1468, KP1_RS09390	PTS fructose transporter subunit IIC	5 (13.1)	30 (39.4)	**0.006209**
KPK_RS16825, VK055_1470, KP1_RS09380	Formate acetyltransferase	5 (13.1)	30 (39.4)	**0.006209**
KPK_RS16830, VK055_1471, KP1_RS09375	Pyruvate formate lyase II activase	5 (13.1)	30 (39.4)	**0.006209**
KPK_RS16840, VK055_1473, KP1_RS09365	Phytanoyl-CoA dioxygenase	5 (13.1)	30 (39.4)	**0.006209**
KPK_RS16845, VK055_1476, KP1_RS09360	AraC family transcriptional regulator	5 (13.1)	30 (39.4)	**0.006209**
KPK_RS16850, VK055_1475, KP1_RS09355	Membrane protein	5 (13.1)	30 (39.4)	**0.006209**

KPN_01782	Deoxygluconate dehydrogenase	26 (68.4)	30 (39.4)	**0.007139**
KPK_RS16865, VK055_1479, KP1_RS09335	Iron ABC transporter substrate-binding protein	5 (13.1)	30 (39.4)	**0.007219**
KPK_RS16820, VK055_1469, KP1_RS09385	PTS fructose transporter subunit IIB	6 (15.8)	32 (42.1)	**0.008326**
KPK_RS16835, VK055_1472, KP1_RS09370	PTS fructose transporter subunit IIB	6 (15.8)	32 (42.1)	**0.008326**
KPNJ1_RS13410, KPN_01702, KP1_RS12840	Deoxyribose-phosphate aldolase	31 (81.5)	43 (56.6)	**0.009493**
KPNJ1_RS13415, KPN_01701, KP1_RS12835	Alkaline phosphatase/deoR	31 (81.5)	43 (56.6)	**0.009493**
KPNJ1_RS13420, KPN_01700, KP1_RS12830	Allulose-6-phosphate 3-epimerase	31 (81.5)	43 (56.6)	**0.009493**
KPNJ1_RS13425, KPN_01699, KP1_RS12825	Carbohydrate kinase (ribokinase)	31 (81.5)	43 (56.6)	**0.009493**
KPNJ1_RS13430, KPN_01698, KP1_RS12820	Fucose permease	31 (81.5)	43 (56.6)	**0.009493**
KPK_RS27195, KP1_RS26720	Tellurium resistance protein TerF	10 (26.3)	6 (7.9)	**0.012875**
KPK_RS27215, KP1_RS26700	Tellurium resistance protein TerB	10 (26.3)	6 (7.9)	**0.012875**
KPK_RS27260, KP1_RS26655	Tellurium resistance protein TerW	10 (26.3)	6 (7.9)	**0.012875**
KPK_RS27270, KP1_RS26650	Tellurium resistance protein TerY	10 (26.3)	6 (7.9)	**0.012875**
KPNJ1_01715	Hypothetical protein	13 (34.2)	10 (13.2)	**0.013956**

Genes in operons were collapsed into one representative gene. representative potential protective factors are in green, & potential risk factors in red.

To determine which loci were significantly and independent associated with infection in our cohort of patients, each representative gene was individually adjusted for the five patient-level variables from the clinical model. The five potential pathogenicity-associated loci (PALs) were significantly associated with *K. pneumoniae* infection (Table 3.5), independent of patient-level variables that differed between cases and controls. These five PALs are annotated as a tellurium resistance protein, part of an operon involved in metalloid resistance; a DeoR family protein and a deoxyribose-phosphate aldolase, a regulator and enzyme both seemingly part of the same sugar metabolism operon; a hypothetical protein with unknown function; and a putative deoxygluconate dehydrogenase that may act as an oxidoreductase (Figure 3.6). In contrast, the five loci that were potentially protective on unadjusted analysis were not associated with whether or not a patient was infected or asymptomatically colonized after adjustment for patient-level variables.

Table 3.5: Bacterial genes significantly and independently associated with clinical infection.

Reference Genome Gene Locus Tags	Gene	Odds Ratio	95% CI	*P* value
KP1_RS26720	**Tellurium resistance protein TerF**	11.3	1.6-80	.015
KP1_RS12850	**DeoR family protein/alkaline phosphatase**	10.5	1.62-67.9	.014
KP1_RS12840	**Deoxyribose-phosphate aldolase**	5.79	1.19-28.2	.03
KPNJ1_01715	**Hypothetical Protein**	5.08	1.27-20.2	.021
KPN_01782	**Putative Deoxygluconate dehydrogenase**	4.5	1.34-15	.015

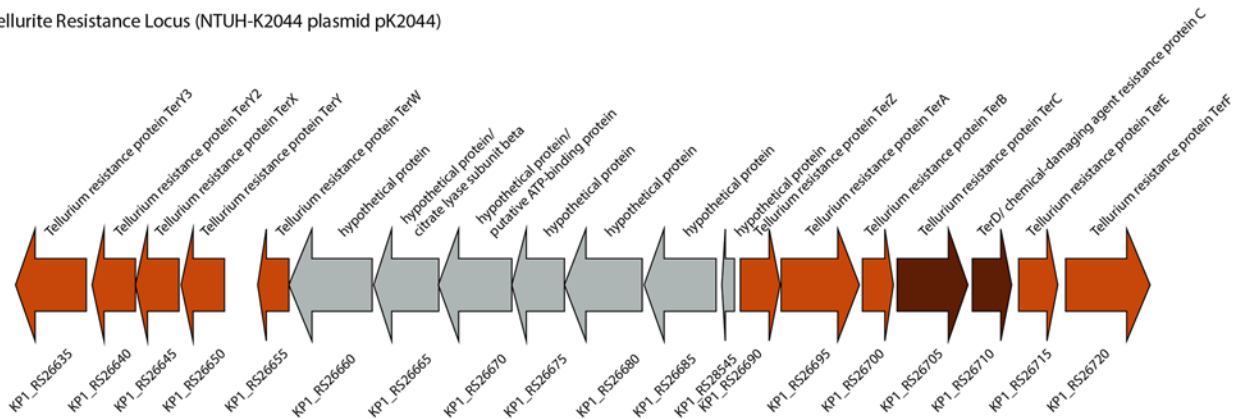Adjusted for the clinical variables in Table 3.2.
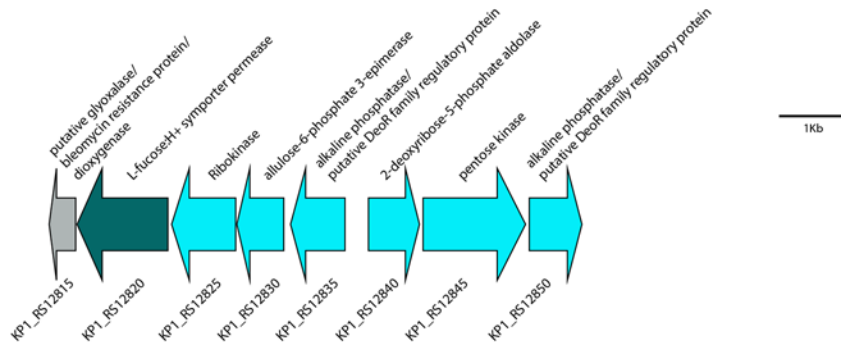
Figure 3.6: Pathogenicity-Associated Loci (PALs).

Two gene operons identified by PAL-Seq are shown, the putative sugar metabolism locus and the tellurite resistance locus. Genes not identified in PAL-Seq are shown in grey. Genes deleted for characterization are darkened.

3.2.5 Combining bacterial and host variables improves prediction of infection

We next wanted to determine if the five PALs identified by PAL-Seq could improve our previous multivariable model (Table 3.2 & Figure 3.1A). Adding in these five PALs to the clinical model followed by backwards elimination retained three of them in the final, combined model (Table 3.6). Central line was not retained in this final model as it only slightly improved the performance of the initial model with a borderline significance ($P = 0.54$ for likelihood ratio test). This model identified the representative genes for tellurite resistance, deoxygluconate dehydrogenase, and the hypothetical protein as highly predictive for disease when combining with patient variables from the previous model (Table 3.6). This revised model had significantly improved accuracy compared to the model based only on patient factors (AUC = 0.95 vs 0.88; P = .011) (Figure 3.1B). In this model, tellurite resistance genes had the greatest association with infection (OR 157; 95% CI 3.34-7350; P = .01) followed by the presence of a fluid & electrolyte disorder (OR 22.9; 95% CI 1.6-329; P = .021), deoxygluconate dehydrogenase (OR 17.8; 95% CI 2.2-143; P = .007), and the hypothetical protein gene (OR 16.9; 95% CI 1.59-179; P = .019).

Table 3.6: Multivariable model of factors predictive for patient infection.

Patient Variable	Odds Ratio	95% CI	*P* value
Fluid & Electrolyte Disorders	22.9	1.6-329	.021
Minimum Serum Glucose (mg/dL)	1.08	1.01-1.16	.026
Body Mass Index Baseline (kg/m^2)	0.69	0.5-0.95	.022
White Race	0.03	<0.01-0.75	.032
Peripheral Vascular Disease	0.001	<0.01-3290	.376
Bacterial Variable			
Tellurium Resistance Protein	157	3.34-7350	.01
Deoxygluconate Dehydrogenase	17.8	2.2-143	.007
Hypothetical Protein	16.9	1.59-179	.019

3.2.6 The tellurite resistance locus is associated with infection independent of phylogenetic

*lineage, but is not required in a murine pneumonia model*

The *ter* locus has been associated with specific hypervirulent sequence types (STs), suggesting it may be a marker of lineage and not an independent predictor of clinical *K. pneumoniae* infection [177]. To determine if the *ter* locus is associated with certain lineages in this cohort, we performed multilocus sequence typing (MLST) on all of our bacterial isolates. We then generated a phylogenetic tree based on MLST, and assessed which isolates possessed the *ter* locus and if they cluster together. We found that the *ter* locus is present across multiple STs and does not appear to cluster among any one ST or any closely related group of STs (Figure 3.7A). We also generated a phylogenetic tree based on sequence variants identified by whole-genome sequencing (WGS). In both trees, three distinct branches corresponding with KpI, KpII, and KpIII were present. However, individual isolates did not branch identically between the two trees, suggesting that evolutionary relationships are elucidated differently by these two methods. Also in both trees, Kp1637 and Kp2007 share the same node, indicating they are the same strain. WGS verified that the *ter* locus did not cluster in any one branch or among closely related branches, which confirmed that possession of the *ter* locus is not lineage-associated (Figure 3.7B). This locus was also not associated with any one *Klebsiella* species, since it was found in *K. pneumoniae*, *K. quasipneumoniae*, and *K. variicola* isolates.

A. *Phylogenetic tree based on multilocus sequence typing*
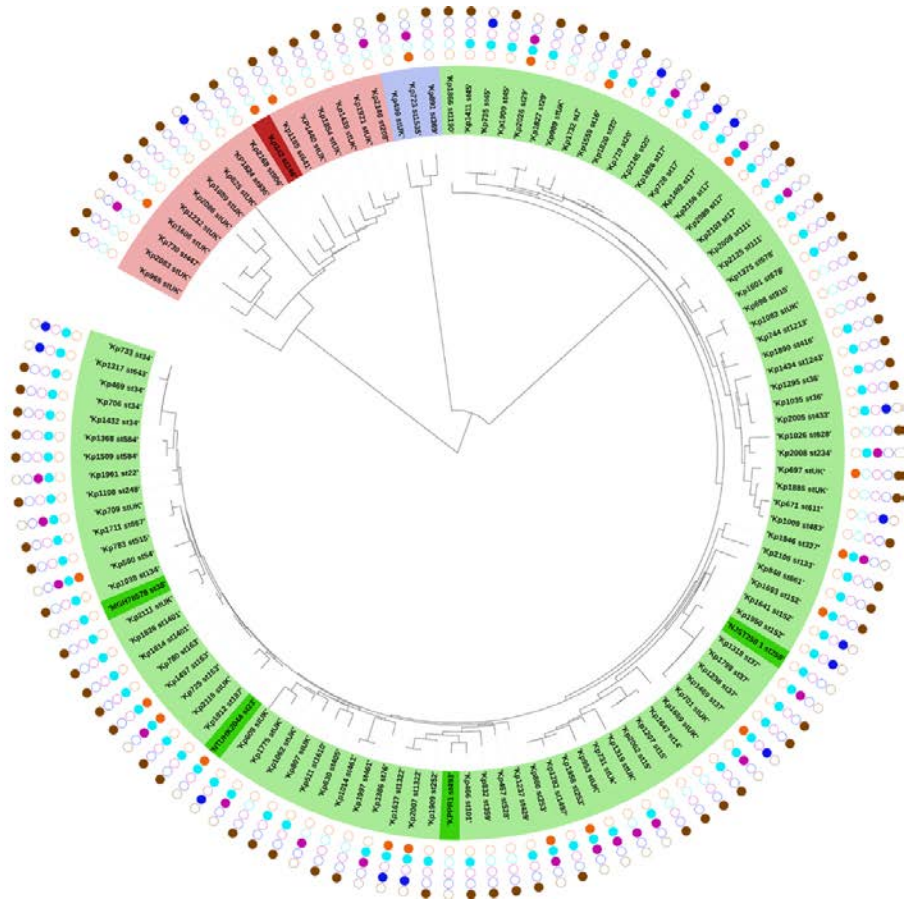
B. *Phylogenetic tree based on whole genome variants*
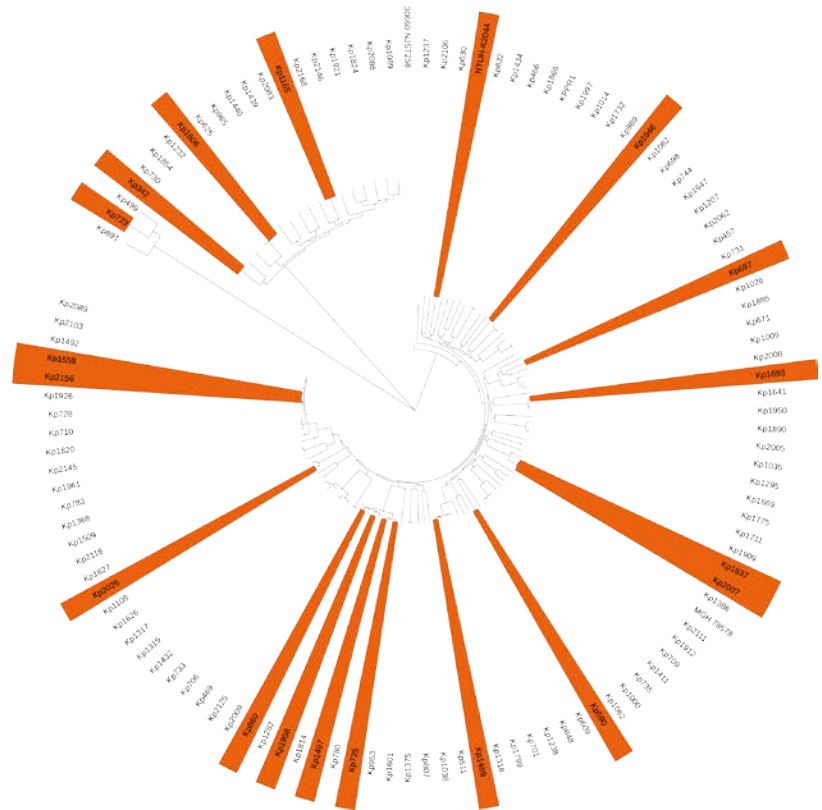
Figure 3.7: Phylogenetic analysis of clinical isolates.

Multilocus sequence typing (MLST) was performed on all clinical isolates plus all five reference strains (A). The three *Klebsiella* species are distinguished by color: *K. variicola* is pink, *K. quasipneumoniae* is blue, and *pneumoniae* is green. Reference isolates are indicated by dark green or dark red. Whole genome sequencing was used to assess variants among isolates and a phylogenetic tree was generated, with tellurite resistance loci presence indicated in orange (B). Tellurite resistance loci presence is spread throughout all three species and does not appear to cluster in any one sequence type or branch, indicating that its presence is not sequence type- or lineage-associated.

To characterize the role of the *ter* locus in virulence, we performed targeted mutagenesis of *terC* and *terD.* Interruption of either *terC* or *terD* on the IncHI2 plasmidR478 leads to an elimination of the tellurite resistant phenotype in *E. coli* [178]. Using lambda red recombination, we deleted these two genes individually on the pK2044 plasmid in the hypervirulent *K. pneumoniae* strain NTUH-K2044 that was part of our pan-genome and has been previously shown to cause pneumonia in an animal model [32]. Overnight growth in Luria-Bertani (LB) broth demonstrated no growth defect following removal of *terC* (Figure 3.8). To determine if deletion of either gene leads to a tellurite sensitive phenotype, we assessed growth of mutants on MacConkey-inositol-potassium tellurite (MCIK) agar [179]. We found that removal of *terC* but not *terD* led to a tellurite sensitive phenotype as determined by growth on MCIK agar (Figure 3.9A). Based on quantitative culture, removal of *terC* leads to a 0.00072% recovery on MCIK agar compared to MAC agar (Figure 3.9B). We next wanted to determine if removal of *terC* led to an *in vivo* fitness defect. To determine this, we infected mice with a 1:1 ratio of mutant to wild type (WT) bacteria using a pneumonia model of infection (see materials and methods) and calculated a competitive index. No significant fitness defect was seen in the lungs at 24 (Figure 3.10A) or 48 (Figure 3.10B) hours post infection.

Figure 3.8: Deletion of terC does not affect growth in nutrient rich broth.

LB broth was inoculated with 2.6 x 10^6 CFU/mL of either NTUH-K2044 (WT) or *ΔterC* bacteria. Growth was recorded using optical density (600nm) overnight up to 16 hours.

A.
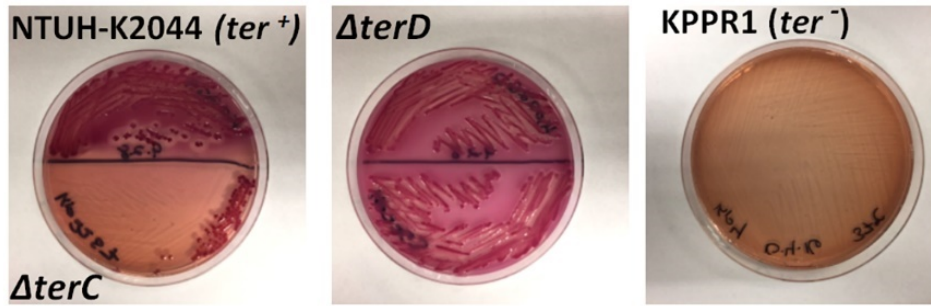
B.
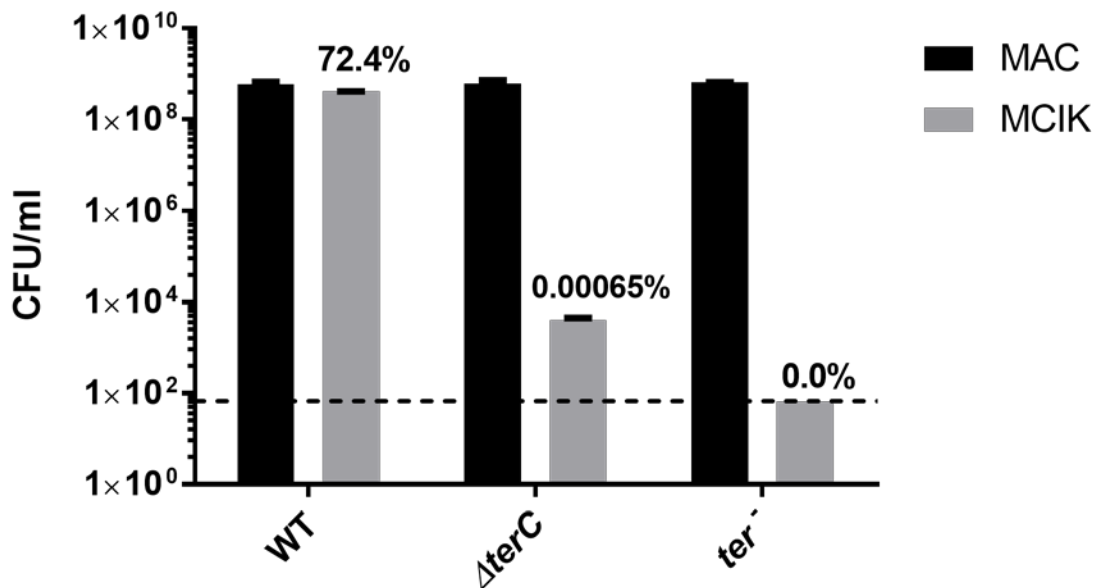
Figure 3.9: Deletion of terC confers a tellurite sensitive phenotype.

NTUH-K2044 (WT), *ΔterC*, *ΔterD*, and KPPR1 (ter ⁻) were plated to MacConkey-inositol-potassium tellurite agar (MCIK) (3µg/mL potassium tellurite) and visualized for inhibition of growth (a). Since *ΔterC* showed an apparent partial inhibition, quantification of growth on MCIK compared to growth on MacConkey (MAC) agar was assessed (b).
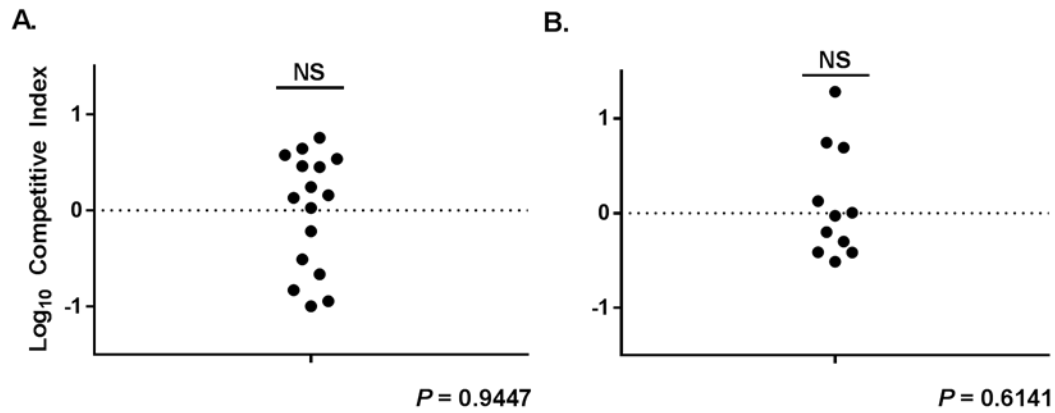
Figure 3.10: Competitive index of terC mutant compared to WT.

A co-challenge infection was performed on C57BL/6 mice to determine if there is an *in vivo* fitness defect in the *ΔterC* strain. Calculation of a competitive index indicates that there is no significant fitness defect at 24 hours ($P = 0.9447$) (A) and 48 hours ($P = 0.6141$) (B) post infection. Data at 24 hours represents three separate experiments, while data at 48 hours represents two experiments.

3.2.7 Sugar utilization locus is required for growth on psicose and contributes to fitness in a

murine pneumonia model

To characterize the function of the sugar transport locus associated with infection, we constructed a deletion mutation of the putative sugar permease gene (*ΔKP1_RS12820*). Several mutant clones were generated and used for characterization of the locus. Deletion of this gene did not affect growth in nutrient rich media (LB) (Figure 3.11). Growth of all mutant isolates was comparable to growth of WT when grown in M9 minimal media with or without glucose (Figure 3.12A & 3.12B). To identify potential sugar substrates of this locus, the WT and *ΔKP1_RS12820* (clone #3) were grown in minimal media with and without addition of sugars. Since *K. pneumoniae* is a facultative anaerobe, growth was measured in both aerobic and anaerobic conditions. As expected, anaerobic conditions resulted in a lower final bacterial density (Figures 3.13 & 3.14). Growth profiles were comparable between aerobic and anaerobic conditions, showing that *K. pneumoniae* is able to grow on a diverse set of carbon sources. The most prominent difference in growth patterns between WT and *ΔKP1_RS12820* was seen on psicose (Figure 3.15A). All mutant clones had a growth defect on psicose compared to WT, with most clones showing delayed growth and clone #3 showing a lower slope. Complementation of the mutant using a plasmid encoding *KP1_RS12820* resulted in partial rescue of the growth defect in clone #3 (Figure 3.15B). Mean optical density at 20 hours was significantly different between the mutant and WT ($P = 0.0041$), while there was no significant difference in mean optical density between the complemented mutant and WT at 20 hours (Figure 3.15C). Our results indicate that psicose is likely a specific substrate of this sugar metabolism locus.
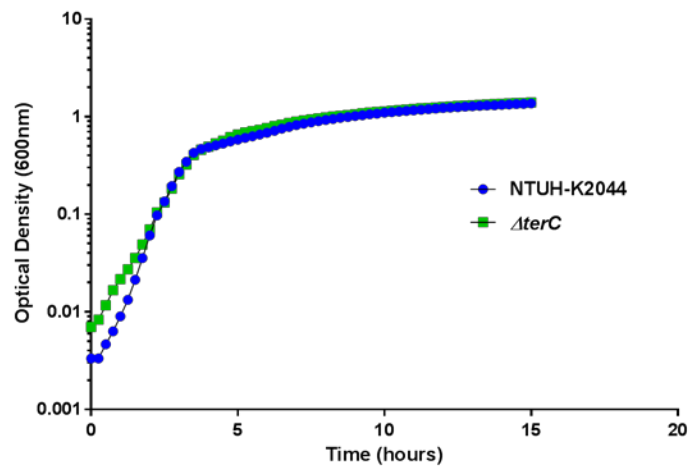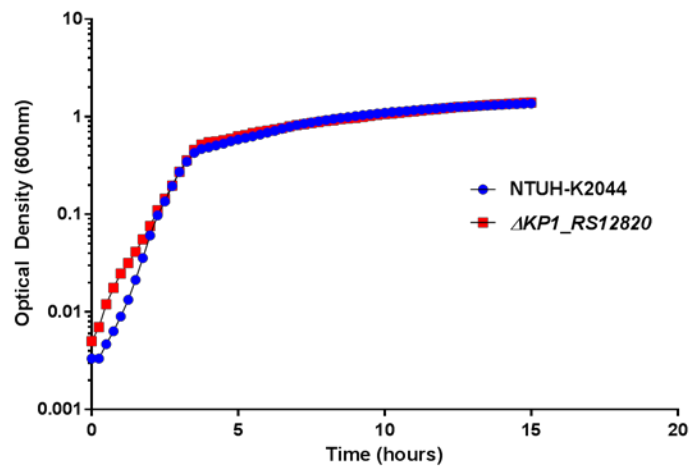
Figure 3.11: Deletion of KP1_RS12820 does not affect growth in nutrient rich broth.

LB broth was inoculated with 2.6 x 10^6 CFU/mL of either NTUH-K2044 (WT) or *ΔKP1_RS12820* clone #3 bacteria. Growth was recorded using optical density (600nm) overnight up to 16 hours. Data represents technical replicates from one experiment.

Figure 3.12: Deletion of KP1_RS12820 does not affect growth in glucose.

Overnight broth of either NTUH-K2044 (WT) or *ΔKP1_RS12820* bacteria was washed in 2X minimal media and then diluted 1:50 in the same. Each well contained either no carbon source (A) or glucose at a final concentration of 5mg/ml (B). Diluted inoculum was added to each well 1:1. Growth was recorded using optical density (600nm) over 24 hours. Data points shown are the mean of three technical replicates, and are representative of 1 to 2 experiments.

Figure showing Optical Density (600nm) versus Time (hours) growth curves for various carbon sources: Arabinose, Cellobiose, Fructose, N-acetylglucosamine, Galactose, Galacturonic Acid, Glucosamine, Glucuronic Acid, Mannose, N-acetylgalactosamine, Heparin, N-acetylneuraminic Acid, Rhamnose, Ribose, Xylose, Chondroitin Sulfate, α-Mannon, Neutral PGM O-glycans, Laminarin, Alginate, Dextran, Arabinan, Arabinogalactan, Pectic Galactan (Potato), Pectic Galactan (Lupin), Polygalacturonate, Rhamnogalacturonan I, Arabinoxylan, Galactomannan, Glucomannan, Xyloglucan, Fructooligosaccharides, Levan, Pullulan, Glycogen, Glucose, Psicose, Inulin.

Figure 3.13: Aerobic metabolic profiles of NTUH-K2044 and ΔKP1_RS12820 clone #3.

Overnight broth was washed in 2X minimal media and then diluted 1:50 in the same. Each well contained a carbohydrate at 2X concentration (10-20mg/ml) [180]. Diluted inoculum was added to each well 1:1. Growth was recorded using optical density (600nm) for up to 48 hours. Data points shown are the mean of two to three technical replicates, and are representative of 1 to 3 experiments.

Glucosamine · Galacturonic Acid · Galactose · Fucose · Fructose · Cellobiose · Arabinose · Ribose · Rhamnose · N-acetylneuraminic acid · N-acetylglucosamine · N-acetylgalactosamine · Mannose · Glucoronic Acid · Alginate · Laminarin · Neutral PGM O-glycans · α-Mannan · Heparin · Chondroitin Sulfate · Xylose · Rhamnogalacturonan I · Polygalacturonate · Pectic Galactan (lupin) · Pectic Galactan (potato) · Arabinogalactan · Arabinan · Dextran · Levan · Fructooligosaccharides · Inulin · Xyloglucan · Glucomannan · Galactomannan · Arabinoxylan · Psicose · Glucose · Lyxose · Glycogen · Pullulan

Optical Density (600nm)

Time (hours)

94

Figure 3.14: Anaerobic metabolic profiles of NTUH-K2044 and ΔKP1_RS12820 clone #3.

Overnight broth (grown anaerobically) was washed in anaerobic-conditioned 2X minimal media and then diluted 1:50 in the same. Each well contained a carbohydrate at 2X concentration (10-20mg/ml) [180]. Diluted inoculum was added to each well 1:1. All steps were performed anaerobically. Growth was recorded using optical density (600nm) for up to 48 hours. Data points shown are the mean of two technical replicates, and are from one experiment.

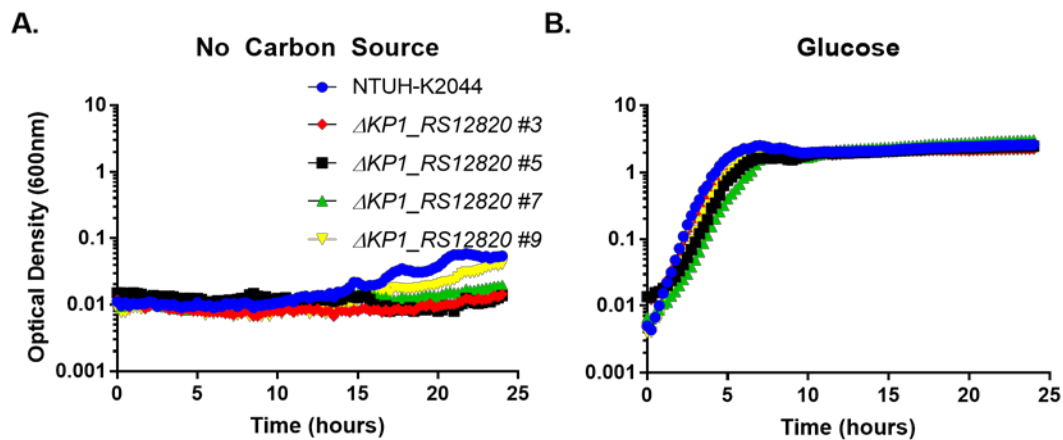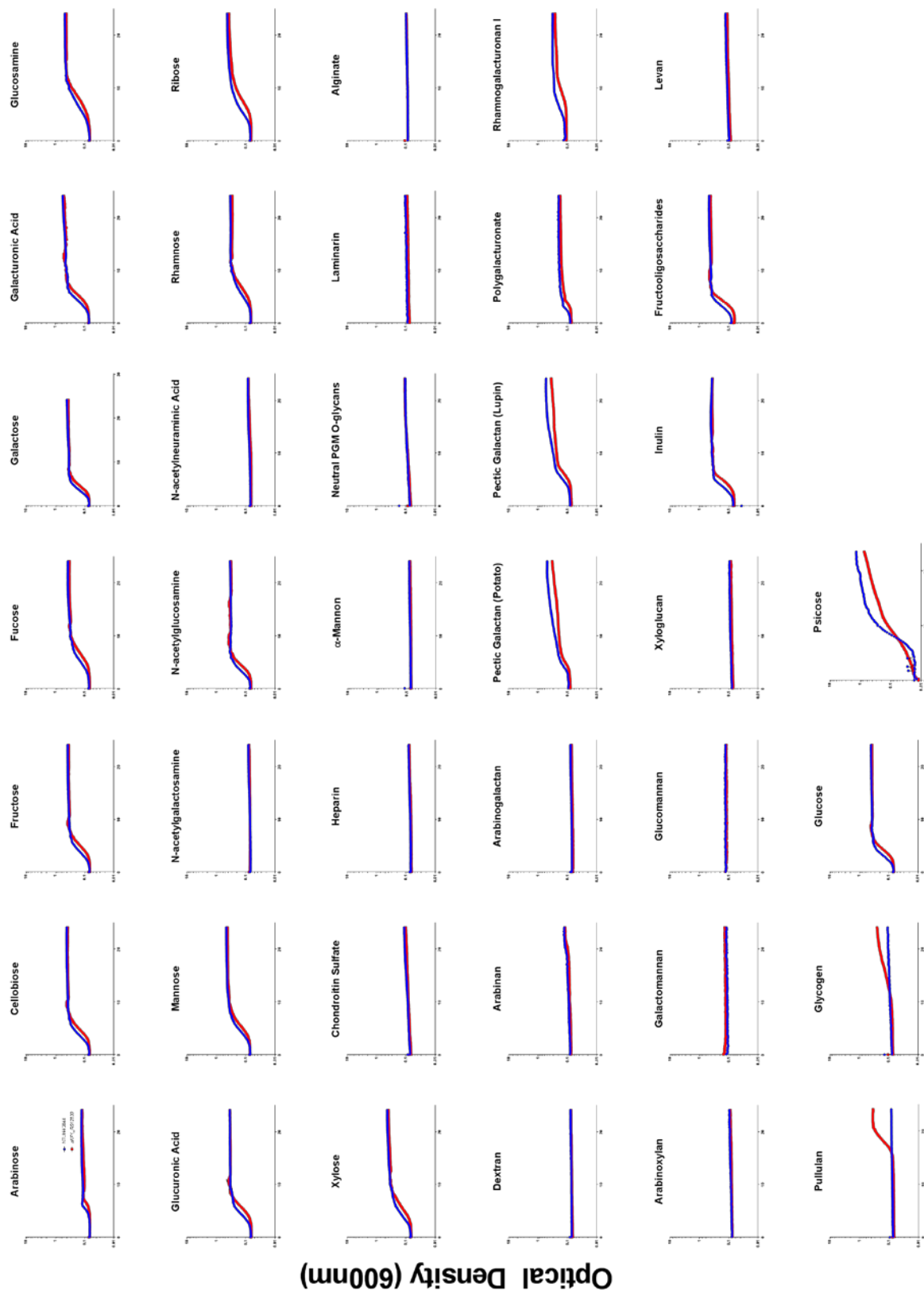*Figure 3.15: Deletion of KP1_RS12820 affects growth in psicose.*

Overnight broth of either NTUH-K2044 (WT) or *ΔKP1_RS12820* bacteria (A), WT and empty vector, *ΔKP1_RS12820* clone #3 and empty vector, or the complemented mutant (B) was washed in 2X minimal media and then diluted 1:50 in the same. Each well contained psicose at a final concentration of 5mg/ml. Diluted inoculum was added to each well 1:1. Growth was recorded using optical density (600nm) over 24 hours. Data points shown a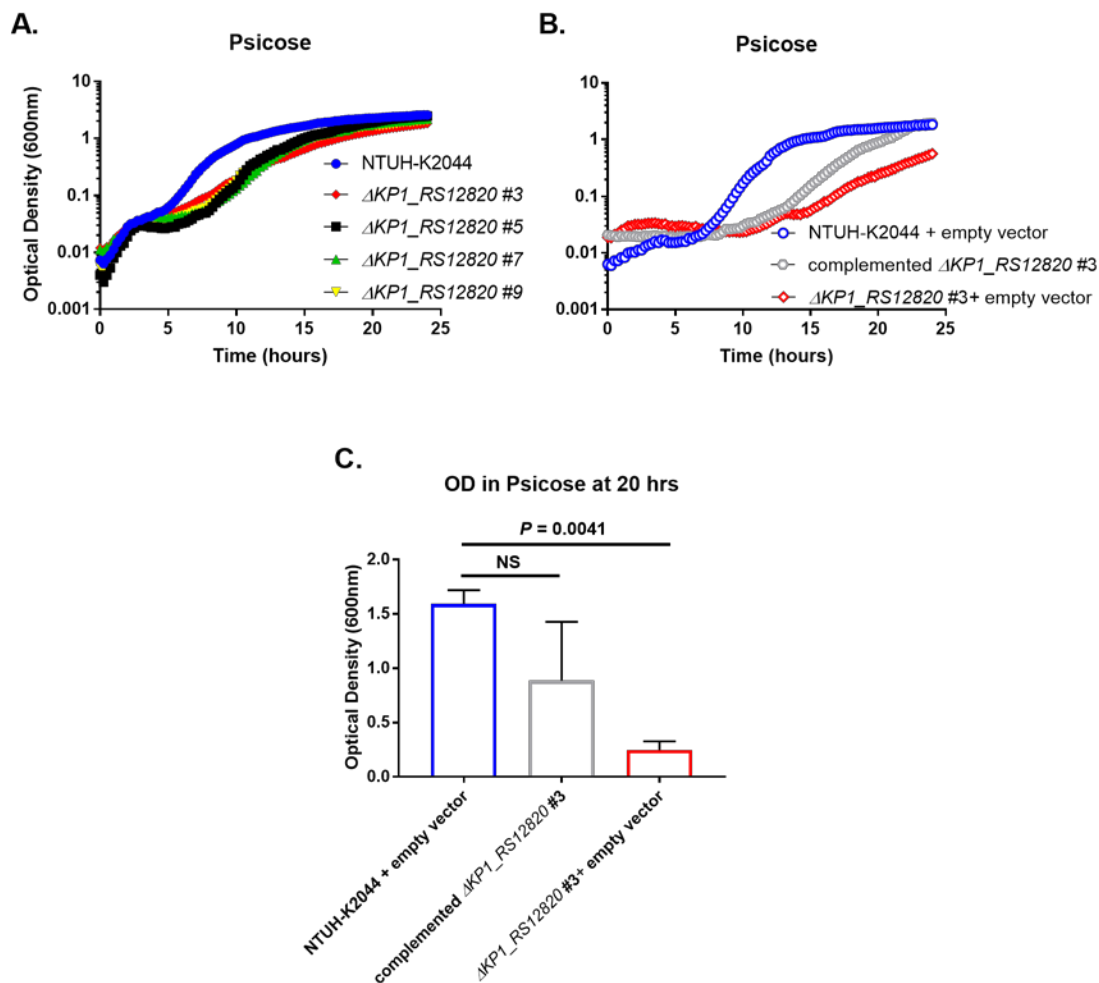re the mean of three technical replicates, and are representative of 1 to 2 experiments. Optical densities at 20 hours are shown for the complementation experiment (C). Statistics were calculated using one-way ANOVA.

Since the identified locus metabolizes a carbon source, we hypothesized that one physiological source of the substrate during infection may be human serum. To determine if the substrate can be found in human serum, we compared growth of WT and *ΔKP1_RS12820* bacteria when grown in minimal media with 10% heat inactivated human serum. There was no difference in growth of mutant isolate clone #3 compared to WT during 24 hours (Figure 3.16), indicating that the substrate is either not present in human serum or is not the only carbon source available in human serum that the bacteria is able to use.

To determine if deletion of *KP1_RS12820* leads to an *in vivo* fitness defect, we infected mice with a 1:1 ratio of mutant to wild type (WT) bacteria using a pneumonia model of infection. After 24 hours, bacterial density of each strain in the lung was assessed and a competitive index was calculated. The *ΔKP1_RS12820* mutant clones #3 and #5 had a significant *in vivo* fitness defect in a pneumonia model compared to the wild type. Mutant clone #4 did not have a significant *in vivo* fitness defect, though it was trending toward significance. However, complementation of mutant clone #3 did not rescue the fitness defect (Figure 3.17). Whole genome sequence variant analysis of the clone #3 compared to WT identified a frameshift deletion mutation in *cyoD*, cytochrome O ubiquinol oxidase subunit IV (Table 3.7). This gene is an oxidoreductase and appears to be part of an operon that is responsible for aerobic respiration. Mutation of this gene may contribute to the *in vivo* fitness defect and confound complementation. However, the significant fitness defect also seen in clone #5 indicates *KP1_RS12820* likely plays a role in bacterial fitness (Figure 3.17).
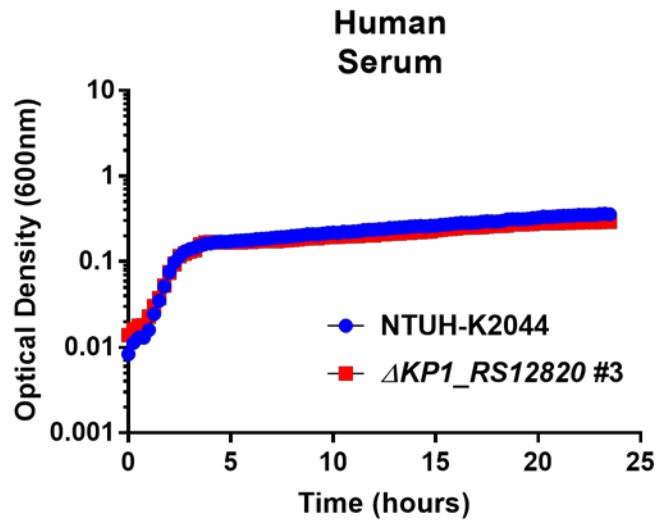
Figure 3.16: Deletion of KP1_RS12820 does not affect growth in human serum.

M9 minimal media with 10% human serum was inoculated with 2.6 x 10^6 CFU/mL of either NTUH-K2044 (WT) or *ΔKP1_RS12820* clone #3 bacteria. Growth was recorded using optical density (600nm) over 24 hours.
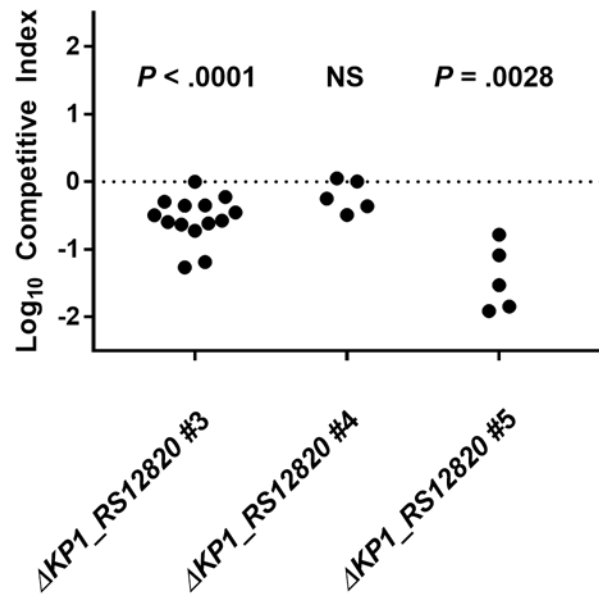
Figure 3.17: Deletion of putative sugar permease KP1_RS12820 leads to an in vivo fitness defect.

C57BL/6 mice were inoculated intrapharyngeally with a 1:1 ratio of NTUH-K2044 (WT) and *ΔKP1_RS12820* bacteria (1 x 10^4 CFU/mouse total). Three mutant clones are shown. At 24 hours post infection, mice were euthanized and their lungs removed. A competitive index was calculated based on WT and mutant input CFU and output CFU.

Table 3.7: Variant analysis of WT and mutant strains.

| Samples	Ref	Var	Type	Ref_nt	Var_nt	Ref_aa	Var_aa	Frameshift	Locus_tag	Gene_name	Function
ΔKP1_RS12820 #3	ACGATCTGCACC	AC	Deletion	gtggtgcagatcgtg	gtGTg	VVQIV	V	yes	KP1_1259	cyoD	Cytochrome O ubiquinol oxidase subunit IV (EC 1.10.3.-)
ΔKP1_RS12820 #3, ΔterC	GTTTTTTTTTA	GTTTTTTTTTTA	Deletion								
WT, ΔKP1_RS12820 #3, ΔterC	TAC	TAAC	Insertion	ttacct	tTAACct	LP	LT	yes	KP1_1655		Outer membrane porin for chitooligosaccharides ChiP
WT, ΔKP1_RS12820 #3, ΔterC	C	G									LysR-family transcriptional regulator STM0952
WT, ΔKP1_RS12820 #3, ΔterC	G	A	Nonsyn	gcc	Acc	A	T		KP1_2042	acT	
WT, ΔKP1_RS12820 #3, ΔterC	T	A	Nonsyn	ctg	cAg	L	Q		KP1_2559		Beta-ketoadipate enol-lactone hydrolase (EC 3.1.1.24)
WT, ΔKP1_RS12820 #3, ΔterC	G	A									
WT, ΔKP1_RS12820 #3, ΔterC	AGC	AGGC	Insertion	cagctc	cAGGCtc	QL	QA	yes	KP1_3064	rnt	Ribonuclease T
WT, ΔKP1_RS12820 #3, ΔterC	C	T	Nonsyn	gcg	gTg	A	V		KP1_3438	dsbB	Periplasmic thiol:disulfide oxidoreductase DsbB, required for DsbA reoxidation
WT, ΔKP1_RS12820 #3, ΔterC	TCC	TC	Deletion	atccgt	aTCgt	IR	I	yes	KP1_1267	bolA	Cell division protein BolA
WT, ΔKP1_RS12820 #3, ΔterC	C	A							KP1_6039		tRNA-Leu
WT, ΔKP1_RS12820 #3, ΔterC	T	G									
WT, ΔKP1_RS12820 #3, ΔterC	CG	CTG	Insertion	gcgcgt	gcCAGgct	AA	AS	yes	KP1_p130	pbrA	Lead, cadmium, zinc and mercury transporting ATPase (EC 3.6.3.3) (EC 3.6.3.5)
WT, ΔKP1_RS12820 #3, ΔterC	A	G	Nonsyn	att	aCt	I	T		KP1_p158	hnh	Hnh endonuclease
WT, ΔKP1_RS12820 #3, ΔterC	CC	GG	Nonsyn	aacctg	aaGGtg	NL	KV		KP1_p166	rsd	Resolvase
WT, ΔKP1_RS12820 #3, ΔterC	G	C	Nonsyn	gtc	Ctc	V	L		KP1_p166	rsd	Resolvase
WT, ΔKP1_RS12820 #3, ΔterC	T	G	Nonsyn	gtc	gGc	V	G		KP1_p089		hypothetical protein
WT, ΔKP1_RS12820 #3, ΔterC	C	G	Synon	gggg	ggC	G	G		KP1_3579	yeeO	Uncharacterized transporter YeeO

The sugar transport locus was present in 77.3% (n=75/97) of *K. pneumoniae* clinical isolates in our cohort, but not in *K. quasipneumoniae* or *K. variicola* (Figure 3.7A), suggesting this locus may a marker of *K. pneumoniae* as a more pathogenic species, and not a pathogenicity locus *per se*. However, multivariable analysis to control for bacterial species showed that presence of the psicose locus remains associated with infection independent of species (Table 3.8).

Table 3.8: Multivariable model for psicose metabolism locus association with infection.

Variable	Odds Ratio	95% CI	*P* value
Psicose metabolism locus	11.73	1.25-110	.031
Fluid & Electrolyte Disorders	2.95	0.82-10.6	.098
K. pneumoniae species	1.22	0.05-30.5	.903
Minimum Serum Glucose (mg/dL)	1.04	1.01-1.08	.012
Body Mass Index Baseline (kg/m^2)	0.83	0.72-0.97	.015
White Race	0.09	0.01-0.64	.016
Peripheral Vascular Disease	0.02	<0.01-1.01	.051

3.3 Discussion

3.3.1 Summary of findings

Klebsiella pneumoniae is a frequent colonizer of hospitalized patients, as well as the third most common cause of hospital-acquired infections in the United States [33]. In recent years there has been an emergence of both antibiotic-resistant and hypervirulent community-acquired strains of *K. pneumoniae*. As these infections worsen and become increasingly difficult to treat, understanding virulence and pathogenesis mechanisms of these bacteria becomes even more important. Therefore, the objective of this study was to identify genes in the *K. pneumoniae* accessory genome that are associated with clinical infection, and determine if a set of host and bacterial factors can predict infection in patients. Using a case-control study design and a novel comparative genomics method (Pathogenicity-associated locus sequencing; PAL-Seq), we were able to identify five genetic loci that are significantly and independently associated with infection in our sample set. We also identified five patient factors that are significantly associated with infection. These five patient factors along with three of the associated genes were found to be highly predictive for infection in this sample set. Presence of the tellurite locus had the strongest association with infection, significantly greater than any patient factor. The sugar transport locus was associated with human infection and transports psicose. Our results suggest that genes in the accessory genome of *K. pneumoniae* are associated with HAIs and could provide a basis for novel predictive diagnostic approaches to identify colonized patients at risk of subsequent *K. pneumoniae* infection.

3.3.2 Patient factors are associated with clinical infection

The purpose of the initial clinical model was to enable identification of bacterial genes independently associated with infection by normalizing for any differences in patient factors

between cases and controls. After matching for age, sex and date of collection, our initial bivariable analysis (Table 3.1) did not show any significant associations with infection except for white race, which was inversely associated. However, some trends toward association were seen. It should be noted that case and control patients were not from the same patient population. Therefore, patients may be exposed to different strains of *K. pneumoniae* depending on whether they were located in an ICU or elsewhere. Our final multivariable model identified fluid and electrolyte disorders, consistent with our findings in a previous study [168], but other factors differed. The differences in associated patient factors between these two studies may be due to the smaller sample size of the current study, which had just over 100 patients compared to 1765 patients in our previous study. It is also worth noting that in this study our case patients were from across the hospital, while the control patients were solely from a screening program for vancomycin resistant *Enterococci* (VRE).

3.3.3 Bacterial genes associated with clinical infection

A plasmid-encoded tellurite resistance locus was identified as associated with infection by our PAL-Seq analysis. The antibacterial properties of tellurite has long been known [181]. Plasmid-mediated bacterial resistance to tellurite is observed in several *Enterobacteriaceae*, including *K. pneumoniae.* In these bacteria, tellurite resistance is dependent on the *ter* operon. In *K. pneumoniae* and some other species, this locus consists of two operons (*terZABCDEF* and *terWXY*) separated by seven uncharacterized putative open reading frames. The mechanism of bacterial tellurite resistance is unclear, though it is suspected to be linked to resistance to superoxide and other reactive oxygen species. The *ter* operon has been associated with hypervirulent *K. pneumoniae* clonal groups [177]. This operon has also been found to contribute to pathogenicity in mammalian models of infection with *Bacillus anthracis* [182]. Multiple genes

in the operon have also been associated with phage inhibition and colicin resistance [178]. Inclusion of this operon in the multivariable model significantly improved the predictability of the model (Figure 3.1B). This locus also appears to be associated with infection independent of *K. pneumoniae* lineage (Figure 3.7). Although not required for a hypervirulent strain to cause murine pneumonia, it may directly contribute to pathogenesis at other sites of infections. Alternatively, it may be a robust genotypic marker that is strongly linked genetically to virulence genes on transmissible plasmids. Regardless, the corresponding tellurite resistance phenotype can be easily screened for in patient samples using either pre-formulated media or through simple modifications of MacConkey agar routinely used in clinical laboratories [177, 179].

Of the top ten representative genes independently associated with infection, both *KP1_RS12850* and *KP1_RS12840* appear to be part of the same putative sugar metabolism locus with slightly different frequencies. These genes were independently associated with infection in our cohort (Table 3.5 and Figure 3.1). We determined that deletion of the permease gene in our sugar metabolism locus impaired growth on D-psicose, a carbon substrate (Figure 3.15). This is consistent with a recent study that associated the same sugar metabolism locus with psicose uptake and degradation in *K. pneumoniae* strains using comparative genomics [183]. We further identified an *in vivo* fitness defect attributable to the permease gene using multiple mutant clones, further confirming the association of this locus to infection (Figure 3.17). Complementation of *ΔKP1_RS12820* in mutant clone #3 resulted in partial rescue of the growth defect on psicose (Figure 3.15B), providing functional evidence for the role of this gene in utilizing psicose. However, complementation in this clone did not show a significant difference in fitness compared to WT *in vivo* (Figure 3.17). As noted, we did identify a second mutation in a different gene in *ΔKP1_RS12820* clone #3 (Table 3.7). It is possible that the *in vivo* fitness defect is attributable to

this second mutation. However, a second mutant clone also displayed an *in vivo* fitness defect, and a third clone was trending toward significance (Figure 3.17). Further investigation is needed to more clearly define the role of the sugar metabolism locus in infection *in vivo*.

D-psicose is also known as allulose and is a C-3 epimer of D-fructose. D-psicose/allulose is a rare sugar, occurring only in small quantities in nature. It is, however, encountered as a natural sweetener for food and drink [184]. Upon consumption, D-psicose is absorbed in the small intestine and excreted in the urine [185], providing two potential sites where colonizing *K. pneumoniae* could encounter this substrate. Though *K. pneumoniae* is a frequent colonizer of the large intestine in humans, it has been commonly identified in cases of small intestine bacterial overgrowth (SIBO) [186, 187]. It is possible that presence of D-psicose could enhance intestinal *K. pneumoniae* colonization, although this has not been investigated. If true, this could identify D-psicose as a risk factor for progression to disease, since we have already identified an association between intestinal colonization and infection. Consumption of allulose has been associated with improvements in insulin resistance and glucose tolerance in patience with type two diabetes mellitus (T2D) and has been posited as a weight loss aid, which could further help control T2D [188]. Infection with *K. pneumoniae* is already associated with diabetes [25, 26, 38, 40]. It is therefore possible that therapeutic consumption of allulose may increase risk of infection in an already susceptible population.

Among the top genes that were significantly and independently associated with clinical infection were a putative deoxygluconate dehydrogenase (*KPN_RS09590*) and a hypothetical protein (*KPNJ1_01715*). Deoxygluconate dehydrogenases are enzymes categorized under EC: 1.1.1.125 and are a class of oxidoreductases that are thought to play a role in pentose and glucuronate interconversions [189]. Amongst the strains in our reference pan-genome, this gene

was present only in MGH 78578, which caused pneumonia. It was also present in 56 (49.1%) of our patient isolates (30 colonizing [39.4%] vs 26 infecting [68.4%]). The hypothetical protein nucleotide sequence *KPNJ1_01715* is less than 100bp and was only annotated in NJST258_1 (a *Klebsiella pneumoniae* carbapenemase-producer) in the reference pan-genome. A BLAST query against the non-redundant database identified 100% amino acid homology to sequences in two other *K. pneumoniae* genomes (NJST258_2 and NTUH K2044). *KP1_3870* in NTUH K2044 has been discontinued as an open reading frame in GenBank and was not part of our pan-genome reference strain. At the time of this publication, *KPNJ1_01715* has also been removed as an open reading frame from GenBank. In both of our reference genomes, the former hypothetical protein nucleotide sequence sits between two divergently transcribed genes, phosphoporin PhoE (*KPNJ1_RS08390* and *KP1_RS18000*) and phosphotransferase RcsD (*KPNJ1_RS08380* and *KP1_RS18005*). It sits 131bp from PhoE, and 542bp from RcsD in NJST258_1, and 543bp from RcsD and 131bp from PhoE in NTUH-K2044. It is possible that this represents a regulatory region of these genes.

3.3.4 Limitations

Though we were able to successfully identify genes independently associated with *K. pneumoniae* infection, there are some limitations to our PAL-Seq approach. Our primary limitation is use of only five strains to make our reference strain. With the capabilities of whole genome sequencing we are beginning to understand that the *K. pneumoniae* pangenome (conserved and accessory genes) is open, indicating that new genes will continue to be identified [30]. Since we only included five strains in our reference strain, it is probable that we did not cover the breadth of genes represented among clinical isolates. Of the approximately 30,000 protein coding sequences identified to date, we sampled just over 8,000 in our study. Similarly, we did not include

a *K. quasipneumoniae* strain in our reference strain, though we identified three patient isolates as *quasipneumoniae* species. PAL-Seq analysis showed that these mapped poorly to our reference strain. Whole genome sequencing of all three species has identified a 3-4% nucleotide divergence (across core genes) between phylogroups, compared to ~0.5% divergence within phylogroups, indicating that allelic differences may account for poor mapping of our *K. quasipneumoniae* isolates [30]. It is possible that this affected our gene frequency calculations, though unlikely that it greatly affected the overall outcome since only three isolates were identified as *quasipneumoniae* species. Due to these issues, it is likely that there are pathogenicity-associated genes among our isolates that were present but that we failed to identify. However, we chose to use well-characterized, genetically tractable strains that can be used in animal models for our reference pan-genome. This approach facilitated translation from association with human infection to phenotypic characterization *in vitro* and in animal models of infection.

3.3.5 Conclusions

Overall, we were able to identify patient and bacterial factors that are independently and significantly association with, as well as highly predictive of, *K. pneumoniae* infection in patients using our PAL-Seq method. This bacterial GWAS provides a proof of principle in using clinical isolates to gain insight into bacterial pathogenesis. Our findings can be translated back to the clinical microbiology lab to screen for at-risk patients, or translated into basic science of the mechanisms of bacterial pathogenesis.

3.4 Experimental procedures

Patient population and setting. The study was conducted at the University of Michigan Health System (UMHS), a tertiary care hospital with more than 1,000 beds, in Ann Arbor,

Michigan. Approval for this study was granted by the Institutional Review Board of the University of Michigan Medical School. During a three-month period from July 30 to October 31, 2014, rectal swabs from 1800 adult ($\geq$18 years) patients from the intensive care unit (ICU) or hematology/oncology wards were screened for *K. pneumoniae*. Concurrently, extra-intestinal *K. pneumoniae* isolates were obtained from the clinical microbiology lab for approximately 300 adult patients from all units within UMHS. Patient demographic characteristics and clinical information was obtained through the electronic medical record (EMR).

Bacterial identification and growth conditions. Rectal swabs were collected during the course of clinical care (upon unit admission, weekly, and at discharge) and were transported and stored in the ESwab Transport System (BD, Franklin Lakes, NJ) at room temperature. Within 24 hours of receipt, 1μl of inoculated ESwab media was plated to MacConkey agar (Remel, Lenexa, KS) streaked for quantification and incubated for 18–24 hours at 35ºC. The expected analytical sensitivity is 10^3 CFU/mL of ESwab media. To ensure collection of the dominant clone in each sample, three mucoid lactose fermenting (MLF) colonies were isolated as potential *K. pneumoniae* and subcultured onto blood agar plates (BAP) (Remel, Lenexa, KS) [158]. If fewer than three MLF colonies were present in a particular sample, all were subcultured. Only one isolate per patient was used for analysis unless otherwise indicated. Isolates for each patient were usually the first isolate of the three collected. Bacterial identification was performed using matrix-assisted laser desorption/ionization time of flight (MALDI-TOF). Isolates were stored at −80ºC in Luria-Bertani (LB) Broth containing 20% glycerol, and were grown on either BAP or LB plates at 30ºC overnight unless otherwise indicated.

Bacterial strains and media. Klebsiella pneumoniae NTUH-K2044 and mutants were cultured at 30°C on LB agar supplemented with kanamycin (50μg/ml), spectinomycin (50μg/ml), chloramphenicol (20μg/ml), or gentamicin (6-10μg/ml) as indicated. Isolates were also cultured in LB broth at 37°C with shaking.

Definitions. In patients without *K. pneumoniae* infection, *K. pneumoniae* colonization was defined as a positive rectal swab culture for *K. pneumoniae* at any point during the hospital admission. Patient EMRs were reviewed for any positive culture for *K. pneumoniae* within 90 days of rectal swab culture. Bloodstream infection was defined as any positive blood culture for *K. pneumoniae*. Pneumonia was defined based on a positive *K. pneumoniae* respiratory culture and other Infectious Diseases Society of America (IDSA) diagnostic criteria [159]. Controls are patients rectally colonized with *K. pneumoniae* but with no positive *K. pneumoniae* extra-intestinal cultures within 90 days post-colonization. Controls were matched 2:1 with cases based on sex, ± 10 years age range, and ± 3 week collect date range. Control pools were generated using SAS and final matches were randomly selected using Microsoft Excel.

Statistical analysis for clinical models. The clinical and multivariable modeling was conducted in R version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria). Unless otherwise specified, a significance threshold of *P* <.05 was used for all analyses. Initial analyses included descriptive statistics and exploring various variable constructions for continuous variables and categorical variables with more than two levels. Bivariable analyses for the outcome of invasive infection (cases) were conducted using conditional logistic regression via the *survival* package, version 2.38 [190]. Where different possible variable constructions existed, the ones with

best fit by *P* value on bivariable analyses were carried forward. Variables with $P < .2$ on these initial bivariable analyses were eligible for inclusion in a final multivariable model. The final clinical model was constructed through backwards elimination with a cutoff α of .05 for the likelihood ratio test. Interactions among variables in the final model were assessed and included if significant. Candidate genes from the PAL-Seq analysis (detailed below) that associated with invasive infection were then adjusted individually for the variables in the clinical model. The candidate genes that remained significant after this initial adjustment were all added to the clinical model, and backwards elimination was again conducted to arrive at the final model that incorporated both patient-level and bacterial genetic features. Interaction testing proceeded as before. The overall performance of the multivariable models was assessed through construction of receiver operator characteristic curves (ROCs) and calculation of the area under the curve. Bootstrapped confidence intervals for specificity at each level of sensitivity were calculated and plotted for the ROCs via the package *pROC* version 1.8 [191]. Multivariable model AUCs were compared using the DeLong method [192].

Statistical analysis for PAL-Seq. Conditional logistic regression models were used to identify genes that are significantly different between cases and controls. This model takes into account the matched nature of the data and was used for each gene to obtain a *P* value to indicate significance of that gene. Finally, *P* values for all genes were adjusted to *Q* values [193] to control the false discovery rate.

Whole genome sequencing. Bacterial genomic DNA (gDNA) was isolated using the PUREGENE ® DNA Purification Kit (PUREGENE®, Minneapolis, MN). Purified gDNA was

sent to the University of Michigan DNA Sequencing Core where it was sheared (200bp) and prepared as a multiplex library with unique barcodes for each sample. Whole genome sequencing was performed using the HiSeq 4000 sequencing system (Illumina, San Diego, CA).

Pathogenicity-associated locus sequencing. In generating the reference pan-genome, we kept all orthologous genes present in order to account for potential allelic differences amongst our sample isolates. BWA (version 0.7.5a-r405) was used to align the reads in its default mode [194]. For 50 nucleotide reads, it allows up to three mismatches, and it outputs up to one optimal alignment for each read. If multiple optimal matches are found, it reports one randomly. Because we concatenated closely related *Klebsiella* strains in the pan-genome, we expect to have duplicates of identical sequences due to homologous genes, and therefore random alignment among these genes. Since our gene-bin approach groups homologous genes in one bin we took the sum of alignment counts in the bin to represent the gene coverage. In each sample, the normalized count sum values from all bins display a bimodal distribution. A K-means clustering approach [195] was applied to binarize the data into "0" or "1" indicating "absence" or "presence" of genes respectively. The per-gene data was matched with gene bins to produce per-bin count data. For each bin, raw counts were normalized by dividing the raw counts by million of deduped alignments in that sample, and then by kilobase of length for that gene.

Construction of mutants. Lambda Red mutagenesis was performed as previously described [196, 197] with the following modifications. Electrocompetent cells were prepared by culture in LB broth containing a final concentration of 0.5μM EDTA at 37ºC with shaking until an optical density at 600nm (OD$_{600}$) between 0.5 and 0.6. The culture was placed on ice for 45 minutes and

centrifuged in sterile cold bottles at 8,000 RPM for 15 min at 4°C. The supernatant was decanted and bacteria were washed and centrifuged in ice-cold sterile volumes of 25mL 1mM HEPES, 25mL distilled water, and 10 mL of 10% glycerol. Pellets were brought to a final density of 2 x 10^{10} CFU/mL and stored in 50μL aliquots at -80°C. A modified pKD46 plasmid, encoding spectinomycin resistance was electroporated into NTUH-K2044 using a 0.1-cm-gap cuvette at 1.8 kV, 400Ω, 25μF, with a Bio-Rad Micropulser. Cells were recovered in room temperature SOC media and incubated overnight at 30°C with shaking. Electrocompetent cells containing the pKD46 plasmid were prepared as above, except cultures were grown for approximately 4 hours at 30°C in LB broth containing spectinomycin, 50mM L-arabinose, and 0.5μM EDTA.

Complementation of mutants. To complement the *KP1_RS12820* mutants, PCR products containing the open reading frame were inserted into pCR 2.1 using TOPO TA cloning (Life Technologies, Carlsbad, CA) and subsequently ligated into pBBR1MCS-5 following digestion with HindIII and Xho1. The complementation plasmid was electroporated into WT and mutant *K. pneumoniae*.

Murine pneumonia model. Six- to nine-weeks old mice C57BL/6 mice were anesthetized with isoflurane and then inoculated intrapharyngeally with ~1 x 10^4 CFU of bacteria per mouse (1:1 mixture of NTUH-K2044 (WT) and mutant bacteria, either *ΔKP1_RS12820* or *ΔterC*, or WT and complemented mutant). After 24 hours, mice were euthanized by CO_2 asphyxiation and lungs and spleens were removed, and homogenized in 1mL PBS, and cultured on both LB agar and LB agar supplemented with kanamycin. The competitive index was calculated as (mutant lung CFU/WT lung CFU)/(mutant inoculum CFU/WT inoculum CFU).

Growth curves. Bacterial strains were cultured overnight in LB broth. On the following day, cultures were incubated in LB or heat inactivated human serum supplementation at a starting density of 2.6×10^6 CFU/mL in LB. Cultures were also incubated in M9 (Life Technologies) media with or without glucose. Culture concentrations were normalized and washed in 2X minimal media and then diluted to a final concentration of 1.4×10^7 CFU/mL in the same. Each well contained glucose or psicose at a final concentration of 5mg/ml. Diluted inoculum was added to each well 1:1. Strains were cultured for 24 hours at 37°C. Absorbance readings at 600nm were taken every 15 minutes using an Eon microplate spectrophotometer with Gen5 software (BioTek, Winooski, VT).

Bioscreen assay. Bacterial strains were cultured overnight in LB broth either aerobically or in a vinyl anaerobic chamber (Coy Laboratory Products, Grass Lake, MI). On the following day, cultures were incubated in M9 (Life Technologies) media with or without various carbon supplementation as described elsewhere [180, 198]. Strains were cultured for 24-196 hours at 37°C either aerobically or in the anaerobic chamber respectively. Absorbance readings at 600nm were taken every 15 minutes using an Eon microplate spectrophotometer with Gen5 software (BioTek, Winooski, VT).

Multilocus sequence typing (MLST) and phylogenetic tree. Raw sequencing reads (Illumina single end reads) for all 117 patient isolates, as well as FASTA files from GenBank for each reference isolate, were run through the Center for Genomic Epidemiology MLST typing scheme (https://cge.cbs.dtu.dk/services/MLST/). KpI, KpII, and KpIII group isolate sequence

types were obtained from Maatallah et al [135], and MLST sequences for each Kp group isolate were obtained using http://bigsdb.web.pasteur.fr. Resulting sequences for each of the seven genes were concatenated for each isolate. Concatenated MLST sequences were aligned using ClustalX 2.1. The phylogenetic tree was constructed using MEGA6 [162] based on the neighbor-joining method (500 bootstrap replicates) and Jukes-Cantor distance. The tree was then uploaded into the Interactive Tree of Life database for annotation (http://itol.embl.de/). Gene presence or absence was determined using the Normalized Sum Counts from our PAL-Seq data.

Single nucleotide variant identification. Quality of raw reads was assessed with Fastqc [199], and Trimmomatic [163] was used for trimming adapter sequences and low quality bases. Variants were identified by: 1) mapping filtered reads to the finished KPNIH1 reference genome (GenBank accession no. CP008827) using the Burrows–Wheeler short-read aligner (BWA), 2) discarding PCR duplicates with Picard, and 3) calling variants with SAMtools and bcftools. Variants were filtered from raw results using GATK's VariantFiltration (QUAL > 100, MQ > 50, > 10 reads supporting variant, FQ <0.025). In addition, a custom python script was used to filter out single nucleotide variants that were: 1) <5 bp in proximity to indels 2) <10 bp in proximity to another variant, or 3) not present in the core genome. Lastly, for phylogenetic analyses, Gubbins was applied to filter out recombinant variants [200].

WGS phylogenetic analysis. Maximum likelihood trees were constructed in RAxML*(24)* wherein variants were modeled with a general-time reversible model. Bootstrap analysis was performed with the number of bootstrap replicates determined using the bootstrap convergence

test and the autoMRE convergence criteria (-N autoMRE). Bootstrap support values were overlaid on the best scoring tree identified during rapid bootstrap analysis (-f a).

Variant analysis of mutant strains. Genomic DNA from NTUH-K2044, *ΔKP1_RS12820*, and *ΔterC* was purified using the DNeasy Blood and Tissue kit (QIAGEN Inc., Germantown, MD). DNA was sequenced using the Illumina NexteraXT kit on the Illumina MiSeq using a 2x250bp V2 kit. Whole genome sequences were assembled using the Genome Assembly service from PATRIC (www.patricbrc.org). Sequence variants were identified using the PATRIC Variation Analysis service. The NTUH-K2044 genome from GenBank was used as a reference genome.

3.5 Notes

CHAPTER IV

Discussion

4.1 Summary of thesis

Klebsiella pneumoniae has traditionally been considered an endemic hospital-acquired pathogen, although it is increasingly being recognized as an epidemic antibiotic-resistant and emergent hypervirulent pathogen. The accessory genome is likely responsible for the differences in virulence between these types of *K. pneumoniae*. Is it also likely that patient colonization is an important step in epidemiology of infection for the distinct disease presentations. The focus of this thesis dissertation is defining the risk factors for infection with *K. pneumoniae* in hospitalized patients. The central hypothesis of this work is that colonized patients are more likely to develop infection, and that genes in the accessory genome promote clinical infection.

The first part of this hypothesis was explored in Chapter II using a clinical model to address the association between colonization and infection with *K. pneumoniae*. We determined that 23% of adult intensive care unit (ICU) and hematology/oncology patients were colonized at some point during their stay in the University of Michigan Health System. A recent study corroborates these findings, demonstrating a 19% gastrointestinal colonization rate [18]. It was determined that colonization with *K. pneumoniae* is significantly and independently associated with subsequent *K. pneumoniae* infection. This was the strongest independent association identified among patient variables. The concordance of colonizing and infecting isolates within patients was also addressed using multiple molecular techniques, which identified that patients are frequently infected with

their colonizing strain. These studies identified colonization as a key step in progression to clinical infection among hospitalized patients. The implications of these findings for identification of at-risk patients will be discussed in this chapter.

Chapter III addressed the second part of this hypothesis through clinical models, whole genome sequencing, and a novel comparative genomics technique. These methods identified bacterial genes significantly and independently associated with clinical infection in our sample set. Additionally, *in silico*, *in vitro*, and *in vivo* studies further characterized the tellurite resistance loci and the sugar metabolism locus that were independently associated with human infection. Specifically, phylogenetic analysis showed that the tellurite resistance locus likely has a true association with infection rather than being a marker for a specific lineage. Growth on various carbon substrates identified psicose as the likely substrate for the sugar metabolism locus. This chapter will examine the potential diagnostic value of these findings.

4.2 Host factors contribute to epidemiology of infection

Chapter II demonstrated that there is a significant and independent association between colonization with *K. pneumoniae* and infection with *K. pneumoniae*. Despite the association between colonization and infection, we and others have identified patients who either remain asymptomatically colonized, or who are never colonized and still develop infection. The former likely represents a balance of immunocompetent patients able to defend against bacterial infection on the one hand, and bacterial virulence – or avirulence – on the other. Conventionally, infection with *K. pneumoniae* is recognized as a disease that occurs in an already immunocompromised host [17]. In keeping with this paradigm, we did identify several host factors that contribute to our most predictive model of infection (Tables 2.4 & Table 3.6). Though colonization with *K. pneumoniae*

showed the strongest association with infection, other host factors were also identified as associated with and predictive of infection (Table 2.4). Interestingly, Fluid & Electrolyte Disorders contributed to our predictive model in studies from both Chapter II and Chapter III (Tables 2.4 & 3.6). Fluid and electrolyte disorders is a broad comorbidity category defined by the Elixhauser score and encompasses several disorders that present as an imbalance in fluids and/or electrolytes [150]. This association is potentially a marker for general severe underlying disease, since several comorbidities result in fluid and electrolyte imbalances due to diarrhea or vomiting. It is also possible that it points to specific comorbidities associated with *K. pneumoniae* infection. Fluid and electrolyte disorders are often seen in patients with diabetes mellitus, as well as in burn patients [201, 202]. Infection due to *Klebsiella* species has been shown to be associated with both disease states [25, 26, 38, 40, 203, 204]. If *K. pneumoniae* infection is a disease of opportunity, it is possible that patients in our study who were colonized and never progressed to infection were better able to combat colonizing strains due to more effective immune responses. Alternatively, patients may be colonized with an avirulent strain of *K. pneumoniae.* It is known that some strains are unable to cause disease in murine models of infection [32]. It is therefore possible that some of these patients were colonized with a non-pathogenic strain that never progressed to infection. Further studies assessing these nuances of the association between colonization are needed.

Patients who did not present as colonized with *K. pneumoniae* but still developed infection may represent a limitation of our approach. It is possible that our screening method was not sensitive enough to pick up all colonized patients, particularly those colonized with lower numbers of *K. pneumoniae*. A more sensitive approach may include molecular-based methods to identify presence or absence of specific organisms, or development of more selective media to obtain isolates. Alternatively, patients may have been colonized in the nasopharynx prior to infection,

particularly those with respiratory infections [13]. The association of nasopharyngeal colonization and infection was not assessed in this study. It is further possible that some patients were truly never colonized and still developed infection. If true, these patients likely encountered other sources of acquisition, such as the environment, medical equipment, or healthcare workers hands [20, 21]. It is important to recognize that while colonization is an important step in the progression to infection, there are multiple potential reservoirs for infection.

4.3 Identifying at-risk patients may help guide empiric therapy

Understanding how *K. pneumoniae* progresses to infection is important for effective clinical care. Ideally, identifying and characterizing *K. pneumoniae* colonizing isolates from hospitalized patients would help guide empiric treatment. Having determined which patients are colonized with *K. pneumoniae*, and therefore identifying patients at risk for extra-intestinal infection, characterization of the colonizing isolate to pre-emptively assess potential infecting isolates makes sense. Supporting this is the knowledge that accurate identification of the causative agent for serious infections, such as bloodstream infections, and initiation of appropriate antimicrobial therapy results in decreased mortality [205, 206]. With this in mind we carried out antimicrobial susceptibility testing (AST) on our concordant isolates to determine if they showed the same resistance profiles. However, our isolates were highly susceptible (Table 2.5) and many had identical patterns, meaning we were unable to demonstrate statistically significant concordance of AST profiles. Of the four colonizing-infecting isolate pairs that showed intermediate or resistant profiles, two displayed concordant results. The third had discrepancies that were not reproducible, and the fourth had a reproducible discrepancy. This may not be significant, but it raises a potential issue that may be encountered in the clinical microbiology

laboratory; specifically, that colonizing isolates may develop mutations that lead to antibiotic-resistance or acquire genes that confer antibiotic-resistance during the intervening time between colonization and infection. Within the gastrointestinal environment the potential for horizontal gene transfer, and therefore the acquisition of genes conferring antibiotic-resistant genes, is possible [207]. With greater than 10^{10} bacterial organisms per gram of feces in the gastrointestinal tract, representing over 500 bacterial species, the reservoir for gene transfer is large. Furthermore, as discussed earlier, mutations in chromosomal genes can lead to antibiotic-resistant phenotypes [91, 208]. Therefore, even if the antimicrobial susceptibility patterns are determined for colonizing isolates, it is likely that re-verification of AST would be required for the subsequent infecting isolate to ensure accurate results.

4.4 Identifying colonization may aid in infection prevention

It is likely that the association between colonization and infection with *K. pneumoniae* can be effectively used to identify patients at risk for infection. Many institutions currently have screening protocols in effect to identify carriers of organism that pose potentially serious infection control issues, such as methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococci* (VRE). It is therefore possible to implement screening protocols to identify patients colonized with *K. pneumoniae*. Identification of carriers could trigger interventions such as contact precautions, environmental measures, or decolonization procedures, as with MRSA and VRE [209].

Screening for nosocomial pathogens such as MRSA and VRE is considered necessary due to the increased morbidity, mortality, and cost associated with infections from these organisms. Between 2006 and 2007, one study identified a 77% increase in infections due to MRSA [210].

Another study identified a disease rate for MRSA infections of 31.8/100,000 with an almost 20% annual mortality rate [211]. Screening for MRSA has been demonstrated to decrease MRSA acquisition, pointing to the success of these screening programs in infection prevention [212]. VRE have similarly gained prominence in recent years as a cause of healthcare-associated infections. These bacteria have a high colonization rate and a low infection rate. Despite the low infection rate severely immunocompromised patients are at increased risk for infection with VRE, which has prompted surveillance measures in susceptible populations [213]. VRE are particularly problematic due to increased prevalence, gastrointestinal colonization, prolonged survival in the hospital environment, transmissibility between patients, ability to efficiently transfer antibiotic-resistant genes, and high mortality rates in severely immunocompromised patients [214]. For these reasons, as well as the difficulty treating these frequently multidrug-resistant organisms, screening for VRE to reduce infections and transmission has been implemented at several healthcare facilities, specifically in units where patients are immunocompromised and thus at high risk for infection [213, 215]. Active surveillance is shown to be associated with a lower rates of VRE bacteremia in high risk units compared to no surveillance [216].

K. pneumoniae presents similar burdens to healthcare as MRSA and VRE. *Klebsiella* have been identified as the third highest cause of hospital-acquired infections in the United States, causing 9.9% of HAIs, and the second leading cause of Gram-negative bacteremia [33]. Furthermore, these organisms are known to infect immunocompromised patients. There has also been an increase in antibiotic-resistant organisms, including a pan-resistant *K. pneumoniae* clinical isolate [99], as well as the emergence of hypervirulent strains. *K. pneumoniae* are easily transmitted and have been responsible for multiple outbreaks [141, 217]. Taken together, these

data suggest that investigating the implementation of similar infection control procedures for the reduction of *K. pneumoniae* transmission and infection has merit.

There are several infection control methods that could be implemented to aid in combating transmission and infection due to colonizing *K. pneumoniae*. Decolonization by antibiotics has been attempted to eradicate both gastrointestinal MRSA and VRE, and has been explored in antibiotic-resistant ESBL-producing and CRE Gram-negative bacteria. However, efficacy of decolonization by this method is uncertain. In MRSA isolates, there have been reports of higher incidence of subsequent carriage following decolonization [218, 219]. Resistance to the decolonizing agent has also been reported. A recent review determined that efficacy of eradication is limited in attempts to decolonize multidrug-resistant enteric organisms from the gastrointestinal tract [220]. Therefore, alternatives to decolonization should be explored. Intestinal dysbiosis can occur due to various pathological conditions. Understanding the mechanisms behind how intestinal dysbiosis occurs and is maintained, as well as the mechanisms bacteria use to outcompete one another, may also help in controlling colonization with *K. pneumoniae* and subsequent infection [221, 222]. Microcin administration has been demonstrated as an effective way to reduce pathogenic Enterobacteriaceae in the gastrointestinal tract by mediating inter- and intra-species competition among Gram-negative organisms [223]. Fecal transplants have been shown to inhibit gastrointestinal colonization due to antibiotic-resistant organisms including *K. pneumoniae* [224]. Chlorhexidine baths have already been shown to decrease the risk of *K. pneumoniae* CRE colonization and infection and are suggested as a measure to reduce HAIs due to CRE in high risk populations [225, 226].

Further assessment of necessity and effectiveness of some of these measures for controlling transmission by and infection due to *K. pneumoniae* would need to be undertaken before

implementation. Furthermore, the decision whether to implement screening protocols to identify those colonized with *K. pneumoniae* is likely one for individual facilities to make, in conjunction with Infection Control, Infectious Diseases, and Clinical Microbiology, since there are regional and institutional differences in infection and resistance patterns.

4.5 Identification and characterization of pathogenicity-associated loci

Chapter III identified bacterial factors that are associated with *K. pneumoniae* clinical infection, and characterized two pathogenicity-associated loci (PALs). This was done using a case-control study comparing asymptomatically colonized control patients and patients who met case definitions for extra-intestinal infection. Using whole genome sequencing and a novel comparative genomics technique, pathogenicity-associated locus sequencing (PAL-Seq), we identified five PALs significantly and independently associated with infection (Table 3.5). Two PALs were further characterized: the tellurite resistance loci, and the putative sugar metabolism locus. We determined that the tellurite resistance locus did not confer a significant *in vivo* fitness defect in a mammalian pneumonia model. We used a hypervirulent strain for these studies, therefore it is possible that this locus may contribute more significantly to infection in other strains or at other anatomic sites. Alternatively, it may simply be a marker for infection or bacterial lineage rather than directly contributing to it. If this were true, it may still be useful in assessing associations with infection and could still potentially be used to predict risk of infection or assist in tracking outbreaks, although it would not provide direct insight into the pathogenesis of these strains. To investigate whether the tellurite loci are lineage-associated, we looked at the distribution of the loci among strains based on sequence type (ST) and whole genome sequence variation. We determined that the tellurite resistance loci do not cluster in any ST or genetic lineage. Other groups have identified an association between this locus and hypervirulent *K. pneumoniae* (hvKP) isolates

124

[177], so determining the presence of this locus in isolates may still be clinically useful, as discussed above. Tellurite resistance is easily screened for using commercially available media or making minor modifications to MacConkey agar [177, 179], allowing straightforward implementation of this process into clinical microbiology workflow.

There are multiple functional categories that the PALs we identified could represent. First, it is possible that these PALs promote colonization and not virulence directly. These factors could promote intestinal dominance, which in turn leads to infection in the right host. For example, a study assessing allogenic hematopoietic stem cell transplant patients determined that intestinal domination by proteobacteria increased the risk of Gram-negative rod bacteremia five-fold in this population [23]. Similarly, our method may have selected for factors that promote colonization, but that also play a role in infection. Examples of this are seen in bacterial factors that promote nasal colonization by protecting against innate immunity, such as the polysaccharide capsule in *Streptococcus pneumoniae* [227]. The capsule aids in bacterial evasion of complement-mediated phagocytosis and killing, aiding in both colonization and in establishing systemic infection. It is further possible that a PAL may provide bacteria with an advantage that allows them to survive in the bloodstream or other site of infection, but not colonization. Finally, it is possible that some PALs may be strong markers for other genes which have functional effects during colonization and infection. Further work needs to be done to better understand the role the identified PALs play in colonization and infection.

4.6 Species misidentification may impact clinical care

Using MLST and whole genome sequencing, we also confirmed that at our facility there are isolates clinically identified as *K. pneumoniae* that are other *Klebsiella* species, which agrees

with current literature (Figures 3.3-3.5 & 3.7) [132-136]. As already discussed, accurate identification of *Klebsiella* species is potentially important for effective clinical care. These distinct species cause the same spectrum of disease, but more studies are needed to determine if clinically relevant differences exist between them. It has been demonstrated that approximately 50% of *K. quasipneumoniae* are antibiotic-resistant, therefore accurately identifying this species may alert physicians to the possibility of antibiotic-resistance [30]. However, this species represents the smallest percentage of the three that are frequently identified as *K. pneumoniae* species. Therefore, the rate of occurrence of infections due to this species should be assessed as well as confirmation of the high rate of antibiotic-resistant isolates in order to accurately evaluate the necessity of species distinction. Similarly, *K. variicola* has been associated with bloodstream infections that cause a higher mortality [135]. However, it is unclear if this is a regional association or if it is ubiquitous. The sample size of our study was too small to help verify these findings. *K. oxytoca* is another species that is a frequent cause of hospital-acquired infections. It is responsible for fewer infections than *K. pneumoniae*, but *K. oxytoca* has been associated with serious infections in neonates, suggesting a necessity for species identification of this organism [228, 229]. Current bacterial identification systems do not accurately distinguish between *K. pneumoniae, K. variicola*, and *K. quasipneumoniae* species [134]. Therefore, updating system libraries to accurately identify *Klebsiella* isolates at the species level may be needed if species distinction is shown to be clinically relevant.

4.7 Clinical modeling and bacterial whole genome sequencing in patient care

Being able to predict patient infections would be a great benefit to healthcare facilities. Ideally, it would lead to a decrease in patient morbidity and mortality following infection control

measures. Furthermore, the Patient Protection and Affordable Care act requires a reduction of payments to facilities performing poorly with respect to hospital-acquired conditions, including hospital-acquired infections [230]. We have demonstrated a method incorporating clinical modeling and whole genome sequencing to identify factors associated with infection. Confirmation of these bacterial factors is needed at additional sites and with a larger sample size. However, incorporation of screening for *K. pneumoniae* colonization and clinical modeling could be useful for identifying at-risk patients. Colonizing isolates could then be characterized for relevant bacterial factors, such as the tellurite resistance loci and the psicose metabolism locus using simple phenotypic tests or targeted molecular assays. This information, along with patient information, could be used in a clinical model to assess patient risk for infection. Clinical prediction models are used to assess the relationship between baseline health status and future health outcomes, such as specific disease states [231]. These models will frequently incorporate multiple variables. An ideal clinical model will be easy to implement as well as highly accurate, meaning it will predict specific outcomes correctly. Clinical prediction models have been successfully developed to identify predictors of bacterial infections [232-234]. This concept is therefore not novel, but its application to predicting extra-intestinal infection specifically due to *K. pneumoniae* is new. With the looming threat of dual-risk strains, the ability to predict infection is appealing. This approach could similarly be extended to other pathogens or to specific sites of infection. External validation is an important step in creating a successful clinical model. Therefore, this study is currently being repeated at external facilities. Since regional patterns can exist, relevant host and bacterial factors may require independent verification at individual facilities using the methods described here to assess local patient and bacterial patterns.

Incorporating bacterial genetic information into an effective clinical predictive model requires the ability to accurately detect whether bacterial isolates possess the genes of interest. This can be achieved using either phenotypic methods or molecular methods. As discussed, there is commercially available tellurite media, or easily modified recipes for making tellurite media in house, to detect tellurite resistance in bacteria. It is also conceivable that psicose-based selective and/or differential media could be generated to assist in identifying organisms that utilize psicose, as with lactose-fermenting bacteria and MacConkey agar. The main limitation to culture-based methods is the turnaround time. It takes at least 18 hours – and frequently much longer – to receive results from culture-based methods. Molecular methods would also be useful in detecting genes of interest. One easily implemented method is multiplex polymerase chain reaction (PCR), which could identify multiple PALs in a single experiment. Molecular methods also generally have a quicker turnaround time than culture-based methods, requiring only hours instead of days, making them more attractive for rapid diagnostic purposes. Molecular methods are further useful since they identify gene presence or absence with good sensitivity and specificity. The limitation to these methods is that they are often more expensive than culture-based methods which could be prohibitive for some labs, and they frequently require a dedicated clean space which some labs may not have space for. Furthermore, molecular detection of gene presence may not reveal whether the locus is functioning, which could be relevant for pathogenesis and predicting disease epidemiology.

The advent of whole genome sequencing (WGS) opens the door for novel diagnostic approaches in clinical microbiology laboratories. Next-generation sequencing (NGS) is defined as high-throughput, deep DNA sequencing [235]. This encompasses whole genome bacterial sequencing. NGS allows for precise characterization of microbial strains, and is currently most

clinically useful for infection control measures, such as in tracking outbreaks, including those due to *K. pneumoniae* [141, 217, 236, 237]. The approach described here used bacterial WGS in tandem with clinical modeling to identify PALs. This approach may be clinically useful in identifying predictive PALs in other pathogens and guiding development of novel diagnostic approaches. Applications of NGS itself for routine diagnostic use in clinical microbiology laboratories lie further in the future. Clinical laboratories are under strict regulatory and accreditation requirements, meaning any laboratory testing – including NGS applications – require verification and validation, and application of quality control and assurance standards, as well as proficiency testing for medical laboratory scientists performing the testing. Currently, most sequencing methods and subsequent analyses are designed around a specific pathogen. Before WGS and NGS can be successfully implemented for routine diagnostic use, standardization of methods and analysis would be needed [238, 239].

4.8 Conclusions

Once recognized primarily as an opportunistic hospital-acquired pathogen, *Klebsiella pneumoniae* has begun to present as epidemic antibiotic-resistant and emerging hypervirulent strains, complicating treatment of infected patients. As antibiotic-resistance continues to increase, the threat of dual-risk organisms displaying both antibiotic-resistance and hypervirulence is probable. Therefore, understanding how these organisms cause disease and identifying ways to detect at-risk patients is important. The work presented in this thesis dissertation has identified novel risk factors for *Klebsiella pneumoniae* infection in patients. It has determined a significant and independent association between colonization and infection with *K. pneumoniae*, suggesting a window for intervention. This work has further identified and characterized bacterial genes

significantly and independently associated with infection, which could be used to identify at-risk patients. Together this work provides new avenues for research and practical implications for patient care. Initial future work includes further characterization and confirmation of the PALs identified here. The methods developed here could be used to identify additional PALs, both in *K. pneumoniae* as well as in other pathogens. These findings also provide a rationale for healthcare facilities to explore screening protocols to identify patients colonized with *K. pneumoniae*. And finally, this work opens the door for development and implementation of novel diagnostics to assist in predicting and treating infections with *K. pneumoniae* and other pathogens.

REFERENCES

1. Friedlaender, C., *Ueber die Schizomyceten bei der acuten fibrösen Pneumonie.* Archiv für pathologische Anatomie und Physiologie und für klinische Medicin, 1882. **87**(2): p. 319-324.

2. Podschun, R., et al., *Incidence of Klebsiella Species in Surface Waters and Their Expression of Virulence Factors.* Applied and Environmental Microbiology, 2001. **67**(7): p. 3325-3327.

3. Bagley, S.T., *Habitat association of Klebsiella species.* Infection control, 1985. **6**(02): p. 52-58.

4. Matsen, J.M., J.A. Spindler, and R.O. Blosser, *Characterization of Klebsiella isolates from natural receiving waters and comparison with human isolates.* Appl Microbiol, 1974. **28**(4): p. 672-8.

5. Podschun, R., A. Fischer, and U. Ullmann, *Siderophore production of Klebsiella species isolated from different sources.* Zentralbl Bakteriol, 1992. **276**(4): p. 481-6.

6. Podschun, R. and U. Ullmann, *Bacteriocin typing of Klebsiella spp. isolated from different sources.* Zentralbl Hyg Umweltmed, 1996. **198**(3): p. 258-64.

7. Struve, C. and K.A. Krogfelt, *Pathogenic potential of environmental Klebsiella pneumoniae isolates.* Environ Microbiol, 2004. **6**(6): p. 584-90.

8. Podschun, R., *Phenotypic properties of Klebsiella pneumoniae and K. oxytoca isolated from different sources.* Zentralbl Hyg Umweltmed, 1990. **189**(6): p. 527-35.

9. Anderson, D.J., et al., *Seasonal Variation in Klebsiella pneumoniae Bloodstream Infection on 4 Continents.* The Journal of Infectious Diseases, 2008. **197**(5): p. 752-756.

10. Kloos, W.E. and M.S. Musselwhite, *Distribution and persistence of Staphylococcus and Micrococcus species and other aerobic bacteria on human skin.* Appl Microbiol, 1975. **30**(3): p. 381-5.

11. Davis, T.J. and J.M. Matsen, *Prevalence and characteristics of Klebsiella species: relation to association with a hospital environment.* J Infect Dis, 1974. **130**(4): p. 402-405.

12. Wolf, B., et al., *Carriage of gram-negative bacilli in young Brazilian children with community-acquired pneumonia.* International Journal of Infectious Diseases, 2001. **5**(3): p. 155-159.

13. Farida, H., et al., *Nasopharyngeal Carriage of Klebsiella pneumoniae and Other Gram-Negative Bacilli in Pneumonia-Prone Age Groups in Semarang, Indonesia.* Journal of Clinical Microbiology, 2013. **51**(5): p. 1614-1616.

14. Dao, T.T., et al., *Klebsiella pneumoniae oropharyngeal carriage in rural and urban Vietnam and the effect of alcohol consumption.* PLoS One, 2014. **9**(3): p. e91999.

15. Fuxench-Lopez, Z. and C.H. Ramirez-Ronda, *Pharyngeal flora in ambulatory alcoholic patients: prevalence of gram-negative bacilli.* Arch Intern Med, 1978. **138**(12): p. 1815-6.

16. Pollack, M., et al., *Factors influencing colonisation and antibiotic-resistance patterns of gram-negative bacteria in hospital patients.* Lancet, 1972. **2**(7779): p. 668-71.

17. Podschun, R. and U. Ullmann, *Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors.* Clin Microbiol Rev, 1998. **11**(4): p. 589-603.

18. Gorrie, C.L., et al., *Gastrointestinal carriage is a major reservoir of K. pneumoniae infection in intensive care patients.* bioRxiv, 2016.

19.   Rose, H.D. and J. Schreier, *The effect of hospitalization and antibiotic therapy on the gram-negative fecal flora.* Am J Med Sci, 1968. **255**: p. 228-36.

20.   Jarvis, W.R., et al., *The epidemiology of nosocomial infections caused by Klebsiella pneumoniae.* Infect Control, 1985. **6**(2): p. 68-74.

21.   Casewell, M. and I. Phillips, *Hands as route of transmission for Klebsiella species.* Br Med J, 1977. **2**(6098): p. 1315-7.

22.   Donskey, C.J., *The role of the intestinal tract as a reservoir and source for transmission of nosocomial pathogens.* Clin Infect Dis, 2004. **39**(2): p. 219-26.

23.   Taur, Y., et al., *Intestinal Domination and the Risk of Bacteremia in Patients Undergoing Allogeneic Hematopoietic Stem Cell Transplantation.* Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America, 2012. **55**(7): p. 905-914.

24.   Spach, D.H., F.E. Silverstein, and W.E. Stamm, *Transmission of infection by gastrointestinal endoscopy and bronchoscopy.* Ann Intern Med, 1993. **118**(2): p. 117-28.

25.   Tsay, R., et al., *Characteristics of bacteremia between community-acquired and nosocomial klebsiella pneumoniae infection: Risk factor for mortality and the impact of capsular serotypes as a herald for community-acquired infection.* Archives of Internal Medicine, 2002. **162**(9): p. 1021-1027.

26.   Tsai, S.S., et al., *Characteristics of Klebsiella pneumoniae bacteremia in community-acquired and nosocomial infections in diabetic patients.* Chang Gung Med J, 2010. **33**(5): p. 532-9.

27.   Happel, K.I. and S. Nelson, *Alcohol, immunosuppression, and the lung.* Proc Am Thorac Soc, 2005. **2**(5): p. 428-32.

28.   Lau, H.Y., G.B. Huffnagle, and T.A. Moore, *Host and microbiota factors that control Klebsiella pneumoniae mucosal colonization in mice.* Microbes and infection / Institut Pasteur, 2008. **10**(12-13): p. 1283-1290.

29.   Lawlor, M.S., C. O'Connor, and V.L. Miller, *Yersiniabactin Is a Virulence Factor for Klebsiella pneumoniae during Pulmonary Infection.* Infection and Immunity, 2007. **75**(3): p. 1463-1472.

30.   Holt, K.E., et al., *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, 2015. **112**(27): p. E3574-E3581.

31.   Lau, H.Y., S. Clegg, and T.A. Moore, *Identification of Klebsiella pneumoniae genes uniquely expressed in a strain virulent using a murine model of bacterial pneumonia.* Microb Pathog, 2007. **42**(4): p. 148-55.

32.   Fodah, R.A., et al., *Correlation of Klebsiella pneumoniae Comparative Genetic Analyses with Virulence Profiles in a Murine Respiratory Disease Model.* PLOS ONE, 2014. **9**(9): p. e107394.

33.   Magill, S.S., et al., *Multistate Point-Prevalence Survey of Health Care–Associated Infections.* The New England journal of medicine, 2014. **370**(13): p. 1198-1208.

34.   Kalanuria, A.A., W. Zai, and M. Mirski, *Ventilator-associated pneumonia in the ICU.* Critical Care, 2014. **18**(2): p. 208.

35.   Selina, F., et al., *Organisms associated with ventilator associated pneumonia (VAP) in intensive care units (ICU).* 2014, 2014. **22**(2): p. 6.

36.   Richards, M.J., et al., *Nosocomial infections in combined medical-surgical intensive care units in the United States.* Infect Control Hosp Epidemiol, 2000. **21**(8): p. 510-5.

37.   Carpenter, J.L., *Klebsiella pulmonary infections: occurrence at one medical center and review.* Rev Infect Dis, 1990. **12**(4): p. 672-82.

38.   Kang, C.-I., et al., *Community-Acquired versus Nosocomial Klebsiella pneumoniae Bacteremia: Clinical Features, Treatment Outcomes, and Clinical Implication of Antimicrobial Resistance.* Journal of Korean Medical Science, 2006. **21**(5): p. 816-822.

39.   Montgomerie, J.Z. and J.K. Ota, *Klebsiella bacteremia.* Archives of Internal Medicine, 1980. **140**(4): p. 525-527.

40.   Lye, W.C., et al., *Urinary tract infections in patients with diabetes mellitus.* Journal of Infection, 1992. **24**(2): p. 169-174.

41.   Schroll, C., et al., *Role of type 1 and type 3 fimbriae in Klebsiella pneumoniae biofilm formation.* BMC Microbiol, 2010. **10**: p. 179.

42.   Domenico, P., et al., *Polysaccharide capsule-mediated resistance to opsonophagocytosis in Klebsiella pneumoniae.* Infect Immun, 1994. **62**(10): p. 4495-9.

43.   Merino, S., et al., *Mechanisms of Klebsiella pneumoniae resistance to complement-mediated killing.* Infection and Immunity, 1992. **60**(6): p. 2529-2535.

44.   Mizuta, K., et al., *Virulence for mice of Klebsiella strains belonging to the O1 group: relationship to their capsular (K) types.* Infect Immun, 1983. **40**(1): p. 56-61.

45.   Ørskov, I. and F. Ørskov, *4 Serotyping of Klebsiella.* Methods in Microbiology, 1984. **14**: p. 143-164.

46.   Roger, T., et al., *Protection from lethal gram-negative bacterial sepsis by targeting Toll-like receptor 4.* Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2348-52.

47.   Papo, N. and Y. Shai, *A molecular mechanism for lipopolysaccharide protection of Gram-negative bacteria from antimicrobial peptides.* J Biol Chem, 2005. **280**(11): p. 10378-87.

48.   Cheng, Y.H., et al., *Colistin resistance mechanisms in Klebsiella pneumoniae strains from Taiwan.* Antimicrob Agents Chemother, 2015. **59**(5): p. 2909-13.

49.   Griffiths, E., *High-affinity iron uptake systems and bacterial virulence.* Virulence mechanisms of bacterial pathogens, 1988: p. 121-137.

50.   Holden, V.I. and M.A. Bachman, *Diverging roles of bacterial siderophores during infection.* Metallomics, 2015. **7**(6): p. 986-95.

51.   Pollack, J.R. and J.B. Neilands, *Enterobactin, an iron transport compound from Salmonella typhimurium.* Biochem Biophys Res Commun, 1970. **38**(5): p. 989-92.

52.   O'Brien, I.G. and F. Gibson, *The structure of enterochelin and related 2,3-dihydroxy-N-benzoylserine conjugates from Escherichia coli.* Biochim Biophys Acta, 1970. **215**(2): p. 393-402.

53.   Smith, K.D., *Iron metabolism at the host pathogen interface: lipocalin 2 and the pathogen-associated iroA gene cluster.* The international journal of biochemistry & cell biology, 2007. **39**(10): p. 1776-1780.

54.   Hantke, K., et al., *Salmochelins, siderophores of Salmonella enterica and uropathogenic Escherichia coli strains, are recognized by the outer membrane receptor IroN.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(7): p. 3677-3682.

55.   Carniel, E., *The Yersinia high-pathogenicity island: an iron-uptake island.* Microbes and Infection, 2001. **3**(7): p. 561-569.

56.   Bachman, M.A., et al., *Klebsiella pneumoniae yersiniabactin promotes respiratory tract infection through evasion of lipocalin 2.* Infect Immun, 2011. **79**(8): p. 3309-16.

57.   Nassif, X. and P.J. Sansonetti, *Correlation of the virulence of Klebsiella pneumoniae K1 and K2 with the presence of a plasmid encoding aerobactin.* Infection and Immunity, 1986. **54**(3): p. 603-608.

58.   Holden, V.I., et al., *Klebsiella pneumoniae Siderophores Induce Inflammation, Bacterial Dissemination, and HIF-1alpha Stabilization during Pneumonia.* MBio, 2016. **7**(5).

59.   Ofek, I. and R.J. Doyle, *Bacterial adhesion to cells and tissues.* Vol. 735. 1994: Springer.

60.   Andrade, B.G.N., et al., *The genome of a clinical Klebsiella variicola strain reveals virulence-associated traits and a pl9-like plasmid.* FEMS Microbiology Letters, 2014. **360**(1): p. 13-16.

61.   Lery, L.M., et al., *Comparative analysis of Klebsiella pneumoniae genomes identifies a phospholipase D family protein as a novel virulence factor.* BMC Biology, 2014. **12**(1): p. 41.

62.   Murphy, C.N., et al., *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, 2013. **81**(8): p. 3009-3017.

63.   CDC, *Antibiotic Resistance Threats in the United States, 2013.* 2014, Centers for Disease Control and Prevention: Atlanta: U.S. Department of Health and Human Services.

64.   Fleming, A., *On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ.* British journal of experimental pathology, 1929. **10**(3): p. 226-236.

65.   ABRAHAM, E.P. and E. CHAIN, *An Enzyme from Bacteria able to Destroy Penicillin.* Nature, 1940. **146**(3713): p. 837-837.

66.   Bialek-Davenet, S., et al., *Genomic definition of hypervirulent and multidrug-resistant Klebsiella pneumoniae clonal groups.* Emerg Infect Dis, 2014. **20**(11): p. 1812-20.

67.   Babini, G.S. and D.M. Livermore, *Are SHV β-Lactamases Universal in Klebsiella pneumoniae?* Antimicrobial Agents and Chemotherapy, 2000. **44**(8): p. 2230.

68.   Datta, N. and P. Kontomichalou, *Penicillinase synthesis controlled by infectious R factors in Enterobacteriaceae.* Nature, 1965. **208**(5007): p. 239-41.

69.   Jacoby, G.A., *AmpC β-Lactamases.* Clinical Microbiology Reviews, 2009. **22**(1): p. 161-182.

70.   Meroueh, S.O., et al., *Structural Aspects for Evolution of β-Lactamases from Penicillin-Binding Proteins.* Journal of the American Chemical Society, 2003. **125**(32): p. 9612-9618.

71.   Massova, I. and S. Mobashery, *Kinship and Diversification of Bacterial Penicillin-Binding Proteins and β-Lactamases.* Antimicrobial Agents and Chemotherapy, 1998. **42**(1): p. 1-17.

72.   Kelly, J.A., et al., *On the Origin of Bacterial Resistance to Penicillin: Comparison of a β-lactamase and a Penicillin Target.* Science, 1986. **231**(4744): p. 1429-1431.

73.   Medeiros, A.A., *β-LACTAMASES.* British Medical Bulletin, 1984. **40**(1): p. 18-27.

74.   Knothe, H., et al., *Transferable resistance to cefotaxime, cefoxitin, cefamandole and cefuroxime in clinical isolates of Klebsiella pneumoniae and Serratia marcescens.* Infection, 1983. **11**(6): p. 315-7.

75.   Quinn, J.P., et al., *Novel plasmid-mediated beta-lactamase (TEM-10) conferring selective resistance to ceftazidime and aztreonam in clinical isolates of Klebsiella pneumoniae.* Antimicrob Agents Chemother, 1989. **33**(9): p. 1451-6.

76. Bush, K., G.A. Jacoby, and A.A. Medeiros, *A functional classification scheme for beta-lactamases and its correlation with molecular structure.* Antimicrobial Agents and Chemotherapy, 1995. **39**(6): p. 1211-1233.

77. Jacoby, G.A. and L. Sutton, *Properties of plasmids responsible for production of extended-spectrum beta-lactamases.* Antimicrob Agents Chemother, 1991. **35**(1): p. 164-9.

78. Long, S.W., et al., *Population Genomic Analysis of 1,777 Extended-Spectrum Beta-Lactamase-Producing Klebsiella pneumoniae Isolates, Houston, Texas: Unexpected Abundance of Clonal Group 307.* MBio, 2017. **8**(3).

79. Sougakoff, W., S. Goussard, and P. Courvalin, *The TEM-3 β-lactamase, which hydrolyzes broad-spectrum cephalosporins, is derived from the TEM-2 penicillinase by two amino acid substitutions.* FEMS Microbiology Letters, 1988. **56**(3): p. 343-348.

80. Bradford, P.A., *Extended-Spectrum β-Lactamases in the 21st Century: Characterization, Epidemiology, and Detection of This Important Resistance Threat.* Clinical Microbiology Reviews, 2001. **14**(4): p. 933-951.

81. Bush, K. and G.A. Jacoby, *Updated Functional Classification of β-Lactamases.* Antimicrobial Agents and Chemotherapy, 2010. **54**(3): p. 969-976.

82. Samuelsen, O., et al., *Emergence of clonally related Klebsiella pneumoniae isolates of sequence type 258 producing plasmid-mediated KPC carbapenemase in Norway and Sweden.* J Antimicrob Chemother, 2009. **63**(4): p. 654-8.

83. Breurec, S., et al., *Klebsiella pneumoniae resistant to third-generation cephalosporins in five African and two Vietnamese major towns: multiclonal population structure with two major international clonal groups, CG15 and CG258.* Clin Microbiol Infect, 2013. **19**(4): p. 349-55.

84. Nordmann, P., G. Cuzon, and T. Naas, *The real threat of Klebsiella pneumoniae carbapenemase-producing bacteria.* The Lancet Infectious Diseases, 2009. **9**(4): p. 228-236.

85. Yigit, H., et al., *Novel Carbapenem-Hydrolyzing β-Lactamase, KPC-1, from a Carbapenem-Resistant Strain of Klebsiella pneumoniae.* Antimicrobial Agents and Chemotherapy, 2001. **45**(4): p. 1151-1161.

86. Yigit, H., et al., *Novel Carbapenem-Hydrolyzing β-Lactamase, KPC-1, from a Carbapenem-Resistant Strain of Klebsiella pneumoniae.* Antimicrobial Agents and Chemotherapy, 2008. **52**(2): p. 809.

87. Yong, D., et al., *Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India.* Antimicrob Agents Chemother, 2009. **53**(12): p. 5046-54.

88. Kumarasamy, K.K., et al., *Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study.* The Lancet Infectious Diseases, 2010. **10**(9): p. 597-602.

89. Poirel, L., et al., *Emergence of oxacillinase-mediated resistance to imipenem in Klebsiella pneumoniae.* Antimicrob Agents Chemother, 2004. **48**(1): p. 15-22.

90. Jerke, K.H., M.J. Lee, and R.M. Humphries, *Polymyxin Susceptibility Testing: a Cold Case Reopened.* Clinical Microbiology Newsletter, 2016. **38**(9): p. 69-77.

91. Wright, M.S., et al., *Genomic and transcriptomic analyses of colistin-resistant clinical isolates of Klebsiella pneumoniae reveal multiple pathways of resistance.* Antimicrob Agents Chemother, 2015. **59**(1): p. 536-43.

92.	Cannatelli, A., et al., *In vivo emergence of colistin resistance in Klebsiella pneumoniae producing KPC-type carbapenemases mediated by insertional inactivation of the PhoQ/PhoP mgrB regulator.* Antimicrob Agents Chemother, 2013. **57**(11): p. 5521-6.

93.	Poirel, L., et al., *The mgrB gene as a key target for acquired resistance to colistin in Klebsiella pneumoniae.* J Antimicrob Chemother, 2015. **70**(1): p. 75-80.

94.	Olaitan, A.O., et al., *Worldwide emergence of colistin resistance in Klebsiella pneumoniae from healthy humans and patients in Lao PDR, Thailand, Israel, Nigeria and France owing to inactivation of the PhoP/PhoQ regulator mgrB: an epidemiological and molecular study.* Int J Antimicrob Agents, 2014. **44**(6): p. 500-7.

95.	Jayol, A., et al., *Heteroresistance to colistin in Klebsiella pneumoniae associated with alterations in the PhoPQ regulatory system.* Antimicrob Agents Chemother, 2015. **59**(5): p. 2780-4.

96.	Jayol, A., et al., *Resistance to Colistin Associated with a Single Amino Acid Change in Protein PmrB among Klebsiella pneumoniae Isolates of Worldwide Origin.* Antimicrobial Agents and Chemotherapy, 2014. **58**(8): p. 4762-4766.

97.	Liu, Y.-Y., et al., *Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study.* The Lancet Infectious Diseases, 2016. **16**(2): p. 161-168.

98.	McGann, P., et al., *Escherichia coli Harboring mcr-1 and blaCTX-M on a Novel IncF Plasmid: First Report of mcr-1 in the United States.* Antimicrob Agents Chemother, 2016. **60**(7): p. 4420-1.

99.	Chen, L., et al., *Notes from the Field: Pan-Resistant New Delhi Metallo-Beta-Lactamase-Producing Klebsiella pneumoniae - Washoe County, Nevada, 2016.* MMWR Morb Mortal Wkly Rep, 2017. **66**(1): p. 33.

100.	Selden, R., et al., *Nosocomial klebsiella infections: intestinal colonization as a reservoir.* Ann Intern Med, 1971. **74**(5): p. 657-64.

101.	Asensio, A., et al., *Outbreak of a multiresistant Klebsiella pneumoniae strain in an intensive care unit: antibiotic use as risk factor for colonization and infection.* Clin Infect Dis, 2000. **30**(1): p. 55-60.

102.	Nathisuwan, S., D.S. Burgess, and J.S. Lewis, 2nd, *Extended-spectrum beta-lactamases: epidemiology, detection, and treatment.* Pharmacotherapy, 2001. **21**(8): p. 920-8.

103.	Lautenbach, E., et al., *Extended-spectrum beta-lactamase-producing Escherichia coli and Klebsiella pneumoniae: risk factors for infection and impact of resistance on outcomes.* Clin Infect Dis, 2001. **32**(8): p. 1162-71.

104.	Jacoby, G.A., *Editorial Response: Epidemiology of Extended-Spectrum β-Lactamases.* Clinical Infectious Diseases, 1998. **27**(1): p. 81-83.

105.	Kofteridis, D.P., et al., *Risk factors for carbapenem-resistant Klebsiella pneumoniae infection/colonization: a case-case-control study.* J Infect Chemother, 2014. **20**(5): p. 293-7.

106.	Jiao, Y., et al., *Risk factors for carbapenem-resistant Klebsiella pneumoniae infection/colonization and predictors of mortality: a retrospective study.* Pathog Glob Health, 2015. **109**(2): p. 68-74.

107.	Cheng, D.L., et al., *Septic metastatic lesions of pyogenic liver abscess. Their association with Klebsiella pneumoniae bacteremia in diabetic patients.* Arch Intern Med, 1991. **151**(8): p. 1557-9.

108. Wang, J.H., et al., *Primary liver abscess due to Klebsiella pneumoniae in Taiwan.* Clin Infect Dis, 1998. **26**(6): p. 1434-8.

109. Liu, Y.C., D.L. Cheng, and C.L. Lin, *Klebsiella pneumoniae liver abscess associated with septic endophthalmitis.* Arch Intern Med, 1986. **146**(10): p. 1913-6.

110. McCabe, R., L. Lambert, and B. Frazee, *Invasive Klebsiella pneumoniae Infections, California, USA.* Emerging Infectious Diseases, 2010. **16**(9): p. 1490-1491.

111. Fierer, J., L. Walls, and P. Chu, *Recurring Klebsiella pneumoniae pyogenic liver abscesses in a resident of San Diego, California, due to a K1 strain carrying the virulence plasmid.* J Clin Microbiol, 2011. **49**(12): p. 4371-3.

112. Lederman, E.R. and N.F. Crum, *Pyogenic liver abscess with a focus on Klebsiella pneumoniae as a primary pathogen: an emerging disease with unique clinical characteristics.* Am J Gastroenterol, 2005. **100**(2): p. 322-31.

113. Pastagia, M. and V. Arumugam, *Klebsiella pneumoniae liver abscesses in a public hospital in Queens, New York.* Travel Med Infect Dis, 2008. **6**(4): p. 228-33.

114. Nadasy, K.A., R. Domiati-Saad, and M.A. Tribble, *Invasive Klebsiella pneumoniae syndrome in North America.* Clin Infect Dis, 2007. **45**(3): p. e25-8.

115. Frazee, B.W., S. Hansen, and L. Lambert, *Invasive infection with hypermucoviscous Klebsiella pneumoniae: multiple cases presenting to a single emergency department in the United States.* Ann Emerg Med, 2009. **53**(5): p. 639-42.

116. Fang, C.T., et al., *Klebsiella pneumoniae genotype K1: an emerging pathogen that causes septic ocular or central nervous system complications from pyogenic liver abscess.* Clin Infect Dis, 2007. **45**(3): p. 284-93.

117. Johannsen, E.C., C.D. Sifri, and L.C. Madoff, *Pyogenic liver abscesses.* Infect Dis Clin North Am, 2000. **14**(3): p. 547-63, vii.

118. Sng, C.C., et al., *Risk factors for endogenous Klebsiella endophthalmitis in patients with Klebsiella bacteraemia: a case-control study.* Br J Ophthalmol, 2008. **92**(5): p. 673-7.

119. Sheu, S.J., et al., *Risk factors for endogenous endophthalmitis secondary to klebsiella pneumoniae liver abscess: 20-year experience in Southern Taiwan.* Retina, 2011. **31**(10): p. 2026-31.

120. Shon, A.S., R.P. Bajwa, and T.A. Russo, *Hypervirulent (hypermucoviscous) Klebsiella pneumoniae: a new and dangerous breed.* Virulence, 2013. **4**(2): p. 107-18.

121. Hsu, C.R., et al., *The role of Klebsiella pneumoniae rmpA in capsular polysaccharide synthesis and virulence revisited.* Microbiology, 2011. **157**(Pt 12): p. 3446-57.

122. Yu, W.-L., et al., *Association between rmpA and magA Genes and Clinical Syndromes Caused by Klebsiella pneumoniae in Taiwan.* Clinical Infectious Diseases, 2006. **42**(10): p. 1351-1358.

123. Yeh, K.-M., et al., *Capsular Serotype K1 or K2, Rather than magA and rmpA, Is a Major Virulence Determinant for Klebsiella pneumoniae Liver Abscess in Singapore and Taiwan.* Journal of Clinical Microbiology, 2007. **45**(2): p. 466-471.

124. Lee, S.S., et al., *Predictors of septic metastatic infection and mortality among patients with Klebsiella pneumoniae liver abscess.* Clin Infect Dis, 2008. **47**(5): p. 642-50.

125. Siu, L.K., et al., *Klebsiella pneumoniae liver abscess: a new invasive syndrome.* Lancet Infect Dis, 2012. **12**(11): p. 881-7.

126. Chung, D.R., et al., *Emerging invasive liver abscess caused by K1 serotype Klebsiella pneumoniae in Korea.* J Infect, 2007. **54**(6): p. 578-83.

127.   Yu, W.L., et al., *Comparison of prevalence of virulence factors for Klebsiella pneumoniae liver abscesses between isolates with capsular K1/K2 and non-K1/K2 serotypes.* Diagn Microbiol Infect Dis, 2008. **62**(1): p. 1-6.

128.   Fung, C.P., et al., *A global emerging disease of Klebsiella pneumoniae liver abscess: is serotype K1 an important factor for complicated endophthalmitis?* Gut, 2002. **50**(3): p. 420-4.

129.   Liu, Y.M., et al., *Clinical and Molecular Characteristics of Emerging Hypervirulent Klebsiella pneumoniae Bloodstream Infections in Mainland China.* Antimicrobial Agents and Chemotherapy, 2014. **58**(9): p. 5379-5385.

130.   Russo, T.A., et al., *Aerobactin, but not yersiniabactin, salmochelin, or enterobactin, enables the growth/survival of hypervirulent (hypermucoviscous) Klebsiella pneumoniae ex vivo and in vivo.* Infect Immun, 2015. **83**(8): p. 3325-33.

131.   Chung, D.R., et al., *Fecal carriage of serotype K1 Klebsiella pneumoniae ST23 strains closely related to liver abscess isolates in Koreans living in Korea.* Eur J Clin Microbiol Infect Dis, 2012. **31**(4): p. 481-6.

132.   Brisse, S. and J. Verhoef, *Phylogenetic diversity of Klebsiella pneumoniae and Klebsiella oxytoca clinical isolates revealed by randomly amplified polymorphic DNA, gyrA and parC genes sequencing and automated ribotyping.* Int J Syst Evol Microbiol, 2001. **51**(Pt 3): p. 915-24.

133.   Brisse, S., et al., *Development of a rapid identification method for Klebsiella pneumoniae phylogenetic groups and analysis of 420 clinical isolates.* Clin Microbiol Infect, 2004. **10**(10): p. 942-5.

134.   Berry, G.J., M.J. Loeffelholz, and N. Williams-Bouyer, *An Investigation into Laboratory Misidentification of a Bloodstream Klebsiella variicola Infection.* J Clin Microbiol, 2015. **53**(8): p. 2793-4.

135.   Maatallah, M., et al., *Klebsiella variicola is a frequent cause of bloodstream infection in the stockholm area, and associated with higher mortality compared to K. pneumoniae.* PLoS One, 2014. **9**(11): p. e113539.

136.   Brisse, S., V. Passet, and P.A. Grimont, *Description of Klebsiella quasipneumoniae sp. nov., isolated from human infections, with two subspecies, Klebsiella quasipneumoniae subsp. quasipneumoniae subsp. nov. and Klebsiella quasipneumoniae subsp. similipneumoniae subsp. nov., and demonstration that Klebsiella singaporensis is a junior heterotypic synonym of Klebsiella variicola.* Int J Syst Evol Microbiol, 2014. **64**(Pt 9): p. 3146-52.

137.   Rosenblueth, M., et al., *Klebsiella variicola, A Novel Species with Clinical and Plant-Associated Isolates.* Systematic and Applied Microbiology, 2004. **27**(1): p. 27-35.

138.   Fouts, D.E., et al., *Complete genome sequence of the N2-fixing broad host range endophyte Klebsiella pneumoniae 342 and virulence predictions verified in mice.* PLoS Genet, 2008. **4**(7): p. e1000141.

139.   Filius, P.M., et al., *Colonization and resistance dynamics of gram-negative bacteria in patients during and after hospitalization.* Antimicrob Agents Chemother, 2005. **49**(7): p. 2879-86.

140.   Rose, H.D. and J.B. Babcock, *Colonization of intensive care unit patients with gram-negative bacilli.* Am J Epidemiol, 1975. **101**(6): p. 495-501.

141.   Snitkin, E.S., et al., *Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing.* Sci Transl Med, 2012. **4**(148): p. 148ra116.

142. Hasan, C.M., et al., *Dissemination of blaVIM in Greece at the peak of the epidemic of 2005-2006: clonal expansion of Klebsiella pneumoniae clonal complex 147.* Clin Microbiol Infect, 2014. **20**(1): p. 34-7.

143. van Duin, D., et al., *Impact of therapy and strain type on outcomes in urinary tract infections caused by carbapenem-resistant Klebsiella pneumoniae.* J Antimicrob Chemother, 2015. **70**(4): p. 1203-11.

144. Viau, R.A., et al., *"Silent" dissemination of Klebsiella pneumoniae isolates bearing K. pneumoniae carbapenemase in a long-term care facility for children and young adults in Northeast Ohio.* Clin Infect Dis, 2012. **54**(9): p. 1314-21.

145. Diancourt, L., et al., *Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates.* Journal of Clinical Microbiology, 2005. **43**(8): p. 4178-4182.

146. Brisse, S., et al., *wzi Gene Sequencing, a Rapid Method for Determination of Capsular Type for Klebsiella Strains.* Journal of Clinical Microbiology, 2013. **51**(12): p. 4073-4078.

147. Diago-Navarro, E., et al., *Carbapenem-resistant Klebsiella pneumoniae exhibit variability in capsular polysaccharide and capsule associated virulence traits.* J Infect Dis, 2014. **210**(5): p. 803-13.

148. Cubero, M., et al., *Hypervirulent Klebsiella pneumoniae clones causing bacteraemia in adults in a teaching hospital in Barcelona, Spain (2007-2013).* Clin Microbiol Infect, 2016. **22**(2): p. 154-60.

149. Wright, M.S., et al., *Population structure of KPC-producing Klebsiella pneumoniae isolates from midwestern U.S. hospitals.* Antimicrob Agents Chemother, 2014. **58**(8): p. 4961-5.

150. Elixhauser, A., et al., *Comorbidity measures for use with administrative data.* Med Care, 1998. **36**(1): p. 8-27.

151. CLSI, *Performance Standards for Antimicrobial Testing; Twenty-Fifth Informational Supplement.* 2015, Clinical and Laboratory Standards Institute: Wayne, PA.

152. Tascini, C., et al., *KPC-producing Klebsiella pneumoniae rectal colonization is a risk factor for mortality in patients with diabetic foot infections.* Clin Microbiol Infect, 2015. **21**(8): p. 790.e1-3.

153. Dautzenberg, M.J., et al., *The association between colonization with carbapenemase-producing enterobacteriaceae and overall ICU mortality: an observational cohort study.* Crit Care Med, 2015. **43**(6): p. 1170-7.

154. Kitchel, B., et al., *Molecular epidemiology of KPC-producing Klebsiella pneumoniae isolates in the United States: clonal expansion of multilocus sequence type 258.* Antimicrobial agents and chemotherapy, 2009. **53**(8): p. 3365-3370.

155. Endimiani, A., et al., *Characterization of blaKPC-containing Klebsiella pneumoniae isolates detected in different institutions in the Eastern USA.* J Antimicrob Chemother, 2009. **63**(3): p. 427-37.

156. Henderson, J.P., et al., *Quantitative metabolomics reveals an epigenetic blueprint for iron acquisition in uropathogenic Escherichia coli.* PLoS Pathog, 2009. **5**(2): p. e1000305.

157. Hayden, M.K., et al., *Prevention of Colonization and Infection by Klebsiella pneumoniae Carbapenemase–Producing Enterobacteriaceae in Long-term Acute-Care Hospitals.* Clinical Infectious Diseases, 2015. **60**(8): p. 1153-1161.

158. Lidin-Janson, G., et al., *The homogeneity of the faecal coliform flora of normal school-girls, characterized by serological and biochemical properties.* Med Microbiol Immunol, 1978. **164**(4): p. 247-53.

159. ATS, *Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia.* Am J Respir Crit Care Med, 2005. **171**(4): p. 388-416.

160. CDC *Urinary Tract Infection (Catheter-Associated Urinary Tract Infection [CAUTI] and Non-Catheter-Associated Urinary Tract Infection [UTI]) and Other Urinary System Infection [USI]) Events.*

161. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-2948.

162. Tamura, K., et al., *MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0.* Molecular Biology and Evolution, 2013. **30**(12): p. 2725-2729.

163. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: A flexible trimmer for Illumina Sequence Data.* Bioinformatics, 2014.

164. Xu, H., et al., *FastUniq: a fast de novo duplicates removal tool for paired short reads.* PLoS One, 2012. **7**(12): p. e52249.

165. Luo, R., et al., *SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler.* Gigascience, 2012. **1**(1): p. 18.

166. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs.* Genome Res, 2008. **18**(5): p. 821-9.

167. Jolley, K.A. and M.C. Maiden, *BIGSdb: Scalable analysis of bacterial genome variation at the population level.* BMC Bioinformatics, 2010. **11**(1): p. 1-11.

168. Martin, R.M., et al., *Molecular Epidemiology of Colonizing and Infecting Isolates of Klebsiella pneumoniae.* mSphere, 2016. **1**(5).

169. Subashchandrabose, S. and H.L.T. Mobley, *Virulence and Fitness Determinants of Uropathogenic Escherichia coli.* Microbiology spectrum, 2015. **3**(4): p. 10.1128/microbiolspec.UTI-0015-2012.

170. Sheppard, S.K., et al., *Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter.* Proc Natl Acad Sci U S A, 2013. **110**.

171. Laabei, M., et al., *Predicting the virulence of MRSA from its genome sequence.* Genome Res, 2014. **24**.

172. Alam, M.T., et al., *Dissecting vancomycin intermediate resistance in Staphylococcus aureususing genome-wide association.* Genome Biol Evol, 2014. **6**.

173. Read, T.D. and R.C. Massey, *Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology.* Genome Medicine, 2014. **6**(11): p. 109.

174. Broberg, C.A., et al., *Complete Genome Sequence of Klebsiella pneumoniae Strain ATCC 43816 KPPR1, a Rifampin-Resistant Mutant Commonly Used in Animal, Genetic, and Molecular Biology Studies.* Genome Announc, 2014. **2**(5).

175. Wu, K.M., et al., *Genome sequencing and comparative analysis of Klebsiella pneumoniae NTUH-K2044, a strain causing liver abscess and meningitis.* J Bacteriol, 2009. **191**(14): p. 4492-501.

176. Fang, C.-T., et al., *A Novel Virulence Gene in Klebsiella pneumoniae Strains Causing Primary Liver Abscess and Septic Metastatic Complications.* The Journal of Experimental Medicine, 2004. **199**(5): p. 697-705.

177. Passet, V. and S. Brisse, *Association of tellurite resistance with hypervirulent clonal groups of Klebsiella pneumoniae.* J Clin Microbiol, 2015. **53**(4): p. 1380-2.

178. Whelan, K.F., E. Colleran, and D.E. Taylor, *Phage inhibition, colicin resistance, and tellurite resistance are encoded by a single cluster of genes on the IncHI2 plasmid R478.* Journal of Bacteriology, 1995. **177**(17): p. 5016-5027.

179. Tomas, J.M., B. Ciurana, and J.T. Jofre, *New, simple medium for selective, differential recovery of Klebsiella spp.* Appl Environ Microbiol, 1986. **51**(6): p. 1361-3.

180. Martens, E.C., et al., *Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts.* PLOS Biology, 2011. **9**(12): p. e1001221.

181. Taylor, D.E., *Bacterial tellurite resistance.* Trends in Microbiology. **7**(3): p. 111-115.

182. Franks, S.E., et al., *Novel role for the yceGH tellurite resistance genes in the pathogenesis of Bacillus anthracis.* Infect Immun, 2014. **82**(3): p. 1132-40.

183. Blin, C., et al., *Metabolic diversity of the emerging pathogenic lineages of Klebsiella pneumoniae.* Environ Microbiol, 2017.

184. Chattopadhyay, S., U. Raychaudhuri, and R. Chakraborty, *Artificial sweeteners – a review.* Journal of Food Science and Technology, 2014. **51**(4): p. 611-621.

185. Iida, T., et al., *Failure of d-psicose absorbed in the small intestine to metabolize into energy and its low large intestinal fermentability in humans.* Metabolism, 2010. **59**(2): p. 206-14.

186. Pyleris, E., et al., *The prevalence of overgrowth by aerobic bacteria in the small intestine by small bowel culture: relationship with irritable bowel syndrome.* Dig Dis Sci, 2012. **57**(5): p. 1321-9.

187. Bouhnik, Y., et al., *Bacterial populations contaminating the upper gut in patients with small intestinal bacterial overgrowth syndrome.* Am J Gastroenterol, 1999. **94**(5): p. 1327-31.

188. Hossain, A., et al., *Rare sugar D-allulose: Potential role and therapeutic monitoring in maintaining obesity and type 2 diabetes mellitus.* Pharmacol Ther, 2015. **155**: p. 49-59.

189. Kanehisa, M., et al., *KEGG: new perspectives on genomes, pathways, diseases and drugs.* Nucleic Acids Res, 2017. **45**(D1): p. D353-d361.

190. Therneau, T.M. and P.M. Grambsch, *Modeling Survival Data: Extending the Cox Model.* 2000, New York: Springer.

191. Robin, X., et al., *pROC: an open-source package for R and S+ to analyze and compare ROC curves.* BMC Bioinformatics, 2011. **12**(1): p. 77.

192. DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson, *Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach.* Biometrics, 1988. **44**(3): p. 837-845.

193. Storey, J.D., *A direct approach to false discovery rates.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2002. **64**(3): p. 479-498.

194. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-1760.

195. Lloyd, S., *Least squares quantization in PCM.* IEEE Transactions on Information Theory, 1982. **28**(2): p. 129-137.

196. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc Natl Acad Sci U S A, 2000. **97**(12): p. 6640-5.

197. Bachman, M.A., et al., *Genome-Wide Identification of Klebsiella pneumoniae Fitness Genes during Lung Infection.* mBio, 2015. **6**(3).

198. Desai, M.S., et al., *A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility.* Cell. **167**(5): p. 1339-1353.e21.

199. S., A. *FastQC: a quality control tool for high throughput sequence data.* http://www.bioinformatics.babraham.ac.uk/projects/fastqc. 2010 May 31, 2017].

200. Croucher, N.J., et al., *Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins.* Nucleic Acids Res, 2015. **43**(3): p. e15.

201. Palmer, B.F. and D.J. Clegg, *Electrolyte and Acid–Base Disturbances in Patients with Diabetes Mellitus.* New England Journal of Medicine, 2015. **373**(6): p. 548-559.

202. Namdar, T., et al., *Transdermal fluid loss in severely burned patients.* GMS German Medical Science, 2010. **8**: p. Doc28.

203. Pruitt, B.A., Jr., *Infections caused by Pseudomonas species in patients with burns and in other surgical patients.* J Infect Dis, 1974. **130 Suppl**(0): p. S8-13.

204. Jones, R.J., E.A. Roe, and J.L. Gupta, *CONTROLLED TRIALS OF A POLYVALENT PSEUDOMONAS VACCINE IN BURNS.* The Lancet, 1979. **314**(8150): p. 977-983.

205. Ibrahim, E.H., et al., *The influence of inadequate antimicrobial treatment of bloodstream infections on patient outcomes in the ICU setting.* Chest, 2000. **118**(1): p. 146-55.

206. Hanon, F.X., et al., *Survival of patients with bacteraemia in relation to initial empirical antimicrobial treatment.* Scand J Infect Dis, 2002. **34**(7): p. 520-8.

207. Huddleston, J.R., *Horizontal gene transfer in the human gastrointestinal tract: potential spread of antibiotic resistance genes.* Infection and Drug Resistance, 2014. **7**: p. 167-176.

208. Sheng, Z.-K., et al., *Mechanisms of Tigecycline Resistance among Klebsiella pneumoniae Clinical Isolates.* Antimicrobial Agents and Chemotherapy, 2014. **58**(11): p. 6982-6985.

209. Siegel, J., et al., *Management of Multidrug-Resistant Organisms In Healthcare Settings, 2006*, CDC, Editor. 2006.

210. Delorme, T., et al., *Epidemiology and susceptibilies of methicillin-resistant Staphylococcus aureus in Northeastern Ohio.* Am J Clin Pathol, 2009. **132**(5): p. 668-77.

211. Klevens, R.M., et al., *Invasive methicillin-resistant Staphylococcus aureus infections in the United States.* Jama, 2007. **298**(15): p. 1763-71.

212. Tacconelli, E., et al., *Rapid screening tests for meticillin-resistant Staphylococcus aureus at hospital admission: systematic review and meta-analysis.* Lancet Infect Dis, 2009. **9**(9): p. 546-54.

213. Mutters, N.T., et al., *Control of the Spread of Vancomycin-Resistant Enterococci in Hospitals: Epidemiology and Clinical Relevance.* Deutsches Ärzteblatt International, 2013. **110**(43): p. 725-731.

214. Zirakzadeh, A. and R. Patel, *Vancomycin-Resistant Enterococci: Colonization, Infection, Detection, and Treatment.* Mayo Clinic Proceedings. **81**(4): p. 529-536.

215. O'Driscoll, T. and C.W. Crank, *Vancomycin-resistant enterococcal infections: epidemiology, clinical manifestations, and optimal management.* Infection and Drug Resistance, 2015. **8**: p. 217-230.

216. Ho, C., et al., *Screening, Isolation, and Decolonization Strategies for Vancomycin-Resistant Enterococci or Extended Spectrum Beta-Lactamase-Producing Organisms: A Systematic Review of the Clinical Evidence and Health Services Impact.* CADTH Technology Overviews, 2013. **3**(1): p. e3102.

217. Yang, S., et al., *Evolution and Transmission of Carbapenem-Resistant Klebsiella pneumoniae Expressing the blaOXA-232 Gene During an Institutional Outbreak Associated With Endoscopic Retrograde Cholangiopancreatography.* Clinical Infectious Diseases, 2017. **64**(7): p. 894-901.

218. Kauffman, C.A., et al., *Attempts to eradicate methicillin-resistant Staphylococcus aureus from a long-term-care facility with the use of mupirocin ointment.* Am J Med, 1993. **94**(4): p. 371-8.

219. Strausbaugh, L.J., et al., *Antimicrobial therapy for methicillin-resistant Staphylococcus aureus colonization in residents and staff of a Veterans Affairs nursing home care unit.* Infect Control Hosp Epidemiol, 1992. **13**(3): p. 151-9.

220. Rieg, S., et al., *Intestinal decolonization of Enterobacteriaceae producing extended-spectrum β-lactamases (ESBL): a retrospective observational study in patients at risk for infection and a brief review of the literature.* BMC Infectious Diseases, 2015. **15**: p. 475.

221. Winter, S.E., et al., *Host-derived nitrate boosts growth of E. coli in the inflamed gut.* Science (New York, N.Y.), 2013. **339**(6120): p. 708-711.

222. Hughes, E.R., et al., *Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis.* Cell Host & Microbe. **21**(2): p. 208-219.

223. Sassone-Corsi, M., et al., *Microcins mediate competition among Enterobacteriaceae in the inflamed gut.* Nature, 2016. **540**(7632): p. 280-283.

224. Biliński, J., et al., *Fecal Microbiota Transplantation Inhibits Multidrug-Resistant Gut Pathogens: Preliminary Report Performed in an Immunocompromised Host.* Archivum Immunologiae et Therapiae Experimentalis, 2016. **64**: p. 255-258.

225. Donskey, C.J. and A. Deshpande, *Effect of chlorhexidine bathing in preventing infections and reducing skin burden and environmental contamination: A review of the literature.* American Journal of Infection Control, 2016. **44**(5): p. e17-e21.

226. CDC, *Facility Guidance for Control of Carbapenem-Resistant Enterobacteriaceae (CRE).* 2015: CDC.

227. Nelson, A.L., et al., *Capsule Enhances Pneumococcal Colonization by Limiting Mucus-Mediated Clearance.* Infection and Immunity, 2007. **75**(1): p. 83-90.

228. Sahly, H. and R. Podschun, *Clinical, bacteriological, and serological aspects of Klebsiella infections and their spondylarthropathic sequelae.* Clinical and Diagnostic Laboratory Immunology, 1997. **4**(4): p. 393-399.

229. Morgan, M.E., C.A. Hart, and R.W. Cooke, *Klebsiella infection in a neonatal intensive care unit: role of bacteriological surveillance.* J Hosp Infect, 1984. **5**(4): p. 377-85.

230. *Patient Protection and Affordable Care Act*, in *42*. 2010: United States.

231. Lee, Y.-h., H. Bang, and D.J. Kim, *How to Establish Clinical Prediction Models.* Endocrinology and Metabolism, 2016. **31**(1): p. 38-44.

232. Nijman, R.G., et al., *Clinical prediction model to aid emergency doctors managing febrile children at risk of serious bacterial infections: diagnostic study.* Bmj, 2013. **346**: p. f1706.

233. Bearman, G.M., et al., *A Clinical Predictive Model for Catheter Related Bloodstream Infections from the Electronic Medical Record.* Open Epidemiology Journal, 2010. **3**: p. 24-28.

234. Chang, Y.-J., et al., *Predicting Hospital-Acquired Infections by Scoring System with Simple Parameters.* PLoS ONE, 2011. **6**(8): p. e23137.

235. Behjati, S. and P.S. Tarpey, *What is next generation sequencing?* Archives of Disease in Childhood. Education and Practice Edition, 2013. **98**(6): p. 236-238.

236. Dekker, J.P. and K.M. Frank, *Next-Generation Epidemiology: Using Real-Time Core Genome Multilocus Sequence Typing To Support Infection Control Policy.* Journal of Clinical Microbiology, 2016. **54**(12): p. 2850-2853.

237. Gwinn, M., D.R. MacCannell, and R.F. Khabbaz, *Integrating Advanced Molecular Technologies into Public Health.* Journal of Clinical Microbiology, 2017. **55**(3): p. 703-714.
238. Kwong, J.C., et al., *Whole genome sequencing in clinical and public health microbiology.* Pathology, 2015. **47**(3): p. 199-210.
239. Gargis, A.S., L. Kalman, and I.M. Lubin, *Assuring the Quality of Next-Generation Sequencing in Clinical Microbiology and Public Health Laboratories.* J Clin Microbiol, 2016. **54**(12): p. 2857-2865.