MONASH University

Combating Gram negative multi-drug resistance with insights from novel bioinformatics approaches

Nenad Macesic
MBBS MA FRACP

A thesis submitted for the degree of Doctor of Philosophy at
Monash University in 2020
Department of Infectious Diseases, Central Clinical School

Copyright notice

I certify that I have made all reasonable efforts to secure copyright permissions for third-party content included in this thesis and have not knowingly added copyright content to my work without the owner's permission.

Abstract

Antimicrobial resistance (AMR) is an urgent global health challenge that threatens many of the advances of modern medicine. Multi-drug resistant Gram negative bacteria (MDR-GNB) comprise much of this threat. Infections caused by MDR-GNB, such as carbapenem-resistant Enterobacteriaceae (CRE), have a high morbidity and mortality ranging from ~10–50%. In the face of this threat, novel tools such as whole genome sequencing (WGS) are providing crucial insights into how AMR develops and spreads. WGS makes it possible to elucidate detailed phylogenetic relationships and is likely to play an increasing role in the diagnostics of infectious diseases. In parallel, sophisticated bioinformatics analyses such as genome-wide association studies (GWAS) and machine learning (ML) approaches are being developed to maximise the potential of genomics for clinical medicine.

Given the urgent needs posed by MDR-GNB and the promise of these technological developments, this PhD aimed to address how we can use these technologies to better understand, prevent and treat infections caused by these formidable pathogens. More specifically, the aims of my PhD included: 1) to use WGS to determine MDR-GNB colonization and infection dynamics; 2) to define the genomic epidemiology of emerging forms of MDR-GNB (*mcr-1* and *Klebsiella pneumoniae* sequence type [ST]307); 3) to determine the clinical and molecular epidemiology of resistance to polymyxins, treatments of last-resort for MDR-GNB, through use of WGS; 4) to discover novel genetic determinants of polymyxin resistance (PR) through use of comparative genomics and

GWAS; and 5) to use ML approaches to predict phenotypic PR using *K. pneumoniae* clonal group (CG) 258 genomic data.

With this in mind, in a prospective longitudinal cohort of liver transplant recipients, I demonstrated unexpected heterogeneity of CRE colonizing isolates, previously unrecognized transmission spanning cephalosporin- and carbapenem-resistant Enterobacteriaceae phenotypes, and a cluster of *mcr-1*-producing isolates. Two of these findings were then expanded. Firstly, I investigated the *mcr-1* cluster and identified same-day endoscopy as a possible means of transmission. I then performed an analysis of *K. pneumoniae* ST307 in the context of the first detection of a CRE isolate in the Dominican Republic and found similarities between isolates of Caribbean origin that affirm the role of *K. pneumoniae* ST307 as a new MDR clone.

With PR being a worrisome finding in the liver transplant study, a major focus of my PhD was to use WGS and novel bioinformatics techniques to study resistance to this class of last-line antibiotics. Firstly, I conducted a retrospective cohort study using carbapenem-resistant *K. pneumoniae* isolates from a single institution and noted an impressive diversity of variants leading to PR with exposure to polymyxins playing a key role. I then conducted a comparative genomic analysis of closely related *K. pneumoniae* isolates and a GWAS of PR, which ultimately identified the *barA/uvrY* two-component system as a novel determinant of PR that likely causes capsular changes to confer PR, as opposed to the current understanding that focuses on lipid A modifications. Finally, I demonstrated

that ML approaches can successfully use WGS data to predict phenotypic PR, serving as an important proof-of-principle for the role of ML in predicting complex forms of AMR.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature:

Print Name: Nenad Macesic

Date: January 30, 2020

Publications during enrolment

A list of published/accepted and submitted papers during my PhD candidature is presented below. Papers which are included in the thesis are indicated by an asterisk.

Published papers

1. ***Macesic N**, Nelson B, McConville TH, Giddins MJ, Green DA, Stump S, Gomez-Simmonds A, Annavajhala MK, Uhlemann AC. Emergence of Polymyxin Resistance in Clinical *Klebsiella pneumoniae* Through Diverse Genetic Adaptations: A Genomic, Retrospective Cohort Study. *Clin Infect Dis*. 2019 Sep 12; [Epub ahead of print]
2. ***Macesic N**, Khan S, Giddins MJ, Freedberg DE, Whittier S, Green DA, Furuya EY, Verna EC, Annavajhala MK, Gomez-Simmonds A, Uhlemann AC. *Escherichia coli* Harboring mcr-1 in a Cluster of Liver Transplant Recipients: Detection through Active Surveillance and Whole-Genome Sequencing. *Antimicrob Agents Chemother*. 2019 Jun;63(6)
3. ***Macesic N**, Gomez-Simmonds A, Sullivan SB, Giddins MJ, Ferguson SA, Korakavi G, Leeds D, Park S, Shim K, Sowash MG, Hofbauer M, Finkel R, Hu Y, West J, Toussaint NC, Greendyke WG, Miko BA, Pereira MR, Whittier S, Verna EC, Uhlemann AC. Genomic Surveillance Reveals Diversity of Multidrug-Resistant Organism Colonization and Infection: A Prospective Cohort Study in Liver Transplant Recipients. *Clin Infect Dis.* 2018 Aug 31;67(6):905-912.

4. Rojas R, **Macesic N**, Tolari G, Guzman A, Uhlemann AC. Multidrug-Resistant *Klebsiella pneumoniae* ST307 in Traveler Returning from Puerto Rico to Dominican Republic. *Emerg Infect Dis*. 2019 Aug;25(8):1583-1585.
5. Giddins MJ, **Macesic N**, Annavajhala MK, Stump S, Khan S, McConville TH, Mehta M, Gomez-Simmonds A, Uhlemann AC. Successive Emergence of Ceftazidime-Avibactam Resistance through Distinct Genomic Adaptations in blaKPC-2-Harboring *Klebsiella pneumoniae* Sequence Type 307 Isolates. *Antimicrob Agents Chemother*. 2018 Mar;62(3).
6. **Macesic N**, Nelson B, Uhlemann AC. Colistin Resistance in Carbapenem-Resistant *Klebsiella pneumoniae*: De Novo or Drug Exposure?. *Clin Infect Dis*. 2017 Aug 15;65(4):702-703.
7. Humphries RM, Green DA, Schuetz AN, Bergman Y, Lewis S, Yee R, Stump S, Lopez M, **Macesic N**, Uhlemann AC, Kohner P, Cole N, Simner PJ. Multicenter Evaluation of Colistin Broth Disk Elution and Colistin Agar Test: a Report from the Clinical and Laboratory Standards Institute. *J Clin Microbiol*. 2019 Nov;57(11).
8. Green DA, **Macesic N,** Uhlemann AC, Lopez M, Stump S, Whittier S, Schuetz AN, Simner PJ, Humphries RM. Evaluation of Calcium-Enhanced Media for Colistin Susceptibility Testing by Gradient Agar Diffusion and Broth Microdilution. *J Clin Microbiol*. 2019 Nov 20. [Epub ahead of print]
9. Annavajhala MK, Gomez-Simmonds A, **Macesic N**, Sullivan SB, Kress A, Khan SD, Giddins MJ, Stump S, Kim GI, Narain R, Verna EC, Uhlemann AC. Colonizing multidrug-resistant bacteria and the longitudinal evolution of the intestinal microbiome after liver transplantation. *Nat Commun*. 2019 Oct 17;10(1):4715.

10. Gomez-Simmonds A, Annavajhala MK, Wang Z, **Macesic N**, Hu Y, Giddins MJ, O'Malley A, Toussaint NC, Whittier S, Torres VJ, Uhlemann AC. Genomic and Geographic Context for the Evolution of High-Risk Carbapenem-Resistant *Enterobacter cloacae* Complex Clones ST171 and ST78. *mBio*. 2018 May 29;9(3).

Manuscript in submission

1. ***Macesic N**, Bear Don't Walk IV OJ, Pe'er I, Tatonetti NP, Peleg AY, Uhlemann AC. Predicting phenotypic polymyxin resistance in *Klebsiella pneumoniae* through machine learning analysis of genomic data. *mSystems*. [returned for revision]

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes 3 original papers published in peer reviewed journals and 1 submitted publication. Chapter 3 also includes a published paper that is significant to the thesis where I contributed as a second author. Chapter 5 contains work that pertains to the PhD and will be submitted shortly. The core theme of the thesis is the use of whole genome sequencing and novel bioinformatics approaches for studying multi-drug resistant Gram negative bacteria. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Department of Infectious Diseases, Monash University under the supervision of Prof. Anton Peleg. All laboratory work, analysis and manuscript writing was done within the Division of Infectious Diseases, Columbia University under the supervision of A/Prof Anne-Catrin Uhlemann.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of chapters 2, 3, 4, 5 and 6 my contribution to the work involved the following:

Thesis Chapter	**Publication Title**	**Status** *(published, in press, accepted or returned for revision, submitted)*	**Nature and % of student contribution**	**Co-author name(s) Nature and % of Co-author's contribution***	**Co-author(s), Monash student Y/N***
2	Genomic Surveillance Reveals Diversity of Multidrug-Resistant Organism Colonization and Infection: A Prospective Cohort Study in Liver Transplant Recipients.	Published	50%. Concept, collecting data, conducting genomics analyses, interpreting data, writing manuscript and performing all revisions	1) A. Gomez-Simmonds. 15%. Collecting data, interpreting data, writing manuscript 2) AC Uhlemann. 15%. Concept, collecting data, writing manuscript. 3) Other authors. 20%. See chapter for specifics.	No No No No
3	*Escherichia coli* Harboring	Published	60%. Concept, ,	1) S Khan. 15%. *Mcr-1* PCR testing	No

	mcr-1 in a Cluster of Liver Transplant Recipients: Detection through Active Surveillance and Whole-Genome Sequencing		collecting data, conducting genomics analyses, interpreting data, writing manuscript and performing all revisions.	of stored isolates. Review of manuscript. 2) MJ Giddins. 10%. *Mcr-1* PCR testing of stored isolates. Review of manuscript. 3) AC Uhlemann. 10%. Concept, writing manuscript. 4) Other authors. 5%. See chapter for specifics.	No No No
4	Emergence of Polymyxin Resistance in Clinical *Klebsiella pneumoniae* Through Diverse Genetic Adaptations: A Genomic, Retrospective Cohort Study	Published	70%. Concept. Isolate collection, susceptibility testing, collecting clinical data, genomics and phylogenetic analyses, writing manuscript and performing all revisions.	1) B Nelson. 10%. Collecting clinical data and performing analyses. 2) AC Uhlemann. 15%. Concept, assisting with genomics and phylogenetic analysis, writing manuscript. 3) Other authors. 5%. See chapter for specifics.	No No No
6	Predicting phenotypic polymyxin resistance in *Klebsiella pneumoniae* through machine learning analysis of genomic data	Returned for revision	62.5%. Concept, genomics analyses, machine learning pipeline coding, machine learning pipeline execution in AWS, manuscript writing and revisions	1) OJ Bear Don't Walk IV. 25%. Writing machine learning code. Writing manuscript. 2) I Pe'er. 2.5%. Reviewing manuscript. 3) NP Tatonetti. 2.5%. Reviewing manuscript. 4) AY Peleg. 2.5%. Reviewing manuscript. 5) AC Uhlemann. 5%. Concept,	No No No No

				reviewing manuscript.	

I have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student name: Nenad Macesic

Student signature: **Date:** 30/1/2020

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor name: Anton Peleg

Main Supervisor signature: **Date:** 30/1/2020

Acknowledgements

I would like to acknowledge and thank my wife, Emma Watts for the incredible support throughout this PhD, including moving to the United States for an extended period so that I could conduct my research. I also would like to acknowledge my supervisors, Prof. Anton Peleg (Australian supervisor, Central Clinical School, Monash University, Melbourne) and Assoc. Prof. Anne-Catrin Uhlemann (American supervisor, Division of Infectious Diseases, Columbia University Irving Medical Center, New York) for their mentorship and support throughout this PhD. I would like to acknowledge my colleagues in the Uhlemann Laboratory (Thomas McConville, Angela Gomez-Simmonds, Medini Annavajhala and Marla Giddins). Much of the research in this PhD was made possible through their contributions and assistance. I also wanted to thank Profs. Scott Hammer and Magdalena Sobieszczyk from Columbia University Irving Medical Center for the opportunity to come to the United States.

This research was supported by the National Health and Medical Research Council Postgraduate Scholarship (APP1169514), the Columbia University Interdisciplinary Research Initiatives Seed Fund Grant, Columbia University Infectious Diseases Divisional Funds, the National Institute of Allergy and Infectious Diseases at the National Institutes of Health (grant number AI116939 to Anne-Catrin Uhlemann) and the Irving Scholarship from Columbia University (to Anne-Catrin Uhlemann).

Table of Contents

List of Figures and Tables

NB: Tables and Figures embedded within published and submitted papers do not appear in this list

List of Abbreviations

AMR: Antimicrobial resistance

AST: Antimicrobial susceptibility testing

AUROC: Area under receiver operator curve

CDC: Centers for Disease Control and Prevention

CP: Carbapenemase producing

CRE: Carbapenem-resistant Enterobacteriaceae

CRKP: Carbapenem Resistant *Klebsiella pneumoniae*

ESBL: Extended-spectrum beta-lactamases

GWAS: Genome-wide association study

IMP: Imipenem resistance metallo-beta-lactamase

IS: Insertion sequence

KPC: *Klebsiella pneumoniae* carbapenemase

L-Ara4N: 4-amino-4-deoxy-l-arabinose

LPS: Lipopolysaccharide

MBL: Metallo-beta-lactamase

MDR: Multi-drug resistance

MDR-GNB: Multi-drug resistant Gram negative bacteria

MDRO: Multi-drug resistant organism

MGE: Mobile genetic element

MIC: Minimum inhibitory concentration

ML: Machine learning

MLST: Multi-locus sequence type

MRSA: Methicillin-resistant *Staphylococcus aureus*

NDM: New Delhi metallo-beta-lactamase

pEtN: Phosphethanolamine

PR: Polymyxin resistance

SNV: Single nucleotide variant

SOT: Solid organ transplant

ST: Sequence type

VIM: Verona integron-encoded metallo-beta-lactamase

WGS: Whole genome sequencing

NB: Abbreviations embedded within published and submitted papers do not appear in this list

Chapter 1:

Introduction, Background and Literature Review

Introduction

Antimicrobial resistance (AMR) is an urgent global health challenge that threatens our healthcare system and many of the advances of modern medicine. Current projections estimate that in the absence of major interventions over 10 million people could die of AMR infections yearly, exceeding cancer-related deaths [1]. Australia is positioned within a hot-bed of AMR, with our nearest neighbours having the greatest burden worldwide [2]. Improving the prevention, detection and management of AMR is an urgent priority.

Multi-drug resistant Gram negative bacilli (MDR-GNB) comprise much of this threat. Since the turn of the millennium, these pathogens have become a leading cause of hospital-acquired infections [3]. Infections caused by MDR-GNB such as carbapenem-resistant Enterobacteriaceae (CRE) have a high morbidity and mortality ranging from ~10–50% [4]. Multiple factors have contributed to this expansion including the increasing complexity of medical care and the ability of these bacteria to readily transmit AMR and virulence determinants via mobile genetic elements (MGEs) such as plasmids [3]. While multi-drug resistance has typically been defined as resistance to three or more classes of antimicrobials [5], the problem continues to worsen with resistance to agents of last resort such as carbapenems, polymyxins and even novel beta-lactam/beta-lactamase inhibitors being increasingly noted [6, 7]. This makes the threat of pan-drug resistance in Gram negative bacteria increasingly a reality [8, 9].

In the face of this threat, novel tools such as whole genome sequencing (WGS) are providing crucial insights into how AMR develops and spreads [10]. Using WGS has made it possible to elucidate detailed phylogenetic relationships between organisms and track the development of genetic resistance determinants and their transmission between different bacteria. WGS is likely to play an increasing role in the diagnostics of infectious diseases and genomic data can be used as an input for sophisticated bioinformatics analysis such as machine learning (ML) [11, 12]. This has led to increasing interest in the possibility of using genomic data and ML approaches to predict AMR phenotypes [12].

In addition to these technological advances, there has been increasing recognition of the need to expand the therapeutic armamentarium for MDR-GNB infections. This led to initiatives such as 10 x ‘20 [13], which aimed to develop 10 new systemic antibacterial drugs by 2020. In the last five years, multiple new agents have reached clinical use including novel beta-lactam/beta-lactamase inhibitor combinations [14-16], tetracyclines [17, 18], aminoglycosides [19] and fluoroquinolones [20]. Several more agents remain in development and are currently in Phase 2/3 trials (e.g. cefiderocol) [21]. One emerging theme is the importance of the underlying mechanism of action, as their activity is targeted at specific resistance-causing enzymes rather than having broad multi-class effects. Genome sequencing technologies are likely to play a central role in helping us use these agents effectively by allowing better detection and understanding of these underlying mechanisms of resistance [11].

Given the urgent needs posed by MDR-GNB and the promise of WGS and associated novel bioinformatics approaches, this PhD aims to address how we can use these technological tools to better understand, prevent and treat infections caused by these formidable pathogens. It focuses on organisms resistant to last-line agents, particularly CRE and polymyxin-resistant (PR) Enterobacteriaceae due to their extensive drug resistance and increasing endemicity in healthcare settings globally. As immunocompromised patients such as transplant recipients are disproportionately affected by MDR-GNB generally and CRE in particular [22-26], they are also a focus of this work.

Background and Literature Review

Definitions of multi-drug resistance

Due to the evolutionary pressures of increasing antimicrobial use, resistance to multiple classes of drugs has become a prominent public health threat. How to define categories of drug resistant organisms remains a subject of active debate [27]. In 2012, the Centers for Disease Control and Prevention (CDC) and European Centre for Disease Control proposed definitions, which aimed to create epidemiologically meaningful categories around commonly used/tested antimicrobials [5]. They proposed three categories. Multi-drug resistance (MDR) was defined as acquired non-susceptibility to at least one agent in three or more antimicrobial categories. Extensive drug resistance was defined as non-susceptibility to at least one agent in all but two or fewer antimicrobial categories (i.e., bacterial isolates remain susceptible to only one or two categories). Pan-drug resistance was defined as non-susceptibility to all agents in all antimicrobial categories.

In addition to these widely used categories, other categories that aim to be more clinically relevant have also been proposed. Difficult-to-treat resistance has been defined as treatment-limiting resistance to all first-line agents, that is, all beta-lactams, including carbapenems and beta-lactamase inhibitor combinations (not including novel combinations), and fluoroquinolones [28]. This distinguishes low-toxicity first-line agents from agents with higher toxicity and/or less efficacy such as aminoglycosides, polymyxins and tigecycline. Usual drug resistance is a similar category that denotes isolates that aren't fully susceptible wild-type strains but can nonetheless be readily treated with standard therapies [27, 29]. It was developed with a view to use in non-inferiority clinical trials for novel antimicrobial agents.

While useful, these definitions have inherent limitations. The first is that new developments in the form of novel antimicrobial agents or mechanisms of resistance mandate ongoing revision of these categories to accommodate the different spectra of action of these agents. Since the publication of this guideline a number of agents have reached clinical practice, including beta-lactam/beta-lactamase inhibitor combinations such as ceftazidime-avibactam. The second is that resistance to a single class of antimicrobials may have significant public health implications, regardless of co-resistance. The detection of *mcr-1*, a plasmid-mediated form of resistance to polymyxins (PR), in 2015 presents one such case [30]. While most isolates carrying *mcr-1* retain susceptibility to multiple agents, the ongoing spread of *mcr-1* presents a significant epidemiological threat. Thirdly, any discrete categorization ignores that AMR often exists on a continuum with stepwise increases in resistance. For example, while most

carbapenem resistance in Enterobacteriaceae results from acquisition of carbapenemase genes, it also emerges when bacteria with extended spectrum beta-lactamases (ESBLs) develop membrane-based mutations [6].

Finally, the widespread availability of molecular data will likely impact how resistance categories are defined. Bacterial WGS is now starting to reach clinical use and provides unprecedented resolution with which to characterise isolates, ranging from clonal typing to detection of a wide range of MGEs that determine AMR and virulence. This greater resolution will thus entail new schema for classification of resistance, in a similar way that we have had to reconsider taxonomic classification on the basis of genomic data. This will have attendant clinical or infection prevention implications. For example, infections caused by carbapenemase-producing CRE (CP CRE) have been noted to have higher mortality versus non-carbapenemase producing CRE (non-CP CRE) [31] and presence of a carbapenemase allows an organism to be classified as CRE according to recent CDC guidelines, regardless of phenotypic susceptibility testing results [32]. From the perspective of PR, plasmid mediated *mcr-1* carries higher transmission risks than chromosomal forms of PR and could arguably fit into a separate category. The impact of WGS on our understanding of Gram negative resistance will be a key focus of this PhD.

Molecular mechanisms and genomic epidemiology of resistance to last line agents in Gram negatives

Carbapenem resistance

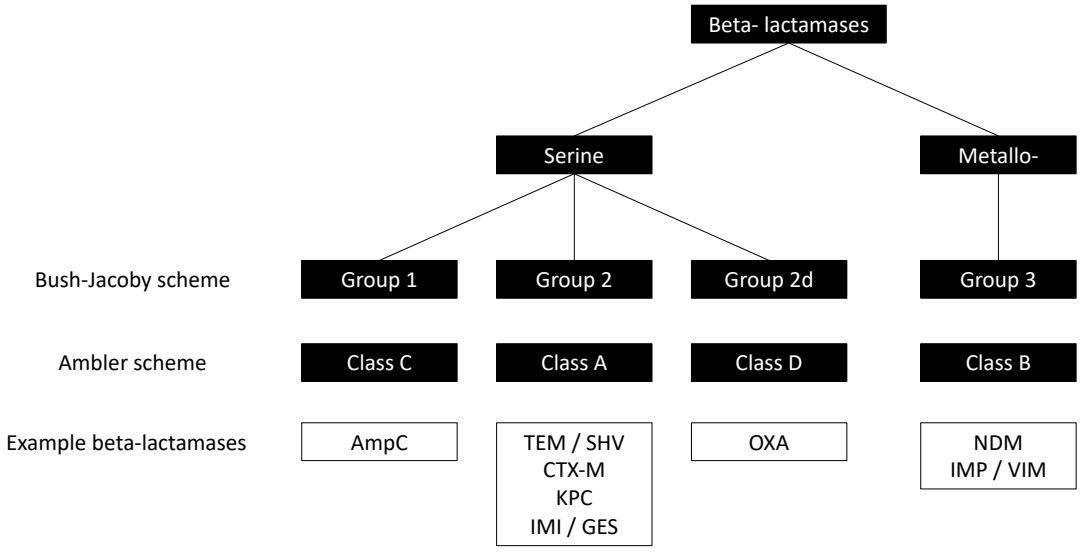
Chromosomally based carbapenemase genes were first identified in Gram-positive organisms [33]. In the 1980s similar genes were then noted in Gram negative organisms, first “metalloenzymes” (now termed metallo-beta-lactamases [MBLs]) in Gram negative lactose non-fermenting organisms, followed by serine carbapenemases in Enterobacteriaceae [33]. Carbapenem resistance in clinical Enterobacteriaceae isolates appears to have been uncommon prior to the 1990s [34]. However the first decade of the 2000s saw significant increases in CRE rates in the US, particularly in *Klebsiella pneumoniae,* with meropenem resistance increasing from 0.6% of isolates tested in 2004 to 5.6% in 2008 [35] in one survey and carbapenem resistance of 10.8% being noted in another [36]. Data from surveillance studies suggested that New York City was at the centre of this emerging epidemic [37-39], and the epidemiology of CRE in New York City and its links to Latin America will constitute a central part of this PhD. This led to increasing recognition of CRE as a clinical problem, both in terms of rising incidence and the high mortality associated with CRE infection.

Phenotypic carbapenem resistance in Enterobacteriaceae can be classified into two groups of underlying mechanisms: those involving carbapenemase production, and those where no carbapenemase is produced. We will address these now.

Carbapenemase production

There are two key classification schemes of carbapenemases, the Ambler scheme, which classifies carbapenemases on the basis of their amino acid sequences [40], and the Bush-Jacoby scheme, a functional classification [41]. The Ambler scheme will be used as a framework in this review (Figure 1.1).

Figure 1.1 – Classification of beta-lactamases



Beta-lactamases can be classified using either the Bush-Jacoby scheme, a functional classification, or the Ambler scheme, which classifies carbapenemases on the basis of their amino acid sequences. This figure demonstrates the corresponding groups or classes in these schema and gives examples of the more common beta-lactamases in each of those specific groups.

Serine carbapenemases

This group requires serine at the active site of the carbapenemase and comprises of Ambler classes A and D. Class A contains numerous important carbapenemases (Figure 1.1), amongst them *K. pneumoniae* carbapenemases (KPC), imipenem-resistant (IMI) carbapenemases and Guiana extended spectrum (GES) carbapenemases.

KPC carbapenemases (including subtypes KPC-2 to KPC-13) have been significant contributors to the global spread of CRE and confer resistance to all beta-lactams [42]. They are typically located on plasmids, which will often carry resistance determinants for other classes of commonly used drugs such as aminoglycosides, trimethoprim/sulfamethoxazole, fluoroquinolones and fosfomycin, thus rendering isolates carrying these plasmids multi-drug resistant. KPC carbapenemases have a typical association with carriage on a Tn3-related transposon that allows mobilization into diverse genetic environments. This has led to KPC carbapenemases being noted in a variety of plasmid Inc types, including FII, L/M, N, R and pKpQIL [43-45]. International KPC spread has been attributed to *K. pneumoniae* clonal group (CG)258 (including multi-locus sequence types [MLST] ST258, ST11 and ST512). However, KPC-carrying plasmids are now noted in a wide variety of not only *K. pneumoniae* sequence types [46] but other Enterobacteriaceae species, with *Enterobacter* being particularly noted [47, 48].

Members of CG258 have geographic associations: ST258 in the US, Americas and Israel, ST11 in Asia and ST512 in Southern Europe and Israel [49]. The reasons for the epidemic spread of this clonal group remain unclear. Capsular variation within ST258 has been proposed as a method of diversification that allows evasion of host immune responses [49] and ST258 has been classified into two clades on the basis of variation in a 215 kb area that included a region for capsule polysaccharide biosynthesis (*cps)* [50]. ST258 may also have a predilection for uptake of foreign genetic material through the presence of a type IV pilus and type 3 restriction-modification system [51, 52], however data on increased virulence in CG258 remain mixed [49].

Regardless of the underlying mechanisms, the successful spread of this high-risk clonal complex is indisputable with several regions now being endemic for KPC. The US was the origin of the first reports of KPC and to date all 50 states have reported cases [53]. However, there appear to be significant regional differences in KPC prevalence with the Northeast and Midwest continuing to be 'hot spots' [46]. Israel is also a KPC-endemic region, where the rapid expansion of KPC was originally linked to importation from the US [54]. In Europe, Greece and Italy report the highest CRE rates [55, 56]. MBLs initially predominated, however since the expansion of epidemic *K. pneumoniae* CG258 in the mid-2000s these regions are now KPC endemic with 61.9% of Greek *K. pneumoniae* isolates being carbapenem-resistant in 2015, the majority due to carriage of KPC [56, 57]. Latin America is another KPC-endemic region, with Colombia initially reporting KPC isolates in 2005 [58]. KPC isolates were noted across the region subsequently, with Brazil becoming a significant focus of KPC endemicity [59].

Class D beta-lactamases are referred to as the OXA class due to their oxacillin-hydrolysing abilities and were initially noted in *Acinetobacter* spp. These included OXA-23, OXA-24/40 and OXA-58 and had weak carbapenemase activity [60] but in 2001 OXA-48, a novel enzyme that significantly hydrolysed carbapenems was noted in a *K. pneumoniae* isolate from Turkey [61]. OXA-48 enzymes have subsequently been increasingly found in other Enterobacteriaceae including *E. coli* and *Citrobacter freundii* and have been noted to be widely geographically spread [62-65]. OXA-48 appears to be endemic in Turkey and Malta with outbreaks and some sporadic spread in the Balkans and North Africa [6, 66]. Relative to Class A carbapenemases, OXA enzymes typically

have lower carbapenem minimum inhibitory concentrations (MICs), sometimes even below the MIC breakpoint defining carbapenem resistance [64]. The CDC CRE surveillance definition was altered in 2015 to include isolates resistant to ertapenem or possessing a carbapenemase gene, which will increase sensitivity for OXA enzymes. Similar to KPC enzymes, OXA enzymes are also typically associated with a transposon (Tn1999) and particular Inc plasmid type (Inc L/M) [62-64].

Metallo-beta-lactamases

These enzymes require zinc for hydrolysis of beta-lactams and are therefore inhibited by metal-chelating agents such as ethylenediaminetetraacetic acid (EDTA). They have activity against all beta-lactams with the exception of monobactams such as aztreonam and are not affected by beta-lactamase inhibitors. Historically, they were amongst the earliest transmissible carbapenemases identified with reports from Japan, Taiwan and Singapore in the 1990s [67, 68]. Key members of the group include New Delhi metallo-beta-lactamase (NDM), imipenem resistance metallo-beta-lactamase (IMP) and Verona integron-encoded metallo-beta-lactamase (VIM). Each of these appears to have a distinct geographic distribution.

NDMs were first identified in a Swedish patient who had received care in India in 2009 [69] and have subsequently been found to be endemic in the Indian subcontinent [70], which serves as the centre of the epidemic. These MBLs have now had global dissemination and have been noted in Enterobacteriaceae at rates of ~5% in the Middle East and the Balkans, with sporadic cases occurring elsewhere usually due to importation

from these higher prevalence regions [71]. A particular concern has also been the detection of NDM-harbouring bacteria in environmental samples such as drinking water and effluent in countries in the Indian subcontinent, typically with concurrent detection of other beta-lactamases such as CTX-M [72, 73]. To date, 24 variants have been noted [74]. In contrast to KPC carbapenemases, which initially appeared to be confined to limited species and clonal backgrounds, NDMs have been detected in diverse bacterial species including *Klebsiella* spp., *E. coli*, *Enterobacter* spp., *Citrobacter* spp., as well as lactose non-fermenting organisms such as *Acinetobacter* spp. and *Pseudomonas* spp. In *K. pneumoniae* and *E. coli,* NDMs have been distributed across more than 40 STs [74].

While the global spread of NDMs and importation into numerous countries has been a particular concern, IMP and VIM MBLs remain endemic in parts of Asia (Japan, Taiwan) and Europe (Italy, Greece, Spain and Hungary), respectively. Following their detection in 2002, IMP MBLs have been repeatedly detected in Australian CRE isolates and prior to recent KPC outbreaks were the most common contributor to CRE in Australia [75, 76]. In *K. pneumoniae*, IMP and VIM MBLs are typically associated as gene cassettes within class I integrons but they have also been identified in other enterobacterial species such as *E. coli*, *Enterobacter* spp., *Serratia* spp. and *Proteus* [45].

Non-carbapenemase mechanisms of carbapenem resistance

Although carbapenemase-producing organisms have accounted for the majority of CRE, non-CP mechanisms remain important causes of carbapenem resistance. These organisms typically produce other beta-lactamases such as ESBLs (e.g. CTX-M) and/or

AmpCs. Although these enzymes have little activity against carbapenems, gene duplication or increased gene expression through mutations in promoter sequences or upstream insertion sequence (IS) elements may lead to higher beta-lactamase levels and more activity against carbapenems [77]. In addition, there are typically structural changes that affect the permeability of the cell membrane to beta-lactams, thus effectively reducing beta-lactam levels [78]. Reduced membrane permeability has been described due to alterations in porins encoded by *ompK35* and *ompK36*. Changes in *ompK35* and *ompK36* have been noted in up to 84% and 34% of clinical carbapenem-resistant *K. pneumoniae* (CRKP) isolates, respectively [79]. Increased drug efflux is another contributory mechanism and there are several classes of MDR efflux pumps, with the resistance-nodulation-division family being the most important [80]. The AcrAB-TolC complex is a member of this family and has been the most widely studied due to its activity against almost all types of antibacterial agents, including carbapenems [80].

The prevalence of non-CP CRE is currently unclear as most clinical laboratories perform phenotypic susceptibility testing but do not routinely perform molecular testing to distinguish CP CRE from non-CP CRE [78]. Published rates have varied widely, ranging from 15.2% to 100% of CRE isolates depending on the study setting [55, 81]. Distinguishing between CP CRE and non-CP CRE may have implications for patient outcomes, treatment and infection prevention, as non-CP CRE have been associated with lower mortality and may be less likely to be transmitted to other patients [31, 78]. This PhD will attempt to address some of these unanswered questions regarding non-CP CRE epidemiology.

Polymyxin resistance

The lack of options for treating CRE infections has led to increased use of the polymyxin class of antibiotics (colistin and polymyxin B). Despite their toxicity, they are widely used as a last resort. Polymyxins are cationic antimicrobial peptides that destabilize Gram negative lipopolysaccharide (LPS), resulting in leakage of cytoplasmic content and ultimately cell death [82]. PR mechanisms in Enterobacteriaceae can be classified into chromosomal and plasmid-mediated determinants of resistance.

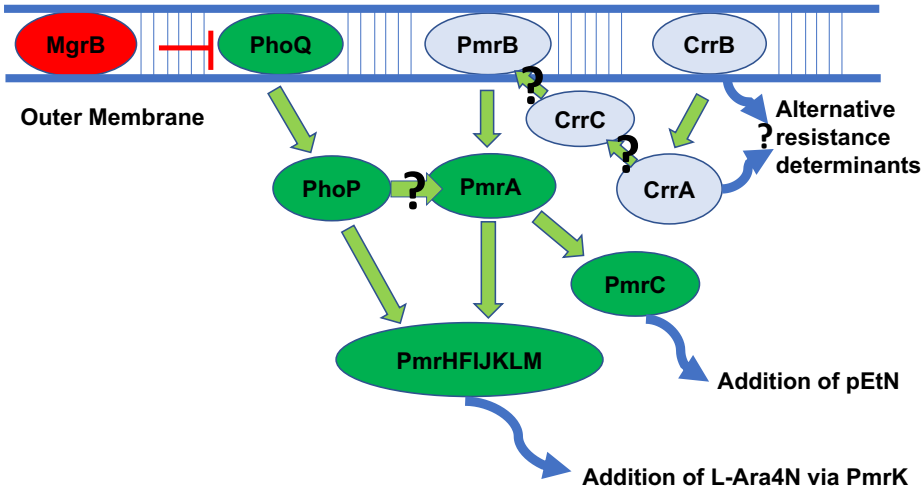
Chromosomal polymyxin resistance

Current knowledge of PR in Enterobacteriaceae has focused on LPS modifications resulting from addition of cationic groups (4-amino-4-deoxy-l-arabinose [L-Ara4N] and phosphethanolamine [pEtN]) that affect interactions of polymyxins with the bacterial cell membrane. In some Enterobacteriaceae including *Proteus* spp., *Serratia spp.* and *Morganella spp.*, the *arnBCADTEF* operon and/or the *eptB* gene are constitutively expressed leading to these alterations and intrinsic PR [83-85].

Other Enterobacteriaceae (e.g. *Klebsiella* spp., *E. coli*, *Enterobacter* spp., *Salmonella* spp.) are capable of acquiring resistance through convergence on a similar mechanism from multiple different pathways that typically include two-component systems. These have been summarised in Figure 1.2 and will be referred to as ‘canonical’ PR genes. In addition, to these more established loci, the literature has implicated numerous other genes that may constitute a ‘secondary resistome’ and these will be addressed briefly.

Figure 1.2 – Schematic representation of genetic mechanisms leading to polymyxin resistance in *K. pneumoniae*

(figure courtesy of Dr. T. H. McConville)



Chromosomal PR appears to be rare in general surveys of Enterobacteriaceae with rates of <0.1% to 1.5% [86]. However, rates amongst CRE isolates are notably higher. Indeed, detection of chromosomal PR globally seems to correlate with presence of CRE in those regions. In the US, rates of 6.7% to 18% have been noted in carbapenem-resistant isolates [87-90]. Southern Europe was noted to have particularly high rates of PR in CRE ranging from 20% to 43% in one countrywide Italian study [91]. Similarly, Greece has reported rates >20% [92, 93]. Other regions including Latin America, the Middle East and Asia have reported lower rates of 2-15% [86, 94-97]. In Australia, PR appears to be rare with 2.1% of Enterobacteriaceae isolates from 2007 – 2016 having PR [98]. Multiple outbreaks of PR have been described [99-107], however molecular data were often lacking in these reports, making it difficult to ascertain whether PR arose *de novo* following transmission of closely-related isolates. It is also unclear to what extent transmission of PR contributes to PR in non-outbreak settings. This PhD will attempt to use WGS to

address some of these gaps in our knowledge regarding the epidemiology and mechanisms of resistance of PR.

Canonical genes

mgrB

mgrB is a 47 amino acid transmembrane protein that acts as a negative regulator of the *PhoPQ* two component system. Inactivation of *mgrB* results in overexpression of *PhoPQ*, subsequently causing activation of the *pmrHFIJKLM* operon and L-Ara4N addition to the bacterial cell wall to confer PR. Inactivation may result from single nucleotide variants (SNVs) in the gene itself such as missense or nonsense mutations, truncations or frameshifts, from large scale changes including full deletions of *mgrB*, or from disruption by IS elements [108-112]. In addition to disruption within the gene, insertion of IS elements in putative regulatory regions upstream of *mgrB* also appear to play a role [113]. Several families of IS elements have been implicated with IS5-like elements being noted in multiple reports [108, 109, 113-116] and the possibility of transfer of IS elements from plasmids into the chromosome being raised [116]. To date, alterations in *mgrB* have been described only in *Klebsiella* spp. *MgrB* appears to be one of the most commonly affected genes to confer PR but there have not been broad molecular surveys of isolates with PR to determine its overall contribution.

crrAB operon

In 2015, Wright et al. noted a novel operon which was demonstrated to contribute to PR in a subset of clinical isolates composed of the regulatory protein *crrA* and the sensor

protein kinase *crrB* [117]. It has been hypothesized that mutations in *crrB* confer PR through subsequent activation of the *pmrHFIJKLM* operon and *pmrC.* Subsequent work has shown that this action is mediated through the increased expression of the newly named *crrC* acting through *pmrAB* [118]. In addition, following construction of a transposon mutant library, loci named H239_3063, H239_3064 and H239_3065 were shown to be co-transcribed, thus forming an operon [119]. H239_3064 was shown to contribute to PR through its actions as a resistance-nodulation-division efflux pump, rather than solely through activation of the *pmrHFIJKLM* operon.

phoPQ two-component system

This operon is composed of the regulator protein *phoP* and the sensor protein kinase *phoQ*, which are normally involved in responding to environmental stimuli associated with bacterial cell stress such as low pH, low magnesium or macrophage phagosomes [120, 121]. This operon controls a variety of genes for physiological function in Enterobacteriaceae including activation of the *pmrHFIJKLM* operon in response to presence of cationic antimicrobial peptides or removal of magnesium from the periplasm [120, 121].

pmrAB two-component system

Similar to the *phoPQ* operon, the *pmrAB* operon is activated under situations of cell stress. However, the specific stimuli seem to differ with aluminium and iron being implicated rather than magnesium, as well as low pH and macrophage phagosomes [120, 122]. *PmrB* activation leads to *pmrA* activation and subsequent transcription not only of

the *pmrHFIJKLM* operon leading to L-Ara4N addition, but also of *pmrE* resulting in pEtN addition (see Figure 1.2). Several mutations have been associated with PR, however there may be clonal polymorphisms that do not have a functional impact. Variants in the *pmrAB* operon have been implicated in PR in a variety of enterobacterial species including *K. pneumoniae, Enterobacter spp*., *Salmonella* spp. and *E. coli* [111, 123-127].

Secondary resistome

Numerous other loci have been proposed as contributors to PR and I have tried to summarise these genes, their actions and the species they have been described in in Table 1.1. Notable loci include *ramA*, which is found in *Klebsiella, Enterobacter*, *Salmonella* and *Citrobacter* spp. and is controlled by the *ramR* regulator. *RamA* has been implicated in PR, as well as tigecycline resistance [128]. In *Enterobacter* spp., heteroresistance to polymyxins has been associated with overexpression of the *acrAB-tolC* efflux pump, a resistance-nodulation-division pump [129]. This overactivity can be induced by *soxRS,* which appears to be a transcriptional regulator of *acrAB-tolC.* In *Acinetobacter* spp., alterations in lipid A synthesis genes *lpxA, lpxC* and *lpxD* have been associated with PR due to loss of LPS but this mechanism has not been noted in Enterobacteriaceae [130]. *LpxO* mutations have been associated with PR in experimental *K. pneumoniae* mutants [131], but to date have not been noted in clinical isolates.

Table 1.1 – Genes implicated in polymyxin secondary resistome in Enterobacteriaceae

Gene	Reference
H239_3063	[119]
H239_3064	[119]
crrC	[118]
arnA	[132]
arnB	[132]
arnC	[132]
arnD	[132]
arnE	[132]
arnF	[132]
arnT	[132]
pmrC	[122]
pmrD	[117]
pmrE	[132]
ompA	[133, 134]
ompC	[133, 134]
ompK35	[133, 134]
ompN	[133, 134]
ompW	[133, 134]
ompX	[133, 134]
soxR	[129]
soxS	[129]
acrB	[129]
tolC	[129]
ramA	[128]

Plasmid-mediated polymyxin resistance

In 2015, Liu et al. described *mcr-1,* a plasmid-mediated form of PR [30]. An increase in PR was noted during surveillance of *E. coli* isolates from Chinese food animals leading to detection of this novel resistance determinant. The protein encoded by *mcr-1* appears to be a pEtN transferase that enables addition of pEtN to lipid A. *Mcr-1* appears to be mobilized by an *ISApl1* composite transposon [135, 136] and may have arisen from *Moraxella* spp. [137, 138]. It has now been noted in numerous additional enterobacterial

species including *Klebsiella, Salmonella, Enterobacter, Citrobacter* and *Shigella* spp. [139]. Although in *E. coli mcr-1* has been found in 13 different plasmid incompatibility types, >90% are accounted for by IncI2, IncX4 and IncHI2 plasmids [140]. *Mcr-1* has been noted in isolates carrying both ESBLs and carbapenemases [141-145]. Polymyxin MICs of isolates with *mcr-1* mediated resistance are typically 2 – 4 mcg/ml and are lower than those noted with chromosomal PR [139].

Since being noted in 2015, *mcr-1* harbouring isolates have been found in Asia, Africa, Europe, North America, South America and Oceania. Worryingly, the prevalence of *mcr-1* carriage in humans in certain regions also appears to be increasing. Zhong et al. noted an increase from 0.4% of faecal samples in 2011 to 26.2% in 2016 in Guangzhou, China [146]. Concerns regarding the possible role of colistin use in animals in contributing to the spread of *mcr-1* have led to recent bans in several countries including China, India, Brazil and Malaysia [147, 148]. Colistin had already been banned as a growth promoter in Europe in 2006 and had never been approved for veterinary use in the US and Canada [147].

In addition to *mcr-1*, to date 9 additional *mcr* homologues have now been described (*mcr-1* to *-9*) [149], as well as several minor variants that only differ to *mcr-1* by 1-2 amino acids [144]. This has led to a proposal for *mcr* gene nomenclature similar to that applied to beta-lactamases [150]. These additional *mcr* homologues appear to be much less widespread than *mcr-1* [139].

Multi-drug resistant Gram negative bacteria in healthcare settings

CRE

The epidemiology of CRE depends on the region and type of carbapenemase present, with KPCs primarily associated with healthcare contact, while NDMs have had substantial spread into the environment and community [6]. The most data are available for KPC-producing *K. pneumoniae,* and healthcare contact and antimicrobial exposure have been consistently identified as risk factors [6, 34, 45, 46]. Healthcare risk factors have included duration of hospitalization, invasive medical devices, multiple comorbidities and poor functional status [6]. Antimicrobial exposure may contribute to CRE infection and colonization through multiple mechanisms including through evolutionary pressures on bacteria, ecological effects on the microbiota and selection of resistant bacteria that may already be present in smaller quantities [78]. Almost every class of antibiotic has been associated with selection of CRE, with total cumulative exposure possibly having more influence than specific antibiotic class [45]. Residence in long-term acute care hospitals also appears to be an important risk factor for CRE colonization [151]. Long-term acute care hospitals may serve as important reservoirs for transmission of CRE due to high rates of CRE colonization and frequent transfers of patients to multiple facilities [152-154].

CRE has also been associated with solid organ and hematopoietic stem cell transplantation [155-157]. For solid organ transplant (SOT) recipients, CRE infections appear to occur early after transplant with most studies reporting infections occurring within 50 days of transplant [26]. Data are mostly limited to liver and kidney recipients with higher rates being reported in liver transplant recipients [26]. Hematopoietic stem cell

transplant recipients are also at particular risk of Gram negative infections due to chemotherapy-related mucositis and neutropenia. There is therefore a significant risk of CRE infection in CRE endemic regions such as the US and Italy [158, 159]. Transplant recipients are disproportionately affected by CRE (and multi-drug resistant organisms [MDROs] more broadly) and infections with these organisms cause substantial morbidity and mortality. The epidemiology of CRE in transplant recipients will therefore be a central part of this PhD.

Patients colonized with CRE constitute an important reservoir for healthcare-associated CRE transmission. This has led to increasing interest in understanding the natural history of CRE colonization. Overall colonization appears prolonged. A recent meta-analysis found that pooled ESBL/CRE colonization was 76.7% 1 month after initial detection and 35% at 12 months, with decolonization interventions not providing a clear benefit [160]. Factors associated with persistent carriage have included readmission, duration of hospitalization, positive clinical culture, carbapenem use and diabetes mellitus [161, 162].

Polymyxin resistant Enterobacteriaceae

Chromosomal mediated polymyxin resistance

In regions where the majority of PR arises due to chromosomal mutations in CRE, risk factors for PR appear to be similar to those for CRE. Exposure to polymyxins appears to be a consistent risk factor [163-167]. In a recent retrospective cohort study of PR isolates at a single institution from 2011 – 2016 neurologic disease, residence in a skilled nursing facility prior to admission, receipt of carbapenems in the last 90 days, prior infection with

a carbapenem-resistant organism, and receipt of ventilatory support were risk factors for PR in a multivariate model [168]. This is consistent with prior studies that showed multiple similar risk factors that are typically associated with healthcare exposure including recent hospitalization, colonization with carbapenem resistant isolates, presence of co-morbidities, older age, length of hospitalization and indwelling urinary catheter use [88, 97, 99, 166, 167, 169, 170].

Two studies have directly compared the epidemiology of polymyxin susceptible CRE to PR-CRE and noted few significant differences [88, 166]. Rojas et al. noted that time from first admission to first positive culture was shorter in patients with PR, as was the total length of stay. There was a trend to patients with PR being more likely to be admitted from long-term acute care hospitals or community, but this did not reach statistical significance [88]. There were no obvious differences in antimicrobial exposure between the groups in this study, however Giacobbe et al. noted that previous colistin therapy, previous KPC-producing *K. pneumoniae* colonization and Charlson score >3 were associated with PR [166].

Multiple healthcare-associated outbreaks of PR have been noted, predominantly in the US [106, 107] and Southern Europe [91, 99, 101-103, 171]. However, outbreaks have also been noted in Latin America [105] and other parts of Europe [97, 104, 172]. In most of these studies clonality was established on the basis of pulse-field gel electrophoresis, however with the exception of the study by Giani et al., the molecular mechanisms of PR were not identified. In that study, clonal expansion of a *K. pneumoniae* ST512 mutant with

deletion of *mgrB* was noted [99]. There are few data to suggest the proportion of PR that occurs due to clonal spread versus *de novo* resistance from drug exposure and this PhD will address this question through use of WGS.

Plasmid-mediated polymyxin resistance

Livestock appear to be the main reservoir of plasmid-mediated PR, due to use of polymyxins as growth promoters, as well as prophylaxis and treatment [173]. In keeping with this, most epidemiological studies have been conducted in China and the highest rates have been observed there [139]. Although there have been several studies exploring the prevalence of *mcr-1* [142, 174-176], few have focused specifically on associations with healthcare. In one study, 80% of patients with *mcr-1* faecal carriage were hospitalized but no further epidemiological data were provided [146]. A recent study from the US did note an outbreak of *mcr-1* due to use of contaminated duodenoscopes [177], therefore the potential for silent dissemination of *mcr-1* in healthcare settings will be studied as part of this PhD.

Multi-drug resistant Gram negative bacteria in Australia

Carbapenem resistance in Australia

CRE have been present in Australia for some decades and have now become endemic in certain high-risk settings such as intensive care units and burns units [178]. The overall prevalence remains low however, with carbapenemases being detected in <0.1% of *E. coli* and 0.3% of *K. pneumoniae* during surveillance of bloodstream infections in 2016 [179, 180].

The predominant carbapenemase has been *bla*$_{IMP-4}$, which was initially found in *A. baumanni* but has subsequently spread to a variety of Gram negative organisms including *E. cloacae* complex, *K. pneumoniae, Klebsiella oxytoca, Citrobacter* spp. and *Serratia marcescens* [75, 181-185]. Recent studies have noted that this spread may be occurring due to horizontal transfer on specific plasmid backgrounds such as Inc L/M and IncHI2 [182, 184].

In addition to this, other carbapenemases have been increasingly noted. This firstly relates to acquisition of carbapenemases during overseas travel [186, 187] and has led to outbreaks. The most significant of these was an extensive outbreak of *bla*$_{KPC-2}$ in Victoria [188, 189]. This has led to enhanced surveillance in Victoria with CRE becoming a notifiable public health condition and subsequent WGS of all CRE isolates [190].

Polymyxin resistance in Australia

Few data are available regarding prevalence of PR in Australia. In a large prospective survey of CRE in Victoria, only 3.1% of isolates were colistin non-susceptible [187] but the mechanisms underpinning PR in this setting have not been defined. In another Australian study, 96/4555 (2.1%) of CRE isolates tested were non-susceptible to colistin [98]. Plasmid-mediated PR was noted in this collection with *mcr-1* being detected in 2 isolates [98].

Role of WGS in studying the emergence and spread of MDR-GNB

WGS provides an unprecedented opportunity to study and intervene against AMR. These possibilities can be divided into its use in studying resistance mechanisms in MDR-GNB and prevention of AMR through better understanding of genomic epidemiology. Both of these will be employed throughout the course of this PhD.

WGS for studying the emergence of resistance

Understanding the underlying genetic mechanisms of resistance plays a central role in developing novel diagnostic tools and treatments for MDR-GNB infections, while also providing insights into how these mechanisms arise and spread. The advent of treatments targeted at specific resistance mechanisms (e.g. beta-lactamase inhibitors targeting serine beta-lactamases) adds further impetus to elucidating resistance mechanisms.

AMR arising in Gram negatives can broadly be classified into two categories, mutational changes that occur due to evolutionary pressure from ongoing antimicrobial exposure, and horizontal gene transfer [191]. Previously, it was only possible to track this evolution towards resistance in experimental models [192] but WGS has made it possible to detect single nucleotide changes that may lead to resistance. A widely used methodology to detect mutational changes leading to AMR has been to track the genomic changes of serial isolates from single patients where phenotypic differences emerge over time. The assumption is that any changes between two closely related but phenotypically different isolates are likely to be causative of resistance and/or due to host immune selection pressure [193]. This approach has been used to discover novel resistance mechanisms

[194-196] and may be particularly applicable for changes that arise in the bacterial chromosome. This is the case for PR and will therefore be used as an approach in this PhD.

In addition to the mutational changes described above, horizontal gene transfer is another crucial way that Gram negative organisms acquire resistance. Transfer of plasmids between bacteria is the most obvious example [197] and the use of long-read sequencing has made it increasingly possible to dissect plasmid dynamics both within individual patients [198, 199] and amongst larger populations of organisms, such as in the context of outbreaks [189, 199, 200]. These studies have shown the complexity of these interactions. Plasmids may transfer between bacterial species [189, 199, 200] and they have different propensity towards transfer [197]. While the plasmid may be the large overall unit responsible for horizontal gene transfer of resistance determinants, within that setting they are frequently mobilized by smaller level MGEs such as IS elements or transposons [201]. Two prominent examples of this are *bla*$_{KPC}$ and *mcr-1. Bla*$_{KPC}$ is almost invariably associated with segments of the Tn4401 transposon, which allow spread of KPC-encoding sequences [45]. In the original description, *mcr-1* was flanked by an IS*Apl*1 transposon upstream and this association with IS*Apl*1 has been confirmed in multiple plasmid backgrounds [30, 201-203]. Subsequent experiments have shown that IS*Apl*1 likely contributes to the mobilization of *mcr-1* [136].

WGS for hospital epidemiology of MDR-GNBs

Traditional hospital epidemiology centres on surveillance for important nosocomial pathogens (e.g. MDR pathogens) and epidemiological investigation when multiple patients are found positive for the pathogen of interest [204]. The available information (e.g. phenotypic antimicrobial susceptibility testing [AST], ward locations, medical interventions) then informs whether a possible outbreak may have occurred and allows additional infection prevention measures to be instituted. In addition to this 'shoe leather' epidemiology, bacterial typing approaches (e.g. pulse-field gel electrophoresis or MLST) have been used to provide additional data about the relatedness of isolates. However, there are multiple shortcomings of these approaches including significant use of resources to investigate outbreaks, possible delays in initiating infection prevention interventions while analysis is being conducted and limited resolution to distinguish between isolates that belong to the same clonal lineage. WGS provides the means to overcome some of these difficulties and is likely to play an increasing role as cost and turnaround time decrease.

The utility of WGS for outbreak detection was initially demonstrated in several early studies of methicillin-resistant *Staphylococcus aureus* (MRSA) outbreaks including a retrospective investigation of an MRSA outbreak in a neonatal intensive care unit [205] and a subsequent retrospective/prospective investigation in a special care baby unit [206]. Inevitably, studies in MDR-GNB followed with its use in an outbreak of KPC-producing *K. pneumoniae* at the National Institutes of Health being a prominent example [207]. In this study, WGS was applied to investigate why the outbreak persisted despite

early initiation of infection prevention measures. The findings underscored that integrating genetic data with epidemiological data provided the best means to elucidate transmission networks. In addition to detecting previously unrecognized relatedness, WGS also plays an important role in ruling out outbreaks where clusters may temporally occur by chance but are subsequently found to be unrelated [204].

Subsequently WGS has been used in a variety of settings to track nosocomial outbreaks of MDR-GNB [104, 119, 189, 200, 207-216]. Several important points emerge from these studies. Firstly, several have highlighted the importance of silent carriers in propagating outbreaks [189, 207], allowing an argument to be made for active surveillance to detect these carriers and stop spread. Secondly, while most transmission occurs in the setting of patient overlap in time and space [204], the environment may also serve as a reservoir for ongoing transmission of MDR-GNB. This has been recognized for some time for organisms such as *Pseudomonas* and *Acinetobacter* spp. [214, 217, 218], is now increasingly recognized in the spread of CRE (e.g. KPC-producing *E. coli* and *K. pneumoniae* [219-222]) and has also identified emerging pathogens such as *Sphingomonas* [210]. Thirdly, while traditional epidemiological approaches have focused on organism species, use of WGS has highlighted the potential role of the 'mobilome' or spread of MGEs such as plasmids [189, 199, 200, 223, 224] that can occur between species to propagate MDR-GNB outbreaks. Finally, these studies have illustrated the importance of interfacility transmission for causing importation events that can then lead to outbreaks [189, 207, 223].

Use of WGS for hospital epidemiology has several limitations, however. Prolonged colonization may lead to significant bacterial within-host diversity, thus affecting inference about transmission patterns [189, 192, 199, 225]. Patients may also be susceptible to multiple acquisition events over time, leading to colonization with diverse and different strains and once again complicating transmission inference [225, 226]. When studying potential transmission, it can also be difficult to define standard cut-offs to determine if a transmission event has occurred [193, 227]. These may vary depending on the type of organism, the length of colonization and availability of samples to determine within-host versus inter-host variability. Due to these issues, WGS may often provide only an assessment of the probability of transmission rather than a definitive answer [204].

Use of WGS with novel bioinformatics approaches to study MDR-GNB

The increasing availability of bacterial WGS data now makes it possible to use it as an input for various novel bioinformatics approaches that require large datasets. The two that will be discussed include use of bacterial GWAS and use of ML for genotype-phenotype prediction of AST. Both of these approaches will be applied to study PR as part of this PhD.

Bacterial genome-wide association studies

GWAS study genetic variants that occur within a population of individuals and attempt to detect associations with a given phenotype [228]. This approach was based on the 'common disease common variant' hypothesis that the majority of human illness was due to highly prevalent variants that had small effects [229]. The first human GWAS was

published in 2005 in the era of microarrays [230] and has subsequently led to numerous studies that have established thousands of associations between SNVs and human disease. Due to initial technical difficulties and high upfront costs associated with creating a microarray for bacterial species [228], conducting a bacterial GWAS became feasible once bacterial WGS became more widespread. Here, I will address some of the methodology of conducting bacterial GWAS and summarize relevant studies in the field and tools that have been developed.

Methodology of conducting GWAS

Genotype representation

The first challenge in conducting a GWAS is the issue of how to represent genotypic data. While human studies initially used SNVs derived from microarrays, the use of WGS to characterize the genotype in bacterial GWAS opens several possibilities. The approach that most closely mirrors human studies uses SNVs derived through variant calling with a widely used reference. This may be most applicable to species where there is more limited recombination and clonal variation, such as *Mycobacterium tuberculosis,* but many bacterial species have a significant genome plasticity with a highly variable number of accessory genes. An alternative approach is therefore to attempt to test for gene presence or absence in addition to SNVs and integrate that information into the genotype representation, analogous to approaches in human GWAS that attempt to address copy number variation by assigning different weighting depending on the predicted functional impact [231, 232]. In addition, the functional impact of SNVs may be represented in several ways including a binary representation stating whether a gene contains a

mutation with a predicted significant functional impact, or whether there is a predicted change at the protein level. This form of grouping of individual variants (e.g. SNVs, insertions and deletions) allows an association between that genotype and the phenotype of interest to be identified when the individual variant has low penetrance, while at the same time reducing multiple testing and increasing statistical power [233].

One way to circumvent some of these issues is to use a reference-free approach that deals directly with genomic data. Approaches that rely on detecting variants in comparison with a reference will only detect changes in regions that are present in the reference. This may therefore ignore accessory genes or mobile genetic elements present in the isolate studied but not present in the reference, which may nevertheless have an impact on the phenotype. Using unprocessed genomic data rather than variants may thus allow study of both the core and accessory genome and analysis of isolates for which an appropriate reference may not be available [234]. In practical terms, this has usually involved detecting the presence or absence of k-mers, which are strings of nucleotides of *k* length. As opposed to human GWAS, this approach is more feasible in bacteria due to the smaller size of bacterial genomes and the fact that there are very few non-coding regions.

Prior to representing the genotype, the genomic data also needs to undergo quality control in order to remove low-quality data. This typically includes removing variants with low minor allele frequency (e.g. <1-5%), while also doing filtering as necessary for any genomic project such as filtering isolates with low coverage.

Phenotype representation

Similar to the challenges of representing a genotype for GWAS, representing a phenotype also involves making several choices. Obviously, the type of assay that was conducted has a significant impact. Taking phenotypic AST as an example, the assay determines an MIC that is a continuous variable. This can then be divided into discrete categories, as is done by using susceptibility breakpoints recommended to classify an organism as 'susceptible', 'intermediate' or 'resistant'. This can then be further simplified into a binary representation of 'susceptible' and 'non-susceptible'. Bacterial genomic data are becoming increasingly publicly available and several databases have been established to attempt to match this with accurate phenotypic metadata [235-238]. While this makes it possible to integrate these data into a bacterial GWAS, it is important to ensure the accuracy of the phenotypic information (e.g. AST method, reproducibility) prior to its use.

Population structure

The success of GWAS partially depends on the way that the genetic variants being tested are distributed through the population. In human genetics, segments of the genome are co-inherited over generations, leading to a non-random association of alleles at two or more loci that are close together, a phenomenon known as linkage disequilibrium. Linkage disequilibrium typically decreases with increasing distance on the chromosome [228], however in bacteria there is widespread linkage disequilibrium that is further interrupted by homologous recombination.

Bacteria also undergo clonal reproduction, thus creating less opportunity for genetic exchange. However, other mechanisms exist for creating genetic diversity including horizontal gene transfer, homologous recombination and recurrent mutations [239]. Each of these processes may give rise to homoplasy, a genetic trait arising separately in different genetic lineages as a result of evolution. Homoplasy indicates convergent evolution and is a marker of positive selection. It can therefore be used in association tests with phenotypes of interest [234]. Population structure can affect the outcome of GWAS as there may be a higher frequency of the allele of interest in a particular lineage that is a result of clonal relatedness. This may also be true for phenotype, which may cluster according to clonal lineage. In addition, the population structure of the isolates studied may be the result of positive selection (e.g. antimicrobial administration), thus further confounding the GWAS [231]. These findings may therefore lead to false positive results if population structure is not accounted for.

Statistical considerations

Linear and logistic regression have been the tests of association typically used in human GWAS [231]. Early bacterial GWAS studies used tools from human GWAS and thus also used regression approaches [228]. Multiple testing needs to be accounted for in order to avoid false positive results. While the standard cut-off for an association to be statistically significant is P=0.05 (also known as the α), the number of false positives will rise as more and more tests are performed. Therefore several techniques have been developed to correct for this [240]. The Bonferroni correction is the simplest of these, where the cut-off

for association becomes adjusted as α/n where n represents the number of tests performed. Other approaches include estimating a false discovery rate or a q value.

Validation

Once possible associations have been detected, they require validation. In human GWAS, this has been done through validation in an independent cohort [231]. This can also be done in bacterial GWAS studies, however the possibility of conducting *in vitro* experiments provides an important alternative for validation. Validation is done through mutagenesis experiments and recent developments in genome editing such as the use of CRISPR-Cas9 systems [241] have made this increasingly feasible. CRISPR-Cas9 will be used for functional validation of novel determinants of resistance as part of this PhD.

Published bacterial GWAS and tools for conducting bacterial GWAS

The first bacterial GWAS was published in 2013 [242] and subsequently numerous studies have been conducted, looking at a wide variety of pathogens and phenotypes (Table 1.2). AMR and toxicity have been common phenotypes studied. While early studies relied on tools that had been used for human studies, in recent years several tools specifically for bacterial GWAS have been developed [233, 243-245]. These have largely attempted to address the problems of population structure discussed above. Earle et al. formulated a k-mer based approach that incorporated linear mixed models to adjust for population structure, with putative advantages over other phylogenetic approaches as differences between strains can account for much of the genetic and phenotypic variability [244]. Collins et al. also attempted to use a phylogenetic approach to GWAS by using

Table 1.2 – Summary of selected bacterial genome-wide association studies

Reference	Year	Organisms	Phenotype	No. of genomes	Approach/ Software used	Findings
Sheppard et al. [242]	2013	*Campylobacter jejuni, C. coli*	Host adaptation	192	K-mers	Vitamin B5 synthesis plays central role
Farhat et al. [246]	2013	*Mycobacterium tuberculosis*	Antibiotic resistance	123	PhyC	39 novel resistance loci identified
Alam et al. [247]	2014	*Staphylococcus aureus*	Antibiotic resistance	75	ROADTRIPS	Novel SNV in *rpoB* gene associated with vancomycin intermediate susceptibility
Laabei et al. [248]	2014	*S. aureus*	Virulence	90	PLINK	121 novel loci associated with virulence
Chewapreecha et al. [249]	2014	*Streptococcus pneumoniae*	Antibiotic resistance	3701	PLINK	Multiple novel resistance loci
Salipante et al. [250]	2015	*Escherichia coli*	Antibiotic resistance	312	Gene presence / absence	No new loci identified
Holt et al. [251]	2015	*Klebsiella pneumoniae*	Virulence	288	Pangenome	*rmpA/2*, siderophore genes and iron metabolism genes important for virulence
Weinert et al. [252]	2015	*Streptococcus suis*	Virulence	191	PLINK / DAPC	No associations with human disease
Earle et al. [244]	2016	*M. tuberculosis, S. aureus, E. coli, K. pneumoniae*	Antibiotic resistance	3144	bugWAS	Detected genes and genetic variants underlying resistance to 17 antimicrobials
Brynildsrud et al. [253]	2016	*Staphylococcus epidermidis, S. pneumoniae*	Antibiotic resistance	21, 3085	Scoary	Correct prediction of genes associated with linezolid and erythromycin resistance
Collins et al. [243]	2018	*Neisseria meningitidis*	Antibiotic resistance / Virulence	171, 129	treeWAS	Confirmed known penicillin resistance loci, identified known and novel virulence determinants
Jaillard et al. [245]	2018	*M. tuberculosis, S. aureus, Pseudomonas aeruginosa*	Antibiotic resistance	1302, 992, 282	DBGWAS	Applied De Bruijn graph-based approach (unitigs) to recover known resistance determinants
Lees et al. [254, 255]	2016, 2018	*S. pneumoniae, Streptococcus pyogenes*	Antibiotic resistance / Virulence	3701, 3069	SEER, pySeer	Applied variable length k-mer approaches in SEER, pySeer also allows unitigs

the input dataset to conduct data simulations. The results of these were used to establish whether associations found are truly significant or arise by chance due to confounding factors found in the dataset [243]. Jaillard et al. used a De Bruijn graph-based approach for genotype representation in the GWAS by taking k-mers and constructing De Bruijn graphs of overlapping k-mers, thus attempting to reduce the redundancy of k-mer only approaches where multiple k-mers may ultimately map to the same locus [245]. Lees et al. initially developed SEER, a C+ package that uses variable length k-mers to represent the genome [255], which has subsequently been re-implemented in Python and now allows use of Elastic Net regularization [254].

Machine learning for genotype-phenotype prediction of AMR

The decreasing cost of bacterial WGS has led to its widespread use in research, with increasing interest in applying it in clinical settings [10]. Traditional culture techniques and AST can be associated with significant delays that impact patient care. This has led to active investigation into using bacterial WGS data to predict AMR phenotype [11, 256, 257]. Early work in this field focused on using rule-based approaches [258-260] that would predict susceptibility through detection of known resistance determinants (e.g. beta-lactamase genes) or known resistance mutations in housekeeping genes (e.g. *rpoB* conferring rifampin resistance in *S. aureus*).

The performance of these approaches has varied depending on the organism and antimicrobial tested (Table 1.3) but multiple limitations remain. Firstly, the approach assumes all determinants of resistance are known and is therefore unable to detect

previously uncharacterized determinants such as *mcr-1* [30]. Secondly, these rule-based models struggle to account for complex interactions between variants in multiple loci. In order to move beyond these limitations, several studies have attempted to use ML methods to predict phenotypic antimicrobial susceptibility.

Overview of machine learning

ML is a subfield of artificial intelligence described as giving "computers the ability to learn without being explicitly programmed" [261]. This is achieved by 'training' agnostic algorithms (also referred to as 'models') to identify patterns in data associated with a particular outcome of interest. Once trained, these models can then be applied to other datasets to predict the outcome measure correctly [262]. This differs to rule-based systems, which require explicitly pre-programmed algorithms. One prior limitation to the application of ML was then need for large datasets in order to successfully train ML models, however this has been overcome with the advent of big data and the emergence of such data-rich fields as "omics" disciplines.

ML can broadly assume two forms: supervised or unsupervised learning. In supervised learning, the prediction model is built using a training 'set' of data that specifies the outcome of interest [262]. This model takes an outcome measurement (referred to as a 'label', e.g., a patient developing a healthcare-associated infection) and tries to use available data (referred to as 'features' e.g. clinical information regarding the patient) to predict the label [263]. Once trained, the performance of the model is tested by using it on new, unseen data to predict the label. In order to arrive at the model that performs the

best predictions, the researcher iteratively tests a variety of algorithms with different individual settings. In contrast, unsupervised learning does not provide the model labels, only a dataset of features. The aim of unsupervised learning is to ascertain the underlying organization or structure of the data, with clustering being one of the most common unsupervised learning methods. This approach is commonly used in microbiome analyses where genetic data from a sample are used to cluster organisms based on genetic similarity into operational taxonomic units [264].

Methodology of ML for genotype-phenotype prediction of AMR

The methodology of GWAS and ML approaches share some similarities, particularly regarding representation of genotype and phenotype data. However, the statistical underpinnings differ, and these issues will now be discussed.

Feature engineering

Features represent the predictor variables in a dataset that are used as an input for ML methods. Feature engineering is the process of creating features that optimize the performance of ML algorithms. This is done through the ways that features are represented or are selected.

Feature representation

Genotype information can be represented in several ways, leading to much scope for creating different features. These challenges are similar to those described for GWAS methodology. Broadly, the representations could be grouped into two main categories,

reference-based and reference-free approaches. For reference-based approaches, there is typically variant calling using a reference organism. Individual SNVs can then be used as features, or grouped together as non-synonymous gene changes, gene presence or absence, or changes at the protein level. Hypothetically, it would also be possible to classify changes at the pathway level, although this has not been done for AST genotype-phenotype prediction to date. Depending on the choice of approach, the features can then be presented as individual alleles or in a binary matrix.

Reference-free approaches typically take k-mers as inputs. This offers multiple advantages because it allows for study of the core and accessory genome, incorporates complex genetic changes such as large insertions or deletions and can be used in situations where no well-curated reference is available. The potential drawback of this approach is that it increases the computational complexity of the analyses.

Feature selection

Feature selection is highly important for ML algorithms to achieve optimum performance. This is because it allows redundant or irrelevant data to be removed, reduces dimensionality of datasets and thus makes running ML approaches more computationally feasible, as well as making the model simpler to understand and explain because fewer features are involved in the way that it makes its predictions [265]. Feature selection approaches are divided into filtering, using embedded modules or wrapper methods.

Filtering approaches use intrinsic properties of features to select out the most important ones. In genomics, several studies have done this through using p-values from a previously performed GWAS [266-268] but other statistical measures such as chi-square tests, correlation coefficients or information gain could also be used. Embedded modules choose features as part of the process of training a model, allowing those features which contribute to the accuracy of the model to be incorporated into the model. Regularization methods are the most commonly used embedded modules and include algorithms such as LASSO, Elastic Net and Ridge Regression [269].

Regularization methods also introduce penalties while the predictive model is being optimized that intentionally bias the process to lower complexity (i.e. fewer features). This also has the flow-on effect of avoiding overfitting. Overfitting describes a phenomenon where the model performs extremely well on the training dataset but cannot be generalized to other datasets [265]. Wrapper methods are stand-alone methods that take different combinations of features and compare how well a standard model performs using those different combinations. As opposed to embedded modules, wrappers are implemented before formal training of the model begins.

Supervised learning algorithms

Genotype-phenotype prediction is a classification problem which fits into supervised learning, hence only supervised learning algorithms will be addressed here. These can broadly be divided into categories based on their functions. The most commonly

encountered for classification problems include regression, regularization, decision tree and ensemble algorithms [265, 270].

Regression algorithms co-opt standard statistical approaches such as linear regression and logistic regression. These model relationships between variables in order to make predictions based on the model function that has been optimized to reduce error. Regularization algorithms are essentially extensions of regression methods which penalize models based on complexity and their use in feature selection has already been mentioned. Decision tree algorithms classify instances using a series of rules based on feature attributes (e.g. is this k-mer present?). They have a tree-like structure and continue to fork until a prediction is made. This makes the underlying steps that an algorithm makes to arrive at a decision interpretable to a human and has contributed to the popularity of these algorithms. Classification and Regression Trees are a common example. Ensemble algorithms concurrently run multiple weaker algorithms, and then combine the results of these to reach a consensus about an overall prediction. Random Forest is a common ensemble algorithm that uses the outputs from multiple decision trees as votes to make a final decision. Other examples include AdaBoost and Gradient Boosted Regression Trees.

Published AST genotype-phenotype prediction studies

In parallel to the increasing availability of WGS data, in the last five years there has been a profusion of studies using ML for AST genotype-phenotype prediction. Available studies are summarized in Table 1.3 and reveal a diversity of organisms, antimicrobials

Table 1.3 – Summary of selected studies using machine learning for antimicrobial susceptibility genotype-phenotype prediction

Reference	Organism	No. of genomes	Input data	Algorithm	Performance metric	Phenotype prediction type
Alam et al. [247]	*Staphylococcus aureus*	75	SNVs	Random Forest	Sensitivity, Specificity	Susceptibility category
Chen et al. [271]	*Mycobacterium tuberculosis*	3601	SNVs in 28 known drug resistance loci	Deep Neural Networks	Sensitivity, Specificity	Susceptibility category
Davis et al. [272]	*M. tuberculosis, Streptococcus pneumoniae, Acinetobacter baumannii*	232 - 3067	K-mers	AdaBoost	Area under receiver operator curve (AUROC), F1 Score, Accuracy	Susceptibility category
Drouin et al. [273]	*Clostridium difficile, M. tuberculosis, Pseudomonas aeruginosa, S. pneumoniae*	111 - 556	K-mers	Set Covering Machine	Error Rate	Susceptibility category
Drouin et al. [274]	12 pathogens	120 - 5022	K-mers	Set Covering Machine	Accuracy	Susceptibility category
Eyre et al. [275]	*Neisseria gonorrheae*	681	SNVs in 17 known drug resistance loci	Multivariate Linear Regression	Accuracy	MIC
Her et al. [276]	*Escherichia coli*	59	Pangenome	Support Vector Machine, Naïve Bayes, AdaBoost, Random Forest	AUROC	Susceptibility category

Hicks et al. [277]	*N. gonorrheae, A. baumannii, K. pneumoniae*	204 - 1560	K-mers	Set Covering Machine, Random Forest, Random Forest Regression	Balanced Accuracy	Susceptibility category
Kavvas et al. [278]	*M. tuberculosis*	1595	Pangenome	Support Vector Machine	Mutual Information	Susceptibility category
Kouchaki et al. [279]	*M. tuberculosis*	13402	SNVs in 23 known drug resistance loci	Support Vector Machine, Logistic Regression, Product of Marginals	Accuracy, Sensitivity, Specificity, F1 Score, AUROC	Susceptibility category
Li et al. [280]	*S. pneumoniae*	4309	Amino acids variants in 3 drug resistance loci	Random Forest, Mode MIC	Agreement with phenotypic testing using Food and Drug Administration criteria	MIC
Long et al. [281]	*K. pneumoniae*	1777	K-mers	AdaBoost	AUROC, F1 Score	Susceptibility category
MacFadden et al. [282]	*E. coli*	414	Presence of known resistance loci, ST, epidemiologic data	Logistic Regression	AUROC	Susceptibility category
Mahe et al. [283]	*M. tuberculosis, Staphylococcus aureus*	2896, 971	K-mers	Logistic Regression with LASSO	AUROC	Susceptibility category
Mahe et al. [284]	*M. tuberculosis*	6574	K-mers	Logistic Regression with LASSO	AUROC	Susceptibility category

Moradigaravand et al. [285]	*E. coli*	1936	SNVs, gene content, population structure, isolation year	L2-penalized logistic regression, Random Forest, Gradient Boosted Descent Trees, Deep neural networks	Accuracy, F1 score, Precision, Recall	Susceptibility category
Nguyen et al. [286]	*K. pneumoniae*	1668	K-mers	XG Boost	Accuracy	MIC
Nguyen et al. [287]	Non-typhoidal *Salmonella* spp.	5278	K-mers	XG Boost	Accuracy	MIC
Pesesky et al. [288]	*Enterobacteriaceae spp.*	78	Annotated genes	Logistic Regression	Agreement with phenotypic testing	Susceptibility category
Rishishwar et al. [289]	*S. aureus*	25	SNVs in 14 loci	Logistic Regression, multiple others	Accuracy	Susceptibility category
Yang et al. [290]	*M. tuberculosis*	1839	SNVs in 23 known drug resistance loci	Multiple algorithms used	AUROC	Susceptibility category

and methodologies. This underlies the fact that there does not at present appear to be a clear consensus, with a possibility that the methodology is perhaps organism-, antimicrobial- and even dataset-dependent [277]. We will now discuss some of the issues underlying these studies.

ML has been applied to a wide variety of organisms, with *M. tuberculosis* and organisms causing hospital-acquired infections being a key focus (Table 1.3). The choice of organism often seems to be largely dictated by the availability of accurate genotypic data linked to appropriately curated AST phenotypic data, but the organisms studied vary substantially in terms of the composition of their genome. On one end of the spectrum is *M. tuberculosis* with a highly conserved genome and most resistance arising through mutations in core housekeeping genes. *K. pneumoniae* on the other hand has a large accessory genome including many MGEs that serve as important determinants of AMR. These differences impact methodology, in particular how genotypes should be represented. For *M. tuberculosis* it would be more feasible to use a reference-based approach involving variant calling, whereas for *K. pneumoniae* a reference-free approach using k-mers would allow for better assessment of the accessory genome. Organism population structure may also confound findings of AST genotype-phenotype prediction studies, similar to bacterial GWAS, but few studies have attempted to account for this [247, 285, 291].

While many studies have used similar methods to test multiple antimicrobials, there are limitations to this ‘one size fits all’ approach due to differences in underlying mechanisms

of resistance to that antimicrobial (which may also be organism-specific). Carbapenem resistance is a good example: while in Gram negative bacteria many forms of resistance are due to acquisition of carbapenemases, other factors such as porin mutations and efflux pumps play an important role, particularly in lactose non-fermenting organisms such as *Pseudomonas* spp*.* Furthermore, resistance to some antimicrobials can be complex requiring multiple concurrent changes to confer resistance, rather than a change in a single gene or acquisition of a single resistance determinant and this may affect performance of ML algorithms and require larger training datasets [286, 287, 290]. Vancomycin resistance in *S. aureus* is one such example and has been the subject of several studies [247, 289].

The choice of algorithm has also varied between studies, likely as a reflection of these algorithms being available in toolkits such as Sci-kit learn or Caret [292, 293]. However, the algorithms differ substantially, thus impacting how applicable they are for genotype-phenotype prediction. Many ML algorithms such as support vector machine classifiers and artificial neural networks abstract original data in ways that are difficult for humans to interpret, resulting in predictions seemingly generated by a 'black box'. This lack of transparency to clinicians may lead to problems with their uptake in clinical settings [294, 295]. Furthermore, one of the advantages of using ML approaches for AST genotype-phenotype prediction is the possibility of detecting novel resistance determinants. In order to do this, there has to be a degree of model interpretability. To this end, Drouin et al. elected to use an algorithm with interpretability in mind (the set covering machine) and implemented it for AST genotype-phenotype prediction [273, 274].

Finally, the goal of any approach utilizing ML for AST genotype-phenotype prediction is to train a model that has the best performance. However, there is not a single established metric used to assess performance as indicated by the variability in the studies listed in Table 1.3. While sensitivity and specificity are used to determine performance of diagnostic tests, commonly used metrics in the ML literature include accuracy, F1 score, precision (also known as positive predictive value), recall (equivalent to sensitivity) and area-under-receiver operator curve (AUROC). Hicks et al. noted that model performance for AST genotype-phenotype prediction varied substantially depending on the metric used [277]. The future application of the ML approach may therefore dictate the metrics used. For example, if the focus is to act as a screening test whose findings will then be confirmed using phenotypic methods, then sensitivity/recall would be an appropriate metric for measuring performance. Future work in ML for AST genotype-phenotype prediction may need to explicitly acknowledge the trade-offs between different performance metrics. There is currently a paucity of literature on AST genotype-phenotype prediction methodology. As a result, this topic will form a key component of my PhD.

Chapter 2:

Characterization of multi-drug resistant organism colonization and infection by genomic surveillance in liver transplant recipients

Work from this chapter was published in:

1) **Macesic N**, Gomez-Simmonds A, Sullivan SB, et al. Genomic Surveillance Reveals Diversity of Multidrug-Resistant Organism Colonization and Infection: A Prospective Cohort Study in Liver Transplant Recipients. *Clin Infect Dis* 2018; 67:905–912. [1]

The following contributions were made by specific co-authors to the work. No co-authors were Monash University students.

Nature of contribution	Co-authors
Study design	Uhlemann AC, Gomez-Simmonds A, Sullivan SB, Verna EC
Patient recruitment	Gomez-Simmonds A, Sullivan SB, Giddins MJ, Korakavi G, Leeds D, Park S, Shim K, Sowash MG, Hofbauer M, Finkel R, Hu Y, West J, Uhlemann AC
Bacterial culture	Gomez-Simmonds A, Sullivan SB, Giddins MJ, Ferguson SA, Korakavi G, Leeds D, Park S, Shim K, Sowash MG, Hofbauer M, Finkel R, Hu Y, West J, Uhlemann AC
DNA extraction and sequencing	Giddins MJ, Korakavi G
Sequence analysis	Toussaint NC
Patient data collection	Gomez-Simmonds A, Uhlemann AC

Introduction

The increasing availability of WGS has the potential to provide important insights into MDR-GNB colonization and infection. This is likely to happen through the increased resolution offered by WGS that allows precise identification of bacterial strains and mechanisms of resistance in ways that are difficult to perform with other molecular or phenotypic methods. As discussed in Chapter 1, this will allow a greater appreciation of MDR-GNB epidemiology by determining strains and mechanisms, but also by allowing relatedness between isolates to be quantified.

In order to assess how WGS could be used in a clinically impactful manner, I focused on SOT recipients. This is due to the extensive morbidity and mortality caused by MDRO infections in SOT recipients [22, 296, 297], but also because of the high rates of MDRO colonization in SOT recipients. Due to their extensive healthcare contact, SOT recipients colonized with MDROs may also serve as important reservoirs of MDRO transmission [298]. Therefore understanding MDRO colonization, infection, and transmission in SOT recipients is crucial to preventing healthcare-associated transmission of these organisms. Liver transplant recipients have also been increasingly affected by Gram negative infections [157], particularly CRE [296, 297, 299], making them a focus of this PhD.

My interest in this field began as a Transplant Infectious Diseases clinician following my arrival to the US in 2015. The Northeast had been the epicentre of the CRE epidemic in the US. During this time I became involved with a large project initiated by A/Prof. Uhlemann and her team studying the burden of MDRO infections in liver transplant

recipients. The project then evolved into several sub-studies over time, with me being in charge of the study that focused on characterizing the epidemiology of MDRO colonization and infection in this patient group through use of WGS. Our understanding of the temporal evolution of MDRO colonization, clearance, and infection, including interaction between different MDROs, was limited. In particular, previous studies often focused on a single MDRO class and lacked WGS data.

I therefore used active surveillance to characterize MDRO colonization and infection, then applied bacterial WGS to track the evolution and diversity of Gram negative colonization in liver transplant recipients.

Clinical Infectious Diseases

MAJOR ARTICLE

Genomic Surveillance Reveals Diversity of Multidrug-Resistant Organism Colonization and Infection: A Prospective Cohort Study in Liver Transplant Recipients

Nenad Macesic,[1,2] Angela Gomez-Simmonds,[1] Sean B. Sullivan,[1,3] Marla J. Giddins,[1,3] Samantha A. Ferguson,[1] Gautam Korakavi,[1] David Leeds,[1] Sarah Park,[1] Kevin Shim,[1] Madeleine G. Sowash,[1] Melanie Hofbauer,[1] Ryan Finkel,[1] Yue Hu,[1] Jared West,[1] Nora C. Toussaint,[4,a] William G. Greendyke,[1] Benjamin A. Miko,[1] Marcus R. Pereira,[1] Susan Whittier,[5] Elizabeth C. Verna,[6] and Anne-Catrin Uhlemann[1,3]

[1]Division of Infectious Diseases, Columbia University Medical Center, New York, New York; [2]Central Clinical School, Monash University, Melbourne, Victoria, Australia; and [3]Microbiome and Pathogen Genomics Core, Department of Medicine, Columbia University Medical Center, [4]New York Genome Center, [5]Clinical Microbiology Laboratory, Department of Pathology and Cell Biology, and [6]Division of Digestive and Liver Disease, Columbia University Medical Center, New York, New York

Background. Multidrug-resistant organisms (MDROs) are an important cause of morbidity and mortality after solid organ transplantation. We aimed to characterize MDRO colonization dynamics and infection in liver transplant (LT) recipients through innovative use of active surveillance and whole-genome sequencing (WGS).

Methods. We prospectively enrolled consecutive adult patients undergoing LT from March 2014 to March 2016. Fecal samples were collected at multiple timepoints from time of enrollment to 12 months posttransplant. Samples were screened for carbapenem-resistant Enterobacteriaceae (CRE), Enterobacteriaceae resistant to third-generation cephalosporins (Ceph-RE), and vancomycin-resistant enterococci. We performed WGS of CRE and selected Ceph-RE isolates. We also collected clinical data including demographics, transplant characteristics, and infection data.

Results. We collected 998 stool samples and 119 rectal swabs from 128 patients. MDRO colonization was detected in 86 (67%) patients at least once and was significantly associated with subsequent MDRO infection (0 vs 19.8%, $P = .002$). Child-Turcotte-Pugh score at LT and duration of post-LT hospitalization were independent predictors of both MDRO colonization and infection. Temporal dynamics differed between MDROs with respect to onset of colonization, clearance, and infections. We detected an unexpected diversity of CRE colonizing isolates and previously unrecognized transmission that spanned Ceph-RE and CRE phenotypes, as well as a cluster of *mcr-1*–producing isolates.

Conclusions. Active surveillance and WGS showed that MDRO colonization is a highly dynamic and complex process after LT. Understanding that complexity is crucial for informing decisions regarding MDRO infection control, use of therapeutic decolonization, and empiric treatment regimens.

Keywords. multidrug-resistant organism colonization; multidrug-resistant organism infection; carbapenem-resistant Enterobacteriaceae; whole-genome sequencing; liver transplantation.

Multidrug-resistant organisms (MDROs) represent an important cause of morbidity and mortality after solid organ transplantation (SOT) [1–3]. SOT recipients colonized with MDROs may serve as important reservoirs of MDRO transmission across the healthcare setting [4]. Elucidating mechanisms of MDRO colonization, infection, and transmission in SOT recipients is thus an urgent priority. In liver transplantation (LT), there has been a notable shift toward gram-negative infections [5], particularly carbapenem-resistant Enterobacteriaceae (CRE) [2, 3, 6]. Endogenous intestinal colonization with CRE appears to be an important risk factor for subsequent infection in LT patients [7, 8] and diverse SOT populations [9].

Our understanding of temporal evolution of MDRO colonization, clearance, and infection including interaction between different MDROs is limited. Previous studies often focused on a single MDRO class and were largely conducted during outbreaks or over limited time periods [7, 10–16]. Moreover, these studies lacked whole-genome sequencing (WGS) data and did not assess evolving resistance in MDROs or identify important interactions between organisms. For example, a subset of CRE isolates may result from stepwise mutations in extended-spectrum β-lactamase–producing organisms or uptake of resistance conferring plasmids [17], whereas others may result from new transmission events.

We used active surveillance to characterize MDRO colonization and infection including carbapenem-resistant Enterobacteriaceae (CRE), Enterobacteriaceae resistant to third-generation

Received 20 November 2017; editorial decision 23 February 2018; accepted 22 March 2018; published online April 28, 2018.

[a]Present affiliation: NEXUS Personalized Health Technologies, ETH Zürich, and Swiss Institute of Bioinformatics, Zurich, Switzerland.

Correspondence: A. -C. Uhlemann, Columbia University Medical Center, 630 W 168th St, New York, NY 10032 (au2110@columbia.edu).

Clinical Infectious Diseases® 2018;67(6):905–12

DOI: 10.1093/cid/ciy199

cephalosporins (Ceph-RE), and vancomycin-resistant enterococci (VRE), in a prospective, longitudinal cohort of LT recipients. We then applied bacterial WGS to track the evolution and diversity of gram-negative colonization in LT patients. Our central hypothesis was that patients experience a previously unrecognized diversity of MDRO colonization, both in terms of organism diversity and temporal dynamics, with important infection prevention implications.

METHODS

Study Population and Clinical Data

We prospectively enrolled consecutive adult patients (aged >18 years) undergoing LT at a tertiary care hospital from March 2014 to March 2016. Patients listed for LT or within 1 week posttransplant were eligible for participation. Fecal samples and/or rectal swabs were collected at the time of pretransplant enrollment, weekly during transplant hospitalization, and at months 2, 3, 6, 9, and 12 posttransplant. Subjects were excluded if they had not provided at least 1 sample at 12 months posttransplant. Approval was granted by the Columbia University Medical Center Institutional Review Board (IRB-AAAM7704).

Clinical data were collected including demographics, comorbidities, transplant indication, liver disease severity at time of LT (laboratory Model for End-Stage Liver Disease [MELD]; Child-Turcotte-Pugh [CTP] score), and prior MDRO infection or colonization. Donor characteristics were collected from the United Network for Organ Sharing. We also collected data on LT complications from enrollment until 1 year posttransplant (full details are provided in the Supplementary Methods).

Sample Collection and Specimen Analysis

Fecal samples were screened for CRE, Ceph-RE, and VRE with selective chromogenic agar (DRG International). Cultured isolates underwent identification and susceptibility testing with VITEK-2 (bioMérieux). Minimum inhibitory concentrations (MICs) were interpreted using Clinical and Laboratory Standards Institute criteria [18]. Isolates were classified as CRE according to 2015 Centers for Disease Control and Prevention (CDC) definitions (imipenem, meropenem, or doripenem MIC ≥4 μg/mL, ertapenem MIC ≥2 μg/mL) or detection of a carbapenemase as described below [19]. Enterobacteriaceae phenotypically resistant to any third-generation cephalosporin, with the exception of species with inducible cephalosporin resistance (eg, *Enterobacter* species) were designated Ceph-RE, according to institutional infection control guidelines. *Enterococcus* species isolates nonsusceptible to vancomycin (MIC ≥8 μg/mL) were defined as VRE. Any colonization detected in clinical microbiology records from 3 months pretransplant to end of follow-up was included in the analysis. We also collected all available clinical isolates during enrollment.

Definitions

Intestinal MDRO colonization was defined as isolation of VRE, CRE, or Ceph-RE from a rectal swab or stool sample. Colonization clearance was defined as collection of 2 successive negative fecal samples >1 week apart after the initial sample demonstrating colonization [16]. Colonization persistence was defined as failure to clear colonization. All positive bacterial cultures identified from clinical microbiology records were reviewed to determine whether they reflected colonization or infection. Infectious episodes were independently assessed by 2 infectious diseases physicians (N. M., A. G. S.) using National Healthcare and Safety Network criteria [20] and refereed by a third infectious diseases physician (A.-C. U.) in cases of disagreement.

Genomic Analysis

We performed WGS for at least 1 CRE isolate per timepoint from each CRE-positive LT patient. This included multiple CRE colonies (up to 5) per patient isolated simultaneously to estimate the clonal diversity of CRE colonization. We sequenced Ceph-RE isolates if a patient was persistently colonized with a Ceph-RE isolate, had a transition from Ceph-RE to CRE isolates of the same species, or had a colonization event after day 60.

Genomic DNA was extracted from overnight cultures, and index-tagged libraries were generated and sequenced using the HiSeq 2500 and MiSeq instruments (lllumina). We used SRST2 to identify multilocus sequence types (MLSTs) and determine antimicrobial resistance gene profiles including carbapenemase genes [21]. Detection of *mcr-1* by SRST2 was confirmed by polymerase chain reaction [22]. Further comparative genomic analysis was conducted on isolates from the same sequence type (ST) that were cultured from >2 patients. Isolates from nonstudy hospitalized patients during the study period were included for genomic context in comparative analyses and were available for *Klebsiella pneumoniae* clonal complex (CC) 258 (n = 88), *K. pneumoniae* ST17 (n = 5), and *Escherichia coli* ST117 (n = 2). Full details of genomic analyses are given in the Supplementary Methods.

Statistical Analysis

The primary outcome was MDRO colonization during the first year post-LT, and the secondary outcome was MDRO infection during the first year post-LT. In univariable analyses, categorical variables were compared using χ^2 or Fisher exact tests and continuous variables were compared using Student *t* test or Mann-Whitney-Wilcoxon test, as appropriate. Covariates with a *P* value <.05 in univariable analyses were considered for inclusion in multivariable models, which were constructed using stepwise model selection and manually curated. Statistical tests were 2-tailed with a threshold for statistical significance of $P < .05$. Statistical analyses were performed using SAS (version 9.4) and R (version 3.4.0) software.

RESULTS

We enrolled 142 of 180 patients who underwent LT over the 2-year study period, with 130 completing 1-year follow-up. Of these, 128 provided sufficient fecal samples and were included in the analysis (Supplementary Figure 1). The majority of patients were male (60%) and the median age was 60.4 years (Table 1). Hepatitis C infection was the most common reason for transplant (43%), followed by nonalcoholic fatty liver disease (17%). At time of transplant the median MELD score was 25 (interquartile range [IQR], 17–31), and 48 (38%) patients had hepatocellular carcinoma. During the 12-month follow-up period, 8 (6%) patients died. There were no significant differences between enrolled and nonenrolled patients with respect to demographics, baseline clinical characteristics, MDRO infections, and transplant outcomes (data not shown).

Table 1. Clinical Characteristics of Cohort (N = 128)

Characteristic	No. (%)
Demographic data	
Age, y, median (IQR)	60.4 (54.8–64.5)
Sex, male	77 (60)
UNOS race/ethnicity	
White	74 (58)
Hispanic/Latino	32 (25)
Asian	13 (10.1)
Black/African American	8 (6.2)
Multiracial	1 (0.7)
Underlying liver disease	
Hepatitis C infection	49 (38)
NAFLD	18 (14)
Autoimmune hepatitis/PSC/PBC	16 (13)
Alcoholic liver disease	14 (11)
Hepatitis B infection	9 (7)
Fulminant hepatic failure	2 (2)
Other	20 (15)
Other comorbidities	
Coexisting hepatocellular carcinoma	48 (38)
Diabetes mellitus	45 (35)
BMI at transplant, kg/m^2, median (IQR)	27.7 (23.5–31.0)
Transplant characteristics	
MELD score at transplant, median (IQR)	25 (17–31)
Child-Turcotte-Pugh score, median (IQR)	9.5 (7–11)
Living donor	18 (14)
Cold ischemic time, min, median (IQR)	341.5 (259–488)
Warm ischemic time, min, median (IQR)	35 (31–42)
PRBC transfusion required	30 (23)
PRBC units transfused, median (IQR)	5 (1–10)
Roux-en-Y	17 (13)
Postoperative complications	
Biliary stricture	17 (13)
Bleeding (<14 d)	31 (24)
Renal replacement therapy	16 (13)
Biliary leak	11 (9)
Hepatic artery thrombosis	2 (2)
Graft complication	
Any rejection episode	34 (26)
Primary graft failure	4 (3)
Outcome	
Days of initial ICU stay, median (IQR)	4 (2–6)
Days of initial hospital stay, median (IQR)	12 (8–17)
365 day post-LT mortality	8 (6)

Data are presented as No. (%) unless otherwise indicated.

Abbreviations: BMI, body mass index; ICU, intensive care unit; IQR, interquartile range; LT, liver transplant; MELD, Model for End-Stage Liver Disease; NAFLD, nonalcoholic fatty liver disease; PBC, primary biliary cirrhosis; PRBC, packed red blood cell; PSC, primary sclerosing cholangitis; UNOS, United Network for Organ Sharing.

MDRO Colonization and Infection

We screened 998 stool samples and 119 rectal swabs (median, 8 [IQR, 4–12] samples per patient). Rectal swabs were taken predominantly in the early posttransplant period and were associated with less MDRO recovery than stool specimens (32/119 [27%] vs 366/998 [37%], respectively, P = .035). MDRO colonization was detected in 86 (67%) patients. This included 25 (20%) patients with CRE, 53 (41%) patients with Ceph-RE, and 66 (52%) with VRE. Approximately half of MDRO-colonized patients (42/86 [48%]) harbored >1 class of MDRO.

In the cohort, 23 of 97 (23%) bacterial infection episodes were due to MDRO (Supplementary Table 1). Five CRE infections occurred in 3 patients, 13 Ceph-RE infections in 10 patients, and 8 VRE infections in 8 patients. MDRO infections were most frequently intra-abdominal (53%) and urinary tract infections (30%). Three intra-abdominal infections (caused by VRE, Ceph-RE, and VRE/CRE, respectively) were complicated by bacteremia, and 1 patient died.

MDRO-colonized patients vs noncolonized patients were significantly more likely to develop subsequent infection for all 3 categories of MDRO (CRE: 3/25 vs 0/103, P = .007; Ceph-RE: 9/52 vs 1/76, P = .001; and VRE: 8/66 vs 0/62, P = .006). Infection occurred a median of 31 days after colonization for CRE, 71 days for Ceph-RE, and 17.5 days for VRE. One patient had infection with Ceph-RE without prior detection of colonization.

Clinical Factors Associated With MDRO Colonization and Infection

Univariable and multivariable analyses of clinical factors associated with MDRO colonization and infection are shown in Supplementary Tables 2 and 3. MDRO-colonized patients, compared with noncolonized patients, had significantly higher CTP scores (P = .0004) and were more likely than noncolonized patients to have postoperative complications including prolonged mechanical ventilation (P = .003), bleeding (P = .002), need for reoperation (P = .04), prolonged postoperative hospital stays (P = .0002), and rehospitalization (P = .002) (Supplementary Table 2). In a multivariable model, CTP score (odds ratio [OR], 1.3 [95% confidence interval {CI}, 1.1–1.6], P = .002) and duration of posttransplant hospitalization (OR, 1.1 [95% CI, 1.0–1.2], P = .009) were independently associated with MDRO colonization (Supplementary Table 3). Patients with MDRO infection similarly had higher CTP scores (P = .003), and MDRO infection was associated with postoperative

complications (Supplementary Table 2). Overall, MDRO colonization was significantly associated with subsequent MDRO infection (100% vs 62%, $P = .002$). Multivariable analysis identified CTP score (OR, 1.5 [95% CI, 1.1–2.0], $P = .01$), bleeding complications (OR, 7.5 [95% CI, 2.2–25.1], $P = .001$), and need for rehospitalization (OR, 11.1 [95% CI, 1.3–98.2], $P = .03$) as independent predictors of MDRO infection (Supplementary Table 3). MDRO colonization could not be included in this model due to failure to converge.

Dynamics of MDRO Colonization Posttransplant and Relationship to Infection

At time of transplant, fewer patients were colonized with CRE compared with Ceph-RE and VRE (2% vs 14% and 29%, respectively, $P < .001$). CRE colonization was detected a median of 24 (IQR, 11–42) days post-LT vs 7 (IQR, 0–86) days for Ceph-RE and 0.5 (IQR, –3 to 19) days for VRE (overall $P = .05$; Figure 1). However, a greater proportion of Ceph-RE initial or recurrent colonization occurred after day 60 (22/53 [42%] vs 7/25 [28%] for CRE and 8/66 [12%] for VRE, $P = .0012$). While all patients with late CRE colonization (7/7) had recently been hospitalized, had invasive procedures, or had new antibiotic use, patients with late Ceph-RE colonization were less frequently hospitalized in the previous 30 days (12/22 [55%], $P = .045$).

Most patients cleared MDRO colonization (CRE: 15/18 [83%]; Ceph-RE: 24/34 [71%]; and VRE: 38/46 [83%]). Median times to clearance are shown in Table 2 and ranged from 22 to 58 days. However, 9 of 15 patients with CRE colonization achieved clearance through reversion from CRE to a Ceph-RE phenotype of the same species. Colonization persistence was not associated with infection (CRE: 2/3 [67%] patients, $P = .07$; Ceph-RE: 3/10 [30%] patients, $P = .34$; VRE: 2/8 patients [25%], $P = .59$). Colonization recurrence occurred in 7%–25% of patients (Table 2).

Genomic Epidemiology of Gram-Negative Isolates

We sequenced 95 CRE, including 80 colonizing and 15 infectious isolates. Fifty-seven were *K. pneumoniae*, 20 *E. coli*, 15 *Enterobacter cloacae* complex, 2 *Citrobacter freundii*, and 1 *Klebsiella oxytoca*. Through comparative genomic analyses within each species and ST, we identified 26 CRE clades. *Klebsiella pneumoniae* ST258 was the most common (39%), followed by *E. cloacae* complex ST252 (8%), *K. pneumoniae* ST17 (7%), and *K. pneumoniae* ST307 (5%) (Figure 2). We detected Bla_{KPC}-encoded carbapenemases in 63 of the 96 (66%) isolates, which was presumed to be the most frequent mechanism of carbapenem resistance. We found Bla_{KPC} in all CRE infection isolates, and 3 of 9 patients colonized with carbapenemase-producing CRE (CP-CRE) developed infection, compared with 0 of 16 patients with non–CP-CRE colonization ($P = .037$). Preceding or concurrent colonizing CRE isolates were the same ST as CRE infection isolates. However, 7 patients had colonization with multiple CREs (7/25 [28%]) as indicated by species and ST, including 1 patient colonized with at least 7 different CREs (*K. pneumoniae* ST17, ST34, and ST258; *E. cloacae* complex ST252 and ST454; *E. coli* ST167 and ST2585). Isolates from this patient had 3 different possible mechanisms of carbapenem resistance (bla_{KPC-2}, bla_{KPC-3}, and carbapenemase-negative).

Genomic comparisons of Ceph-RE (n = 40) and CRE isolates showed complex relationships. Nine patients demonstrated intraspecies transition from Ceph-RE to CRE. For 6 of 8 patients with serial isolates available, CRE was of the same ST as the preceding Ceph-RE isolate and in 5 of 6 patients

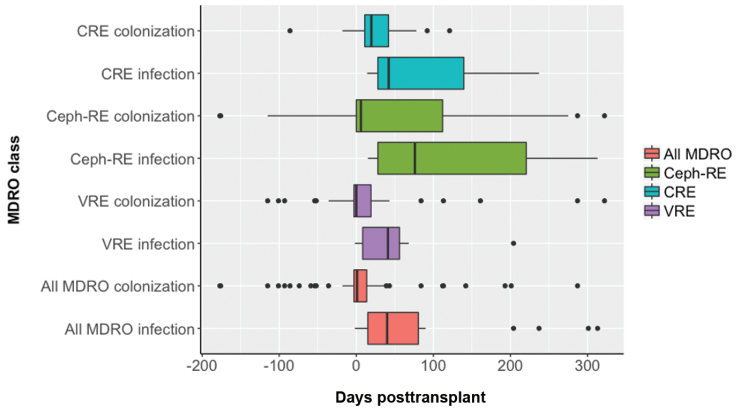


Figure 1. Timing of multidrug-resistant organism (MDRO) colonization and infection by organism class. Timing of first colonization and infection between different MDRO classes, as well as for all MDROs as a combined category, were plotted. Differences were noted between MDRO classes as well as between colonization and infection for individual MDRO classes. Abbreviations: Ceph-RE, Enterobacteriaceae resistant to third-generation cephalosporins; CRE, carbapenem-resistant Enterobacteriaceae; VRE, vancomycin-resistant enterococci.

Table 2. Colonization Trajectory by Multidrug-Resistant Organism Class

Colonization Trajectory	CRE	Ceph-RE	VRE	*P* Value
Time to first colonization, d, median (IQR)	24 (11–42)	7 (0–86)	0.5 (–3 to 19)	.05
Persistence	3/18 (17)	10/34 (29)	8/46 (17)	
Clearance	15/18 (83)	24/34 (71)	38/46 (83)	.42
Time for clearance, d, median (IQR)	22 (8–65)	53 (12.5–143.2)	36.5 (14–79.8)	.069
Colonization recurrence	1/15 (7)	8/24 (25)	5/38 (13)	.098
Colonization episode after day 60 (patients)	7/25 (28)	22/53 (42)	8/66 (12)	.0012
Hospitalization in previous 30 d	7/7 (100)	12/22 (55)	7/8 (88)	.045
Invasive procedure in previous 30 d	7/7 (100)	15/22 (68)	5/8 (63)	.22
New antibiotic use in previous 30 d	7/7 (100)	11/22 (50)	5/8 (63)	.062

Data are presented as no./No. (%) unless otherwise indicated.

Abbreviations: Ceph-RE, Enterobacteriaceae resistant to third-generation cephalosporins; CRE, carbapenem-resistant Enterobacteriaceae; IQR, interquartile range; VRE, vancomycin-resistant enterococci.

was non-CP CRE. However, CRE to Ceph-RE transition was associated with a change in ST of the Ceph-RE isolate in 5 of 9 patients. Analysis of isolates causing colonization persistence showed persistence of the same ST in 2 of 3 patients with CRE and 9 of 10 Ceph-RE.

We then used comparative genomics to assess relatedness between isolates of the same clonal background (Table 3). The analysis for *K. pneumoniae* CC258 indicated that median pairwise single-nucleotide polymorphism (SNP) distance was 1 (range, 0–7) between isolates occurring within single patients, compared with 381 (range, 0–2066) between isolates from different patients. We applied the maximum pairwise SNP distance within individual patients as a cutoff for putative transmission clusters between patients. We identified 11 clusters of closely related isolates shared by at least 2 patients. Five clusters involved LT patients, including 4 clusters containing colonizing isolates and 1 cluster containing isolates resulting from infections in LT patients (Supplementary Figure 2). No cluster was limited to LT patients. Likewise, *K. pneumoniae* ST17 isolates were similar between LT and non-LT patients. With the exception of *K. pneumoniae* ST307, all analyses also showed evidence of interpatient transmission.

An unexpected finding was the detection of *mcr-1* in Ceph-RE *E. coli* isolates from 3 patients, all of whom were noted to have a new Ceph-RE colonization event >60 days posttransplant. The isolates were *E. coli* ST117, as was a previously reported *mcr-1*–producing case [23]. Comparative analyses found a high level of relatedness between the isolates (3–24 SNPs).

DISCUSSION

In this prospective longitudinal cohort of LT patients, we identified high rates of MDRO colonization (67%) and significant association between MDRO colonization and subsequent infection. Using a combination of active surveillance and WGS, we demonstrated unexpected heterogeneity of CRE colonizing isolates, previously unrecognized transmission spanning Ceph-RE and CRE phenotypes, and a cluster of *mcr-1*–producing isolates. This approach allowed us to characterize temporal dynamics of MDRO colonization, revealing differences in onset of colonization, clearance, and infections between the different MDROs and complex relationships between Ceph-RE and CRE colonization.

We detected MDRO colonization in 67% of patients during the 1-year study period. Consistent with previous literature, MDRO colonization was associated with subsequent infection [8, 24–26]. Posttransplant complications (including need for

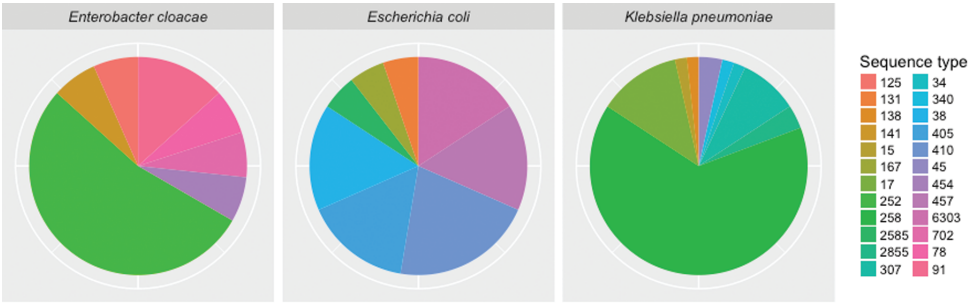


Figure 2. Multilocus sequence types of common carbapenem-resistant Enterobacteriaceae species. Carbapenem-resistant Enterobacteriaceae isolates were categorized by species, then sequence type, showing a diversity of carbapenem-resistant Enterobacteriaceae clades.

Table 3. Comparative Genomic Analysis of Carbapenem-Resistant Enterobacteriaceae and Enterobacteriaceae Resistant to Third-Generation Cephalosporin Isolates of the Same Clonal Background

Species and MLST	Reference	Accession	No. of Study Patient Isolates	No. of Nonstudy Isolates	Median Within-Patient Pairwise SNP Distance (Range)	Median Between-Patient Pairwise SNP Distance (Range)	Susceptibility Phenotype
Klebsiella pneumoniae ST258	NJST258_1	CP006923.1	22	88	1 (0–7)	381 (0–2066)	CRE only
K. pneumoniae ST17	MGH_60	NZ_JMYY00000000.1	8	5	44 (10–92)	2078.5 (90–2244)	CRE and Ceph-RE
K. pneumoniae ST307	KPN11	NZ_CTN01000001.1	5	0	4–35	64–201	CRE only
Escherichia coli ST117	MDR_56	CP019903.1	4	2	3–17	14 (7–24)	Ceph-RE only
E. coli ST131	JJ1886	CP006784.1	9	0	49 (2–96)	80 (21–234)	CRE and Ceph-RE
E. coli ST405	BIDMC106	GCA_001030595.1	7	0	2 (2–4)	117.5 (5–972)	CRE and Ceph-RE

Abbreviations: Ceph-RE, Enterobacteriaceae resistant to third-generation cephalosporins; CRE, carbapenem-resistant Enterobacteriaceae; MLST, multilocus sequence type; SNP, single-nucleotide polymorphism; ST, sequence type.

reoperation and prolonged hospital stays) were associated both with MDRO colonization and infection, suggesting that in-hospital acquisition of MDRO in the early posttransplant period played an important role. This was particularly true for VRE, with a median time to first colonization of 0.5 days. However, Ceph-RE colonization often occurred after day 60 and was not associated with healthcare exposure, raising the possibility of community- rather than healthcare-associated acquisition [27].

The high MDRO colonization burden noted in our study may be partially explained by active surveillance. Without surveillance, many colonization events would have been undetected despite serving as drivers for MDRO transmission. For example, our data suggest that for every CRE-infected patient, approximately 7 others are colonized with CRE, with similar ratios for Ceph-RE (1:4) and VRE (1:7). This previously occult MDRO colonization has important infection control implications, as illustrated by the cluster of *mcr-1*–producing *E. coli* isolates: All were colonizing isolates with possible silent transmission.

Active surveillance and WGS provided a high-resolution viewpoint on MDRO colonization dynamics. Despite previous attempts to define its duration [10–15], MDRO colonization does not appear to be a simple "on–off" phenomenon. Patients follow complex trajectories (Figure 3). In select cases, phenotypic findings suggested persistent colonization, but WGS

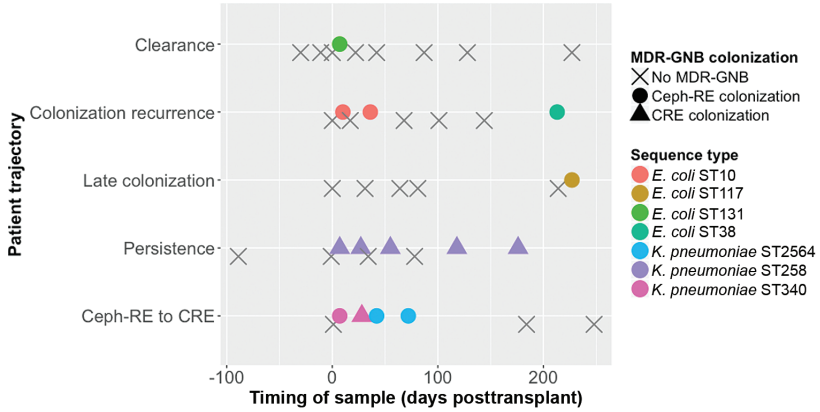


Figure 3. Schematic figure of dynamics of multidrug-resistant organism (MDRO) colonization. MDRO colonization has complex dynamics, with patients following different trajectories. In this schematic figure, 5 patients with typical colonization trajectories are shown. In the first example, the patient had rapid clearance after early posttransplant colonization with Enterobacteriaceae resistant to third-generation cephalosporins (Ceph-RE). In the second example, the patient clearance of initial colonization with *Escherichia coli* sequence type (ST) 10 but subsequent recolonization with *E. coli* ST38. The third example is of late colonization, in this case with *mcr-1*–producing *E. coli* ST117, likely due to healthcare-associated transmission. The fourth example shows a patient persistently colonized with *Klebsiella pneumoniae* ST258. The final example shows a patient colonized with Ceph-RE *K. pneumoniae* ST340, with transition to colonization with carbapenem-resistant Enterobacteriaceae *K. pneumoniae* ST340. Reversion to Ceph-RE colonization is associated with a change in sequence type to *K. pneumoniae* ST2564. Abbreviations: CRE, carbapenem-resistant Enterobacteriaceae; MDR-GNB, multidrug-resistant gram-negative bacteria.

revealed a change in colonizing organism ST. Conversely, WGS demonstrated that isolates phenotypically classified as CRE and Ceph-RE were in fact related. These findings can inform infection control interventions by detecting new transmission events and excluding others due to within-host evolution (eg, change from Ceph-RE to CRE).

Pathogen-related factors detected by WGS may help predict future infection and allow us to prioritize therapeutic interventions for MDRO colonization [28, 29]. In our study, detection of carbapenemase genes was one such factor. CRE infection developed in 33% of patients colonized with CP-CRE but in none with non–CP-CRE colonization only. These findings are in keeping with previously reported CRE infection rates from CP-CRE outbreak settings [7, 8] and with recent literature suggesting that patient outcomes differ between CP-CRE and non–CP-CRE infections [30].

Our WGS data demonstrated an unexpected diversity of multidrug-resistant gram-negative bacteria isolates, particularly CRE. While the most common CRE clone was *K. pneumoniae* ST258, consistent with US epidemiologic studies [31–33], we noted multiple non-ST258 *K. pneumoniae* clonal types including the emerging ST307 clone [34]. *Enterobacter* species were also prominent and are an increasingly recognized cause of CRE infections [35]. This diversity of CRE isolates supports CRE endemicity rather than detection within outbreak settings reported previously [36–38]. Through WGS, we noted that approximately 25% of CRE-colonized patients had multiple CRE clones. The presence of diverse mechanisms of carbapenem resistance in several patients suggests repeated and independent acquisitions of CREs. This finding underscores the high exposure of SOT recipients to MDRO colonization, both by virtue of their immunosuppressed state and through repeated healthcare-associated exposures.

Our analyses of isolates sharing the same ST showed several instances of horizontal transmission between patients. This relationship between liver transplant recipients and other hospital patients highlights the possibility of common modes of transmission. Three of 6 analyses included isolates that were CRE and Ceph-RE, demonstrating the continuum between the 2 categories. Our genomic data therefore highlighted the limitations of using susceptibility phenotypes to investigate isolate relatedness and possible horizontal transmission.

Several challenges for use of active surveillance and WGS in clinical practice remain. Despite their lower MDRO detection rate, use of rectal swabs may be an appropriate compromise to improve sample collection. Our center currently collects rectal swabs on all LT patients immediately following their transplant surgery. Isolation precautions are applied according to CDC guidelines [39]. Currently, we do not systematically collect samples at other time points; however, it would be feasible to collect rectal swabs during regular outpatient follow-up that recipients receive as standard care. The second challenge is use of WGS, both due to cost and technical feasibility. This challenge is part of a broader discussion of how genomic data should be integrated into clinical management of infectious diseases [40]. We are hopeful that with the decreasing cost of sequencing and increased availability of automated analysis platforms, WGS will be integrated into clinical care in the foreseeable future. Our study shows the possible advantages of having these data.

Our study had several limitations. First, our data came from a single center with a focus on one patient population. However, the data were collected prospectively over several years and for parts of our analysis we conducted comparisons with the broader hospital. We also faced significant challenges in conducting a real-world study involving many patients over such a long duration. Approximately 30% of patients receiving transplants during the study period were not enrolled. During the course of our study, there were disparities in the completeness of longitudinal stool sampling between patients. Our study relied on traditional culture methods for initial MDRO detection and may not have detected isolates present in low concentrations. Conversely, to minimize the impact of possible false negatives, we therefore defined clearance only when 2 samples were negative. Last, due to logistical constraints, our efforts at WGS were limited to CRE and selected gram-negative isolates.

In conclusion, MDRO colonization is a highly dynamic and complex process after LT. However, better understanding of MDRO colonization may allow us to make more precise decisions for our patients. Possible advances include identifying patients at risk of infection who would benefit from therapeutic decolonization through fecal microbiota transplantation, tailoring isolation precautions by identifying patients who truly present MDRO transmission risks, and limiting antimicrobial use to patients suspected of having developed infection. It is only through acknowledging the protean nature of MDRO colonization and infection using novel tools such as WGS that we can effectively understand and combat the spread of antimicrobial resistance and improve patient outcomes.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. We thank the team of the curators of the Institut Pasteur multilocus sequence types system (Paris, France) for importing novel profiles at http://bigsdb.web.pasteur.fr.

Financial support. This work was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (grant numbers AI116939 to A.-C. U., S1-AI116939 to A.-C. U. and A. G. S., and 5T32AI100852 to A. G. S.) and the Irving Scholarship from Columbia University (to A.-C. U.).

Potential conflicts of interest. A.-C. U. has received funding from Merck. All other authors report no potential conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. de Gouvêa EF, Martins IS, Halpern M, et al. The influence of carbapenem resistance on mortality in solid organ transplant recipients with *Acinetobacter baumannii* infection. BMC Infect Dis **2012**; 12:351.
2. Kalpoe JS, Sonnenberg E, Factor SH, et al. Mortality associated with carbapenem-resistant *Klebsiella pneumoniae* infections in liver transplant recipients. Liver Transpl **2012**; 18:468–74.
3. Pereira MR, Scully BF, Pouch SM, et al. Risk factors and outcomes of carbapenem-resistant *Klebsiella pneumoniae* infections in liver transplant recipients. Liver Transpl **2015**; 21:1511–9.
4. Pereira MR, Uhlemann AC. The evolving threat of carbapenem-resistant infections after liver transplantation: the case of *Acinetobacter baumannii*. Liver Transpl **2016**; 22:579–81.
5. Singh N, Wagener MM, Obman A, Cacciarelli TV, de Vera ME, Gayowski T. Bacteremias in liver transplant recipients: shift toward gram-negative bacteria as predominant pathogens. Liver Transpl **2004**; 10:844–9.
6. Lübbert C, Hau HM, Rodloff A, et al. Clinical impact of infections with carbapenem-resistant Enterobacteriaceae in liver transplant recipients. Z Gastroenterol **2015**; 53:1276–87.
7. Lübbert C, Becker-Rux D, Rodloff AC, et al. Colonization of liver transplant recipients with KPC-producing *Klebsiella pneumoniae* is associated with high infection rates and excess mortality: a case-control analysis. Infection **2014**; 42:309–16.
8. Giannella M, Bartoletti M, Morelli MC, et al. Risk factors for infection with carbapenem-resistant *Klebsiella pneumoniae* after liver transplantation: the importance of pre- and posttransplant colonization. Am J Transplant **2015**; 15:1708–15.
9. Freire MP, Abdala E, Moura ML, et al. Risk factors and outcome of infections with *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae* in kidney transplant recipients. Infection **2015**; 43:315–23.
10. Lübbert C, Lippmann N, Busch T, et al. Long-term carriage of *Klebsiella pneumoniae* carbapenemase-2–producing *K. pneumoniae* after a large single-center outbreak in Germany. Am J Infect Control **2014**; 42:376–80.
11. Zimmerman FS, Assous MV, Bdolah-Abram T, Lachish T, Yinnon AM, Wiener-Well Y. Duration of carriage of carbapenem-resistant Enterobacteriaceae following hospital discharge. Am J Infect Control **2013**; 41:190–4.
12. Patel R, Allen SL, Manahan JM, et al. Natural history of vancomycin-resistant enterococcal colonization in liver and kidney transplant recipients. Liver Transpl **2001**; 7:27–31.
13. Haverkate MR, Weiner S, Lolans K, et al. Duration of colonization with *Klebsiella pneumoniae* carbapenemase-producing bacteria at long-term acute care hospitals in Chicago, Illinois. Open Forum Infect Dis **2016**; 3:ofw178.
14. Apisarnthanarak A, Bailey TC, Fraser VJ. Duration of stool colonization in patients infected with extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*. Clin Infect Dis **2008**; 46:1322–3.
15. Birgand G, Armand-Lefevre L, Lolom I, Ruppe E, Andremont A, Lucet JC. Duration of colonization by extended-spectrum β-lactamase-producing Enterobacteriaceae after hospital discharge. Am J Infect Control **2013**; 41:443–7.
16. Feldman N, Adler A, Molshatzki N, et al. Gastrointestinal colonization by KPC-producing *Klebsiella pneumoniae* following hospital discharge: duration of carriage and risk factors for persistent carriage. Clin Microbiol Infect **2013**; 19:E190–6.
17. Goodman KE, Simner PJ, Tamma PD, Milstone AM. Infection control implications of heterogeneous resistance mechanisms in carbapenem-resistant Enterobacteriaceae (CRE). Expert Rev Anti Infect Ther **2016**; 14:95–108.
18. Clinical and Laboratory Standards Institute. Performance standards for antimicrobial susceptibility testing. 25th informational supplement, M100-S25. Wayne, PA: CLSI, **2015**.
19. Centers for Disease Control and Prevention. Facility guidance for control of carbapenem-resistant Enterobacteriaceae (CRE). Atlanta, GA; CDC, **2015**.
20. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. Am J Infect Control **2008**; 36:309–32.
21. Inouye M, Dashnow H, Raven LA, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med **2014**; 6:90.
22. Liu YY, Wang Y, Walsh TR, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis **2016**; 16:161–8.
23. Macesic N, Green D, Wang Z, et al. Detection of *mcr-1*-carrying *Escherichia coli* causing bloodstream infection in a New York City hospital: avian origins, human concerns? Open Forum Infect Dis **2017**; 4:ofx115.
24. Tischendorf J, de Avila RA, Safdar N. Risk of infection following colonization with carbapenem-resistant Enterobactericeae: a systematic review. Am J Infect Control **2016**; 44:539–43.
25. Vehreschild MJ, Hamprecht A, Peterson L, et al. A multicentre cohort study on colonization and infection with ESBL-producing Enterobacteriaceae in high-risk patients with haematological malignancies. J Antimicrob Chemother **2014**; 69:3387–92.
26. Alevizakos M, Gaitanidis A, Nasioudis D, Tori K, Flokas ME, Mylonakis E. Colonization with vancomycin-resistant enterococci and risk for bloodstream infection among patients with malignancy: a systematic review and meta-analysis. Open Forum Infect Dis **2017**; 4:ofw246.
27. Doi Y, Park YS, Rivera JI, et al. Community-associated extended-spectrum β-lactamase-producing *Escherichia coli* infection in the United States. Clin Infect Dis **2013**; 56:641–8.
28. Millan B, Park H, Hotte N, et al. Fecal microbial transplants reduce antibiotic-resistant genes in patients with recurrent *Clostridium difficile* infection. Clin Infect Dis **2016**; 62:1479–86.
29. Bilinski J, Grzesiowski P, Sorensen N, et al. Fecal microbiota transplantation in patients with blood disorders inhibits gut colonization with antibiotic-resistant bacteria: results of a prospective, single-center study. Clin Infect Dis **2017**; 65:364–70.
30. Tamma PD, Goodman KE, Harris AD, et al. Comparing the outcomes of patients with carbapenemase-producing and non-carbapenemase-producing carbapenem-resistant Enterobacteriaceae bacteremia. Clin Infect Dis **2017**; 64:257–64.
31. Munoz-Price LS, Poirel L, Bonomo RA, et al. Clinical epidemiology of the global expansion of *Klebsiella pneumoniae* carbapenemases. Lancet Infect Dis **2013**; 13:785–96.
32. Logan LK, Weinstein RA. The epidemiology of carbapenem-resistant Enterobacteriaceae: the impact and evolution of a global menace. J Infect Dis **2017**; 215:28–36.
33. Gupta N, Limbago BM, Patel JB, Kallen AJ. Carbapenem-resistant Enterobacteriaceae: epidemiology and prevention. Clin Infect Dis **2011**; 53:60–7.
34. Long SW, Olsen RJ, Eagar TN, et al. Population genomic analysis of 1777 extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* isolates, Houston, Texas: unexpected abundance of clonal group 307. MBio **2017**; 8. doi:10.1128/mBio.00489-17.
35. Gomez-Simmonds A, Hu Y, Sullivan SB, Wang Z, Whittier S, Uhlemann AC. Evidence from a New York City hospital of rising incidence of genetically diverse carbapenem-resistant *Enterobacter cloacae* and dominance of ST171, 2007–14. J Antimicrob Chemother **2016**; 71:2351–3.
36. Marchaim D, Chopra T, Pogue JM, et al. Outbreak of colistin-resistant, carbapenem-resistant *Klebsiella pneumoniae* in metropolitan Detroit, Michigan. Antimicrob Agents Chemother **2011**; 55:593–9.
37. Epson EE, Pisney LM, Wendt JM, et al. Carbapenem-resistant *Klebsiella pneumoniae* producing New Delhi metallo-beta-lactamase at an acute care hospital, Colorado, 2012. Infect Control Hosp Epidemiol **2014**; 35:390–7.
38. Snitkin ES, Zelazny AM, Thomas PJ, et al. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. Sci Transl Med **2012**; 4:148ra16.
39. Siegel JD, Rhinehart E, Jackson M; Healthcare Infection Control Practices Advisory Committee. 2007 guideline for isolation precautions: preventing transmission of infectious agents in healthcare settings. Available at: https://www.cdc.gov/infectioncontrol/guidelines/isolation/index.html. Accessed 20 February 2018.
40. Pak TR, Kasarskis A. How next-generation sequencing and multiscale data analysis will transform infectious disease management. Clin Infect Dis **2015**; 61:1695–702.

Conclusion

In conclusion, I showed that MDRO colonization is a highly dynamic and complex process after liver transplant. The use of active surveillance and WGS led to several unexpected findings in this chapter, and these findings formed the basis for further studies that are addressed in more detail in subsequent chapters.

The first finding was the possible healthcare-associated transmission of *mcr-1*, which had gone undetected clinically as all isolates were colonising isolates collected as part of this study. The second was the unexpected emergence of novel MDR *K. pneumoniae* strains, in particular *K. pneumoniae* ST307. Shortly prior to the publication of this manuscript, another group had similarly noted its emergence [281] and hence this formed the focus of further work. Finally, among CRE isolates included in this study, I noted numerous PR isolates. This led me to hypothesize that there was an underappreciated burden of PR in CRE isolates from our institution and paved the way for the work on PR in Chapters 4 to 6.

Chapter 3:

Genomic epidemiology of emerging forms of Gram negative resistance: mcr-1 and *Klebsiella pneumoniae* ST307

Work from this chapter was published in:

1) Macesic N, Khan S, Giddins MJ, et al. Escherichia coli Harboring mcr-1 in a Cluster of Liver Transplant Recipients: Detection through Active Surveillance and Whole-Genome Sequencing. *Antimicrobial Agents and Chemotherapy* 2019; 63:161. [2]

2) Rojas R, **Macesic N**, Tolari G, Guzman A, Uhlemann A-C. Multidrug-Resistant *Klebsiella pneumoniae* ST307 in Traveler Returning from Puerto Rico to Dominican Republic. *Emerg. Infect. Dis.* 2019; 25:1583–1585.

The following contributions were made by specific co-authors to the first published work. No co-authors were Monash University students.

Nature of contribution	Co-authors
Concept	Uhlemann AC
Sample collection	Khan S, Giddins MJ, Green DA, Whittier S, Freedberg DE
PCR analyses	Khan S, Giddins MJ
DNA extraction and sequencing	Khan S, Giddins MJ
Sequence analysis	Annavajhala M
Manuscript writing	Uhlemann AC

Introduction

The work in the previous chapter led to several observations regarding emerging forms of Gram negative resistance, both in terms of mechanisms and novel clonal backgrounds, which were deemed of sufficient importance to form publications of their own.

In the first of these, I address the unexpected detection of *mcr-1,* a plasmid-mediated form of PR, in stool isolates from liver transplant recipients enrolled in the prospective study in Chapter 2. Fortunately, all isolates were colonizing isolates and the patients did not experience infections. However, our group had previously reported an index case of bloodstream and biliary fluid infection with *mcr-1*-carrying *E. coli* ST117 in 2015 [300]. I therefore conducted an epidemiological analysis in order to discern possible transmission events. In retrospect, this formed the earliest known healthcare-associated transmission of *mcr-1* in the US.

In the second, I focus on *K. pneumoniae* ST307, a rapidly emerging MDR clone [281, 301]. We noted several *K. pneumoniae* ST307 isolates during the study detailed in Chapter 2 and subsequently have continued to note this as an endemic clone at Columbia University Irving Medical Center. We collaborate with a group in the Dominican Republic, where prior to this report there had been no documented cases of CRE. Upon detection of a CRKP by our collaborators, we conducted WGS of the isolate and noted it was *K. pneumoniae* ST307. I therefore used isolates from Columbia University Irving Medical Center and publicly available ST307 isolates to provide context for this case and

understand the genomic epidemiology of this MDR clone, with a particular focus on the Western Hemisphere.

AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy®

EPIDEMIOLOGY AND SURVEILLANCE

Escherichia coli Harboring *mcr-1* in a Cluster of Liver Transplant Recipients: Detection through Active Surveillance and Whole-Genome Sequencing

Nenad Macesic,[a,b] Sabrina Khan,[a] Marla J. Giddins,[a] Daniel E. Freedberg,[c] Susan Whittier,[d] Daniel A. Green,[d] E. Yoko Furuya,[a] Elizabeth C. Verna,[c] Medini K. Annavajhala,[a] Angela Gomez-Simmonds,[a] Anne-Catrin Uhlemann[a]

[a]Division of Infectious Diseases, Columbia University Irving Medical Center, New York, New York, USA
[b]Central Clinical School, Monash University, Melbourne, Victoria, Australia
[c]Division of Digestive and Liver Diseases, Columbia University Irving Medical Center, New York, New York, USA
[d]Clinical Microbiology Laboratory Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, New York, USA

ABSTRACT *mcr-1*, a plasmid-associated gene for colistin resistance, was first described in China in 2015, but its spread in the United States is unknown. We report detection of *mcr-1*-carrying *Escherichia coli* ST117 in a cluster of three liver transplant recipients.

KEYWORDS antimicrobial resistance, colistin, *mcr-1*, polymyxin B, whole-genome sequencing

Polymyxins are treatments of "last resort" for multidrug-resistant Gram-negative bacterial (MDR-GNB) infections. *mcr-1*, a plasmid-associated gene for polymyxin resistance, was first described in China in 2015, with widespread dissemination noted (1, 2), but only 53 U.S. cases have been reported to date (https://www.cdc.gov/drugresistance/biggest-threats/tracking/mcr.html). We report a cluster of *Escherichia coli* sequence type (ST) 117 harboring *mcr-1* in three liver transplant recipients after an initial case of infection in 2015 (3).

The index case (patient 1) had bloodstream and biliary fluid infection with *mcr-1*-carrying *E. coli* (3). The remaining three patients carried *mcr-1*-positive *E. coli* in stool samples collected during the prospective liver transplant (PLT) study but did not experience infections (4). We conducted an epidemiological analysis in order to discern possible transmission events (Fig. 1).

Patient 1 underwent elective biliary tree excision and Roux-en-Y hepaticojejunostomy in May 2015 for recurrent episodes of cholangitis on a background of sickle-cell anemia and cholecystectomy. Intraoperative biliary cultures grew *mcr-1*-positive *E. coli*. The patient became febrile the day after surgery, and blood cultures also grew *mcr-1*-positive *E. coli*. The patient was successfully treated with meropenem. Patients 2 to 4 had undergone liver transplant: two patients in November 2014 and one in January 2015. Patients 2 to 4 regularly provided fecal samples as part of the PLT study (Fig. 1) and were noted to have new onset MDR *E. coli* colonization at days 241, 275, and 322 posttransplant, respectively, prompting WGS and retrospective detection of *mcr-1*.

Patient 2 underwent same-day endoscopic retrograde cholangiopancreatography (ERCP), as did patient 1, in March 2015, and the first available subsequent stool sample was positive for *mcr-1* 6 months later (Fig. 1). Patient 3 underwent gastroduodenoscopy the same day patient 2 underwent ERCP in September 2015. Of note, patient 3 had a stool sample negative for *mcr-1* 6 days prior to the endoscopy, followed by detection of *mcr-1* 7 days following the endoscopy, suggesting acquisition around the time of endoscopy. Patient 4 had no obvious epidemiologic links to explain *mcr-1* acquisition.

Citation Macesic N, Khan S, Giddins MJ, Freedberg DE, Whittier S, Green DA, Furuya EY, Verna EC, Annavajhala MK, Gomez-Simmonds A, Uhlemann A-C. 2019. *Escherichia coli* harboring *mcr-1* in a cluster of liver transplant recipients: detection through active surveillance and whole-genome sequencing. Antimicrob Agents Chemother 63:e02680-18. https://doi.org/10.1128/AAC.02680-18.

Address correspondence to Nenad Macesic, nm2891@cumc.columbia.edu, or Anne-Catrin Uhlemann, au2110@columbia.edu.

Received 27 December 2018
Returned for modification 27 January 2019
Accepted 19 March 2019
Accepted manuscript posted online 8 April 2019
Published 24 May 2019

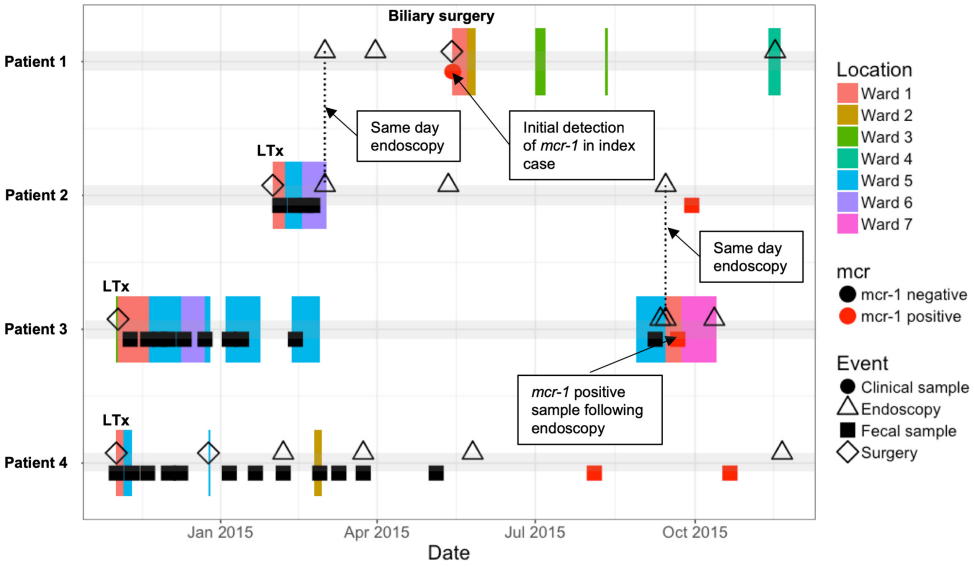


FIG 1 Epidemiological analysis of cluster of patients colonized and infected with health care-associated *mcr-1*-carrying *E. coli*. Patient 1 was the index case and had infection with *mcr-1*-carrying *E. coli* in biliary fluid and blood isolates following elective biliary surgery. Subsequently, closely related *mcr-1*-carrying *E. coli* was detected in fecal samples of patients 2 to 4, who had all undergone liver transplantation. Endoscopy performed on the same day appeared to be a common epidemiological link for patients 1, 2, and 4 (indicated with dotted lines). No such link was found for patient 3. LTx, liver transplant.

No overlap was noted between this patient and other cluster patients with respect to endoscopy procedures, outpatient visits, or radiology investigations. The patients all lived in different zip codes. During 2.5 years of follow-up, patients 2 to 4 did not develop any *mcr-1*-associated infections.

Multilocus sequence typing indicated that all isolates were ST117. Patients 1 and 4 had two isolates each available for analysis, while patients 2 and 3 had single isolates. Cultured isolates underwent identification and susceptibility testing with Vitek-2 (bioMérieux). We performed MIC determination for colistin and polymyxin B with broth microdilution according to established guidelines (5). Tigecycline susceptibility testing was performed with Etest (bioMérieux). Details of whole-genome sequencing (WGS) are provided in the supplemental material.

The following MIC ranges were noted on a Vitek-2 instrument (bioMérieux, Durham, NC): ceftriaxone, 2 to 8 μg/ml; cefepime, $<$1 μg/ml; aztreonam, $>$64 μg/ml; ertapenem, $<$0.5 μg/ml; meropenem, $<$0.25 μg/ml; amikacin, $<$2 μg/ml; tobramycin, $<$1 μg/ml; gentamicin, $<$1 μg/ml; levofloxacin, $>$8 μg/ml; and tigecycline, 0.38 to 0.75 μg/ml. In addition to *mcr-1*, isolates harbored resistance genes to beta-lactams (bla_{SHV-12} and bla_{TEM-1D}), aminoglycosides [*aph(3′)-IIa*, *aph(6)-Ic*, *aadA2*, *aph(3′)-Ia*, *strA*, and *strB*], sulfonamides (*sul1* and *sul2*), trimethoprim (*dfrA12*), chloramphenicol (*catA1*), macrolides (*mphA*), and tetracycline [*tet(A)*]. Of note, bla_{SHV-12} codes for an extended-spectrum beta-lactamase (6). Full isolate antimicrobial susceptibilities, resistance genes, and plasmid replicons are shown in Tables S2 to S4 in the supplemental material. Within-patient pairwise single nucleotide variant (SNV) distances of patient 1 and patient 4 isolates were 15 and 3 SNV (Table S1). Pairwise SNV distances of isolates from different patients ranged from 7 to 28 SNVs. Phylogenetic relationships are shown in Fig. 2. The *mcr-1* gene was carried on an IncX4 plasmid in all isolates. The plasmid replicon content was similar for the isolates.

We then conducted a retrospective molecular survey using *mcr-1* PCR to evaluate for possible spread of *mcr-1* (1). First, we tested all 197 *E. coli* isolates resistant to third-generation cephalosporins cultured from PLT study patients from 2015 to 2017.

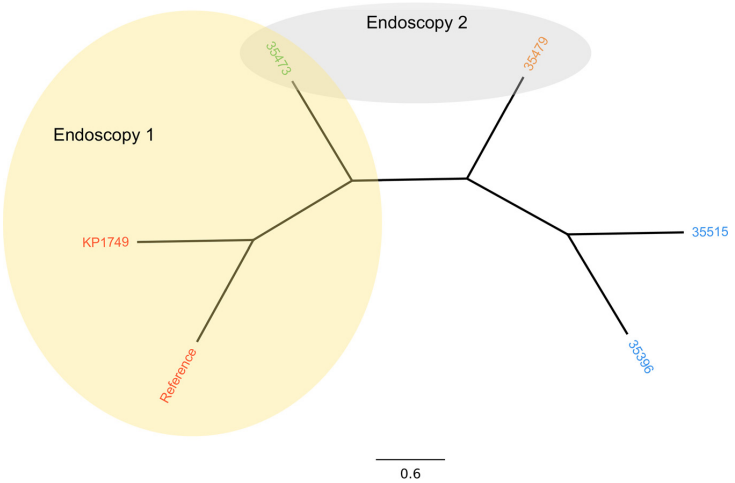


FIG 2 Phylogenetic tree of *mcr-1*-carrying *E. coli* isolates. The genome of isolate NR2148 (GenBank accession no. CP019903.1) was the reference. Different patients are represented by colored nodes (red, patient 1; green, patient 2; orange, patient 3; blue, patient 4). The epidemiological associations for the shared endoscopy location are indicated by the colored fields.

We then tested 187 discarded surveillance rectal swabs of consecutive patients admitted to medical and surgical intensive care units (ICUs) in 2017. Finally, we tested 123 stored clinical *E. coli* isolates resistant to third-generation cephalosporins collected from 2011 to 2016. No further cases of *mcr-1*-carrying isolates were detected ($n = 507$).

The detection of this cluster demonstrates the potential for silent dissemination of *mcr-1* in a U.S. hospital setting through asymptomatic colonization and suggests a possible role for same-day endoscopy-related transmission, independent of using the same endoscope. This represents the earliest documented health care-associated cluster of *mcr-1* in the United States and predates a recent report from 2017 (7). Both reports implicate endoscopy as a potential route for transmission. The current report also highlights the difficulty of instituting surveillance measures for *mcr-1*-harboring isolates when endemicity is low. While infective episodes have the greatest impact on patient outcomes, detection of silent colonization may play an important role in stemming the spread of *mcr-1*.

Our epidemiological analysis showed the possible role of endoscopy, with three patients having same-day endoscopies. However, the same endoscopes were not used during same-day procedures. Other mechanisms, such as a common environmental source, may have contributed. A previous outbreak of KPC-2 *Klebsiella pneumoniae* implicated a positioning pillow (8). Although endoscopy location was the clearest association, there were multiple other possible contributors since these patients had frequent radiological investigations, were cared for by the same caregivers, and had admissions to the same hospital wards. This may explain the acquisition in patient 4. Unfortunately, *mcr-1* was only detected several years after sample collection, limiting our ability to conduct a real-time investigation. Our isolate collections also focused only on *E. coli* isolates resistant to third-generation cephalosporins. While *mcr-1* colonization was found in these patients due to participation in a prospective study, colonization in other patients may have gone undetected. Nevertheless, our molecular survey of over 500 samples did not detect additional cases, either before or since (through 2017), demonstrating that *mcr-1* has not yet become widespread. Furthermore, colonization in all three liver transplant recipients was not associated with clinical infection.

Genomic analysis confirmed that isolates belonged to ST117, were related, and carried *mcr-1* on near-identical IncX4 plasmids. The distance of 15 SNVs between

isolates from the index patient was high and suggests possible within-host evolution through long-term colonization. This relatively high SNV distance has direct implications for defining related clusters and spread. Conversely, the reference isolate was less distant from isolate 35479 from patient 2 (7 SNVs), suggesting transmission.

Our findings highlight the need for ongoing surveillance of *mcr-1* and other forms of transferrable resistance to polymyxins in the United States. Our analysis suggested endoscopy location as an epidemiologic link but did not implicate endoscopes directly. Although we documented acquisition of *mcr-1* colonization in several liver transplant recipients through active surveillance of fecal MDR-GNB carriage and WGS, this is a highly resource-intensive approach and is not applicable on a larger scale. A more viable alternative may be to implement PCR screening of discarded specimens (e.g., stool samples or rectal swabs) (9). In order to stop further spread of emerging forms of polymyxin resistance, future surveillance approaches need to recognize that clinical isolates may only represent the "tip of the iceberg" formed by the burden of asymptomatic colonization.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AAC.02680-18.

SUPPLEMENTAL FILE 1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

This study was supported by the National Institute of Allergy and Infectious Diseases at the National Institutes of Health (grants AI116939 to A.-C.U. and S1-AI116939 to A.-C.U. and A.G.-S.) and by an Irving Scholarship from Columbia University (A.-C.U.).

All authors report no conflicts of interest relevant to the manuscript.

REFERENCES

1. Liu Y-Y, Wang Y, Walsh TR, Yi L-X, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu L-F, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu J-H, Shen J. 2016. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis 16:161–168. https://doi.org/10.1016/S1473-3099(15)00424-7.
2. Zhong LL, Phan HTT, Shen C, Doris-Vihta K, Sheppard AE, Huang X, Zeng KJ, Li HY, Zhang XF, Patil S, Crook DW, Walker AS, Xing Y, Lin JL, Feng LQ, Doi Y, Xia Y, Stoesser N, Tian GB. 2017. High rates of human fecal carriage of *mcr-1*-positive multidrug resistant *Enterobacteriaceae* isolates emerge in China in association with successful plasmid families. Clin Infect Dis 66:676–685. https://doi.org/10.1093/cid/cix885.
3. Macesic N, Green D, Wang Z, Sullivan SB, Shim K, Park S, Whittier S, Furuya EY, Gomez-Simmonds A, Uhlemann AC. 2017. Detection of *mcr-1*-carrying *Escherichia coli* causing bloodstream infection in a New York City hospital: avian origins, human concerns? Open Forum Infect Dis 4:ofx115. https://doi.org/10.1093/ofid/ofx115.
4. Macesic N, Gomez-Simmonds A, Sullivan SB, Giddins MJ, Ferguson SA, Korakavi G, Leeds D, Park S, Shim K, Sowash MG, Hofbauer M, Finkel R, Hu Y, West J, Toussaint NC, Greendyke WG, Miko BA, Pereira MR, Whittier S, Verna EC, Uhlemann AC. 2018. Genomic surveillance reveals diversity of multidrug-resistant organism colonization and infection: a prospective cohort study in liver transplant recipients. Clin Infect Dis 67:905–912. https://doi.org/10.1093/cid/ciy199.
5. CLSI. 2015. Performance standards for antimicrobial susceptibility testing; 25th informational supplement. Document M100-S25. Clinical and Laboratory Standards Institute, Wayne, PA.
6. Nüesch-Inderbinen MT, Kayser FH, Hächler H. 1997. Survey and molecular genetics of SHV beta-lactamases in *Enterobacteriaceae* in Switzerland: two novel enzymes, SHV-11 and SHV-12. Antimicrob Agents Chemother 41:943–949. https://doi.org/10.1128/AAC.41.5.943.
7. Shenoy ES, Pierce VM, Walters MS, Moulton-Meissner H, Lawsin A, Lonsway D, Shugart A, McAllister G, Halpin AL, Zambrano-Gonzalez A, Ryan EE, Suslak D, DeJesus A, Barton K, Madoff LC, McHale E, DeMaria A, Hooper DC. 2018. Transmission of mobile colistin resistance (*mcr-1*) by duodenoscope. Clin Infect Dis https://doi.org/10.1093/cid/ciy683.
8. Kaiser T, Finstermeier K, Hantzsch M, Faucheux S, Kaase M, Eckmanns T, Bercker S, Kaisers UX, Lippmann N, Rodloff AC, Thiery J, Lubbert C. 2018. Stalking a lethal superbug by whole-genome sequencing and phylogenetics: influence on unraveling a major hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae*. Am J Infect Control 46:54–59. https://doi.org/10.1016/j.ajic.2017.07.022.
9. Chan WS, Au CH, Ho DN, Chan TL, Ma ES, Tang BS. 2018. Prospective study on human fecal carriage of *Enterobacteriaceae* possessing *mcr-1* and *mcr-2* genes in a regional hospital in Hong Kong. BMC Infect Dis 18:81. https://doi.org/10.1186/s12879-018-2987-y.

Supplementary Methods

Retrospective molecular survey

In order to assess possible undetected spread of *mcr-1*, we conducted a retrospective molecular survey. We tested isolates from three sources for the presence of *mcr-1* by polymerase chain reaction (1). The first source was the Prospective Liver Transplant (PLT) study with 197 *Escherichia coli* isolates resistant to third-generation cephalosporins collected from 2015 – 2017 (2). In this study, fecal samples were collected at regular intervals from pre-liver transplant to one year post-transplant and screened for third-generation cephalosporin-resistant *Enterobacteriaceae* with selective chromogenic agar (DRG International). The second source were 187 discarded surveillance rectal swabs of consecutive patients admitted to medical and surgical ICUs in 2017, similarly cultured and screened for third-generation cephalosporin-resistant *Enterobacteriaceae* with selective chromogenic agar (DRG International). Finally, we accessed an institutional collection of clinical multi-drug resistant *Enterobacteriaceae*. *E. coli* isolates resistant to third-generation cephalosporins are routinely stored and we tested all 123 *E. coli* isolates not tested during a previous study (3).

Whole genome sequencing

Six *mcr-1* positive isolates were available for analysis: two infecting isolates from Patient 1 and four colonizing isolates from PLT study patients.

We previously published whole genome sequencing (WGS) for the *mcr-1*-positive bloodstream isolate from Patient 1 (NCBI Sequence Read Archive accession no. SRR5224401) (3). As part of ongoing surveillance efforts, the Centers for Disease Control and Prevention also performed long-read sequencing of Patient 1's biliary fluid isolate (GenBank accession no. CP019903.1) (4). We then performed WGS for the remaining four *mcr-1* positive isolates from three patients.

DNA was extracted on an EpMotion liquid handling workstation using the Qiagen UltraClean Microbial DNA Isolation Kit (Hilden, Germany). Libraries were prepared using the Nextera XT DNA Library Prep Kit and sequenced on the Illumina MiSeq (Illumina, San Diego, CA).

SRST2 analysis was performed for multilocus sequence typing (MLST), characterization of resistance determinants using the ArgANNOT database, and characterization of plasmid replicons using the PlasmidFinder database (5-7). Detection of *mcr-1* by SRST2 was confirmed by polymerase chain reaction (PCR) (8). For comparative sequence analyses, Illumina reads were mapped against the long-read assembly of the biliary fluid isolate from Patient 1 (GenBank accession no. CP019903.1). Variant calling was performed using Snippy 3.1 after exclusion of mobile genetic elements with PHASTER and IslandViewer 4 (9-11). Plasmid sequences were *de novo* assembled with SPAdes using reads that did not map to the reference bacterial chromosome

47 (12). The resulting contigs were then mapped to plasmids from the reference
48 assembly (GenBank accession nos. CP019904 – CP019910) for comparison. A
49 maximum-likelihood phylogenetic tree was constructed using the RaXML plug-in
50 in Geneious 10.1.3 (13, 14).

S1– Details of patients and isolates implicated in cluster of *mcr-1 E. coli* ST117

Patient	**1**	**2**	**3**	**4**
Demographic details	50 year-old male	50 year-old female	64 year-old female	60 year-old male
Clinical details	Biliary surgery	Liver transplant recipient	Liver transplant recipient	Liver transplant recipient
Available isolates and isolate source	Reference (biliary fluid)	35479 (stool)	35473 (stool)	35396 (stool)
	KP1749 (blood)			35515 (stool)
Putative epidemiological link	Index case	Same day endoscopy	Same day endoscopy	None identified

S2 – Susceptibility testing of *mcr-1* positive isolates

Isolate	Patient	PB MIC	COL MIC	PIP/ TAZ MIC	CRO MIC	CPM MIC	AZT MIC	ETP MIC	MEM MIC	AK MIC	GENT MIC	TOB MIC	LVX MIC	TMP-SMX MIC	TGC MIC
NR2148	1	4	4	<4	2	<1	>64	<0.5	<0.25	<2	<1	<1	>8	>320	0.38
KP1749	1	4	4	<4	2	<1	>64	<0.5	<0.25	<2	<1	<1	>8	>320	0.38
35479	2	4	4	<4	8	<1	>64	<0.5	<0.25	<2	<1	<1	>8	<20	0.38
35473	3	4	4	16	8	<1	>64	<0.5	<0.25	<2	<1	<1	>8	>320	0.75
35396	4	4	4	8	2	<1	>64	<0.5	<0.25	<2	<1	<1	>8	>320	0.75
35515	4	4	4	8	8	<1	>64	<0.5	<0.25	<2	<1	<1	>8	>320	0.75

Abbreviations: MIC – minimum inhibitory concentration (expressed in mcg/ml); PB – polymyxin B (broth microdilution); COL – colistin (broth microdilution); PIP/TAZ – piperacillin / tazobactam (VITEK-2); CRO – ceftriaxone (VITEK-2); CPM – cefepime (VITEK-2); AZT – aztreonam (VITEK-2); ETP – ertapenem (VITEK-2); MEM – meropenem (VITEK-2); AK – amikacin (VITEK-2); GENT – gentamicin (VITEK-2); TOB – tobramycin (VITEK-2); LVX – levofloxacin (VITEK-2); TMP-SMX – trimethoprim-sulfamethoxazole (VITEK-2); TGC – tigecycline (Etest).

S3 - Resistance determinants of *mcr-1* positive isolates

Isolate	AMPH	AadA	AmpC1	AmpC2	Aph3"Ia	AphA2	CatA1	DfrA	Mcr1	MphA	PBP	SHV-OKP-LEN
35396	AMPH_615	-	AmpC1_1670	AmpC2_346	-	-	CatA1_1470	-	Mcr1_1237	-	PBP_Ecoli_836	-
35473	AMPH_615	AadA2_1605	AmpC1_1670	AmpC2_346	Aph3-Ia_1218	AphA2_1492	CatA1_1470	DfrA12_1089	Mcr1_1237	MphA_1663	PBP_Ecoli_836	SHV-12_1288
35479	AMPH_615	AadA2_1605	AmpC1_1670	AmpC2_346	Aph3-Ia_1218	AphA2_1492	CatA1_1470	DfrA12_1089	Mcr1_1237	MphA_1663	PBP_Ecoli_836	SHV-12_1288
35515	AMPH_615	AadA2_1605	AmpC1_1670	AmpC2_346	Aph3-Ia_1218	AphA2_1492	CatA1_1470	DfrA12_1089	Mcr1_1237	MphA_1663	PBP_Ecoli_836	SHV-12_1288
KP1749	AMPH_615	AadA2_1605	AmpC1_1670	AmpC2_346	Aph3-Ia_1218	AphA2_1492	CatA1_1470	DfrA12_1089	Mcr1_1237	MphA_1663	PBP_Ecoli_836	SHV-12_1288
NR2148	AMPH_615	AadA2_1605	AmpC1_1670	AmpC2_346	Aph3-Ia_1218	AphA2_1492	CatA1_1470	DfrA12_1089	Mcr1_1237	MphA_1663	PBP_Ecoli_836	SHV-12_1288

Isolate	StrA	StrB	SulII	SulI	TEM-1D	TetA	TetR
35396	StrA_1501	StrB_1614	SulII_1219	-	TEM-104_955	TetA_1545	TetR_1473
35473	StrA_1501	StrB_1614	SulII_1219	SulI_1616	TEM-76_929	TetA_1545	TetR_1473
35479	StrA_1501	StrB_1614	SulII_1219	SulI_1616	TEM-198_1035	TetA_1545	TetR_1473
35515	StrA_1501	StrB_1614	SulII_1219	SulI_1616	TEM-70_925	TetA_1545	TetR_1473
KP1749	StrA_1501	StrB_1614	SulII_1219	SulI_1616	TEM-198_1035	TetA_1545	TetR_1473
NR2148	StrA_1501	StrB_1614	SulII_1219	SulI_1616	TEM-33_908	TetA_1545	TetR_1473

S4 - Plasmid replicons of *mcr-1* positive isolates

Isolate	Col	ColMG828	ColRNAI_1	ColpVC	FIBAP001918_1	FIC	FIIS	FIIY_ps	FII_1
35396	_Col-plasmid_2_AY929248_ColK_23?	_ColMG828_1_NC_008486	-	_ColpVC_1_JX133088	-	_FIC_1_AP001918	-	-	FII_1_AY458016
35473	_Col-plasmid_1_J01566__22	_ColMG828_1_NC_008486	ColRNAI_1_DQ298019	_ColpVC_1_JX133088	FIBAP001918_1_AP001918	_FIC_1_AP001918	FIIp14_1_p14_JQ418538_295	-	FII_1_AY458016
35479	_Col-plasmid_1_J01566__22	_ColMG828_1_NC_008486	-	_ColpVC_1_JX133088	-	_FIC_1_AP001918	-	-	FII_1_AY458016
35515	_Col-plasmid_1_J01566__22	_ColMG828_1_NC_008486	-	_ColpVC_1_JX133088	-	_FIC_1_AP001918	FIIYp_1_Yersenia_CP000670_294	-	FII_1_AY458016
KP1749	_Col-plasmid_1_J01566__22	_ColMG828_1_NC_008486	ColFST258_1_JN247853_8	_ColpVC_1_JX133088	-	_FIC_1_AP001918	-	-	FII_1_pKP91_(P000966_17
NR2148	_Col-plasmid_1_J01566__22	_ColMG828_1_NC_008486	-	_ColpVC_1_JX133088	-	_FIC_1_AP001918	FIIYp_1_Yersenia_CP000670_294	FIIY_1_ps_CP001049_288	FII_1_AY458016

Isolate	I1_Alpha	N	P_alpha	Q1	X4	X4TaxC
35396	I1_1_Alpha_AP005147_238	_N_1_AY046276	P_1_alpha_L27758_140	_Q1_1_HE654726	_X4_1_CP002895	_X4TaxC_1_CP002895_180
35473	I1_1_Alpha_AP005147_238	_N_1_AY046276	P_1_alpha_L27758_140	_Q1_1_HE654726	_X4_1_CP002895	_X4TaxC_1_CP002895_180
35479	I1_1_Alpha_AP005147_238	_N_1_AY046276	P_1_alpha_L27758_140	_Q1_1_HE654726	_X4_1_CP002895	_X4TaxC_1_CP002895_180
35515	I1_1_Alpha_AP005147_238	_N_1_AY046276	P_1_alpha_L27758_140	_Q1_1_HE654726	_X4_1_CP002895	_X4TaxC_1_CP002895_180
KP1749	I1_1_Alpha_AP005147_238	_N_1_AY046276	P_1_alpha_L27758_140	_Q1_1_HE654726	_X4_1_CP002895	_X4TaxC_1_CP002895_180
NR2148	I1_1_Alpha_AP005147_238	_N_1_AY046276	P_1_alpha_L27758_140	_Q1_1_HE654726	_X4_1_CP002895	_X4TaxC_1_CP002895_180

References

1. Liu YY, Wang Y, Walsh TR, Yi LX, Zhang R, Spencer J, Doi Y, Tian G, Dong B, Huang X, Yu LF, Gu D, Ren H, Chen X, Lv L, He D, Zhou H, Liang Z, Liu JH, Shen J. 2016. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis 16:161-8.
2. Macesic N, Gomez-Simmonds A, Sullivan SB, Giddins MJ, Ferguson SA, Korakavi G, Leeds D, Park S, Shim K, Sowash MG, Hofbauer M, Finkel R, Hu Y, West J, Toussaint NC, Greendyke WG, Miko BA, Pereira MR, Whittier S, Verna EC, Uhlemann AC. 2018. Genomic Surveillance Reveals Diversity of Multidrug-Resistant Organism Colonization and Infection: A Prospective Cohort Study in Liver Transplant Recipients. Clin Infect Dis doi:10.1093/cid/ciy199.
3. Macesic N, Green D, Wang Z, Sullivan SB, Shim K, Park S, Whittier S, Furuya EY, Gomez-Simmonds A, Uhlemann AC. 2017. Detection of mcr-1-Carrying Escherichia coli Causing Bloodstream Infection in a New York City Hospital: Avian Origins, Human Concerns? Open Forum Infect Dis 4:ofx115.
4. Centres for Disease Control and Prevention. 2018. Tracking the *mcr* gene. https://www.cdc.gov/drugresistance/tracking-mcr1.html. Accessed December 26, 2018.
5. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med 6:90.
6. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain JM. 2014. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212-20.
7. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58:3895-903.
8. Feb 2016. 2. 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, 16.161-168. http://linkinghub.elsevier.com/retrieve/pii/S1473309915004247.
9. Seemann T. 2017. Snippy 3.1. https://github.com/tseemann/snippy. Accessed August 15.
10. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16-21.
11. Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing G, Lau BY, Hoad G, Winsor GL, Brinkman FS. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res doi:10.1093/nar/gkx343.
12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-77.
13. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-9.
14. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-3.

9. Kojima Y, Furubayashi K, Kawahata T, Mori H, Komano J. Circulation of distinct *Treponema pallidum* strains in individuals with heterosexual orientation and men who have sex with men. J Clin Microbiol. 2019;57:e01148-18. https://doi.org/10.1128/JCM.01148-18
10. Staudová B, Strouhal M, Zobaníková M, Cejková D, Fulton LL, Chen L, et al. Whole genome sequence of the *Treponema pallidum* subsp. *endemicum* strain Bosnia A: the genome is related to yaws treponemes but contains few loci similar to syphilis treponemes. PLoS Negl Trop Dis. 2014;8:e3261. https://doi.org/10.1371/journal.pntd.0003261

Address for correspondence: Takuya Kawahata, Osaka Institute of Public Health, 1-3-69 Nakamichi, Higashinari-ku, Osaka 537-0025, Japan; email: kawahata@iph.osaka.jp

Multidrug-Resistant *Klebsiella pneumoniae* ST307 in Traveler Returning from Puerto Rico to Dominican Republic

Rita Rojas, Nenad Macesic, Gilda Tolari, Anel Guzman, Anne-Catrin Uhlemann

Author affiliations: Hospital General Plaza de la Salud, Santo Domingo, Dominican Republic (R. Rojas, G. Tolari, A. Guzman); Monash University, Melbourne, Victoria, Australia (N. Macesic); Columbia University Medical Center, New York, New York, USA (N. Macesic, A.-C. Uhlemann)

DOI: https://doi.org/10.3201/eid2508.171730

We report *bla*$_{KPC-2}$–harboring carbapenem-resistant *Klebsiella pneumoniae* in an emerging sequence type 307 lineage in a traveler returning from Puerto Rico to the Dominican Republic. Phylogenetic analyses indicate regional dissemination of this highly drug-resistant clone across the Americas, underscoring the need for adequate surveillance and infection control efforts to prevent further spread.

Carbapenemase-resistant *Enterobacteriaceae* (CRE), in particular carbapenem-resistant *Klebsiella pneumoniae* (CRKp), represent a serious threat to public health (*1*). CRKp infections have been associated with high mortality rates, up to 50% in some studies (*2*). In resource-limited regions, such as the Dominican Republic, multiple challenges hinder efforts to contain CRE infections, including lack of novel antimicrobial drugs, inability to monitor drug levels of potentially toxic treatment regimens, and absence of molecular tools to investigate outbreaks and potential spread.

In fall 2015, a 66-year-old woman with diabetes mellitus, hepatitis C virus infection, and end-stage renal disease on hemodialysis was admitted to a hospital in the Dominican Republic for fever, anorexia, chills, and myalgia. On day 3, her blood culture tested positive for *K. pneumoniae.* She had been admitted to a hospital in Puerto Rico a few months before and had been treated for a multidrug-resistant bacterial infection.

The *K. pneumoniae* isolate from the patient was nonsusceptible to all tested antimicrobial drugs except polymyxins (Appendix Table 1, http://wwwnc.cdc.gov/EID/article/25/8/17-1730-App1.pdf). We began combination therapy with a loading dose of colistin, then 100 mg postdialysis, plus ertapenem (150 mg postdialysis) and fosfomycin (2 g 3×/d). We implemented infection control measures by placing the patient in a single room and using gloves, gowns, masks, and a dedicated stethoscope. Despite initial improvement, the patient died on day 25 after admission.

Whole-genome sequencing revealed that the patient isolate, NR6025, was of the emerging sequence type 307 (ST307) (*3*) and closely related (≤185 SNPs) to several international ST307 isolates of similar phenotype (Figure). Of note, this isolate was most closely related, within 36 SNPs, to an isolate recovered from a patient in New York, NY, USA, who also had been hospitalized in Puerto Rico in 2016 (*4*). This finding raises the possibility that both patients acquired CRE in Puerto Rico and their infections subsequently developed in their home countries.

In silico resistance gene detection demonstrated that *bla*$_{KPC-2}$, on Tn*4401e*, was likely the mechanism of carbapenem resistance for this isolate. Moreover, the meropenem MIC was >32 μg/mL, consistent with high carbapenem MICs observed in the ST307 Tn*4401e* isolates (*4*) from New York, suggesting association with a strong promoter. In addition, the isolate harbored a large repertoire of acquired-resistance genes, including additional β-lactamase genes CTX-M-15, SHV-100, OXA-1, and TEM-1D (Appendix Table 1). The isolate contained IncFIBK, ColRNA1, and IncA/C2 plasmid replicons; IncA/C plasmid encodes for *bla*$_{KPC-2}$, *bla*$_{TEM}$, *sulI*, *aadB*, *aac6*, and *qacE*, which has been implicated in chlorhexidine resistance.

A case of CRKp was described from Medellin, Colombia, in 2005, and subsequent CRKp infections have been reported in Mexico, in South America in Brazil, Argentina, and Venezuela, and in the Caribbean in Cuba, Puerto Rico, and Trinidad and Tobago (*5*–*7*). In many of these studies, CRKp isolates were mainly accounted for

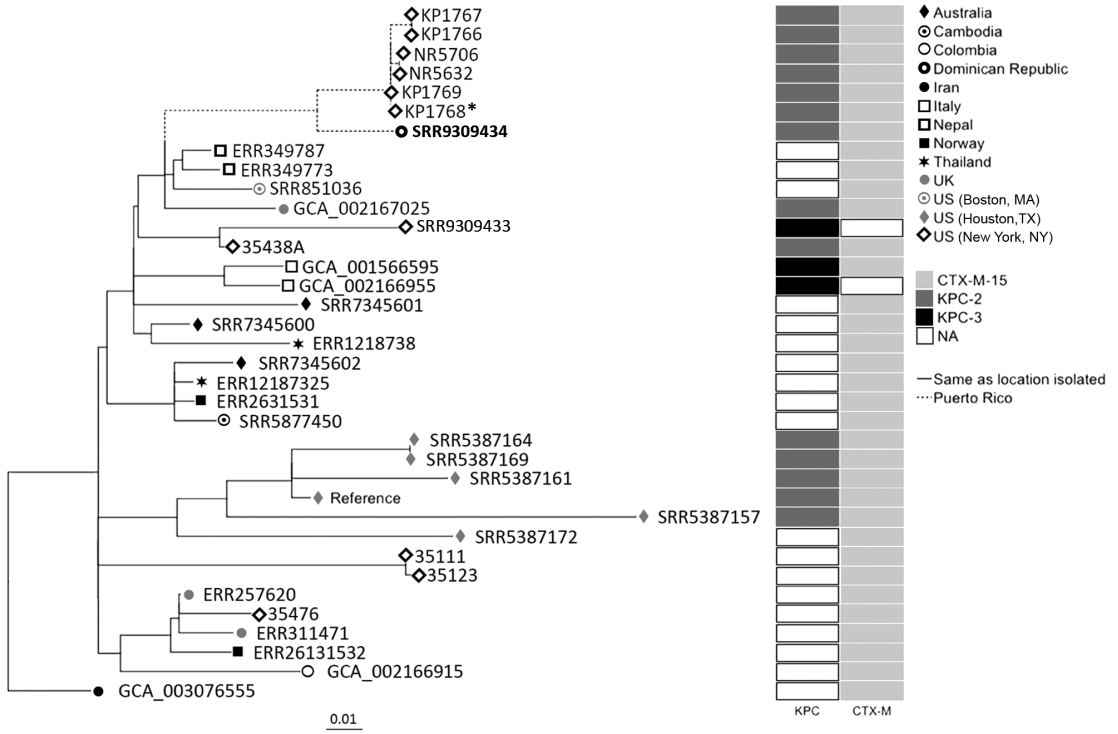


Figure. Maximum-likelihood phylogenetic tree of geographically diverse *Klebsiella pneumoniae* sequence type 307 isolates based on 860 concatenated single-nucleotide polymorphisms, extracted from an alignment length of 5,248,133 bp. Bold indicates isolate from a traveler from Puerto Rico to the Dominican Republic (this study). Asterisk (*) indicates an isolate recovered from a patient admitted to a hospital in Puerto Rico during the same year as the case-patient in this study (*4*). *bla* gene types (KPC, CTX-M) are indicated. Scale bar indicates nucleotide substitutions per site. NA, not applicable.

by ST258 and ST512. The SENTRY Antimicrobial Surveillance Program showed that bla_{KPC-2}-harboring CRE accounted for most CRE infections in Latin America and that the incidence rate has been rising sharply (*8*). These organisms also are prevalent in Puerto Rico, where a 6-month, PCR-based, islandwide hospital surveillance study conducted in 2011 found that 333/2,805 (11.9%) *K. pneumoniae* isolates harbored bla_{KPC} (*9*). However, little is known about CRKp genotypes in Puerto Rico.

Our case highlights the many challenges for controlling CRE infections in resource-limited countries like the Dominican Republic and accentuates the potential for international spread of CRKp through travel, particularly between resource-limited regions. Rapid molecular diagnostic tests for CRKp are not widely available, which can delay treatment. Optimal treatment regimens for CRKp remain controversial, but combination therapy could reduce risk for death compared with monotherapy (*10*). Our facility lacked the resources needed to monitor colistin drug levels, a major concern in particular in patients with underlying renal dysfunction.

Risk factors for acquisition of CRE in resource-limited settings are not well defined, potentially delaying diagnosis and implementation of infection control strategies. In our case, recent travel, healthcare contact, and unspecified exposure to antimicrobial drugs might have played a role in the patient's CRE infection. We did not observe additional CRKp infections at our institution during a 6-month follow-up period after this case. However, we were unable to institute an active molecular surveillance program. We cannot rule out silent transmission and colonization of other hospitalized patients or contacts. Even though 2 cases have now been linked to travel to Puerto Rico, no molecular epidemiologic data are available from that island. Future studies should target active surveillance for CRKp in the Caribbean.

Of note, although ST258 has been the dominant genotype of the CRE epidemic globally, the ST307 clone could be expanding disproportionately in some locations. For example, in Houston, Texas, USA, ST307 now accounts for more *K. pneumoniae* infections than ST258 (*3*). Moreover, ST307 Tn*4401e* bla_{KPC-2} isolates showed high carbapenem

MICs. Taken together, our data suggest that ST307 is highly drug resistant and harbors an extended repertoire of antimicrobial resistance genes, which might have accelerated its recent emergence and wide dissemination.

This work was supported by grants from the National Institutes of Health, National Institute of Allergy and Infectious Diseases R01 AI116939 (A.-C.U.). The funders had no role in study design, data collection, interpretation, or the decision to submit the work for publication. A.-C.U. has received research funding from Merck, unrelated to the current study. All other authors declare no conflict of interest.

About the Author

Dr. Rojas is medical director of Centro Medico Luperon and the hospital epidemiologist at Hospital General de la Plaza de la Salud in Santo Domingo, Dominican Republic. Her research interests include healthcare-associated infections and detection and control of carbapenemase-producing *Enterobacteriaceae*.

References

1. US Centers for Disease Control and Prevention. Antibiotic resistance threats in the United States, 2013. 2013 Apr [cited 2017 Oct 10]. http://www.cdc.gov/drugresistance/threat-report-2013/index.html
2. Zarkotou O, Pournaras S, Tselioti P, Dragoumanos V, Pitiriga V, Ranellou K, et al. Predictors of mortality in patients with bloodstream infections caused by KPC-producing *Klebsiella pneumoniae* and impact of appropriate antimicrobial treatment. Clin Microbiol Infect. 2011;17:1798–803. https://doi.org/10.1111/j.1469-0691.2011.03514.x
3. Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis JJ, et al. Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* isolates, Houston, Texas: unexpected abundance of clonal group 307. MBio. 2017;8:e00489–17. https://doi.org/10.1128/mBio.00489-17
4. Giddins MJ, Macesic N, Annavajhala MK, Stump S, Khan S, McConville TH, et al. Successive emergence of ceftazidime-avibactam resistance through distinct genomic adaptations in bla_{KPC-2}-harboring *Klebsiella pneumoniae* sequence type 307 isolates. Antimicrob Agents Chemother. 2018;62:e02101–17.
5. Escandón-Vargas K, Reyes S, Gutiérrez S, Villegas MV. The epidemiology of carbapenemases in Latin America and the Caribbean. Expert Rev Anti Infect Ther. 2017;15:277–97. https://doi.org/10.1080/14787210.2017.1268918
6. Pasteran FG, Otaegui L, Guerriero L, Radice G, Maggiora R, Rapoport M, et al. *Klebsiella pneumoniae* carbapenemase-2, Buenos Aires, Argentina. Emerg Infect Dis. 2008;14:1178–80. https://doi.org/10.3201/eid1407.070826
7. Falco A, Ramos Y, Franco E, Guzmán A, Takiff H. A cluster of KPC-2 and VIM-2-producing *Klebsiella pneumoniae* ST833 isolates from the pediatric service of a Venezuelan Hospital. BMC Infect Dis. 2016;16:595. https://doi.org/10.1186/s12879-016-1927-y
8. Castanheira M, Costello AJ, Deshpande LM, Jones RN. Expansion of clonal complex 258 KPC-2-producing *Klebsiella pneumoniae* in Latin American hospitals: report of the SENTRY Antimicrobial Surveillance Program. Antimicrob Agents Chemother. 2012; 56:1668–9, author reply 1670–1. https://doi.org/10.1128/AAC.05942-11
9. Robledo IE, Aquino EE, Vázquez GJ. Detection of the KPC gene in *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Acinetobacter baumannii* during a PCR-based nosocomial surveillance study in Puerto Rico. Antimicrob Agents Chemother. 2011;55:2968–70. https://doi.org/10.1128/AAC.01633-10
10. Morrill HJ, Pogue JM, Kaye KS, LaPlante KL. Treatment options for carbapenem-resistant *Enterobacteriaceae* infections. Open Forum Infect Dis. 2015;2:ofv050. https://doi.org/10.1093/ofid/ofv050

Address for correspondence: Anne-Catrin Uhlemann, Columbia University, Department of Medicine, Division of Infectious Diseases, 630 W 168th St, New York, NY 10032, USA; email: au2110@columbia.edu

Feast of Sacrifice and Orf, Milan, Italy, 2015–2018

Stefano Veraldi, Luigi Esposito, Paolo Pontini, Fabrizio Vaira, Gianluca Nazzaro

Author affiliation: Università degli Studi di Milano, Milan, Italy

DOI: https://doi.org/10.3201/eid2508.181063

Orf (ecthyma contagiosum) is an infection of the skin caused by a DNA virus belonging to the genus *Parapoxvirus*. We recently observed 7 cases of orf in Muslim men living in the metropolitan area of Milan, Italy, who acquired the infection after the Feast of Sacrifice.

Orf (ecthyma contagiosum) is an infection of the skin caused by a DNA virus of the genus *Parapoxvirus*, family *Poxviridae*. Skin lesions (e.g., vesicles, blisters, pustules, erosions, ulcers, papules, nodules) occur at sites of inoculation of the virus 3–15 days after infection. Hands are usually affected (*1*). The differential diagnosis for orf includes milker's nodule, anthrax, tularemia, fish tank granuloma, cutaneous leishmaniasis, pyogenic granuloma, and keratoacanthoma (*1*). The disease spontaneously heals within 6 weeks, although pain, bacterial superinfections, and regional lymphadenitis are possible (*1*). Treatment is based on topical antiseptics (*1*).

Orf virus usually infects sheep and goats. Humans are infected by handling infected meat from these animals; orf is considered an occupational disease in shepherds, shearers, veterinarians, butchers, and cooks (*1*).

Article DOI: https://doi.org/10.3201/eid2508.171730

Multidrug-Resistant *Klebsiella pneumoniae* ST307 in Traveler from Puerto Rico to Dominican Republic

Appendix

Detailed Clinical Case

The patient and family did not recall the source of the infection in Puerto Rico or treatment regimen, but endorsed that she had experienced adverse reactions to her antimicrobial treatments. Further workup, including transthoracic echocardiogram, revealed a vegetation of 0.4 X 0.5 cm on the tip of the hemodialysis catheter on the right side of the heart, which was subsequently removed. The patient remained bacteremic for 6 days after initiation of treatment. She clinically improved and her procalcitonin levels decreased from ≥200 ng/mL to 29 ng/mL. We did not have the ability to measure colistin levels in our laboratory. However, on day 11 of treatment, the patient began to experience neuropathy and diarrhea and fosfomycin was discontinued. On day 25 of admission the patient experienced cardiac arrest and died.

Antimicrobial Susceptibility Testing

We performed antimicrobial susceptibility testing by using Vitek2 Compact (bioMérieux, https://www.biomerieux.com) and interpreted susceptibilities according to Clinical and Laboratory Standards Institute (CLSI) guidelines (*1*). The patient's isolate was nonsusceptible to all tested antimicrobial drugs, except polymyxin (Table 1). Fosfomycin susceptibility testing and broth microdilution testing for polymyxin was not initially available at our institution and was performed after the patient died according to CLSI guidelines (*1*). Modified Hodge's test was positive, indicating presence of a carbapenemase.

Whole Genome Sequencing and Bioinformatics Analyses

We extracted DNA from bacteria cultured overnight by using the UltraClean Microbial DNA Isolation Kit (QIAGEN, https://www.qiagen.com). We prepared libraries by using Nextera XT DNA Library Prep Kit and sequenced on MiSeq (Illumina, https://www.illumina.com).

We performed SRST2 analysis (*2*) for multilocus sequence typing and characterization of resistance determinants. For comparative sequence analyses, we mapped Illumina reads against a *K. pneumoniae* ST307 reference genome (GenBank accession no. GCA_002148835.1) and included additional, previously published sequences for comparative analyses (Appendix Table 2) (*3*–*5*). We performed variant calling by using Snippy 3.1 after exclusion of mobile genetic elements with PHASTER and IslandViewer 3 (*6*–*8*).

For phylogenetic analyses, we generated a core chromosomal single nucleotide variant alignment by using Snippy 3.1 (*6*). We used a maximum likelihood approach with RAXML to construct a phylogenetic tree based on 860 concatenated SNPs (*9*). We assessed support for nodes by using 1,000 rapid bootstrap inferences and then by a thorough maximum likelihood search. We estimated free model parameters by RAxML and evaluated and optimized likelihood of the final tree under GAMMA (*10*). We created the phylogenetic tree in R 3.4.3 by using the ggtree R package (*11*). The tree was rooted on isolate GCA_003076555, the earliest isolate in the collection.

To determine the location of resistance genes, we used SPAdes (http://cab.spbu.ru/software/spades) for assembly and mapped contigs to the NYC ST307, isolate KP1768, core chromosome and plasmids. This indicated that the IncA/C plasmid harbored bla_{KPC}, bla_{TEM}, *sulI*, *aadB*, *aac6*, as well as *qacE*, implicated in chlorhexidine resistance. $bla_{\text{OXA-1}}$, $bla_{\text{CTX-M-15}}$ and $bla_{\text{SHV100,}}$ as well as *catB*, *fosA*, *tet*, *aac6*, and *aadB* mapped to the ColR replicon, putatively integrated into the core chromosome. No major resistance genes mapped to the IncFIB(K) plasmid, which contained many elements for encoding resistance to diverse metals.

References

1. Clinical and Laboratory Standards Institute. Performance standards for antimicrobial susceptibility testing. 25th informational supplement. CLSI document M100-S25. Wayne (PA): The Institute; 2015.

2. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014;6:90. PubMed https://doi.org/10.1186/s13073-014-0090-6

3. Cerqueira GC, Earl AM, Ernst CM, Grad YH, Dekker JP, Feldgarden M, et al. Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. Proc Natl Acad Sci U S A. 2017;114:1135–40. PubMed https://doi.org/10.1073/pnas.1616248114

4. Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis JJ, et al. Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* isolates, Houston, Texas: unexpected abundance of clonal group 307. MBio. 2017;8:e00489–17. PubMed https://doi.org/10.1128/mBio.00489-17

5. Wyres KL, Hawkey J, Hetland MAK, Fostervold A, Wick RR, Judd LM, et al. Emergence and rapid global dissemination of CTX-M-15-associated *Klebsiella pneumoniae* strain ST307. 2018 [cited 2018 August 11]. https://www.biorxiv.org/content/early/2018/06/21/352740

6. Seemann T. Snippy: Rapid haploid variant calling and core SNP phylogeny. 2017 [cited 2018 July 23]. https://github.com/tseemann/snippy

7. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 2016;44(W1):W16–21. PubMed https://doi.org/10.1093/nar/gkw387

8. Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, et al. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Res. 2015;43(W1):W104–8. PubMed https://doi.org/10.1093/nar/gkv401

9. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3. PubMed https://doi.org/10.1093/bioinformatics/btu033

10. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web servers. Syst Biol. 2008;57:758–71. PubMed https://doi.org/10.1080/10635150802429642

11. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution. 2017;8:28–36. https://doi.org/10.1111/2041-210X.12628

12. European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 8.1, 2018. [cited 2017 Oct 10] http://www.eucast.org

Table 1. Results of antimicrobial susceptibility testing and molecular mechanisms of resistance of *Klebsiella pneumoniae* in patient in Dominican Republic with recent travel to Puerto Rico*

Antimicrobial drug	Mean inhibitory concentration (μg/mL) and EUCAST interpretation	Molecular mechanism
Meropenem	≥32 R	bla_{KPC-2}
Ceftriaxone	≥64 R	$bla_{CTX-M-15}$, $bla_{SHV-100}$, bla_{OXA-1}, bla_{TEM-1D}
Piperacillin/Tazobactam	≥64 R	
Cefepime	≥64 R	
Ciprofloxacin	≥4 R	*gyrA* Y83I, *parC* S80I
Gentamicin	≥16 R	*aac(3)-Iia, aac(6')-Ib, aac(6')-33, aadB*
Tobramicin	≥16 R	
Amikacin	16 I	
Colistin	<0.5 S	
Polymyxin	1 S	
Fosfomycin	NA R	*fosA3*
Trimethoprim/Sulfamethoxazole	≥320 R	*dfrA14, sulI, sulII*
Tetracycline	≥16 R	*tetD*
Minocycline	≥16 R	
Tigecycline	≥8 R	
Chloramphenicol	32 R	*catB4*

*Susceptibility testing for fosfomycin was performed by using disc diffusion testing; the zone diameter was 19 mm. Polymyxin B susceptibility testing was performed by using broth microdilution and meropenem testing was performed by using Etest (bioMérieux, https://www.biomerieux.com). Clinical and Laboratory Standards Institute breakpoints are not available for intravenous fosfomycin, however EUCAST criteria interpret the isolate as resistant (*12*). EUCAST, European Committee on Antimicrobial Susceptibility Testing; I, intermediate; NA, not available; R, resistant; S, susceptible.

Table 2. ST307 reference genomes from GenBank and metadata for previously published sequences used for comparative analyses of *Klebsiella pneumoniae* in patient in Dominican Republic with recent travel to Puerto Rico*

Isolate	Accession number	Location	Year	bla_{KPC}	bla_{CTX-M}
35111	SRR6892777	U.S. (New York, NY)	2014		CTX-M-15
35123	SRR6892773	U.S. (New York, NY)	2014		CTX-M-15
35476	SRR6892718	U.S. (New York, NY)	2015		CTX-M-15
35438A	SRR6892699	U.S. (New York, NY)	2015	KPC-2	CTX-M-15
ERR1218732	ERR1218732	Thailand	2015		CTX-M-15
ERR1218738	ERR1218738	Thailand	2015		CTX-M-15
ERR257620	ERR257620	UK	2010		CTX-M-15
ERR2631531	ERR2631531	Norway	2013		CTX-M-15
ERR2631532	ERR2631532	Norway	2012		CTX-M-15
ERR311471	ERR311471	UK	2012		CTX-M-15
ERR349773	ERR349773	Nepal	2012		CTX-M-15
ERR349787	ERR349787	Nepal	2012		CTX-M-15
GCA_001566595	GCA_001566595	Italy	2014	KPC-3	CTX-M-15
GCA_002166915	GCA_002166915	Colombia	2013		CTX-M-15
GCA_002166955	GCA_002166955	Italy	2014	KPC-3	
GCA_002167025	GCA_002167025	UK	2015	KPC-2	CTX-M-15
GCA_003076555	GCA_003076555	Iran	2009		CTX-M-15
KP1766	SRR6844958	U.S. (New York, NY)	2016	KPC-2	CTX-M-15
KP1767	SRR6844959	U.S. (New York, NY)	2016	KPC-2	CTX-M-15
KP1768	SRR6845005	U.S. (New York, NY)	2016	KPC-2	CTX-M-15
KP1769	SRR6845004	U.S. (New York, NY)	2016	KPC-2	CTX-M-15
NR0970	SRR9309433	U.S. (New York, NY)	2014	KPC-3	
NR5632	SRR6348596	U.S. (New York, NY)	2016	KPC-2	CTX-M-15
NR5706	SRR6348592	U.S. (New York, NY)	2016	KPC-2	CTX-M-15
NR6025	**SRR9309434**	**Dominican Republic**	**2015**	**KPC-2**	**CTX-M-15**
Reference	GCA_002148835.1	U.S. (Houston, TX)	2011	KPC-2	CTX-M-15
SRR5387157	SRR5387157	U.S. (Houston, TX)	2015	KPC-2	CTX-M-15
SRR5387161	SRR5387161	U.S. (Houston, TX)	2015	KPC-2	CTX-M-15
SRR5387164	SRR5387164	U.S. (Houston, TX)	2015	KPC-2	CTX-M-15
SRR5387169	SRR5387169	U.S. (Houston, TX)	2015	KPC-2	CTX-M-15
SRR5387172	SRR5387172	U.S. (Houston, TX)	2015		CTX-M-15
SRR5877450	SRR5877450	Cambodia	2013		CTX-M-15
SRR7345600	SRR7345600	Australia	2013		CTX-M-15
SRR7345601	SRR7345601	Australia	2013		CTX-M-15
SRR7345602	SRR7345602	Australia	2014		CTX-M-15
SRR851036	SRR851036	U.S. (Boston, MA)	2012		CTX-M-15

*Bold text indicates isolate in this case.

Conclusion

In conclusion, the work presented in this chapter built upon observations from Chapter 2 and used WGS to elucidate the genomic epidemiology of emerging forms of Gram negative resistance: *mcr-1* and *K. pneumoniae* ST307. WGS was able to provide important insights at a single-institution level during the study of *mcr-1.* In this context, it allowed detection of *mcr-1* in isolates that may not routinely have polymyxin susceptibility testing and showed that all *E. coli* isolates carrying *mcr-1* were closely related. This in turn allowed further epidemiological investigation that identified same-day endoscopy as a possible linking epidemiological factor.

In the case of *K. pneumoniae* ST307, it provided regional-level insights by putting the first CRKP isolate identified in the Dominican Republic into the context of CRE within the Western Hemisphere. This highlights that WGS is becoming an indispensable tool for understanding the rise and spread of Gram negative resistance, with a view to actionable interventions to prevent this spread.

Chapter 4:

Emergence of polymyxin resistance in clinical *Klebsiella pneumoniae* through diverse genetic adaptations: a genomic retrospective cohort study

Work from this chapter was published in:

1) Macesic N, Nelson B, McConville TH, et al. Emergence of Polymyxin Resistance in Clinical *Klebsiella pneumoniae* Through Diverse Genetic Adaptations: A Genomic, Retrospective Cohort Study. *Clin Infect Dis* 2019 10.1093/cid/ciz623 [3]

The following contributions were made by specific co-authors to the work. No co-authors were Monash University students.

Nature of contribution	Co-authors
Concept	Uhlemann AC
Isolate collection	McConville TH, Giddins MJ, Green DA, Stump S, Gomez-Simmonds A
Susceptibility testing	McConville TH
Illumina and Nanopore sequencing	Giddins MJ, Gomez-Simmonds A, Annavajhala M
Sequence analysis	Annavajhala M, Uhlemann AC
Manuscript writing	Uhlemann AC

Introduction

This chapter expands on the theme of use of WGS to study MDR-GNB mechanisms of resistance and epidemiology. In the prospective study in Chapter 2 we noted PR in several patients, firstly due to *mcr-1* in cephalosporin-resistant *E. coli* (as detailed in Chapter 3) but also in CRKP isolates. This was a particular concern as polymyxins (including colistin and polymyxin B) are considered antimicrobials of 'last resort' for treatment of CRE infections. Prior surveys of polymyxin susceptibility in CRE have observed high rates of PR ranging from 5 to >40%, with PR *K. pneumoniae* accounting for the majority [82]. Despite these high rates, laboratory PR testing is not routinely performed on all GNB, potentially leading to under-diagnosis and undetected spread of PR [300]. This alarming situation led me to study PR in CRKP more systematically.

Firstly, I wanted to focus on how PR develops and spreads in *K. pneumoniae*. While polymyxin exposure is considered a risk factor, some patients develop PR without exposure [88, 166, 168]. Furthermore, healthcare-associated transmission of PR has been described [102, 104, 106] but it is not known to what extent in-hospital spread contributes to PR outside of outbreak settings. Finally, our understanding of the underlying molecular causes of PR has been limited. In order to address these questions, I therefore conducted an extensive genomic survey of clinical PR *K. pneumoniae* isolates from 2011–2018 (including isolates sequenced as part of the prospective study in Chapter 2).

Clinical Infectious Diseases

MAJOR ARTICLE

Emergence of Polymyxin Resistance in Clinical *Klebsiella pneumoniae* Through Diverse Genetic Adaptations: A Genomic, Retrospective Cohort Study

Nenad Macesic,[1,2] Brian Nelson,[3] Thomas H. Mcconville,[1] Marla J. Giddins,[1,4] Daniel A. Green,[5] Stephania Stump,[1,4] Angela Gomez-Simmonds,[1] Medini K. Annavajhala,[1,4] and Anne-Catrin Uhlemann[1,4]

[1]Division of Infectious Diseases, Columbia University Irving Medical Center, New York City, New York; [2]Central Clinical School, Monash University, Melbourne, Australia; [3]Department of Pharmacy, Columbia University Irving Medical Center/New York Presbyterian Hospital, [4]Microbiome & Pathogen Genomics Core, and [5]Clinical Microbiology Laboratory, Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York City, New York

Background. Polymyxins are antimicrobials of last resort for the treatment of carbapenem-resistant Enterobacteriaceae, but resistance in 5% to >40% isolates has been reported. We conducted a genomic survey of clinical polymyxin-resistant (PR) *Klebsiella pneumoniae* to determine the molecular mechanisms of PR and the role of polymyxin exposure versus transmission in PR emergence.

Methods. We included 88 patients with PR *K. pneumoniae* from 2011–2018 and collected demographic, antimicrobial exposure, and infection data. Whole-genome sequencing was performed on 388 isolates, including 164 PR isolates. Variant calling and insertion sequence detection were performed, focusing on key genes associated with PR (mgrB, crrAB, phoPQ, and pmrAB). We conducted phylogenetic analyses of key *K. pneumoniae* multi-locus sequence types (ST258, ST17, ST307, and ST392).

Results. Polymyxin exposure was documented in 53/88 (60%) patients prior to PR detection. Through an analysis of key PR genes, we detected 129 individual variants and 72 unique variant combinations in PR isolates. This included multiple, distinct changes in 36% of patients with serial PR isolates. Insertion sequence disruption was limited to mgrB ($P < .001$). Polymyxin minimum inhibitory concentrations showed stepwise increases with the number of PR genes affected ($P < .001$). When clusters containing PR isolates in ≥2 patients were analyzed, 10/14 had multiple genetic events leading to PR.

Conclusions. Molecular mechanisms leading to PR in clinical *K. pneumoniae* isolates are remarkably heterogenous, even within clusters or individual patients. Polymyxin exposure with de novo PR emergence led to PR in the majority of patients, rather than transmission. Optimizing polymyxin use should be a key strategy in stopping the spread of PR.

Keywords. polymyxin B; colistin; antimicrobial resistance; multidrug resistance; *Klebsiella pneumoniae*.

Polymyxins (including colistin and polymyxin B) are considered antimicrobials of last resort for the treatment of carbapenem-resistant Enterobacteriaceae (CRE) infections. Despite their toxicity, they have been increasingly used in the last decade, leading to concern about the emergence of polymyxin resistance (PR) [1]. Studies focusing on CRE have observed disturbingly high rates, ranging from 5% to >40%, with PR *Klebsiella pneumoniae* accounting for the majority [1]. Laboratory PR testing is not routinely performed on all gram-negative bacilli, potentially leading to underdiagnoses and the undetected spread of PR [2]. To complicate matters further, plasmid-mediated forms of PR (*mcr* genes) have been noted since 2015 [3]. While the spread of *mcr* to CRE has been limited, it underlines the importance of PR as a public health priority.

Despite this emerging threat, how PR develops and spreads in *K. pneumoniae* remains incompletely understood. Exposure to polymyxins is considered a common risk factor; however, some patients develop PR without exposure [4–6]. Early data on the clinical factors contributing to PR focused on outbreaks [7–9], but it is not known to what extent in-hospital spread contributes to PR outside of outbreak settings. Furthermore, our understanding of the underlying molecular causes of PR has been limited, but has important implications for improving diagnoses and treatments of PR. Whole-genome sequencing (WGS) now provides an unprecedented opportunity to address these questions. To date, previous epidemiologic studies have incorporated limited genomic data [4–6], while genomic studies have included limited numbers of isolates, often with a paucity of clinical metadata [10–15].

In order to address these gaps, we conducted a comprehensive genomic survey of clinical PR *K. pneumoniae* isolates from 2011–2018. Our central hypothesis was that PR in *K. pneumoniae* is a complex genetic trait with a previously unrecognized diversity of PR-conferring variants in clinical isolates. Secondly, we hypothesized that exposure to polymyxins, rather

Received 3 January 2019; editorial decision 5 June 2019; accepted 17 July 2019; published online September 12, 2019.

Correspondence: A.-C. Uhlemann, 630 West 168th Street, New York, NY 10032 (au2110@columbia.edu).

Clinical Infectious Diseases® 2019;XX(XX):1–8

DOI: 10.1093/cid/ciz623

than in-hospital transmission, leads to the development of PR in *K. pneumoniae*.

METHODS

Study Population and Isolate Selection

The study was reviewed and approved by the Columbia University Irving Medical Center Institutional Review Board. Polymyxin B is the polymyxin on formulary at our institution and polymyxin B dosing details are included in the Supplementary Methods. We systematically reviewed all PR *K. pneumoniae* isolates in an institutional collection spanning 2011–2018 at a health-care system with a tertiary care center and an affiliated community hospital. Patients with PR *K. pneumoniae* isolates available in this collection were included and all available *K. pneumoniae* isolates in these patients were analyzed, regardless of polymyxin susceptibility (Supplementary Figure 1).

To provide a genomic context, we included clinical isolates, as well as isolates collected in our study of multidrug-resistant bacteria in liver transplant recipients [16]. These isolates were selected if they belonged to multi-locus sequence types (MLSTs) and their single locus variants were present in PR isolates cultured from ≥2 patients (ST258, ST17, ST307, and ST392).

Clinical data were extracted from the electronic medical record of each patient, including demographics, comorbidities, and antimicrobial exposure. For each patient, the first PR isolate was assessed as being associated with colonization or infection, according to National Healthcare and Safety Network criteria [17]. Crude mortality at 7 and 30 days was recorded and compared to an unmatched cohort of patients from the same time period who had polymyxin-susceptible, carbapenem-resistant, *Klebsiella pneumoniae* infections, as identified in an institutional database.

Susceptibility Testing

All isolates had antimicrobial susceptibility testing using Vitek2 (bioMerieux; 2011–2017) and Microscan Walkaway (2017–2018). Isolates were initially screened with a polymyxin B Etest performed during routine clinical care. We then performed minimum inhibitory concentration (MIC) determinations with broth microdilution on all isolates and used these as the definitive polymyxin susceptibility data in all analyses. Testing was done according to Clinical and Laboratory Standards Institute guidelines [18]. In vitro PR was defined as an MIC >2 mg/L [18, 19].

Genomic Definitions and Analyses

Firstly, we focused on key genes described in multiple previous studies (*crrAB*, *mgrB, phoPQ,* and *pmrAB*) [10–15, 20]. Henceforth, these genes will be referred to as canonical PR genes. We also conducted a search in other genes implicated in the literature, which can be considered a "secondary resistome" [1] (Supplementary Table 1).

We performed WGS on all included isolates, as described previously [16, 21, 22]. Sequences were deposited under accession number PRJNA557275. SRST2 was used for MLST and resistance determinant detection [23]. There was 1 isolate from each key MLST that had MinION (Oxford Nanopore) long-read sequencing and assembly with Unicycler [24]. For details of variant calling, phylogenetic analyses, and insertion sequence detection, see the Supplementary Methods. The functional impact of SNVs on the protein sequence was predicted using PROVEAN [25].

Statistical Analysis

Categorical variables were compared using χ^2 or Fisher's exact tests and continuous variables were compared using the Student's *t*-test or Mann-Whitney-Wilcoxon, as appropriate. An analysis of variance was used for the comparison of means between multiple groups. Statistical analyses were performed in R (v3.4.0).

RESULTS

Study Population and Clinical Characteristics

We identified 665 patients with carbapenem-resistant *K. pneumoniae* and 106 patients with PR *K. pneumoniae* over the 7-year study period. Of these, 88 patients had PR isolates available and were included in the study (Supplementary Figure 1). Approximately half of the patients were admitted from home and 22% were solid-organ transplant recipients (Table 1). The median time from admission to detection of the first PR *K. pneumoniae* isolate was 21 days (interquartile range [IQR] 1–37 days). In 61/88 (69%) patients, the first PR *K. pneumoniae* isolate was associated with a clinical infection (Table 2), most frequently of the respiratory tract (29/61, 48%). Following the detection of PR *K. pneumoniae*, 7-day and 30-day all-cause

Table 1. Clinical Characteristics of Cohort

Clinical characteristic	(N = 88)
Median age (IQR)	61 years (18–90 years)
Male gender	53 (60%)
Origin	
Community	42 (48%)
Hospital	23 (26%)
Nursing facility	23 (26%)
Comorbidities	
Diabetes mellitus	25 (28%)
Pulmonary disease	29 (33%)
Liver disease	10 (11%)
Malignancy	17 (19%)
HIV	4 (4.5%)
Solid organ transplant recipient	19 (21%)
Median Charlson Comorbidity Score (IQR)	4 (2–5)
Outcomes	
Crude 7-day mortality	11 (13%)
Crude 30-day mortality	28 (32%)

Abbreviations: HIV, human immunodeficiency virus; IQR, interquartile range.

Table 2. Details of Infection Associated With First Polymyxin-resistant Isolate

Infection Details	Number of Patients
Patients with infection with first polymyxin resistant isolate	61
Site of infection	
Urinary tract infection	15 (25%)
Respiratory tract infection	16 (26%)
Ventilator-associated pneumonia	13 (21%)
Skin and soft tissue infection	3 (4.9%)
Intra-abdominal infection	7 (12%)
Multiple sites	5 (8%)
Bloodstream infection associated with primary infection	13 (21%)
Outcome	
Death at 7 days	9 (15%)
Death at 30 days	23 (38%)

mortality rates were 13% and 32%, respectively, which increased to 15% and 38%, respectively, if the first PR *K. pneumoniae* isolate was associated with an infection. When compared with an unmatched cohort of 66 patients with polymyxin-susceptible infections, PR infections were associated with all-cause mortality at 7 days (9/61 vs 1/66 patient, respectively; $P = .0069$), but not at 30 days (23/61 vs 18/66 patients, respectively; $P = .29$).

Patients received combination therapy for most PR infections (45/61, 74%; Supplementary Figure 2), with tigecycline-containing regimens predominating (35/61 patients, 57%). Receipt of a polymyxin-containing regimen was associated with significantly lower clinical response rates (13/34 vs 19/24 patients not on a polymyxin-containting regimen; $P = .0048$). There were 5 patients (8%) that received a ceftazidime-avibactam–containing regimen initially and 6 others that received ceftazidime-avibactam during follow-up. Ceftazidime-avibactam resistance subsequently developed in 4/11 patients.

Exposure to Polymyxins

Most patients (53/88, 60%) had polymyxin exposure prior to the detection of PR, receiving intravenous (IV) polymyxin for a median of 12 days (range 1–66). While no patients received nebulized colistin alone preceding PR, 11 patients received it with IV polymyxin for a median of 12 days (range 1–49). There were 3 patients that only had exposure to topical formulations of polymyxin prior to the detection of PR. For the 42 patients admitted from home, 27/42 (64%) had been admitted to our institution in the prior year and 23/42 (55%) had received polymyxins. Following the detection of PR, 53/88 patients (60%) had ongoing polymyxin exposure (median 11 days, range 1–76 days).

Isolate Characteristics

For 88 patients with PR *K. pneumoniae,* we analyzed and sequenced 164 PR *K. pneumoniae* isolates and 97 polymyxin-susceptible isolates (median 2 per patient, IQR 1–4). An additional 127 polymyxin-susceptible, MLST-matched *K. pneumoniae* isolates from 89 patients were included for phylogenetic studies. MICs of PR isolates ranged from 4 to >128 mcg/ml (Supplementary Figure 3). In patients with infections caused by the first PR *K. pneumoniae* isolate, neither 7-day nor 30-day mortality were associated with higher MICs ($P = .95$ and $P = .075$, respectively).

All isolates were resistant to multiple classes of antibiotics (Supplementary Table 2), including near-universal carbapenem resistance (377/388, 97%) with $bla_{\text{KPC-2/3}}$ being the putative cause in 361/377 isolates (97%). We also detected $bla_{\text{NDM-1}}$ (3/377 isolates) and $bla_{\text{OXA-48}}$ (2/377 isolates). Tigecycline resistance was noted in 25/148 (17%) PR isolates.

ST258 was the most common isolate (301/388 isolates in 145 patients, 78%), with ST17, ST307, and ST392 being other key sequence types (Supplementary Table 3). Serial isolates were available in 56 patients. Polymyxin-susceptible and PR isolates were often isolated almost concurrently, being detected within 7 days of each other in 18/56 patients (32%). When they were of same sequence type, the median pairwise SNV distance between these isolates was 2.5 SNVs (IQR 0–22.5 SNVs). However, multiple *K. pneumoniae* sequence types were noted in 8/56 (14%) patients during follow-up, while 6 patients (10.5%) were colonized with different ST258 clades.

Genomic Determinants of PR

Isolates exhibited a remarkable multiplicity of genomic alterations leading to PR. While variants in *mgrB* were most frequently detected, non–*mgrB* mediated PR was present in 67/164 (41%) isolates (Figure 1). In total, we detected 129 individual variants in canonical PR genes, of which 83 were noted in PR isolates, with 67/83 (82%) not described previously (Supplementary Table 4). We detected 72 genetic combinations to PR if each combination of variants in canonical PR genes is considered an individual pathway (Figure 2). On a patient level, 20/56 (36%) patients with serial PR *K. pneumoniae* isolates available were noted to have >1 genetic combination leading to PR. In these patients, ongoing exposure to polymyxin and/or nebulized colistin may have contributed and was noted in 18/20 patients (90%). No isolates harbored *mcr* genes.

Within this diversity, there were important differences between the canonical genes associated with PR. IS were almost exclusively associated with *mgrB* (40/97 isolates with variants in *mgrB* vs 1/79 isolates with variants in other genes; $P < .001$; Figure 1). *CrrB* was an important contributor (28/164 isolates, 17%), but only 8 isolates had variants in *crrB* without any concurrent variants in other PR canonical genes. Variants in multiple canonical PR genes were detected in 39/164 (24%) PR isolates and were associated with a significant difference in the mean MIC when isolates were categorized by the number of PR canonical genes affected (Supplementary Figure 4; $P < .001$). The increasing duration of polymyxin exposure weakly correlated with variants in an increasing number of PR canonical genes (Spearman's $\rho = 0.27$; P

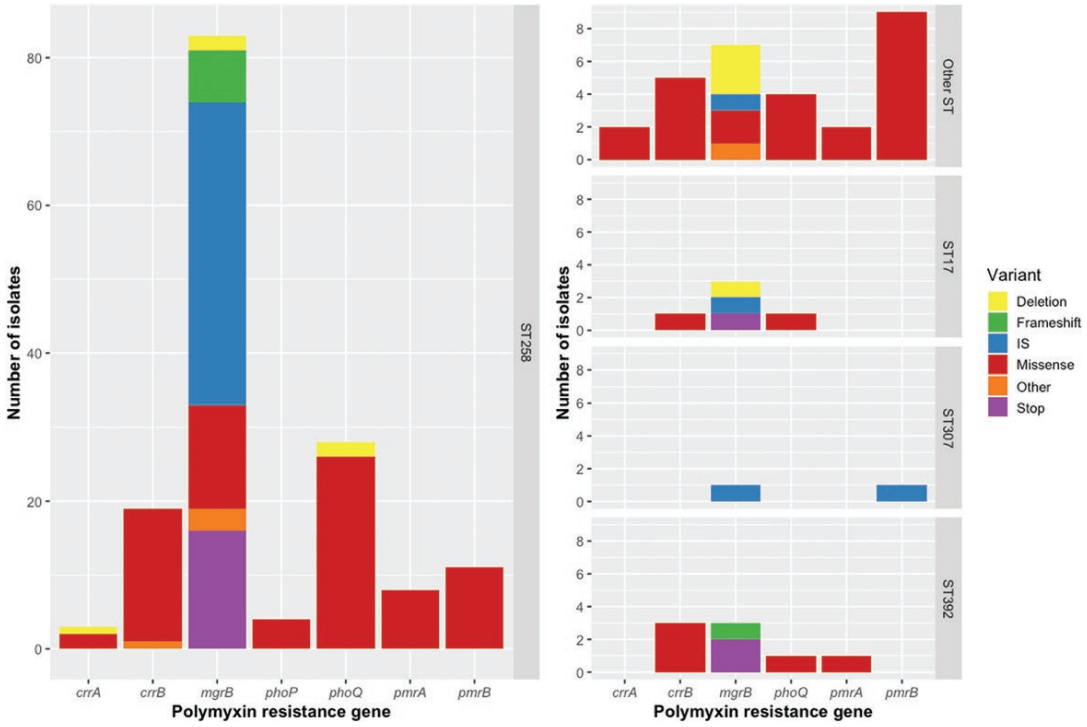


Figure 1. Summary of variants in canonical polymyxin resistance genes in *Klebsiella pneumonia*. The variants in polymyxin resistance canonical genes depended upon the gene affected and clonal background of isolates. Insertion sequences often disrupted *mgrB,* but with the exception of *pmrB* in 1 isolate, did not affect other canonical genes. Abbreviation: IS, insertion sequence.

= .0047), and also with an increasing MIC (Spearman's ρ = 0.26; *P* = .0052). Variants in *crrA, phoP,* and *pmrA* were infrequently observed (5, 4, and 11 isolates, respectively).

In 15/164 (9%) PR isolates, no variants were detected in canonical PR genes. In our screen of the putative secondary resistome [1] (Supplementary Table 5), we noted 2 variants unique to this group: a V82A variant *ramA* and IS disruption of *ompW,* in separate isolates. Despite this screening approach, 13 isolates without an obvious contributor to PR remained.

Variants in PR canonical genes were also noted in 54/218 (25%) polymyxin-susceptible isolates (Supplementary Table 4), highlighting the need for studies to confirm functional impacts. In polymyxin=susceptible isolates, *phoQ* was the gene most frequently affected (36/218 isolates, 17%; *P* < .001; Supplementary Figure 5). Importantly, 15/129 (12%) variants in canonical PR genes were noted in both PR and polymyxin-susceptible isolates, suggesting the restoration of polymyxin susceptibility (eg, in isolates with IS in *mgrB*) or nonfunctional variants with PR resulting from other contributors.

Population Structure of Polymyxin Resistance

We conducted phylogenetic analyses of ST258, ST17, ST307, and ST392 (Figure 3). For ST258, PR isolates were present across the phylogeny, with no predisposition towards a single clade or large clusters suggestive of a generalized PR outbreak. We detected 22 distinct clusters with multiple patients in ST258 and 1 cluster in ST392. There were 17 clusters containing PR isolates, and 14 had ≥2 patients with PR isolates (Supplementary Table 6). In each cluster, we then compared PR genetic combinations between isolates to differentiate the *de novo* emergence of PR from transmission (Figure 4). In 10/14 clusters, there were PR isolates with different genetic combinations leading to PR, suggesting that PR emerged independently within these isolates from the same genomic background. However, 6 clusters had possible transmission with PR isolates from different patients who had the same genetic combination leading to PR, implicating 16 patients and 24 isolates in total, with a median pairwise SNV distance of 6 (IQR 1–9). Of these patients, 9/16 (56%) had polymyxin exposure prior to culture of the PR isolate.

DISCUSSION

Through the integration of genomic and clinical data, our study provides novel insights into how PR develops and spreads in a setting where carbapenem-resistant *K. pneumoniae* is endemic.

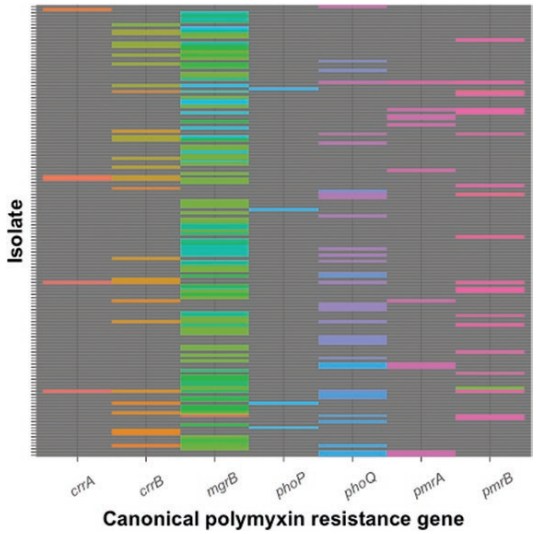


Figure 2. Combinations of genetic alterations leading to polymyxin resistance based on canonical polymyxin resistance gene profile. A heterogeneity of pathways to polymyxin resistance was noted, with 72 unique combinations in 164 polymyxin resistant isolates. To illustrate this, each unique variant in each polymyxin resistance canonical gene was represented by a different color. Specifics of variants encountered can be found in Supplementary Table 4.

Genomics allowed us to detect a multitude of genetic alterations leading to PR. This diversity occurred even within closely related clusters of isolates or individual patients, where we found evidence for the frequent *de novo* emergence of PR. Polymyxin exposure was detected in the majority of patients and may have been a driver of this phenomenon. The transmission of PR was uncommon, suggesting that, in most cases, PR arises either after polymyxin exposure or sporadically, rather than through clonal spread. These data provide a comprehensive overview of the clinical genomics of PR and highlight its heterogeneity, with important implications for diagnostics and infection prevention and for control of these extensively drug-resistant organisms.

While many forms of antimicrobial resistance result from alterations in or the acquisition of single genes [26], we noted not only a remarkable diversity of variants in PR canonical genes, but also complex interplay between them. In at least one-quarter of PR isolates, multiple hits may have contributed to PR, with the acquisition of variants in different PR canonical genes having an additive effect to increase polymyxin MICs. The duration of polymyxin exposure may have contributed to this and weakly correlated both with increasing MICs and the number of PR canonical genes with variants present. Disruptions of *mgrB* play an important role in PR [12]; however, more than 40% of isolates have no alterations in *mgrB*, suggesting additional pathways to PR.

Despite comprehensive genomic screening, no putative genetic cause of PR was identified in approximately 10% of isolates. To address this, we examined genes that have been proposed to constitute a secondary resistome [1] and found potential candidate variants in *ramA* and *ompW*, which require functional validation. Although beyond the scope of this study, future avenues for identifying novel PR determinants include analyzing the plasmid content of these isolates and conducting transcriptomic analyses. Previous studies noted the increased expression of transcripts when traditional WGS did not find mutations present, particularly in heteroresistant *Enterobacter* isolates [27, 28].

Conversely, we noted variants in PR canonical genes in polymyxin-suceptible isolates. Although functional validation

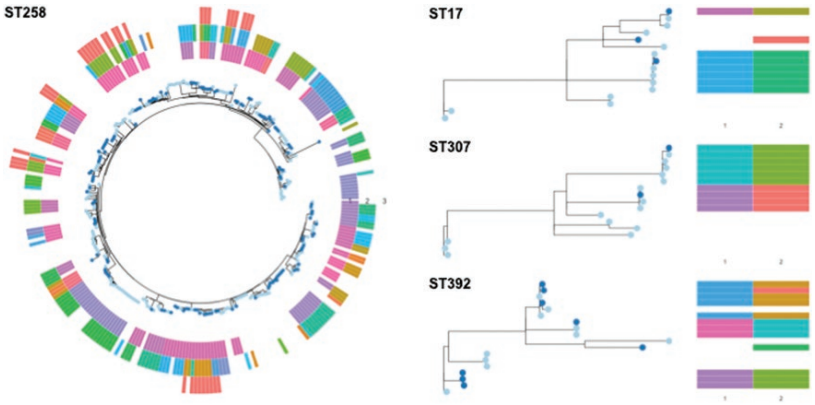


Figure 3. Phylogenetic analysis of polymyxin resistance in key multi-locus sequence types. Polymyxin-susceptible isolates were denoted with a light blue tip point, whereas polymyxin-resistant isolates had dark blue tip points. Other metadata were shown with 3 different bands. Band 1 was used to indicate different clusters of closely related isolates that contained polymyxin-resistant isolates. Band 2 used different colors to represent patients these isolates came from, with each color representing an individual patient. Band 3 showed isolates that were associated with the likely transmission of polymyxin resistance.

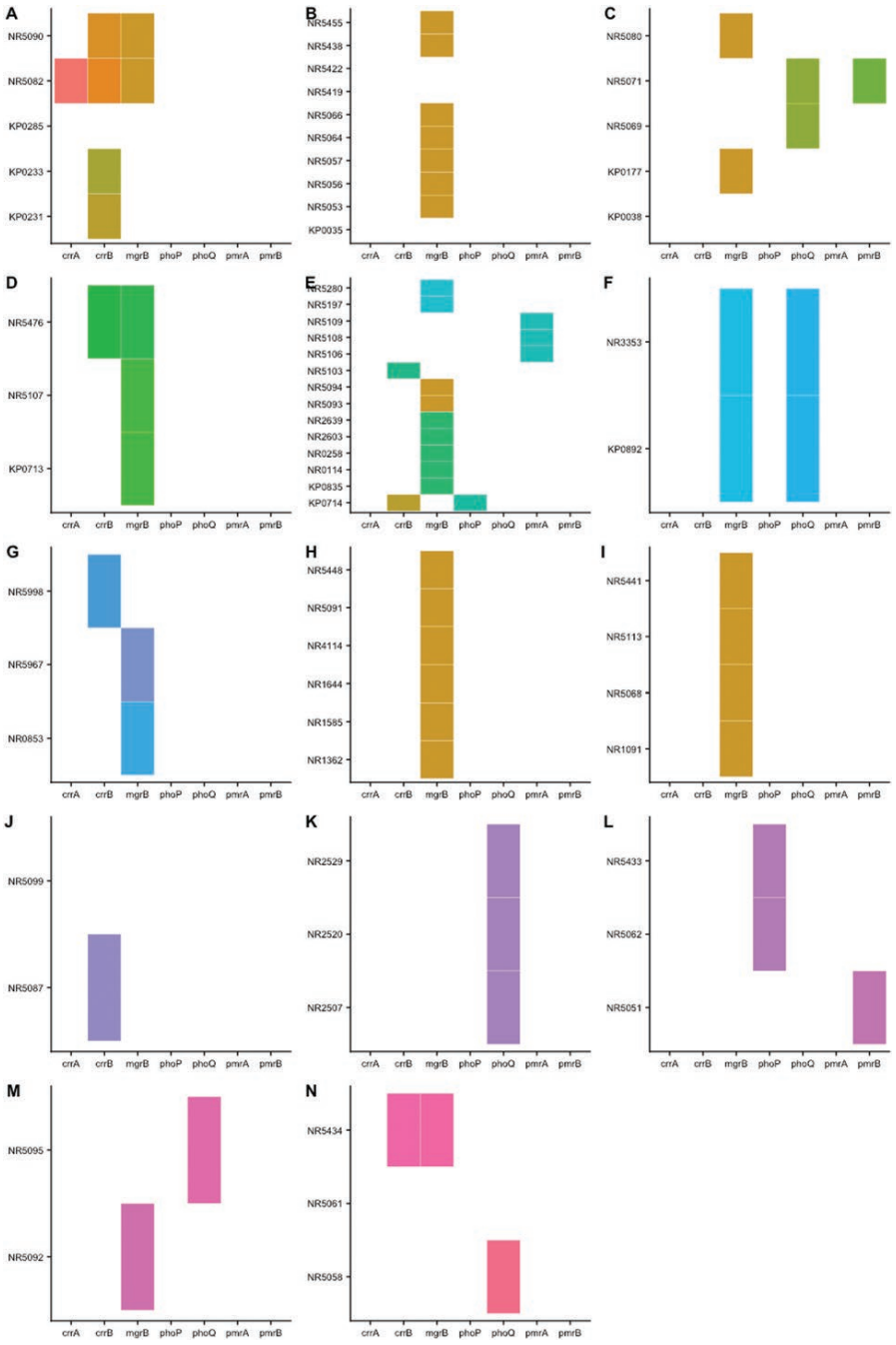


Figure 4. Variants in canonical polymyxin resistance locus according to clusters of closely related isolates. Most clusters of closely related isolates had multiple combinations of genetic alterations leading to polymyxin resistance. Each color indicates an individual variant in the noted canonical polymyxin resistance gene.

is necessary for each variant, a PROVEAN analysis predicted many to not have a deleterious effect on function. In addition, there were instances where the same variants were seen both in PR and polymyxin-susceptible isolates, raising the possibility of additional changes that suppressed PR, as noted previously in *phoPQ* [14].

While our findings confirmed previous work regarding the disruption of *mgrB* by IS [12, 14, 29], it was striking how infrequently other PR genes were affected by IS disruption. The underlying reasons for this predilection of IS for *mgrB* remain to be determined, but may be related to *mgrB*'s role as a constitutive suppressor, with inactivation leading to the upregulation of downstream targets leading to PR [1]. Some of this disruption by IS may be due to the transfer of IS from plasmids, and could offer a novel path to plasmid-mediated PR, in addition to *mcr* genes [30].

While the transmission of PR isolates has been invoked as a possible cause of PR in patients without polymyxin exposure [4, 6], this was an uncommon occurrence in our study: it accounted for PR in only 16/88 patients (18%). This suggests that in the majority of patients, PR arose after polymyxin exposure in isolates from a similar genomic background. The diversity of combinations of PR canonical gene alterations we found within clusters (Figure 4) provides further evidence of this.

Polymyxin exposure was likely a major contributor to the emergence of PR and was noted in the majority of our patients, in contrast to previous reports [4, 6]. Ongoing, selective pressure from polymyxin use may also explain why we saw distinct variants in PR canonical genes arise within clusters and individual patients. Interestingly, non-IV formulations may play a role in PR developing or being maintained. We found only exposure to topical polymyxin in 3 patients prior to the development of PR, and nebulized colistin was commonly used, although usually in conjunction with IV therapy.

There are several limitations to this work. Firstly, it was observational and based at a single center. Polymyxin-susceptibility testing remains a challenge, even when performing broth microdilution [31]. This is of particular concern for isolates with MICs close to the breakpoint, as this impacts susceptibility classification in downstream analyses. Despite our analyses, in 28/88 (32%) patients we did not find evidence of polymyxin exposure or transmission. This may reflect polymyxin exposure outside our institution or, alternately, the undetected transmission of PR due to silent carriers of PR. There are also inherent limitations in genomic analyses: variant calling may have been affected by the selection and quality of our references. IS detection is difficult and our approach relied upon the draft *de novo* assembly of each isolate.

Taken together, our findings reveal PR to be a remarkable challenge on multiple fronts. The detection of PR is a crucial first step, and our study should alert physicians that PR may emerge rapidly or exist concurrently with polymyxin-susceptible isolates, particularly in the setting of polymyxin exposure. The phenotypic diagnosis of PR remains difficult and, while there is interest in predicting antimicrobial resistance phenotypes from genotypes [32], the polygenic nature of PR noted in our study may temper that enthusiasm. This underscores the importance of developing reliable phenotypic assays. Several are in development, including polymyxin broth disk elution, rapid colorimetric tests, and screening agars [33, 34].

Our study also provides a roadmap for preventing the emergence of PR in *K. pneumoniae*: polymyxin exposure and *de novo* emergence of resistance may be the major contributors in a nonoutbreak setting, rather than transmission. We can therefore redouble antimicrobial stewardship efforts. The inappropriate use of polymyxins should obviously be the primary target, including non-IV formulations of polymyxins. This may be of particular relevance in the intensive care unit, where colistin is the most frequently used nebulized antibiotic and topical polymyxins are frequently used for decolonization [35, 36]. Low-dose polymyxin exposure has also previously been linked to the emergence of resistance [20]; therefore, therapeutic drug monitoring of polymyxins offers another opportunity, which is being investigated [37]. While novel β-lactam/β-lactamase inhibitor combinations are a welcome addition in the treatment of CRE infections, they are also limited by the emergence of resistance during treatment [21]. Polymyxins will, therefore, remain essential components of our CRE armamentarium and need to be preserved.

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

Acknowledgments. The authors thank Mia Scholl for assistance in performing *mgrB* polymerase chain reaction.

Disclaimer. The funders had no role in the study design, data collection and interpretation, or decision to submit the work for publication.

Financial support. This work was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (grant numbers AI116939 to A.-C. U., S1-AI116939 to A.-C. U. and A. G.-S., and 5T32AI100852 and K23 AI137316 to A. G.-S.) and the Irving Scholarship from Columbia University (to A.-C. U.).

Potential conflicts of interest. N. M. has received research support from GlaxoSmithKline, unrelated to the current study. A.-C. U. has received research funding from Merck, GlaxoSmithKline, and Allergan, unrelated to the current study. All other authors report no potential conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Poirel L, Jayol A, Nordmann P. Polymyxins: antibacterial activity, susceptibility testing, and resistance mechanisms encoded by plasmids or chromosomes. Clin Microbiol Rev **2017**; 30:557–96.
2. Macesic N, Green D, Wang Z, et al. Detection of *mcr*-1-carrying *Escherichia coli* causing bloodstream infection in a New York city hospital: avian origins, human concerns? Open Forum Infect Dis **2017**; 4:ofx115. doi:10.1093/ofid/ofx115

3. Liu YY, Wang Y, Walsh TR, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis **2016**; 16:161–8.
4. Giacobbe DR, Del Bono V, Trecarichi EM, et al. Risk factors for bloodstream infections due to colistin-resistant KPC-producing *Klebsiella pneumoniae*: results from a multicenter case-control-control study. Clin Microbiol Infect **2015**; 21:1106 e1–8.
5. Richter SE, Miller L, Uslan DZ, et al. Risk factors for colistin resistance among gram-negative rods and *Klebsiella pneumoniae* isolates. J Clin Microbiol **2018**; 56. pii:e00149–18.
6. Rojas LJ, Salim M, Cober E, et al; Antibacterial Resistance Leadership Group. Colistin resistance in carbapenem-resistant *Klebsiella pneumoniae*: laboratory detection and impact on mortality. Clin Infect Dis **2017**; 64:711–8.
7. Antoniadou A, Kontopidou F, Poulakou G, et al. Colistin-resistant isolates of *Klebsiella pneumoniae* emerging in intensive care unit patients: first report of a multiclonal cluster. J Antimicrob Chemother **2007**; 59:786–90.
8. Marchaim D, Chopra T, Pogue JM, et al. Outbreak of colistin-resistant, carbapenem-resistant *Klebsiella pneumoniae* in metropolitan Detroit, Michigan. Antimicrob Agents Chemother **2011**; 55:593–9.
9. Weterings V, Zhou K, Rossen JW, et al. An outbreak of colistin-resistant *Klebsiella pneumoniae* carbapenemase-producing *Klebsiella pneumoniae* in the Netherlands (July to December 2013), with inter-institutional spread. Eur J Clin Microbiol Infect Dis **2015**; 34:1647–55.
10. Aires CA, Pereira PS, Asensi MD, Carvalho-Assef AP. MgrB mutations mediating polymyxin B resistance in *Klebsiella pneumoniae* isolates from rectal surveillance swabs in Brazil. Antimicrob Agents Chemother **2016**; 60:6969–72.
11. Berglund B, Hoang NTB, Tärnberg M, et al. Insertion sequence transpositions and point mutations in mgrB causing colistin resistance in a clinical strain of carbapenem-resistant *Klebsiella pneumoniae* from Vietnam. Int J Antimicrob Agents **2018**; 51:789–93.
12. Cannatelli A, Giani T, D'Andrea MM, et al; COLGRIT Study Group. MgrB inactivation is a common mechanism of colistin resistance in KPC-producing *Klebsiella pneumoniae* of clinical origin. Antimicrob Agents Chemother **2014**; 58:5696–703.
13. Cheng YH, Lin TL, Pan YJ, Wang YP, Lin YT, Wang JT. Colistin resistance mechanisms in *Klebsiella pneumoniae* strains from Taiwan. Antimicrob Agents Chemother **2015**; 59:2909–13.
14. Pitt ME, Elliott AG, Cao MD, et al. Multifactorial chromosomal variants regulate polymyxin resistance in extensively drug-resistant *Klebsiella pneumoniae*. Microb Genom **2018**; 4.
15. Wright MS, Suzuki Y, Jones MB, et al. Genomic and transcriptomic analyses of colistin-resistant clinical isolates of *Klebsiella pneumoniae* reveal multiple pathways of resistance. Antimicrob Agents Chemother **2015**; 59:536–43.
16. Macesic N, Gomez-Simmonds A, Sullivan SB, et al. Genomic surveillance reveals diversity of multidrug-resistant organism colonization and infection: a prospective cohort study in liver transplant recipients. Clin Infect Dis **2018**; 67:905–12.
17. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. Am J Infect Control **2008**; 36:309–32.
18. CLSI. Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Fifth Informational Supplement. CLSI document M100-S25. Wayne, PA: Clinical and Laboratory Standards Institute, **2015**.
19. European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Available at: http://www.eucast.org. Accessed 14 December 2018.
20. Cannatelli A, Di Pilato V, Giani T, et al. In vivo evolution to colistin resistance by PmrB sensor kinase mutation in KPC-producing *Klebsiella pneumoniae* is associated with low-dosage colistin treatment. Antimicrob Agents Chemother **2014**; 58:4399–403.
21. Giddins MJ, Macesic N, Annavajhala MK, et al. Successive emergence of ceftazidime-avibactam resistance through distinct genomic adaptations in blaKPC-2-harboring *Klebsiella pneumoniae* sequence type 307 isolates. Antimicrob Agents Chemother **2018**; 62. pii: e02101–17.
22. Gomez-Simmonds A, Annavajhala MK, Wang Z, et al. Genomic and geographic context for the evolution of high-risk carbapenem-resistant *Enterobacter cloacae* complex clones ST171 and ST78. MBio **2018**; 9: 165–15.
23. Inouye M, Dashnow H, Raven LA, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med **2014**; 6:90.
24. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLOS Comput Biol **2017**; 13:e1005595.
25. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics **2015**; 31:2745–7.
26. Munita JM, Arias CA. Mechanisms of antibiotic resistance. Microbiol Spectr **2016**; 4.
27. Band VI, Crispell EK, Napier BA, et al. Antibiotic failure mediated by a resistant subpopulation in *Enterobacter cloacae*. Nat Microbiol **2016**; 1:16053.
28. Telke AA, Olaitan AO, Morand S, Rolain JM. SoxRS induces colistin heteroresistance in *Enterobacter asburiae* and *Enterobacter cloacae* by regulating the acrAB-tolC efflux pump. J Antimicrob Chemother **2017**; 72:2715–21.
29. Poirel L, Jayol A, Bontron S, et al. The mgrB gene as a key target for acquired resistance to colistin in *Klebsiella pneumoniae*. J Antimicrob Chemother **2015**; 70:75–80.
30. Giordano C, Barnini S, Tsioutis C, et al. Expansion of KPC-producing *Klebsiella pneumoniae* with various mgrB mutations giving rise to colistin resistance: the role of ISL3 on plasmids. Int J Antimicrob Agents **2018**; 51:260–5.
31. Ezadi F, Ardebili A, Mirnejad R. Antimicrobial susceptibility testing for polymyxins: challenges, issues, and recommendations. J Clin Microbiol **2019**; 57. pii: e01390–18.
32. Macesic N, Polubriaginof F, Tatonetti NP. Machine learning: novel bioinformatics approaches for combating antimicrobial resistance. Curr Opin Infect Dis **2017**; 30:511–7.
33. Nordmann P, Jayol A, Poirel L. Rapid detection of polymyxin resistance in Enterobacteriaceae. Emerg Infect Dis **2016**; 22:1038–43.
34. Simner PJ, Bergman Y, Trejo M, et al. Two-site evaluation of the colistin broth disk elution test to determine colistin in vitro activity against gram-negative Bacilli. J Clin Microbiol **2019**; 57. pii: e01163–18.
35. Solé-Lleonart C, Roberts JA, Chastre J, et al; European Study Group for Infections in Critically Ill Patients (ESGCIP) Investigators. Global survey on nebulization of antimicrobial agents in mechanically ventilated patients: a call for international guidelines. Clin Microbiol Infect **2016**; 22:359–64.
36. Hurley JC. Unusually high incidences of *Pseudomonas bacteremias* within topical polymyxin-based decolonization studies of mechanically ventilated patients: benchmarking the literature. Open Forum Infect Dis **2018**; 5:ofy256.
37. Kim ES. Clinical usefulness of therapeutic drug monitoring of colistin in patients treated with colistin. Available at: https://clinicaltrials.gov/ct2/show/NCT03007290. Accessed 17 December 2018.

Conclusion

In this chapter I demonstrated that the genetic basis of PR is more complex and diverse than previously appreciated. In addition to the insights of this study, the data generated through conducting polymyxin susceptibility testing and WGS of isolates in this chapter formed an important resource that allowed several other avenues of inquiry.

Despite extensive efforts, in approximately 10% of isolates no obvious mechanism of PR was identified, leading to the possibility that there are additional determinants of PR. Attempting to detect these will be the focus of the Chapter 5, where WGS data will be used as an input for detecting novel determinants of PR by comparing closely related isolate pairs and by conducting a bacterial GWAS.

As noted in Chapter 1, there is also significant interest in predicting AMR phenotype from genotype. Phenotypic diagnosis of PR remains difficult, therefore a molecular-based approach for detection of PR remains attractive. The unique nature of the dataset resulting from this study gave me an unprecedented opportunity to attempt AST genotype-phenotype prediction of PR in Chapter 6. Given the complex polygenic nature of PR, I will do that through use of machine learning rather than rule-based approaches for prediction.

Chapter 5:

Determining novel determinants of polymyxin resistance in *Klebsiella pneumoniae* clonal complex 258 through genomic analysis and genome-wide association study

Introduction

In the previous chapter I focused on known determinants of PR (*crrAB*, *mgrB, phoPQ, pmrAB*) that I termed canonical PR genes [109, 113-115, 117, 123, 124], and showed that the pathways leading to PR are remarkably diverse but that our knowledge of the mechanisms underlying PR in *K. pneumoniae* remains incomplete. Firstly, canonical genes may be implicated in the majority of isolates, but their functional impact remains unclear. Furthermore, a proportion of PR isolates have no canonical gene alterations. It also remains unclear which genetic changes are sufficient to induce PR alone and which may provide an additive or synergistic effect to increase the polymyxin MIC.

My central hypothesis in this chapter was that PR in *K. pneumoniae* is a complex genetic trait with previously unrecognized PR-conferring variants in clinical isolates. I therefore used several different approaches to identify candidate determinants of PR. Initially, this was done by studying closely related pairs of clinical *K. pneumoniae* CG258 isolates with differing polymyxin susceptibility that had been included in the study in Chapter 4 . I then used the genomic data of isolates from Chapter 4 in combination with publicly available *K. pneumoniae* CG258 genomes with polymyxin susceptibility data to conduct a bacterial GWAS. Finally, Dr. Thomas H. McConville used CRISPR-Cas9 genome editing to

experimentally validate the PR candidate loci identified through these approaches, performed capsular assays and also assessed virulence of mutant strains in a *Galleria mellonella* model. Prof. M. Stephen Trent and colleagues then performed thin layer chromatography to determine lipid A modifications.

Manuscript

Genomic analysis and genome-wide association study reveal role of *barA*/*uvrY* two-component system in polymyxin resistance in *Klebsiella pneumoniae* clonal group 258

Keywords:
Polymyxin B; colistin; CRISPR-Cas9; *Klebsiella pneumoniae*; genome-wide association study

Running title:
Novel genomic determinants of polymyxin resistance

Words: Abstract - 239; Importance – 141; Text – 3702

Tables – 3, Figures – 2, References – 62, Supplementary Tables – 2

Abstract

Polymyxins are treatments of last resort for Gram-negative bacterial infections. Their use over the last decade has led to concerns about increasing polymyxin resistance (PR), particularly in carbapenem-resistant *Klebsiella pneumoniae*. Despite the threat of losing this class of last-line antimicrobials, our knowledge of the mechanisms underlying PR in *K. pneumoniae* remains incomplete. Several key genes have been described in previous studies, but a proportion of PR isolates do not have variants in these genes. We therefore hypothesized that PR in *K. pneumoniae* is a complex genetic trait with previously unrecognized PR-conferring variants. We selected candidate novel determinants of PR, firstly by studying closely related pairs of clinical *K. pneumoniae* clonal group (CG) 258 isolates from our institution with differing polymyxin susceptibility. We then conducted a bacterial GWAS of PR using these genomes and publicly available *K. pneumoniae* CG258 genomes with polymyxin susceptibility data. CRISPR-Cas9 genome editing of PR candidate loci was used to experimentally validate that alterations in the *barA/uvrY* two-component system confer low-level PR (polymyxin minimum inhibitory concentrations of 4 mcg/ml). *BarA/uvrY* have been implicated in regulation of capsular polysaccharide production and we could not find evidence of lipid modifications conferring PR in isogenic *barA/uvrY* mutants. Our findings have important implications for development of future diagnostics and therapeutic agents for PR organisms. The precise mechanism of PR in *barA/uvrY* remains to be determined but may be related to adaptation to a urinary environment and likely implicates the capsule.

Importance

Polymyxins are treatments of last-resort for highly resistant Gram-negative bacteria. With the ongoing crisis in antimicrobial resistance, this class of antibiotics has been increasingly used leading to concerns about emerging polymyxin resistance. Despite this, our knowledge of the genetic mechanisms that cause polymyxin resistance remains limited. We therefore used whole genome sequencing data from *Klebsiella pneumonaie* clonal group 258, a major multi-drug resistant pathogen, to identify new genetic determinants of polymyxin resistance. Our findings

were validated through use of CRISPR-Cas9 genome editing techniques and showed polymyxin resistance conferred through the *barA/uvrY* two-component system that regulates capsular production. While alterations in the cell membrane have been the most common causes of polymyxin resistance, *barA/uvrY* did not appear to mediate their actions through this mechanism. This finding has important implications for development of future diagnostics and treatments for highly drug resistant Gram-negative organisms.

Introduction

Polymyxins (including colistin and polymyxin B) are antimicrobials used as treatments of 'last resort' for carbapenem-resistant Enterobacteriaceae (CRE) infections. In the context of increasing use of polymyxins over the last decade, polymyxin resistance (PR) has become an emerging concern (1). Prior epidemiological studies of polymyxin susceptibility have showed PR rates of 5 to >40% in CRE, with *Klebsiella pneumoniae* accounting for the majority (1). Furthermore, since 2015 plasmid-mediated forms of PR (*mcr* genes) have been noted, underscoring the importance of PR as a public health priority (2).

Despite the threat of losing this class of last-line antimicrobials to antimicrobial resistance, our knowledge of the mechanisms underlying PR in *Klebsiella pneumoniae* remains incomplete. Several key genes have been described in multiple previous studies (*crrAB*, *mgrB, phoPQ, pmrAB*), henceforth referred to as canonical PR genes (3-9). In addition to these, there are numerous other genes implicated in the literature that may constitute a 'secondary resistome' (1, 10). To add further complexity, a variety of genetic changes can lead to PR including non-synonymous single nucleotide variants (SNVs), insertion sequence (IS) disruption and large-scale gene deletions (1, 10). It remains unclear which genetic changes are sufficient to induce PR alone and which may provide an additive or synergistic effect to increase the polymyxin minimum inhibitory concentration (MIC). This remarkably polygenic trait poses significant challenges in the development of novel diagnostics and treatments.

Bacterial whole genome sequencing (WGS) now provides an unprecedented opportunity to address these questions through high-resolution methods to study numerous genetic loci concurrently. Additionally, the increasing availability of numerous publicly available bacterial genomes with associated metadata (including antimicrobial susceptibility testing results) allow application of advanced statistical techniques such as bacterial genome-wide association studies (GWAS) that may provide insights into previously unrecognized resistance determinants (11, 12). The advent of genome editing using CRISPR-Cas9 systems now also makes it possible to

functionally validate the contributions of candidate loci, but its use has been limited in extensively drug-resistant bacteria due to technical issues (13).

Our central hypothesis was that PR in *K. pneumoniae* is a complex genetic trait with previously unrecognized PR-conferring variants. We therefore aimed to detect novel determinants of PR, firstly by studying closely related pairs of clinical *K. pneumoniae* clonal group (CG) 258 isolates from our institution with differing polymyxin susceptibility. We then conducted a bacterial GWAS of PR using these genomes and publicly available *K. pneumoniae* CG258 genomes with polymyxin susceptibility data. Finally, we aimed to use CRISPR-Cas9 genome editing to experimentally validate the PR candidate loci identified through these approaches.

Results

Analysis of closely related K. pneumoniae CC258 isolate pairs suggest several novel polymyxin resistance candidate genes

In order to detect novel determinants of PR, we leveraged our collection of *K. pneumoniae* CC258 isolates and compared SNVs and IS elements in coding regions between pairs of closely related polymyxin-susceptible and PR genomes. These were from clinical isolates collected at our institution from 2011-2018 and included serial isolates from individual patients, as previously described (10). Pairwise SNV distances between genomes were calculated and each PR isolate was paired with its closest polymyxin-susceptible relative, resulting in 131 pairs with median SNV distance between them of 9 (IQR 3 - 23). Differences to the reference present in the PR but not the polymyxin susceptible genome were noted in order to identify candidate genes that may have conferred PR.

In our previous work, we noted a multitude of variants in PR canonical genes (10) and therefore hypothesized that other genes associated with PR would have a similar diversity of variants resulting in PR. We therefore ranked genes according to the number of unique variants present, the total number of PR isolate pairs with variants present and assessed their reported function

(see Table 1 for top ranked genes). The canonical PR genes *mgrB*, *crrAB, phoQ* and *pmrB* were in the ten highest ranked genes. We then identified other high-ranking genes that constituted potential targets for functional testing. These included *sbmA*, an inner membrane protein implicated in transport of antimicrobial peptides (14), and *barA*, which is linked to regulation of capsular polysaccharide (CPS) production (15). Of note, *barA* forms a two-component system with *uvrY*, which had also been identified with this strategy although at a lower frequency. Given the importance of two-component systems in conferring PR amongst canonical genes, *uvrY* was therefore also included as a potential candidate.

GWAS identifies similar candidate determinants of polymyxin resistance

In addition to our collection, we were able to identify a further 306 publicly available *K. pneumoniae* CC258 genomes that had polymyxin susceptibility testing with BMD (Supplementary Table 1). This allowed us to conduct a bacterial GWAS using a total of 619 genomes with PR as a binary outcome ('susceptible' versus 'resistant'), as implemented in the R package *treeWAS* (16). As an input, we created a binary matrix with genomes as rows and coding regions as columns, with presence of SNVs or IS elements in coding regions in the chromosome relative to the reference genome being recorded as a binary outcome. *TreeWAS* was run in 'simultaneous' mode in order to account for population structure.

mgrB and *crrB* were the only genes significantly associated with PR in our analysis (Table 2 outlines the highest ranked genes). However, our study may have been underpowered given our limited sample size and the complex polygenic nature of PR. In order to identify potential novel candidates of PR, we therefore ranked genes according to p-values (Table 2). *BarA,* the candidate identified through study of pairwise differences was also ranked in the top ten genes associated with PR.

Experimental validation of candidate genes using CRISPR-Cas9 confirms role of barA and uvrY in PR

On the basis of the genomics analyses, *barA*/*uvrY* and *sbmA* were selected for experimental validation using CRISPR-Cas9 gene editing. Knockout mutants were created by deleting a 130-150 bp segment close to the start of the coding sequence. The results of polymyxin susceptibility testing of mutants compared to wild type are summarized in Table 3. We noted increases in polymyxin B MIC to the resistant range for *barA* and *uvrY,* but not *sbmA*.

PR is associated with increased capsular polysaccharide production in uvrY but not barA knockout mutants

Previous data on the role of CPS in PR have been conflicting and did not directly assess the genes identified by our analyses (15, 17-19). To clarify whether CPS production was associated with PR, we extracted and quantified CPS as previously described (17). *BarA/uvrY* CRISPR-Cas9 knockout mutants were compared with the wild type isolate in triplicate. CPS was increased in the *uvrY* mutant (mean 47.8 mcg/ml vs 29.7 mcg/ml, *P*=0.0027) but we noted no change in the *barA* mutant (mean 29.2 mcg/ml vs 29.7 mcg/ml, *P*=0.81).

Increased virulence of barA/uvrY mutants in Galleria mellonella model

Given the central role of CPS in *Klebsiella* virulence, we also assessed the virulence of *barA/uvrY* knockout mutants in a *Galleria mellonella* model, as described previously (20). *BarA/uvrY* knockout mutants showed increased virulence compared to wild-type (Figure 1). We also tested *sbmA* but did not note a significant increase in virulence.

No lipid A modifications noted in polymyxin-resistant barA/uvrY mutants

Prior work with *barA* and *uvrY* in *E. coli* indicated that this two-component system may regulate LPS as well as CPS production (21). Given that LPS modification plays a central role in PR (1), we assessed whether there may be concurrent changes in LPS in *barA* and *uvrY* mutant strains that contribute to PR. We performed thin layer chromatography and compared wild type versus

184 *barA/uvrY* knockout mutant strains (Figure 2). No LPS modifications were noted in the mutants
185 compared with wild type.
186
187 *barA/uvrY variants are associated with culture from a urinary source and absence of polymyxin*
188 *exposure*
189 Having identified *barA*/*uvrY* as contributors to low-level PR, we correlated variants in these genes
190 with available clinical data for isolates from our institution. 19/40 (48%) isolates with *barA/uvrY*
191 variants were PR and 26/40 (65%) had concurrent variants in PR canonical genes. We noted
192 multiple variants in *barA/uvrY* for both isolates from our institution and publicly available
193 genomes, which are listed in Supplementary Table 2.
194
195 Prior literature suggested that *barA/uvrY* may play a role in adaptation to new environments and
196 also focused on work in uropathogenic *E. coli* (22, 23). We noted that 31/39 (80%) isolates with
197 *barA/uvrY* variants came from urinary sources, compared with 87/274 (32%) in the remainder of
198 the collection (P<0.0001). Our group has noted an association between polymyxin exposure and
199 emergence of mutations in canonical PR genes (10). However, variants in *barA/uvrY* were
200 associated with an absence of polymyxin exposure (8/39 *barA/uvrY* isolates exposed to
201 polymyxin vs 114/274 non-*barA/uvrY* isolates, P=0.019).
202
203 **Discussion**
204 Through employing extensive genomic analyses and GWAS, we identified several candidate
205 novel determinants of PR and functionally validated them using CRISPR-Cas9 gene editing. *BarA*
206 and *uvrY* both conferred low-level PR, approximately equivalent to that conferred by *mcr-1* (2).
207 These findings may indicate a novel pathway to PR as *barA/uvrY* impact capsular production, in
208 contrast to prior work which demonstrated that PR is mediated primarily through lipid A
209 modifications (1). We also determined that *barA/uvrY*-mediated PR may have a specific clinical
210 milieu, with *barA/uvrY* variants being noted in urinary isolates in the absence of polymyxin
211 exposure.

The current model of PR mechanisms in *K. pneumoniae* focuses on genetic changes in canonical PR genes (*mgrB*, *crrAB, pmrAB* and *phoPQ*) leading to lipid A changes, typically through the addition of 4-amino-4-deoxy-l-arabinose (L-Ara4N) or phosphethanolamine (pEtN). The finding that the *barA/uvrY* two-component system contributes to PR helps refine this model by highlighting that capsular changes may contribute to clinical PR. Capsular changes have been implicated in PR previously (17, 19, 24), but these were *in vitro* studies using non-clinical isolates with limited molecular characterization. Their relevance to clinical isolates therefore remains uncertain. The studies also showed conflicting results. One noted overproduction of capsule, which was thought to subsequently affect the interaction of polymyxins with the bacterial surface and thus their effectiveness (24), while another noted decreased capsular production in a hypermucoviscous *K. pneumoniae* ST23 background (17). The third study postulated the release of anionic capsular polysaccharides that acted as decoys and bound cationic antimicrobial peptides (including polymyxins) (19).

We quantified CPS production in order to assess whether increased production was responsible for PR and noted mixed results. *UvrY* mutants showed significantly higher production than the wild type but *barA* did not. This difference is potentially explained by the greater effect of *uvrY* on *csrB*, the regulator that ultimately mediates the actions of *barA/uvrY* (15). The fact that we found increased CPS production with a *uvrY* knockout is also in contrast to the mechanism hypothesized by Dorman et al., where *barA/uvrY* mutations were thought to lead to a “capsule down” phenotype (15). However, the methods used in that study did not quantify CPS but rather cell density and may therefore not be directly comparable.

In light of the extensive literature on the impact of *barA/uvrY* on capsular production and carbon metabolism (15, 21, 22, 25, 26), we continue to hypothesize that the PR observed is mediated through capsular changes but the exact mechanism remains unclear. Changes in the charge of CPS that affect the activity of the cationic polymyxins remain a possibility and will need to be

investigated further. For example, the decoy mechanism described above relies on the CPS released to be anionic in order to bind cationic antimicrobial peptide (19) and when cationic CPS was tested, there was no change in polymyxin susceptibility. Given that lipid A modifications were significantly implicated in other PR mechanisms, one concern was that *barA/uvrY* may have previously undetected actions on lipid A. However, we specifically tested for and did not find any evidence of lipid A alterations on thin layer chromatography in *barA/uvrY* mutants.

Taken together, these findings have important implications both for development of novel diagnostics and treatments for PR. The use of mass spectrometry has been proposed as a possible modality for detection of PR (27, 28), however our findings of a non-lipid A mechanism of PR show that this approach may have limitations if it focuses solely on lipid A. Similarly, novel treatments targeted at lipid A alterations through addition of L-Ara4N or pEtN would not necessarily have activity against *barA/uvrY*-mediated PR.

In addition to utilizing a different pathway to PR than previously described, isolates with *barA/uvrY* variants also had specific clinical characteristics. *BarA/uvrY* have previously been implicated as virulence factors in *E. coli* urinary tract infection by allowing switching between different carbon sources present in the urine (29). Indeed, in our collection there was a close association between isolates with *barA/uvrY* variants and isolation from a urinary source. While polymyxin exposure has been implicated in the emergence of PR (10, 30-34), this was not the case for isolates with *barA/uvrY* variants with no documented exposure to polymyxins for the majority of them. This suggests that PR could be arising as a by-product of adaptation to the urinary tract environment. Prior studies have noted that a proportion of isolates arise without exposure to polymyxins (35) and it is possible that *barA/uvrY* could be contributors in this setting.

In addition to our findings regarding *barA/uvrY*, there were several novel aspects to the methodology that allowed us to identify candidate novel determinants of PR. Firstly, we used phylogenetic techniques to estimate the ‘nearest neighbours’ for PR isolates. This allowed us to

leverage the entire collection in a setting where isolates may be from different patients but are likely to be closely related and may result from in-hospital transmission (10). Secondly, GWAS has not been applied to study PR previously and our findings add to the literature and highlight several of the issues associated with this approach. Finally, we used CRISPR-Cas9 for functional validation. Due to technical issues, this has been difficult to implement in extensively-drug resistant bacteria but recent work from our group and others indicate that this is now a feasible approach and may be a powerful way to confirm findings from GWAS (13).

Our study had several limitations. While there were multiple novel candidates identified through our screening approach, due to logistical constraints we were able to conduct testing only in highly selected candidates for this study and will continue testing in the future. In particular, *malT, cadC*, and *narL* are candidates worthy of future study. Conducting a GWAS for studying PR was problematic: our dataset ultimately incorporated >600 genomes, but this is orders of magnitude smaller than what is typically done in human GWAS. As a result, it was difficult to achieve sufficient power, especially in the context of PR being highly polygenic. We attempted to adjust for population structure using *treeWAS* but positive selection from antimicrobial administration continues to be a confounding factor (11). Finally, the true mechanism causing PR in *barA/uvrY* remains to be determined. We hypothesized capsular change as a likely mechanism based on prior work with *barA/uvrY* but had mixed results with capsule quantification assays. Nevertheless, we were able to exclude lipid A changes through our thin layer chromatography studies.

In summary, we were able to integrate bacterial WGS, GWAS and functional validation through CRISPR-Cas9 to identify *barA/uvrY* as a novel two-component system implicated in PR. This once again underscores the complexity of PR (36), especially given that the mechanism of action was not mediated through lipid A modification. Our findings have important implications for development of future diagnostics and therapeutic agents for PR organisms. The precise mechanism of PR in *barA/uvrY* remains to be determined but may be related to adaptation to a urinary environment and likely implicates the capsule.

Methods

Isolate selection

The study was reviewed and approved by the Columbia University Irving Medical Center (CUIMC) Institutional Review Board. CUIMC isolates were selected as previously described and included 313 *K. pneumoniae* CC258 isolates from 2011-2018 (10). All CUIMC isolates had MIC determination with BMD according to CLSI guidelines (37). WGS was performed on all included CUIMC isolates as described previously (10, 38-40). Isolates were matched to clinical data regarding anatomical source of isolate and polymyxin exposure collected during a prior study (10).

We also included 306 publicly available *K. pneumoniae* CC258 genomes identified by searching PubMed with the terms “colistin resistance” and “polymyxin resistance”, and also the CDC/FDA Antibiotic Resistance Isolate Bank and PATRIC databases (41, 42). Genome sequence data from NCBI was obtained, with raw sequence data being collected in preference to draft assemblies. Only genomes with BMD phenotypic susceptibility testing were included. Genome details including accession numbers, source datasets and results of phenotypic susceptibility testing are included in Supplementary Table 1.

Dataset preparation

We constructed draft *de novo* assemblies of all CUIMC isolates and publicly available genomes with raw sequence data using the Shovill wrapper for SPAdes, which also utilizes Trimmomatic for read trimming and Pilon for read error correction (43-46). Kleborate was then used for MLST, resistance determinant and virulence factor detection (10). A representative polymyxin susceptible CG258 isolate was selected and used to create a *de novo* hybrid long-read/short-read assembly as described previously (10, 47). This was used as the reference to create a profile of each genome through a combination of variant calling and IS detection. In brief, variant calling was performed using Snippy 3.2 and running Snippy in the ‘-- contig' mode if only an assembly was available (48). ISseeker was used to identify sites with IS elements present (49). In

order to increase sensitivity of detection for key genes, we also used a BLAST database of PR canonical genes to identify if IS disruption or large scale-deletions of these genes may have occurred.

Detection of novel resistance determinants

We used two approaches for detection of novel resistance determinants. For the first approach we used CUIMC genomes to conduct pairwise comparisons between PR isolates and their polymyxin susceptible 'nearest neighbours' (most closely related isolates). This was done by constructing a maximum likelihood phylogeny of these isolates, as described previously (10) and calculating pairwise chromosomal SNV distances to match PR genomes with their closest polymyxin-susceptible neighbours. We conducted a comparison between the PR and polymyxin susceptible isolates to detect differences in SNVs and IS insertions in the bacterial chromosome. We then ranked each gene with variants present according to the total number of unique variants present and the total number of isolate pairs with variants present.

The second approach involved conducting a bacterial GWAS utilizing CUIMC genomes and publicly available genomes. We used variant calling/IS detection as described above to identify SNVs/IS in coding regions relative to the reference genome. Synonymous SNVs in coding regions, intergenic regions and SNVs in known MGEs and repeat regions were not considered. Each of these alleles in each isolate was scored as 1 if it differed from the reference and 0 if it did not. These allele scores were used to create a binary matrix with isolates as rows and coding regions in the reference as columns, which was then used as an input in *treeWAS*. *TreeWAS* is a tool for conducting bacterial GWAS implemented in the R statistical language (16). *TreeWAS* was run in 'simultaneous' mode, which allows input of user-generated phylogenies in order to account for possible bias that arises due to population structure. We therefore constructed a second maximum likelihood phylogeny using all available isolates, as described above and used this as an input. We adjusted for multiple testing by defining a False Discovery Rate of 0.05, as implemented in *treeWAS*. Genes were ranked according to p-values generated by the analysis.

352

353 *Genome editing*

354 For each gene deletion we generated mutation specific CRISPR plasmids. Wild type sequences
355 were analyzed via the CRISPRdirect website to identify appropriate N20 sequences. For gene
356 knockouts (*barA, uvrY, sbmA*) the homologies were engineered to contain a 130-150 bp deletion
357 surrounding the cas9 cut site (50-52). Sequence-confirmed plasmids were inserted into the
358 isolate of interest through electroporation. Transformants were grown at 30°C under zeocin
359 selection and induced with 2% arabinose after 2 hours. Following 6+ hours of induction the
360 cultures were diluted 1:100 and plated on low salt LB + zeocin and arabinose (50-52).
361 Appropriate mutants were identified with colony PCR and verified by Sanger sequencing
362 (Genewiz). Mutants were cured of the CRISPR plasmid with serial passage on non-selective
363 media. To ensure no off-target editing occurred we performed WGS (Illumina MiSeq) as
364 previously described (38).

365

366 *Capsular assays*

367 We performed relative capsule analysis by extracting polysaccharide from mid log phase cultures
368 as described previously (17). Briefly, selected isolates were grown to an OD_{600} of 1.0 and 2 ml of
369 each culture was pelleted. Bacterial pellets were washed five times in 1 ml of 50mM NaCl. Each
370 pellet was then resuspended in 50mM EDTA and incubated at 37°C for one hour. 200 mcl of the
371 supernatant was collected and mixed with 200 mcl of 5% phenol and 1 ml of sulfuric acid. A
372 colorimetric reaction was then allowed to develop for 10 minutes at room temperature and the
373 OD_{490} was read. A standard curve was made with known concentrations of a 1:1 fructose /
374 sucrose solution. Absolute amounts of polysaccharide were defined based on where the optical
375 density read met the standardized line.

376

377

Thin layer chromatography

We performed lipid A analysis with thin layer chromatography. Isolates were grown with ^{32}P labelling and lipid was subsequently extracted utilizing the Caroff and Raetz approach of Bligh-Dyer Extractions and mild acid hydrolysis (53, 54).

Galleria mellonella killing assay

Wild type NR5452 and isogenic mutants were grown in LB broth to mid/late log phase (OD_{600} of 1.0). Each isolate was pelleted and re-suspended in an equal volume of PBS and adjusted to a final OD of 1.0, corresponding approximately to 5×10^8 CFUs/ml. 10 mcl of this resuspension was injected into the pro-leg of 20 *G. mellonella* larvae per isolate (Vanderhost Inc.) (55, 56). Serial dilutions of each isolate were plated to calculate CFUs injected to ensure comparable inocula between isolates. We monitored for larval mortality at 24, 48 and 72 hours by tactile stimulus (55). For each experiment we injected 20 larvae with PBS as a control. Mortality rates were visualized with Kaplan-Meier survival curves and significance was calculated using log rank test with a Bonferroni correction (56).

Statistical analysis

Categorical variables were compared using $\chi 2$ or Fisher's exact tests and continuous variables were compared using Student's *t*-test or Mann-Whitney-Wilcoxon, as appropriate. Statistical analyses were performed in R (v3.4.0).

Financial support

This work was supported by the Australian National Health and Medical Research Council Postgraduate Scholarship (APP1169514) to N.M. and Practitioner Fellowship to A.Y.P (APP1117940), National Institute of Allergy and Infectious Diseases, National Institutes of Health (grant number AI116939 to A-C. U.).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Potential conflicts of interest

NM has received research support from GlaxoSmithKline, unrelated to the current study. A.Y.P has received research support from MSD, unrelated to the current study. A-C.U. has received research funding from Merck, GlaxoSmithKline and Allergan, unrelated to the current study. All other authors declare no conflict of interest.

413 **References**

414 1. Poirel L, Jayol A, Nordmann P. 2017. Polymyxins: Antibacterial Activity, Susceptibility
415 Testing, and Resistance Mechanisms Encoded by Plasmids or Chromosomes. Clin
416 Microbiol Rev 30:557-596.
417 2. Feb 2016. 2. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in
418 animals and human beings in China: a microbiological and molecular biological study.
419 The Lancet Infectious Diseases, 16.161-168.
420 http://linkinghub.elsevier.com/retrieve/pii/S1473309915004247.
421 3. Aires CA, Pereira PS, Asensi MD, Carvalho-Assef AP. 2016. mgrB Mutations Mediating
422 Polymyxin B Resistance in Klebsiella pneumoniae Isolates from Rectal Surveillance
423 Swabs in Brazil. Antimicrob Agents Chemother 60:6969-6972.
424 4. Berglund B, Hoang NTB, Tarnberg M, Le NK, Svartstrom O, Khu DTK, Nilsson M, Le HT,
425 Welander J, Olson L, Larsson M, Nilsson LE, Hanberger H. 2018. Insertion sequence
426 transpositions and point mutations in mgrB causing colistin resistance in a clinical strain
427 of carbapenem-resistant Klebsiella pneumoniae from Vietnam. Int J Antimicrob Agents
428 51:789-793.
429 5. Cannatelli A, Di Pilato V, Giani T, Arena F, Ambretti S, Gaibani P, D'Andrea MM,
430 Rossolini GM. 2014. In vivo evolution to colistin resistance by PmrB sensor kinase
431 mutation in KPC-producing Klebsiella pneumoniae is associated with low-dosage colistin
432 treatment. Antimicrob Agents Chemother 58:4399-403.
433 6. Cannatelli A, Giani T, D'Andrea MM, Di Pilato V, Arena F, Conte V, Tryfinopoulou K,
434 Vatopoulos A, Rossolini GM, Group CS. 2014. MgrB inactivation is a common
435 mechanism of colistin resistance in KPC-producing Klebsiella pneumoniae of clinical
436 origin. Antimicrob Agents Chemother 58:5696-703.
437 7. Cheng YH, Lin TL, Pan YJ, Wang YP, Lin YT, Wang JT. 2015. Colistin resistance
438 mechanisms in Klebsiella pneumoniae strains from Taiwan. Antimicrob Agents
439 Chemother 59:2909-13.

8. Pitt ME, Elliott AG, Cao MD, Ganesamoorthy D, Karaiskos I, Giamarellou H, Abboud CS, Blaskovich MAT, Cooper MA, Coin LJM. 2018. Multifactorial chromosomal variants regulate polymyxin resistance in extensively drug-resistant Klebsiella pneumoniae. Microb Genom 4.
9. Wright MS, Suzuki Y, Jones MB, Marshall SH, Rudin SD, van Duin D, Kaye K, Jacobs MR, Bonomo RA, Adams MD. 2015. Genomic and transcriptomic analyses of colistin-resistant clinical isolates of Klebsiella pneumoniae reveal multiple pathways of resistance. Antimicrob Agents Chemother 59:536-43.
10. Macesic N, Nelson B, McConville TH, Giddins MJ, Green DA, Stump S, Gomez-Simmonds A, Annavajhala MK, Uhlemann AC. 2019. Emergence of Polymyxin Resistance in Clinical Klebsiella pneumoniae Through Diverse Genetic Adaptations: A Genomic, Retrospective Cohort Study. Clin Infect Dis doi:10.1093/cid/ciz623.
11. Power RA, Parkhill J, de Oliveira T. 2017. Microbial genome-wide association studies: lessons from human GWAS. Nature reviews Genetics 18:41-50.
12. Read TD, Massey RC. 2014. Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. Genome Medicine 6:1063-11.
13. Sun Q, Wang Y, Dong N, Shen L, Zhou H, Hu Y, Gu D, Chen S, Zhang R, Ji Q. 2019. Application of CRISPR/Cas9-Based Genome Editing in Studying the Mechanism of Pandrug Resistance in Klebsiella pneumoniae. Antimicrob Agents Chemother 63.
14. Runti G, Lopez Ruiz Mdel C, Stoilova T, Hussain R, Jennions M, Choudhury HG, Benincasa M, Gennaro R, Beis K, Scocchi M. 2013. Functional characterization of SbmA, a bacterial inner membrane transporter required for importing the antimicrobial peptide Bac7(1-35). J Bacteriol 195:5343-51.
15. Dorman MJ, Feltwell T, Goulding DA, Parkhill J, Short FL. 2018. The Capsule Regulatory Network of Klebsiella pneumoniae Defined by density-TraDISort. MBio 9.

16. Collins C, Didelot X. 2018. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comput Biol 14:e1005958.
17. Choi MJ, Ko KS. 2015. Loss of hypermucoviscosity and increased fitness cost in colistin-resistant Klebsiella pneumoniae sequence type 23 strains. Antimicrob Agents Chemother 59:6763-73.
18. Llobet E, Campos MA, Gimenez P, Moranta D, Bengoechea JA. 2011. Analysis of the networks controlling the antimicrobial-peptide-dependent induction of Klebsiella pneumoniae virulence factors. Infect Immun 79:3718-32.
19. Llobet E, Tomas JM, Bengoechea JA. 2008. Capsule polysaccharide is a bacterial decoy for antimicrobial peptides. Microbiology 154:3877-86.
20. Peleg AY, Jara S, Monga D, Eliopoulos GM, Moellering RC, Jr., Mylonakis E. 2009. Galleria mellonella as a model system to study Acinetobacter baumannii pathogenesis and therapeutics. Antimicrob Agents Chemother 53:2605-9.
21. Palaniyandi S, Mitra A, Herren CD, Lockatell CV, Johnson DE, Zhu X, Mukhopadhyay S. 2012. BarA-UvrY two-component system regulates virulence of uropathogenic E. coli CFT073. PLoS One 7:e31348.
22. Pernestig AK, Georgellis D, Romeo T, Suzuki K, Tomenius H, Normark S, Melefors O. 2003. The Escherichia coli BarA-UvrY two-component system is needed for efficient switching between glycolytic and gluconeogenic carbon sources. J Bacteriol 185:843-53.
23. Zhang JP, Normark S. 1996. Induction of gene expression in Escherichia coli after pilus-mediated adherence. Science 273:1234-6.
24. Campos MA, Vargas MA, Regueiro V, Llompart CM, Alberti S, Bengoechea JA. 2004. Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. Infect Immun 72:7107-14.
25. Bhagirath AY, Li Y, Patidar R, Yerex K, Ma X, Kumar A, Duan K. 2019. Two Component Regulatory Systems and Antibiotic Resistance in Gram-Negative Pathogens. Int J Mol Sci 20.

494 26. Suzuki K, Wang X, Weilbacher T, Pernestig AK, Melefors O, Georgellis D, Babitzke P,
495 Romeo T. 2002. Regulatory circuitry of the CsrA/CsrB and BarA/UvrY systems of
496 Escherichia coli. J Bacteriol 184:5130-40.
497 27. Dortet L, Bonnin RA, Pennisi I, Gauthier L, Jousset AB, Dabos L, Furniss RCD, Mavridou
498 DAI, Bogaerts P, Glupczynski Y, Potron A, Plesiat P, Beyrouthy R, Robin F, Bonnet R,
499 Naas T, Filloux A, Larrouy-Maumus G. 2018. Rapid detection and discrimination of
500 chromosome- and MCR-plasmid-mediated resistance to polymyxins by MALDI-TOF MS
501 in Escherichia coli: the MALDIxin test. J Antimicrob Chemother 73:3359-3367.
502 28. Larrouy-Maumus G, Clements A, Filloux A, McCarthy RR, Mostowy S. 2016. Direct
503 detection of lipid A on intact Gram-negative bacteria by MALDI-TOF mass spectrometry.
504 J Microbiol Methods 120:68-71.
505 29. Tomenius H, Pernestig AK, Jonas K, Georgellis D, Mollby R, Normark S, Melefors O.
506 2006. The Escherichia coli BarA-UvrY two-component system is a virulence determinant
507 in the urinary tract. BMC Microbiol 6:27.
508 30. Kontopidou F, Plachouras D, Papadomichelakis E, Koukos G, Galani I, Poulakou G,
509 Dimopoulos G, Antoniadou A, Armaganidis A, Giamarellou H. 2011. Colonization and
510 infection by colistin-resistant Gram-negative bacteria in a cohort of critically ill patients.
511 Clin Microbiol Infect 17:E9-E11.
512 31. Falcone M, Russo A, Iacovelli A, Restuccia G, Ceccarelli G, Giordano A, Farcomeni A,
513 Morelli A, Venditti M. 2016. Predictors of outcome in ICU patients with septic shock
514 caused by Klebsiella pneumoniae carbapenemase-producing K. pneumoniae. Clin
515 Microbiol Infect 22:444-50.
516 32. van Duin D, Doi Y. 2015. Outbreak of Colistin-Resistant, Carbapenemase-Producing
517 Klebsiella pneumoniae: Are We at the End of the Road? J Clin Microbiol 53:3116-7.
518 33. Giacobbe DR, Del Bono V, Trecarichi EM, De Rosa FG, Giannella M, Bassetti M,
519 Bartoloni A, Losito AR, Corcione S, Bartoletti M, Mantengoli E, Saffioti C, Pagani N,
520 Tedeschi S, Spanu T, Rossolini GM, Marchese A, Ambretti S, Cauda R, Viale P, Viscoli
521 C, Tumbarello M, Isgri S. 2015. Risk factors for bloodstream infections due to colistin-

resistant KPC-producing Klebsiella pneumoniae: results from a multicenter case-control-control study. Clin Microbiol Infect 21:1106 e1-8.

34. Giamarellou H. 2016. Epidemiology of infections caused by polymyxin-resistant pathogens. Int J Antimicrob Agents 48:614-621.

35. Rojas LJ, Salim M, Cober E, Richter SS, Perez F, Salata RA, Kalayjian RC, Watkins RR, Marshall S, Rudin SD, Domitrovic TN, Hujer AM, Hujer KM, Doi Y, Kaye KS, Evans S, Fowler VG, Jr., Bonomo RA, van Duin D, Antibacterial Resistance Leadership G. 2017. Colistin Resistance in Carbapenem-Resistant Klebsiella pneumoniae: Laboratory Detection and Impact on Mortality. Clin Infect Dis 64:711-718.

36. Doi Y, Van Duin D. 2019. Polymyxin Resistance in Klebsiella pneumoniae: Complexity at Every Level. Clin Infect Dis doi:10.1093/cid/ciz627.

37. Institute CaLS. 2015. Performance standards for antimicrobial susceptibility testing. , Twenty-fifth informational supplement, Clinical and Laboratory Standards Institute, Wayne, PA, 2015, M100-S25.

38. Giddins MJ, Macesic N, Annavajhala MK, Stump S, Khan S, McConville TH, Mehta M, Gomez-Simmonds A, Uhlemann AC. 2018. Successive Emergence of Ceftazidime-Avibactam Resistance through Distinct Genomic Adaptations in blaKPC-2-Harboring Klebsiella pneumoniae Sequence Type 307 Isolates. Antimicrob Agents Chemother 62.

39. Gomez-Simmonds A, Annavajhala MK, Wang Z, Macesic N, Hu Y, Giddins MJ, O'Malley A, Toussaint NC, Whittier S, Torres VJ, Uhlemann AC. 2018. Genomic and Geographic Context for the Evolution of High-Risk Carbapenem-Resistant Enterobacter cloacae Complex Clones ST171 and ST78. MBio 9.

40. Macesic N, Gomez-Simmonds A, Sullivan SB, Giddins MJ, Ferguson SA, Korakavi G, Leeds D, Park S, Shim K, Sowash MG, Hofbauer M, Finkel R, Hu Y, West J, Toussaint NC, Greendyke WG, Miko BA, Pereira MR, Whittier S, Verna EC, Uhlemann AC. 2018. Genomic Surveillance Reveals Diversity of Multidrug-Resistant Organism Colonization and Infection: A Prospective Cohort Study in Liver Transplant Recipients. Clin Infect Dis 67:905-912.

41. Centers for Disease Control and Prevention. 2019. CDC & FDA Antibiotic Resistance (AR) Isolate Bank. https://www.cdc.gov/drugresistance/resistance-bank/index.html. Accessed August 15, 2019.

42. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res 45:D535-D542.

43. Seemann T. 2018. Shovill: Faster SPAdes (or better SKESA/Megahit/Velvet) assembly of Illumina reads. https://github.com/tseemann/shovill. Accessed July 23, 2018.

44. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-77.

45. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-20.

46. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963.

47. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

48. Seemann T. 2018. Snippy: Rapid haploid variant calling and core SNP phylogeny. https://github.com/tseemann/snippy. Accessed July 23, 2018.

49. Adams MD, Bishop B, Wright MS. 2016. Quantitative assessment of insertion sequence impact on bacterial genome architecture. Microb Genom 2:e000062.

50. Zhao D, Yuan S, Xiong B, Sun H, Ye L, Li J, Zhang X, Bi C. 2016. Development of a fast and easy method for Escherichia coli genome editing with CRISPR/Cas9. Microb Cell Fact 15:205.
51. Jiang Y, Chen B, Duan C, Sun B, Yang J, Yang S. 2015. Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. Appl Environ Microbiol 81:2506-14.
52. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31:233-9.
53. Henderson JC, O'Brien JP, Brodbelt JS, Trent MS. 2013. Isolation and chemical characterization of lipid A from gram-negative bacteria. J Vis Exp doi:10.3791/50623:e50623.
54. Hankins JV, Madsen JA, Needham BD, Brodbelt JS, Trent MS. 2013. The outer membrane of Gram-negative bacteria: lipid A isolation and characterization. Methods Mol Biol 966:239-258.
55. Insua JL, Llobet E, Moranta D, Perez-Gutierrez C, Tomas A, Garmendia J, Bengoechea JA. 2013. Modeling Klebsiella pneumoniae pathogenesis by infection of the wax moth Galleria mellonella. Infect Immun 81:3552-65.
56. Kidd TJ, Mills G, Sa-Pessoa J, Dumigan A, Frank CG, Insua JL, Ingram R, Hobley L, Bengoechea JA. 2017. A Klebsiella pneumoniae antibiotic resistance mechanism that subdues host defences and promotes virulence. EMBO Mol Med 9:430-447.
57. Boos W, Shuman H. 1998. Maltose/maltodextrin system of Escherichia coli: transport, metabolism, and regulation. Microbiol Mol Biol Rev 62:204-29.
58. Cheng YH, Lin TL, Lin YT, Wang JT. 2018. A putative RND-type efflux pump, H239_3064, contributes to colistin resistance through CrrB in Klebsiella pneumoniae. J Antimicrob Chemother 73:1509-1516.
59. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL, Lam MMC, Duchene S, Jenney A, Holt KE. 2019. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of Klebsiella pneumoniae. PLoS Genet 15:e1008114.

60. Hsieh PF, Lin HH, Lin TL, Wang JT. 2010. CadC regulates cad and tdc operons in response to gastrointestinal stresses and enhances intestinal colonization of Klebsiella pneumoniae. J Infect Dis 202:52-64.
61. Lee YH, Kim JH, Bang IS, Park YK. 2008. The membrane-bound transcriptional regulator CadC is activated by proteolytic cleavage in response to acid stress. J Bacteriol 190:5120-6.
62. Adler M, Anjum M, Andersson DI, Sandegren L. 2016. Combinations of mutations in envZ, ftsI, mrdA, acrB and acrR can cause high-level carbapenem resistance in Escherichia coli. J Antimicrob Chemother 71:1188-98.

615 **Tables**

616 **Table 1 – Genes with variants present in polymyxin resistant isolates but not their**

617 **polymyxin susceptible nearest neighbours**

618

Annotation	Total number of unique variants	Total number of isolate pairs	Function and comments	Reference(s)
mgrB	21	65	Known determinant of PR	
crrB	13	22	Known determinant of PR	
phoQ	8	16	Known determinant of PR	
sbmA	8	15	Inner membrane protein involved in the internalization of glycopeptides and antimicrobial peptides	(14)
malT	7	10	Essential for transcription of all *mal* genes except *malI/X/Y* gene cluster. Overexpression of *phoPQ* downregulates *mal* gene expression.	(57)
barA	6	10	Regulation of capsular polysaccharide production	(15)
pmrB	6	7	Known determinant of PR	
H239_3064	5	16	Encodes putative RND-type efflux pump and newly discovered determinant of PR	(58)
galF	5	15	Part of *cps* cluster and mediates capsular polysaccharide production	(15, 59)
crrA	5	9	Known determinant of PR	

619 Abbreviations: PR – polymyxin resistance

620

621 **Table 2 – Results of genome-wide association study of polymyxin resistance**
622

Annotation	*P*	Function	Reference(s)
mgrB	0	Known determinant of PR	
crrB	0	Known determinant of PR	
phoQ	3.28×10^{-5}	Known determinant of PR	
hypothetical protein	1.64×10^{-4}	Unknown	
cadC	1.64×10^{-4}	Involved in transcriptional response to acidic pH	(60, 61)
narL	5.24×10^{-4}	Nitrate sensing and respiration, biofilm formation and motility	(25)
barA	5.24×10^{-4}	Regulation of capsular polysaccharide production	(15)
wza	0.00128	Part of *cps* cluster and mediates capsular polysaccharide production	(15, 59)
malT	0.00128	Essential for transcription of all *mal* genes except *malI/X/Y* gene cluster. Overexpression of *phoPQ* downregulates *mal* gene expression	(57)
mrdA	0.00128	Peptidoglycan biosynthesis	(62)

623 Abbreviations: PR – polymyxin resistance

624

625 **Table 3 – Polymyxin minimum inhibitory concentrations of isogenic mutants of candidate**

626 **loci for polymyxin resistance**

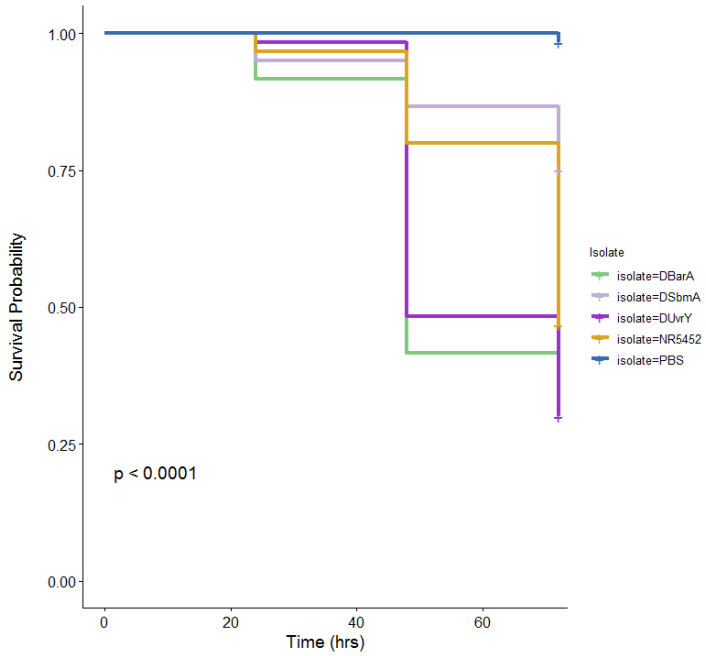
Strain	Polymyxin MIC (mcg/ml)
NR5452 wild type	2
NR5452 Δ*barA*	4
NR5452 Δ*uvrY*	4
NR5452 Δ*sbmA*	2
NR5452 Δ*mgrB*	32
Escherichia coli ATCC 25922	1

627 Abbreviations: MIC – minimum inhibitory concentration

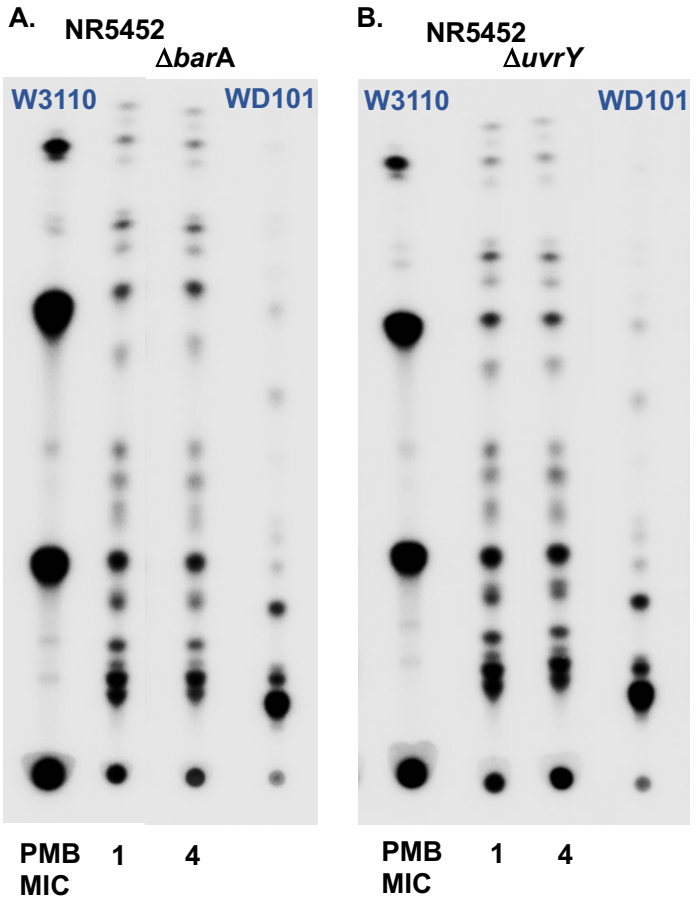
628

Figures

Figure 1 - Kaplan-Meier plots of *Galleria mellonella* survival following infection with isogenic mutants of polymyxin resistance candidate loci



634 **Figure 2 – Thin layer chromatography of *barA/uvrY* isogenic mutants**



635

636 Thin layer chromatography did not show any significant LPS modifications between NR5452 (the

637 wild type background) and the Δ*barA*/Δ*uvrY* mutants. W3110 and WD101 represent *E. coli* and *K.*

638 *pneumoniae* controls, respectively.

Conclusion

In this chapter, I was able to integrate bacterial WGS, GWAS and functional validation through CRISPR-Cas9 to identify a novel two-component system implicated in PR. This work serves as an important proof-of-concept of integrating genomic data, advanced bioinformatics analyses and cutting-edge genome editing techniques to generate and validate hypotheses. The genomic dataset developed for GWAS in this chapter will be used for AST genotype-phenotype prediction of PR in Chapter 6.

The findings from this chapter also challenge the current model of PR, which emphasizes the role of lipid A modifications. Given the importance of the capsule in *K. pneumoniae*, it is perhaps surprising that changes in capsular production have not been implicated in PR in more recent literature. Nevertheless, the precise mechanism of PR in *barA/uvrY* remains to be determined and will be the subject of future work.

Chapter 6:

Predicting phenotypic polymyxin resistance in *Klebsiella pneumoniae* through machine learning analysis of genomic data

Work from this chapter has been submitted:

1) Macesic N, Bear Don't Walk IV, O.J., Pe'er, I. et al. Predicting phenotypic polymyxin resistance in *Klebsiella pneumoniae* through machine learning analysis of genomic data. *mSystems* [returned for revision] [5]

The following contributions were made by specific co-authors to the work. No co-authors were Monash University students.

Nature of contribution	Co-authors
Writing machine learning code	Bear Don't Walk IV OJ
Manuscript writing	Bear Don't Walk IV OJ, Uhlemann AC
Reviewing manuscript	Pe'er I, Tatonetti NP, Peleg AY, Uhlemann AC
Concept	Uhlemann AC

Introduction

In the previous two chapters, I have focused on defining the genomic determinants of PR, firstly by focusing on known determinants (*crrAB*, *mgrB, phoPQ, pmrAB*) [109, 113-115, 117, 123, 124], which I termed canonical genes, then by detecting novel determinants. The work in these chapters highlights the incredible diversity of pathways leading to PR. This makes PR a challenging polygenic trait to diagnose and hampers efforts at rapid molecular detection of PR.

Nevertheless there is an urgent need for improved diagnosis of PR. The gold standard for phenotypic polymyxin susceptibility testing is broth microdilution. However, it is resource intensive and difficult to perform accurately [302], which prohibit its use in most clinical microbiology laboratories. With increasing availability of bacterial WGS data, there has been active investigation into using these data for genotype-phenotype prediction of AST. The rich dataset of genomic data matched with polymyxin phenotypic susceptibility data used to conduct GWAS in Chapter 5 represents a perfect resource for AST genotype-phenotype prediction of PR. Given the incomplete identification of contributing PR mutations and the possible polygenic nature of PR, I hypothesized that ML approaches may be well suited to AST genotype-phenotype prediction in this setting and could ultimately be used to help identify isolates for confirmatory phenotypic testing.

Manuscript

Predicting phenotypic polymyxin resistance in *Klebsiella pneumoniae* through machine learning analysis of genomic data

Nenad Macesic[1, 2], Oliver J. Bear Don't Walk IV[3], Itsik Pe'er[4], Nicholas P. Tatonetti[3], Anton Y. Peleg[2,5], Anne-Catrin Uhlemann[1,6]

Affiliations:

[1] Division of Infectious Diseases, Columbia University Irving Medical Center, New York, NY, United States

[2] Department of Infectious Diseases, The Alfred Hospital and Central Clinical School, Monash University, Melbourne, Australia

[3] Department of Biomedical Informatics, Columbia University, New York, NY, United States

[4] Department of Computer Science, Columbia University, New York, NY, United States

[5] Infection and Immunity Program, Monash Biomedicine Discovery Institute, Department of Microbiology, Monash University, Clayton, VIC, Australia

[6] Microbiome & Pathogen Genomics Core, Columbia University Irving Medical Center, New York, NY, USA

*Corresponding author: A-C. Uhlemann, 630 West 168th Street, New York, NY 10032, au2110@columbia.edu, phone 212-342-4530

Alternate corresponding author: N. Macesic, 630 West 168th Street, New York, NY 10032, nenad.macesic@monash.edu, phone 212-305-3374

Keywords:
Genotype; phenotype; prediction; antimicrobial resistance; machine learning

Running title:
Machine learning prediction of polymyxin resistance

Words: Abstract - 233; Importance – 146; Text – 5054;
Tables – 3, Figures – 3, References – 75, Supplementary Tables – 6

Abstract

Polymyxins are used as treatments of last resort for Gram-negative bacterial infections. Their increased use has led to concerns about emerging polymyxin resistance (PR). Phenotypic polymyxin susceptibility testing is resource-intensive and difficult to perform accurately. The complex polygenic nature of PR and our incomplete understanding of its genetic basis make it difficult to predict PR using detection of resistance determinants. We therefore applied machine learning (ML) on whole genome sequencing data from >600 *Klebsiella pneumoniae* clonal group 258 genomes to predict phenotypic PR. Using a reference-based representation of genomic data with ML outperformed a rule-based approach that detected variants in known PR genes (area under receiver operator curve [AUROC] 0.894 vs 0.791, *P*=0.006). We noted modest increases in performance by using a bacterial genome-wide association study to filter relevant genomic features and by integrating clinical data in the form of prior polymyxin exposure. Conversely, reference-free representation of genomic data as k-mers was associated with decreased performance (AUROC 0.692 vs 0.894, *P*=0.015). When ML models were interpreted to extract genomic features, 6/7 known PR genes were correctly identified by models without prior programming and several genes involved in stress responses and maintenance of the cell membrane were identified as potential novel determinants of PR. These findings are a proof-of-concept that whole genome sequencing data can accurately predict PR in *K. pneumoniae* CG258 and may be applicable to other forms of complex antimicrobial resistance.

Importance

Polymyxins are last-resort antibiotics used to treat highly resistant Gram-negative bacteria. There are increasing reports of polymyxin resistance emerging, raising concerns of a post-antibiotic era. Detecting polymyxin resistance is therefore a significant public health threat but current phenotypic methods are difficult and time-consuming to perform. There have been increasing efforts to use whole genome sequencing for detection of antibiotic resistance, but this has been difficult to apply to polymyxin resistance because of its complex polygenic nature. The significance of our research is that we successfully applied machine learning methods to predict

63 polymyxin resistance in *Klebsiella pneumoniae* clonal group 258, a common healthcare-
64 associated and multi-drug resistant pathogen. Our findings highlight that machine learning can be
65 successfully applied even in complex forms of antibiotic resistance and represent a significant
66 contribution to the literature that could be used to predict resistance in other bacteria and to other
67 antibiotics.
68

Introduction

Carbapenem resistance in Enterobacterales (CRE) is a global health challenge that threatens many medical advances. Due to the lack of antimicrobial options for treating CRE infections, polymyxins (including colistin and polymyxin B) have been revived despite their toxicity and are widely used as treatments of last resort (1). Polymyxin resistance (PR) has now become a growing concern and may critically impair our ability to combat CRE infections. Epidemiological studies indicated PR rates ranging from 5 to >40% in CRE, with *Klebsiella pneumoniae* accounting for the majority (1).

While improving the diagnosis and treatment of PR infections is an urgent priority, there are several challenges. Phenotypic polymyxin susceptibility testing is resource intensive and difficult to perform accurately (2). Broth microdilution (BMD) testing is recommended by the CLSI and EUCAST but is often only available in reference laboratories. Many clinical laboratories have traditionally relied on gradient diffusion methods such as Etest, but significant concerns about the accuracy of these methods have been raised (3, 4). Our understanding of the genetic basis of PR also remains limited. Several important genetic loci of PR have been identified in multiple studies (*crrAB*, *mgrB, phoPQ, pmrAB*) (5-11) and will henceforth be referred to as PR canonical genes. However, PR is noted in isolates not carrying mutations or disruptions in these canonical genes and may result from mutations in multiple genes (12). Conversely, some susceptible isolates have mutations in canonical genes but do not exhibit PR, raising the possibility of compensatory mutations. This makes PR a challenging polygenic trait to diagnose and hampers efforts at rapid molecular detection of PR.

With increasing availability of bacterial whole genome sequencing (WGS) data, there has been active investigation into using these data for genotype-phenotype prediction of antimicrobial susceptibility testing (AST). This was initially in the form of rule-based approaches that would predict susceptibility through detection of known resistance determinants (eg. beta-lactamase genes) or known resistance mutations in housekeeping genes (eg. *rpoB* conferring rifampin

resistance in *Staphylococcus aureus*) (13). The performance of these approaches has varied depending on the organism and antimicrobial tested but multiple limitations remain. Firstly, the approach assumes all determinants of resistance are known and is therefore unable to detect previously uncharacterized determinants. Secondly, these rule-based models struggle to account for complex interactions between variants in multiple loci. In order to move beyond these limitations, machine learning (ML) methods have been used to predict antimicrobial susceptibility (14-17). Given the incomplete identification of contributing PR mutations and the possible polygenic nature of PR, we hypothesize that ML approaches may be well suited to AST genotype-phenotype prediction in this setting (18) and could ultimately be used to help identify isolates for confirmatory phenotypic testing.

We therefore aimed to use ML for genotype-phenotype prediction of PR in *K. pneumoniae* CG258 as this remains the main CRE clone in North America and a key CRE clone globally (19). We aimed to do this using both a reference-based approach that relied on variant calling and insertion sequence (IS) detection, as well as a reference-free approach using detection of k-mers. We compared these ML approaches to a simple rule-based model that relies on detection of variants in canonical PR genes. Finally, we hypothesized that the ML models could be interpreted to confirm biological plausibility and elucidate novel genetic determinants of PR.

Results:

We analyzed 619 previously sequenced *K. pneumoniae* clonal group (CG) 258 genomes (10-12, 20-25). To examine the impact of groups of input genomes on model performance, we conducted analyses on three different datasets: firstly, on all available genomes (193/619 genomes with PR; 31%), then subsets of genomes according to their origins (Columbia University Irving Medical Center [CUIMC] or non-CUIMC, 138/313 [44%] and 55/306 [22%] genomes with PR, respectively). This was done due to differing degrees of clonal relatedness according to dataset used and differing rates of PR, with CUIMC genomes having a high degree of clonal relatedness (with some being serial isolates collected from single patients), as well as higher rates of PR. The

datasets and their polymyxin susceptibility are summarized in Figure 1, including individual non-CUIMC datasets. Full genome details are included in Supplementary Table 1. The study workflow is summarized in Figure 2. Area under receiver operator curve (AUROC) was used as the key performance metric as it is a measure of the tradeoff between the true positive and false positive rate for various decision-making thresholds.

Machine learning approaches outperform rule-based approach for prediction of polymyxin resistance

For our reference-based analyses, we created a binary matrix with genomes as rows and coding regions as columns (referred to as 'instances' and 'features' in ML literature, respectively), with presence of single nucleotide variants (SNVs) or IS elements in coding regions relative to a CG258 reference genome being recorded, as described previously (12). This matrix was then used as an input for a simple rule-based approach which classified isolates as PR if there was a variant present in any of the canonical PR genes (*crrAB, mgrB*, *pmrAB*, *phoPQ*) (5-11). This approach had modest performance, resulting in an AUROC of 0.717–0.832, depending on input genomes.

We then used the same matrix as an input for ML analyses with four different ML algorithms: Logistic Regression, Random Forest, Support Vector Machine Classifier (SVC) and Gradient-Boosted Trees Classifier (GBTC), as implemented in SciKit-Learn (26). The best performing algorithms achieved mean AUROC of 0.885–0.933 (according to input genomes used) (Figure 3A and Table 1). A significant difference in mean performance when different input genome datasets were compared across metrics was noted in accuracy (P=0.0345), but not in any other metrics. The reference-based ML analyses achieved significantly higher AUROC than the rule-based approach (P=0.014 for CUIMC genomes, P=0.006 for non-CUIMC genomes and P=0.006 for all genomes, respectively) For other metrics, ML approaches performed equivalently or better than the rule-based approach, with the exception of recall (Table 1).

Choice of machine learning algorithm did not impact performance across different input genomes and performance metrics

The ML algorithms used (Logistic Regression, Random Forest, SVC and GBTC) differ substantially in their methodology. To assess whether the choice of ML algorithm impacted performance and whether there was an optimal algorithm, we compared ML algorithms according to input genomes and different performance metrics (Figure 3B). No significant differences in performance were noted between algorithms, regardless of the input genomes and performance metric used.

Impact of feature engineering with GWAS filtering and polymyxin exposure data on performance of ML-based prediction

A key challenge in AST genotype-phenotype prediction is the sparsity of the input genomic datasets due to relatively few genomes in datasets compared to the number of genomic features (14). Appropriate feature selection prior to training ML algorithms is a potential solution to this problem. We hypothesized that conducting a bacterial GWAS as a filtering procedure for selecting the most important features by p-value would improve the performance of ML models. We therefore conducted a bacterial GWAS using the R package *treeWAS* in the simultaneous mode (27), which enables correction for population structure. The results of the GWAS are shown in Supplementary Tables 2-4 according to input genomes used. Using these results to prioritize genes resulted in a mean increase of 5.3% in performance when all performance metrics were considered (range -3% to 13.9%), but this moderate rise was not statistically significant for most metrics when standard results were compared with those using GWAS filtering individually (Table 2and Figure 3A).

The widespread availability of electronic medical records has made it possible to rapidly extract clinical data for use in ML approaches (28). We wanted to assess if integrating clinical data would improve performance compared to ML models built only with genomic data. Polymyxin exposure has been recognized as a factor contributing to emergence of PR (1, 12, 29). Data regarding

polymyxin exposure in patients prior to culture of *K. pneumoniae* isolates were available for CUIMC genomes (12). This was used to create an additional binary feature of polymyxin exposure, added to the genome binary matrix and new ML models were trained as per the reference-based approach. Modest increases in all metrics except recall were noted. Integration of polymyxin exposure data resulted in the best performing model when compared with reference-based and GWAS filtering approaches (AUROC 0.923 vs 0.885 and 0.893, respectively) (Table 2). However, these increases did not reach statistical significance.

Reference-based approaches outperform reference-free approaches

Reference-based approaches for creating a genomic feature matrix for ML have inherent limitations. Bacteria such as *K. pneumoniae* with a large accessory genome pose a particular problem as a reference-based approach cannot evaluate novel genomic content present in the test genomes but lacking in the reference (15). Use of reference-free approaches, such as using k-mers (strings of nucleotides of *k* length) as inputs, may therefore be an attractive solution to overcome this limitation. Furthermore, being able to use reference-free approaches may potentially expand the available dataset by including a more diverse collection of genomes and thus improve performance of ML algorithms.

We tested this hypothesis by generating a reference-free binary matrix based on k-mer profiles from the DSK k-mer counting software (30) using k-mers of 31 nucleotides. This approach resulted in much higher numbers of features being incorporated than in reference-based datasets (415,373 vs 3,054 features for all genomes, 124,954 vs 2,278 features for CUIMC genomes, and 348,157 vs 2,350 vs features for non-CUIMC genomes, respectively). This matrix was then used as an input into the same ML pipeline as described previously. Despite only focusing on chromosomal changes relative to the reference, reference-based ML models had higher mean performance than reference-free models across all metrics in all datasets (Table 2), reaching statistical significance when all genomes were used.

In order to assess whether these findings may have been due to the ML algorithms in our pipeline not being specifically designed for k-mer data inputs, we used the same binary matrices as inputs in the KOVER package (31). This was developed for the sparse datasets seen when k-mers are used in AST genotype-phenotype prediction and uses either the Set Covering Machine (SCM) algorithm or Classification and Regression Trees (CART) algorithm. However, performance in KOVER was similar to our pipeline using both SCM and CART (AUROC 0.525–0.770, Supplementary Table 5).

Understanding determinants of PR through machine learning

Being able to interpret how ML algorithms arrive at their conclusions remains a central challenge to applying ML approaches in clinical decision making, including AST genotype-phenotype prediction (18). We therefore wanted to interpret ML models to assess whether they are biologically plausible and incorporate known PR determinants, as well as use them to identify potential novel determinants of PR. We were able to extract a ranking of genes according to their relative importance to the model (model-specific feature importance) for Logistic Regression, Random Forests and GBTC, but not for SVC due to the nature of the model.

We focused on the GBTC model trained on all genomes using a reference-based approach due to the diversity of the genomes and its high performance. The genes ranked highest in feature importance and their quantitative feature importance metric are listed in Table 3. Feature importance results for datasets using CUIMC and non-CUIMC genomes only are included in Supplementary Table 6. With no prior programming, the model identified all canonical PR genes except *crrA* and ranked *mgrB, phoQ, pmrA* and *pmrB* as the four genes with highest feature importance.

In addition to identifying these known determinants, we assessed other high ranked genes as candidate novel determinants of PR by conducting a literature search (Table 3). Several genes have been noted to have potential roles in PR previously including H239_3063, which encodes a

newly-identified putative RND-type efflux pump (32), and *arnA,* which is part of the *arn* operon that attaches arabinose to lipid A to confer PR (33). Many of the other genes are outer membrane proteins or have functions that may affect the cell membrane and thus are possibly involved in interactions with polymyxins including *lpdA*, *ahpF*, *envZ*, *pstB*, *pepN* and *pgpB*.

Discussion:

Our study provides a proof-of-principle demonstrating the utility of using ML for prediction of phenotypic PR from genomic data and raises important methodological issues that have implications for use of ML for AST genotype-phenotype prediction more generally. We focused efforts on PR due to the clinical need posed by resistance to this class of last-line antimicrobials, the difficulties associated with performing phenotypic AST for polymyxins and the complexity of underlying resistance mechanisms that may be uniquely suited to ML approaches. We noted high model performance across a range of metrics by leveraging a large collection of PR isolates from our institution and identifying publicly available genomes with BMD phenotypic susceptibility data. This performance was achieved through use of reference-based input datasets, use of GWAS as a filtering procedure and addition of polymyxin exposure as a clinical variable. We also were able to interpret ML models to confirm biological plausibility and identify additional potential genetic determinants of PR as candidates for functional testing.

Representation and selection of genomic features, a process termed ‘feature engineering’ in the ML literature, was central to the success of resistance prediction and could affect performance both positively and negatively. Firstly, our reference-based approach had high performance (AUROC 0.885–0.933) and outperformed the rule-based approach that we used as a benchmark. Despite PR being more biologically complex than other forms of antimicrobial resistance in *K. pneumoniae,* it was encouraging that our results were similar to those obtained for other antimicrobials with less complex mechanisms of resistance (eg. carbapenem resistance) tested with ML approaches (14, 15, 34-36). The performance of this approach falls below the FDA cutoffs for AST tests (37). However, given the problematic nature of phenotypic polymyxin

susceptibility testing, it may help identify a subset of genomes with a high likelihood of PR for confirmatory phenotypic testing.

In the setting of this high performance with our baseline approach, it was difficult to demonstrate a statistically significant improvement using additional feature engineering. With this in mind, we noted a moderate rise across nearly all metrics using GWAS filtering. The role of GWAS filtering for improving performance in other AST genotype-phenotype prediction settings remains to be determined but may therefore be of greater utility when the initial approach has lower performance. AST genotype-phenotype datasets typically have a small number of genomes with a large number of genomic features, therefore GWAS filtering provides a biologically consistent way of selecting genomic features. This approach has been used successfully in other non-microbiological studies (38-40) but has not previously been applied to AST genotype-phenotype prediction.

In contrast to filtering unnecessary genomic features, we noted that adding a single pertinent clinical feature in the form of polymyxin exposure data may increase performance by a similar amount to the more complex GWAS filtering approach, although the same caveats regarding a lack of statistical significance apply. MacFadden *et al.* used prior antimicrobial exposure data with genotypic data for ML prediction of *Escherichia coli* susceptibility phenotype to three different antimicrobial classes and similarly noted an increase in performance over using genotypic data alone (16). We focused on prior polymyxin exposure as it is a known epidemiological risk factor and as a proof-of-principle because drug administration is represented by highly structured variables in electronic medical records. We envision that the widespread adoption of electronic medical records will allow more extensive and automated integration of such clinical and genotypic data, thus enabling more accurate prediction both of AST phenotypes and other outcomes.

Certain forms of representation of genetic data can also have a negative impact on performance. While using reference-free approaches for feature representation incorporating k-mers has been the focus of much recent work (14, 15, 31, 34, 35, 41-44), we noted that reference-based approaches using variant calling and IS detection had significantly better performance overall. This may reflect the biological basis for PR, with IS elements playing an important role, particularly in association with *mgrB* (6, 8, 12, 45). K-mer based approaches may have difficulty accurately detecting this due to the diversity of IS elements in *Klebsiella* genomes, as well as the possibility that they insert in different sites in individual genes or in regulatory regions upstream (46). Additionally, reference-free approaches take the accessory genome into account. This offers putative benefits in terms of detecting various mobile genetic elements and making it easier to use this approach across diverse genomes. However, these benefits may be offset by the additional 'noise' in the case of organisms with large accessory genomes such as *K. pneumoniae*. Hicks *et al.* also found that ML approaches had worse performance when attempting to predict ciprofloxacin resistance in organisms with larger pan-genomes (*K. pneumoniae* and *Acinetobacter*) than in *Neisseria gonorrheae* (15). While k-mer based approaches may work well when the mechanism largely depends on a single or few resistance determinants (eg. beta-lactamases), our findings suggest that a more curated approach may be needed for more complex forms of resistance such as PR.

Beyond these biological issues, use of reference-free input data also has important computational consequences. Firstly, it increased the number of features by orders of magnitude compared with using reference-based approaches that use genes as inputs. In order to attenuate that effect, we incorporated feature selection hyperparameters in our pipeline but given the limited number of genomes for analysis, the increased dimensionality nevertheless likely negatively affected algorithm performance. The large number of features also requires more computationally intensive analyses, leading to much longer run times and raising questions about how plausible it would be to run these analyses as part of a clinical workflow. In order to address some of these computational issues, Drouin *et al*. developed the Set Covering Machine, a novel algorithm

implemented in the KOVER package (31). When we used our data with KOVER the run time was much lower than our pipeline, but model performance was similarly low.

In contrast to our findings regarding the importance of feature engineering, surprisingly neither the data used to train ML algorithms (in the form of input genomes) nor the choice of specific ML algorithm significantly impacted model performance. With regards to the input genomes used for training, this was in contrast to the sampling bias noted by Hicks et al. (15) and was somewhat surprising as the CUIMC isolates had a high degree of clonal relatedness and rates of PR varied depending on the origin dataset selected (Figure 1). Similarly, Moradigaravand et al. hypothesized that ensemble algorithms such as GBTC or Random Forests may be better suited to AST genotype-phenotype prediction (17) but we did not observe these differences. Our different findings may be explained by the fact that these studies focused on different organisms (*N. gonorrhea* and *E. coli,* respectively) and different antimicrobials. This further underlines prior concerns about attempting to use a 'one size fits all' approach given the diverse biological underpinnings of antimicrobial resistance (15).

The choice of performance metric that should be used in AST genotype-phenotype prediction also remains an open question with no consensus between prior studies (13). Indeed, the choice may be dictated by the expected prevalence of the phenotype of interest and the intended use of the ML model. Our intention was to create a screening method where isolates could be identified for confirmatory phenotypic testing, hence we chose AUROC as it is an aggregative metric that looks at the balance between true positive and false positive rate. It has been used in other AST genotype-phenotype studies (16, 34, 41-43, 47-49) and is a metric commonly used to assess clinical test performance. However, it may not be the optimal metric if there is significant imbalance in the dataset (50) leading Hicks *et al.* to advocate for use of multiple metrics (15). We therefore reported multiple metrics including precision (equivalent to positive predictive value), which we felt was important for our use case as it helps assess how many isolates identified as

resistant by the ML algorithm are phenotypically resistant. Using GWAS filtering we achieved precision of ~90%.

Our best performing ML models were also interpretable, thus allowing us to confirm that predictions were biologically plausible. The models correctly identified canonical PR genes with the exception of *crrA,* which had a low prevalence of variants across all datasets. This is highly encouraging given that this comprises six genes, with variants in multiple canonical genes often occurring in PR isolates (10, 12). In addition, they also confirmed genes that had been associated with PR in more limited settings including H239_3063 and *arnA.* Finally, several potential novel determinants were identified through use of ML, with multiple genes involved in stress responses or maintenance of the cell membrane. Although beyond the scope of this study, these determinants now require functional testing to prove causality and it is possible that exposure to other antimicrobials may have contributed to the observed genetic changes. This is demonstrated by the fact that the ML model incorporating all genomes identified *gyrB*, which has been associated with fluoroquinolone resistance.

Our study had several limitations, particularly related to the use of a reference-based approach. Firstly, the representation of genomes as binary variants in a reference's coding regions is an intentional simplification. An obvious problem is that all variants are treated as equal, whereas our previous work has shown that some variants may not have a functional impact, particularly in *phoQ* (12). However, this appears to be a reasonable trade-off given the limited number of genomes available for analysis. Alternative approaches may include using individual alleles or a weighted score that takes into account likelihood of functional impact based on amino acid changes, rather than a simple binary representation. A second limitation is our use of a single reference chromosome. On the one hand, this limits generalizability to CG258 genomes, which in future work may be overcome through attempting to define a *K. pneumoniae* core genome for analysis. On the other, it also entails that non-chromosomal regions are not incorporated into models. While plasmid-mediated PR is rare in *K. pneumoniae* (and was checked for specifically in

our collection), *mcr* genes or another plasmid-based determinant would not be detected by our reference-based approach. An additional consideration is the impact of population structure. While this has been well described as a potential confounder in bacterial GWAS (51), it has not been addressed in AST genotype-phenotype prediction but may play a similarly confounding role (15).

In summary, our study demonstrated that ML methods can achieve high performance for prediction of phenotypic PR in *K. pneumoniae* CG258*,* even in the face of PR being a remarkably polygenic trait. In contrast to other recent work on AST genotype-phenotype prediction, we noted best performance through use of a reference-based and curated input dataset which may reflect the underlying biological complexity of PR and be applicable to other complex forms of antimicrobial resistance. We noted that incorporating GWAS as a filtering procedure and addition of clinical data on antimicrobial exposure may improve performance, but these findings need confirmation in other settings. The increasing availability of genomic data makes AST genotype-phenotype prediction an important priority and use of ML will need to be tailored to specific organisms and antimicrobial agents.

Methods

Genome selection

The study was reviewed and approved by the CUIMC Institutional Review Board. CUIMC isolates were selected as previously described and comprised of *K. pneumoniae* CG258 isolates spanning 2011-2018 (12). All CUIMC isolates had MIC determination with BMD according to CLSI guidelines (52). We performed WGS on all included CUIMC isolates as described previously (12, 53-55). In total, we included 313 *K. pneumoniae* CG258 genomes. We then identified publicly available *K. pneumoniae* genomes by searching PubMed with the terms “colistin resistance” and “polymyxin resistance”. We also searched the CDC/FDA Antibiotic Resistance Isolate Bank and PATRIC databases (20, 25). When possible, we obtained genome raw sequence data from NCBI in preference to draft assemblies. Genomes with non-BMD

phenotypic susceptibility testing data were excluded due to concerns about accuracy (2). See Figure 1 and Supplementary Table 1 for all accession numbers, source datasets, phenotypic susceptibility and specific multi-locus sequence types. We included 306 publicly available *K. pneumoniae* CG258 genomes. For each analysis, we tested all 619 genomes, then subsets of 313 CUIMC genomes and 306 non-CUIMC genomes.

MIC ranges varied according to susceptibility testing platform (eg. SensiTitre versus manual methods). We therefore used a qualitative definition of polymyxin susceptibility, with isolates considered PR if colistin or polymyxin B MIC was >2 mg/L (52, 56). We constructed draft *de novo* assemblies of all genomes with raw sequence data using the Shovill wrapper for SPAdes (57), then used Kleborate for MLST and resistance determinant (including *mcr* genes) detection (58, 59). No *mcr* genes were detected.

Dataset preparation

A reference-based and a reference-free approach were used to create input matrices for ML algorithms. The outcome of interest ('label' in ML literature) for both was polymyxin susceptibility as a binary outcome (susceptible/resistant). For the reference-based approach, we had selected a representative polymyxin susceptible CG258 isolate and created a *de novo* hybrid assembly (12, 60) that was then used to create a profile of each genome through a combination of variant calling and IS detection, as described previously (12). In brief, variant calling was performed using Snippy 3.2 and running Snippy in the '-- contig' mode if only an assembly was available (61). ISseeker was used to identify sites with IS elements present (46). In order to increase sensitivity of detection for key genes, we also used a BLAST database of PR canonical genes to identify if IS disruption or large scale-deletions of these genes may have occurred.

These analyses allowed us to identify CG258 isolates containing SNVs/IS in coding regions relative to the reference genome. Intergenic regions, synonymous SNVs in coding regions, and SNVs in known mobile genetic elements and repeat regions were not considered. Each of these

alleles in each isolate was scored as 1 if it differed from the reference and 0 if it did not. These allele scores were used to create a binary matrix with isolates as rows and coding regions in the reference as columns. We also integrated clinical data about polymyxin exposure from our prior study for CUIMC isolates and created a binary variable that indicated polymyxin exposure at any time in that patient prior to culture of the isolate (12).

We constructed a reference-free input matrix based on k-mer profiles generated with the DSK k-mer counting software with k=31, a length commonly used in bioinformatics analyses (14, 15, 30, 31, 34, 62, 63). A k-mer presence/absence binary matrix was then created and used as an input for ML analyses.

Rule-based prediction of polymyxin resistance

We used rule-based prediction as a baseline for comparison with ML approaches. The SNV/IS detection results described for reference-based datasets above were used to find variants in any of the canonical PR genes (*crrAB, mgrB*, *pmrAB*, *phoPQ*) that have been established as key contributors to PR (5-11). Isolates would then be classified as PR if there was a variant present in any of these genes.

Use of genome-wide association study for feature filtering

As an additional feature engineering procedure, for the reference-based approach we filtered features using p-values from a GWAS. The reference-based binary matrix was used as input in *treeWAS*, an R package for bacterial GWAS (27). *TreeWAS* was run in 'simultaneous' mode to account for population structure. The resulting p-values were used to rank and select genomic features under a specific p-value threshold, discarding the rest.

Machine learning analyses

We created a bespoke pipeline using Scikit-Learn (26), which performed additional feature selection and hyperparameter tuning. Hyperparameters were tuned with cross validation (CV)

using AUROC as the performance metric. Reference-based datasets without GWAS filtering and reference-free datasets used a linear support vector classifier to extract important features for feature selection, and thus 10-fold CV was used to tune feature selection and model-specific hyperparameters jointly. As GWAS p-values were generated using labeled data, for reference-based datasets with GWAS filtering we used a 75%/25% training-validation split to tune hyperparameters where the GWAS values and models were trained using the 75% split training data. All datasets made use of a linear support vector classifier to filter features, while datasets with GWAS p-values also used p-values to filter features. In the latter process features chosen by either step were all selected. Hyperparameters were tuned for each dataset across the feature selection step and for each model configuration.

We tested four different ML models: Logistic Regression, Random Forests, SVC and GBTC. The performance metrics used were AUROC, balanced accuracy, accuracy, precision, recall and F1 score. Best performing models were chosen based on AUROC as this is an aggregative metric and has been used in prior studies of AST genotype-phenotype prediction (16, 34, 41-43, 47-49).

We then also compared the results of our ML pipeline to the ML implementation in the KOVER package that utilizes the Set Covering Machine and Classification and Regression Trees algorithms (14, 31). For the KOVER implementation, the best conjunctive and/or disjunctive model was selected using five-fold cross-validation, testing the suggested broad range of values for the trade-off hyperparameter of 0.1, 0.178, 0.316, 0.562, 1.0, 1.778, 3.162, 5.623, 10.0, 144 and 999999.0 to determine the optimal rule scoring function with default parameters. The pROC R package was used to calculate AUROC for the predictions made by the rule-based algorithm and KOVER package (64).

Statistical testing

Cross-validation results were used to generate means and 95% confidence intervals for all performance metrics. As models trained using the 75%/25% split with GWAS p-values don’t have

sensical confidence intervals, bootstrapping was performed. Ten bootstrapped training and validation sets were created by pooling the split data and repeatedly bootstrapping (sampling n observations with replacement) a training dataset from the original data, then using the rest of the data as a validation set. The best model and its hyperparameters for each model class (Logistic Regression, SVC, etc.) were chosen based on the original 75%/25% split performance and then re-trained and evaluated on each of the bootstrapped training/validation pairs and metrics were collected. Model performance was evaluated by comparing mean performance metrics using two-tailed t-tests or Mann-Whitney U Tests, as appropriate. Categorical variables were compared using $\chi 2$ or Fisher's exact tests, as appropriate. ANOVA was used for comparison of means between multiple groups. Statistical analyses were performed in R (v3.4.0).

Data availability

All CUIMC genomes have been deposited to the NCBI Sequence Read Archive under NCBI BioProject accession numbers PRJNA557275 and PRJNA445400. For individual accession numbers please see Supplementary Table 1.

Financial support

This work was supported by the Australian National Health and Medical Research Council Postgraduate Scholarship (APP1169514) to N.M. and Practitioner Fellowship to A.Y.P (APP1117940), National Institute of Allergy and Infectious Diseases, National Institutes of Health (grant number AI116939 to A-C. U.).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Potential conflicts of interest

NM has received research support from GlaxoSmithKline, unrelated to the current study. A.Y.P has received research support from MSD, unrelated to the current study. A-C.U. has received research funding from Merck, GlaxoSmithKline and Allergan, unrelated to the current study. All other authors declare no conflict of interest.

References

1. Poirel L, Jayol A, Nordmann P. 2017. Polymyxins: Antibacterial Activity, Susceptibility Testing, and Resistance Mechanisms Encoded by Plasmids or Chromosomes. Clin Microbiol Rev 30:557-596.
2. Hindler JA, Humphries RM. 2013. Colistin MIC variability by method for contemporary clinical isolates of multidrug-resistant Gram-negative bacilli. J Clin Microbiol 51:1678-84.
3. Humphries RM. 2015. Susceptibility testing of the polymyxins: where are we now? Pharmacotherapy 35:22-27.
4. Humphries RM, Green DA, Schuetz AN, Bergman Y, Lewis S, Yee R, Stump S, Lopez M, Macesic N, Uhlemann AC, Kohner P, Cole N, Simner PJ. 2019. Multi-center evaluation of colistin broth disk elution and colistin agar test: a report from the Clinical and Laboratory Standards Institute. J Clin Microbiol doi:10.1128/JCM.01269-19.
5. Aires CA, Pereira PS, Asensi MD, Carvalho-Assef AP. 2016. mgrB Mutations Mediating Polymyxin B Resistance in Klebsiella pneumoniae Isolates from Rectal Surveillance Swabs in Brazil. Antimicrob Agents Chemother 60:6969-6972.
6. Berglund B, Hoang NTB, Tarnberg M, Le NK, Svartstrom O, Khu DTK, Nilsson M, Le HT, Welander J, Olson L, Larsson M, Nilsson LE, Hanberger H. 2018. Insertion sequence transpositions and point mutations in mgrB causing colistin resistance in a clinical strain of carbapenem-resistant Klebsiella pneumoniae from Vietnam. Int J Antimicrob Agents 51:789-793.
7. Cannatelli A, Di Pilato V, Giani T, Arena F, Ambretti S, Gaibani P, D'Andrea MM, Rossolini GM. 2014. In vivo evolution to colistin resistance by PmrB sensor kinase mutation in KPC-producing Klebsiella pneumoniae is associated with low-dosage colistin treatment. Antimicrob Agents Chemother 58:4399-403.
8. Cannatelli A, Giani T, D'Andrea MM, Di Pilato V, Arena F, Conte V, Tryfinopoulou K, Vatopoulos A, Rossolini GM, Group CS. 2014. MgrB inactivation is a common mechanism of colistin resistance in KPC-producing Klebsiella pneumoniae of clinical origin. Antimicrob Agents Chemother 58:5696-703.

9. Cheng YH, Lin TL, Pan YJ, Wang YP, Lin YT, Wang JT. 2015. Colistin resistance mechanisms in Klebsiella pneumoniae strains from Taiwan. Antimicrob Agents Chemother 59:2909-13.
10. Pitt ME, Elliott AG, Cao MD, Ganesamoorthy D, Karaiskos I, Giamarellou H, Abboud CS, Blaskovich MAT, Cooper MA, Coin LJM. 2018. Multifactorial chromosomal variants regulate polymyxin resistance in extensively drug-resistant Klebsiella pneumoniae. Microb Genom 4.
11. Wright MS, Suzuki Y, Jones MB, Marshall SH, Rudin SD, van Duin D, Kaye K, Jacobs MR, Bonomo RA, Adams MD. 2015. Genomic and transcriptomic analyses of colistin-resistant clinical isolates of Klebsiella pneumoniae reveal multiple pathways of resistance. Antimicrob Agents Chemother 59:536-43.
12. Macesic N, Nelson B, McConville TH, Giddins MJ, Green DA, Stump S, Gomez-Simmonds A, Annavajhala MK, Uhlemann AC. 2019. Emergence of Polymyxin Resistance in Clinical Klebsiella pneumoniae Through Diverse Genetic Adaptations: A Genomic, Retrospective Cohort Study. Clin Infect Dis doi:10.1093/cid/ciz623.
13. Su M, Satola SW, Read TD. 2019. Genome-Based Prediction of Bacterial Antibiotic Resistance. Journal of Clinical Microbiology 57:2018-15.
14. Drouin A, Letarte G, Raymond F, Marchand M, Corbeil J, Laviolette F. 2019. Interpretable genotype-to-phenotype classifiers with performance guarantees. Sci Rep 9:4071.
15. Hicks AL, Wheeler N, Sanchez-Buso L, Rakeman JL, Harris SR, Grad YH. 2019. Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. PLoS Comput Biol 15:e1007349.
16. MacFadden DR, Melano RG, Coburn B, Tijet N, Hanage WP, Daneman N. 2019. Comparing Patient Risk Factor-, Sequence Type-, and Resistance Locus Identification-Based Approaches for Predicting Antibiotic Resistance in Escherichia coliBloodstream Infections. Journal of Clinical Microbiology 57:h2346-9.

17. Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts L. 2018. Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data. PLoS Computational Biology 14:e1006258-17.
18. Macesic N, Polubriaginof F, Tatonetti NP. 2017. Machine learning: novel bioinformatics approaches for combating antimicrobial resistance. Curr Opin Infect Dis 30:511-517.
19. Pitout JDD, Nordmann P, Poirel L. 2015. Carbapenemase-Producing Klebsiella pneumoniae, a Key Pathogen Set for Global Nosocomial Dominance. Antimicrobial Agents and Chemotherapy 59:5873-5884.
20. Centers for Disease Control and Prevention. 2019. CDC & FDA Antibiotic Resistance (AR) Isolate Bank. https://www.cdc.gov/drugresistance/resistance-bank/index.html. Accessed August 15, 2019.
21. Deleo FR, Chen L, Porcella SF, Martens CA, Kobayashi SD, Porter AR, Chavda KD, Jacobs MR, Mathema B, Olsen RJ, Bonomo RA, Musser JM, Kreiswirth BN. 2014. Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 Klebsiella pneumoniae. Proc Natl Acad Sci U S A 111:4988-93.
22. Rimoldi SG, Gentile B, Pagani C, Di Gregorio A, Anselmo A, Palozzi AM, Fortunato A, Pittiglio V, Ridolfo AL, Gismondo MR, Rizzardini G, Lista F. 2017. Whole genome sequencing for the molecular characterization of carbapenem-resistant Klebsiella pneumoniae strains isolated at the Italian ASST Fatebenefratelli Sacco Hospital, 2012-2014. BMC Infect Dis 17:666.
23. Stoesser N, Giess A, Batty EM, Sheppard AE, Walker AS, Wilson DJ, Didelot X, Bashir A, Sebra R, Kasarskis A, Sthapit B, Shakya M, Kelly D, Pollard AJ, Peto TE, Crook DW, Donnelly P, Thorson S, Amatya P, Joshi S. 2014. Genome sequencing of an extended series of NDM-producing Klebsiella pneumoniae isolates from neonatal infections in a Nepali hospital characterizes the extent of community- versus hospital-associated transmission in an endemic setting. Antimicrob Agents Chemother 58:7347-57.
24. van Dorp L, Wang Q, Shaw LP, Acman M, Brynildsrud OB, Eldholm V, Wang R, Gao H, Yin Y, Chen H, Ding C, Farrer RA, Didelot X, Balloux F, Wang H. 2019. Rapid phenotypic

evolution in multidrug-resistant Klebsiella pneumoniae hospital outbreak strains. Microb Genom 5.

25. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res 45:D535-D542.

26. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Müller A, Nothman J, Louppe G, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12:2825-2830.

27. Collins C, Didelot X. 2018. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comput Biol 14:e1005958.

28. Rajkomar A, Dean J, Kohane I. 2019. Machine Learning in Medicine. N Engl J Med 380:1347-1358.

29. Giacobbe DR, Del Bono V, Trecarichi EM, De Rosa FG, Giannella M, Bassetti M, Bartoloni A, Losito AR, Corcione S, Bartoletti M, Mantengoli E, Saffioti C, Pagani N, Tedeschi S, Spanu T, Rossolini GM, Marchese A, Ambretti S, Cauda R, Viale P, Viscoli C, Tumbarello M, Isgri S. 2015. Risk factors for bloodstream infections due to colistin-resistant KPC-producing Klebsiella pneumoniae: results from a multicenter case-control-control study. Clin Microbiol Infect 21:1106 e1-8.

30. Rizk G, Lavenier D, Chikhi R. 2013. DSK: k-mer counting with very low memory usage. Bioinformatics 29:652-3.

31. Drouin A, Giguere S, Deraspe M, Marchand M, Tyers M, Loo VG, Bourgault AM, Laviolette F, Corbeil J. 2016. Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons. BMC Genomics 17:754.

32. Cheng YH, Lin TL, Lin YT, Wang JT. 2018. A putative RND-type efflux pump, H239_3064, contributes to colistin resistance through CrrB in Klebsiella pneumoniae. J Antimicrob Chemother 73:1509-1516.
33. Gatzeva-Topalova PZ, May AP, Sousa MC. 2005. Structure and mechanism of ArnA: conformational change implies ordered dehydrogenase mechanism in key enzyme for polymyxin resistance. Structure 13:929-42.
34. Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis JJ, Brettin T, Xia F, Musser JM. 2017. Population Genomic Analysis of 1,777 Extended-Spectrum Beta-Lactamase-Producing Klebsiella pneumoniae Isolates, Houston, Texas: Unexpected Abundance of Clonal Group 307. MBio 8.
35. Nguyen M, Brettin T, Long SW, Musser JM, Olsen RJ, Olson R, Shukla M, Stevens RL, Xia F, Yoo H, Davis JJ. 2017. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. Scientific Reports doi:papers3://publication/doi/10.1038/s41598-017-18972-w:1-11.
36. Pesesky MW, Hussain T, Wallace M, Patel S, Andleeb S, Burnham C-AD, Dantas G. 2016. Evaluation of Machine Learning and Rules-Based Approaches for Predicting Antimicrobial Resistance Profiles in Gram-negative Bacilli from Whole Genome Sequence Data. Frontiers in microbiology 7:414-17.
37. Humphries RM, Ambler J, Mitchell SL, Castanheira M, Dingle T, Hindler JA, Koeth L, Sei K, Development CM, Standardization Working Group of the Subcommittee on Antimicrobial Susceptibility T. 2018. CLSI Methods Development and Standardization Working Group Best Practices for Evaluation of Antimicrobial Susceptibility Tests. J Clin Microbiol 56.
38. Okser S, Pahikkala T, Aittokallio T. 2013. Genetic variants and their interactions in disease risk prediction - machine learning and network perspectives. BioData mining 6:5.
39. Wei Z, Wang K, Qu H-Q, Zhang H, Bradfield J, Kim C, Frackleton E, Hou C, Glessner JT, Chiavacci R, Stanley C, Monos D, Grant SFA, Polychronakos C, Hakonarson H. 2009.

From Disease Association to Risk Assessment: An Optimistic View from Genome-Wide Association Studies on Type 1 Diabetes. PLoS Genetics 5:e1000678-11.

40. Wei Z, Wang W, Bradfield J, Li J, Cardinale C, Frackelton E, Kim C, Mentch F, Van Steen K, Visscher PM, Baldassano RN, Hakonarson H, International IBDGC. 2013. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. American journal of human genetics 92:1008-1012.

41. Davis JJ, Boisvert S, Brettin T, Kenyon RW, Mao C, Olson R, Overbeek R, Santerre J, Shukla M, Wattam AR, Will R, Xia F, Stevens R. 2016. Antimicrobial Resistance Prediction in PATRIC and RAST. Sci Rep 6:27930.

42. Mahé P. 2018. Predicting bacterial resistance from whole-genome sequences using k-mers and stability selection. doi:papers3://publication/doi/10.1186/s12859-018-2403-z:1-11.

43. Mahé P, El Azami M, Barlas P, Tournoud M. 2019. A large scale evaluation of TBProfiler and Mykrobe for antibiotic resistance prediction in Mycobacterium tuberculosis. PeerJ 7:e6857-21.

44. Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, Tyson GH, Zhao S, Davis JJ. 2019. Using Machine Learning To Predict Antimicrobial MICs and Associated Genomic Features for Nontyphoidal Salmonella. Journal of Clinical Microbiology 57:495-15.

45. Arena F, Henrici De Angelis L, Cannatelli A, Di Pilato V, Amorese M, D'Andrea MM, Giani T, Rossolini GM. 2016. Colistin Resistance Caused by Inactivation of the MgrB Regulator Is Not Associated with Decreased Virulence of Sequence Type 258 KPC Carbapenemase-Producing Klebsiella pneumoniae. Antimicrob Agents Chemother 60:2509-12.

46. Adams MD, Bishop B, Wright MS. 2016. Quantitative assessment of insertion sequence impact on bacterial genome architecture. Microb Genom 2:e000062.

47. Her H-L, Wu Y-W. 2018. A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the Escherichia coli strains. Bioinformatics 34:i89-i95.
48. Kouchaki S, Yang Y, Walker TM, Walker AS, Wilson DJ, Peto TEA, Crook DW, Clifton DA, consortium CR. 2018. Application of machine learning techniques to tuberculosis drug resistance analysis. Bioinformatics 35:2276-2282.
49. Yang Y, Niehaus KE, Walker TM, Iqbal Z, Walker AS, Wilson DJ, Peto TEA, Crook DW, Smith EG, Zhu T, Clifton DA. 2018. Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. Bioinformatics 34:1666-1671.
50. Jeni LA, Cohn JF, De La Torre F. 2013. Facing Imbalanced Data Recommendations for the Use of Performance Metrics. Int Conf Affect Comput Intell Interact Workshops 2013:245-251.
51. Read TD, Massey RC. 2014. Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. Genome Medicine 6:1063-11.
52. Institute CaLS. 2015. Performance standards for antimicrobial susceptibility testing. , Twenty-fifth informational supplement, Clinical and Laboratory Standards Institute, Wayne, PA, 2015, M100-S25.
53. Giddins MJ, Macesic N, Annavajhala MK, Stump S, Khan S, McConville TH, Mehta M, Gomez-Simmonds A, Uhlemann AC. 2018. Successive Emergence of Ceftazidime-Avibactam Resistance through Distinct Genomic Adaptations in blaKPC-2-Harboring Klebsiella pneumoniae Sequence Type 307 Isolates. Antimicrob Agents Chemother 62.
54. Gomez-Simmonds A, Annavajhala MK, Wang Z, Macesic N, Hu Y, Giddins MJ, O'Malley A, Toussaint NC, Whittier S, Torres VJ, Uhlemann AC. 2018. Genomic and Geographic Context for the Evolution of High-Risk Carbapenem-Resistant Enterobacter cloacae Complex Clones ST171 and ST78. MBio 9.
55. Macesic N, Gomez-Simmonds A, Sullivan SB, Giddins MJ, Ferguson SA, Korakavi G, Leeds D, Park S, Shim K, Sowash MG, Hofbauer M, Finkel R, Hu Y, West J, Toussaint NC, Greendyke WG, Miko BA, Pereira MR, Whittier S, Verna EC, Uhlemann AC. 2018.

Genomic Surveillance Reveals Diversity of Multidrug-Resistant Organism Colonization and Infection: A Prospective Cohort Study in Liver Transplant Recipients. Clin Infect Dis 67:905-912.

56. European Committee on Antimicrobial Susceptibility Testing. 2016. Breakpoint tables for interpretation of MICs and zone diameters. http://www.eucast.org. Accessed December 14, 2018.

57. Seemann T. 2018. Shovill: Faster SPAdes (or better SKESA/Megahit/Velvet) assembly of Illumina reads. https://github.com/tseemann/shovill. Accessed July 23, 2018.

58. Holt KE. 2019. Kleborate. https://github.com/katholt/Kleborate. Accessed August 15, 2019.

59. Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med 6:90.

60. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

61. Seemann T. 2018. Snippy: Rapid haploid variant calling and core SNP phylogeny. https://github.com/tseemann/snippy. Accessed July 23, 2018.

62. Earle SG, Wu CH, Charlesworth J, Stoesser N, Gordon NC, Walker TM, Spencer CCA, Iqbal Z, Clifton DA, Hopkins KL, Woodford N, Smith EG, Ismail N, Llewelyn MJ, Peto TE, Crook DW, McVean G, Walker AS, Wilson DJ. 2016. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. Nat Microbiol 1:16041.

63. Lees JA, Vehkala M, Valimaki N, Harris SR, Chewapreecha C, Croucher NJ, Marttinen P, Davies MR, Steer AC, Tong SY, Honkela A, Parkhill J, Bentley SD, Corander J. 2016. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. Nat Commun 7:12797.

64. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12:77.
65. Ramos PI, Custodio MG, Quispe Saji GD, Cardoso T, da Silva GL, Braun G, Martins WM, Girardello R, de Vasconcelos AT, Fernandez E, Gales AC, Nicolas MF. 2016. The polymyxin B-induced transcriptomic response of a clinical, multidrug-resistant Klebsiella pneumoniae involves multiple regulatory elements and intracellular targets. BMC Genomics 17:737.
66. Brinkworth AJ, Hammer CH, Olano LR, Kobayashi SD, Chen L, Kreiswirth BN, DeLeo FR. 2015. Identification of Outer Membrane and Exoproteins of Carbapenem-Resistant Multilocus Sequence Type 258 Klebsiella pneumoniae. PLoS One 10:e0123219.
67. Miethke M, Marahiel MA. 2007. Siderophore-based iron acquisition and pathogen control. Microbiol Mol Biol Rev 71:413-51.
68. Bhagirath AY, Li Y, Patidar R, Yerex K, Ma X, Kumar A, Duan K. 2019. Two Component Regulatory Systems and Antibiotic Resistance in Gram-Negative Pathogens. Int J Mol Sci 20.
69. Dorman MJ, Feltwell T, Goulding DA, Parkhill J, Short FL. 2018. The Capsule Regulatory Network of Klebsiella pneumoniae Defined by density-TraDISort. MBio 9.
70. Lamarche MG, Wanner BL, Crepin S, Harel J. 2008. The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. FEMS Microbiol Rev 32:461-73.
71. Nganje CN, Haynes SA, Qabar CM, Lent RC, Bou Ghanem EN, Shainheit MG. 2019. PepN is a non-essential, cell wall-localized protein that contributes to neutrophil elastase-mediated killing of Streptococcus pneumoniae. PLoS One 14:e0211632.
72. Pitt ME, Cao MD, Butler MS, Ramu S, Ganesamoorthy D, Blaskovich MAT, Coin LJM, Cooper MA. 2019. Octapeptin C4 and polymyxin resistance occur via distinct pathways in an epidemic XDR Klebsiella pneumoniae ST258 isolate. J Antimicrob Chemother 74:582-593.

765 73. Hooper DC, Jacoby GA. 2015. Mechanisms of drug resistance: quinolone resistance.
766 Ann N Y Acad Sci 1354:12-31.
767 74. Klein G, Raina S. 2019. Regulated Assembly of LPS, Its Structural Alterations and
768 Cellular Response to LPS Defects. Int J Mol Sci 20.
769 75. Tong S, Lin Y, Lu S, Wang M, Bogdanov M, Zheng L. 2016. Structural Insight into
770 Substrate Selection and Catalysis of Lipid Phosphate Phosphatase PgpB in the Cell
771 Membrane. J Biol Chem 291:18342-52.

772

Tables

Table 1 – Comparison of rule-based vs reference-based approaches for prediction of polymyxin resistance

Genomes used	Metric	Rule-based	Reference-based (95% CI)	Reference-based with GWAS (95% CI)	P value (Ref. based vs rule-based)	P value (GWAS vs rule-based)
CUIMC	AUROC	0.832	0.885 (0.849, 0.92)	0.893 (0.864, 0.922)	0.014*	0.004*
	bACC	0.832	0.789 (0.751, 0.827)	0.841 (0.82, 0.862)	0.049*	0.262
	Accuracy	0.821	0.796 (0.762, 0.83)	0.854 (0.83, 0.878)	0.185	0.052
	F1	0.819	0.755 (0.701, 0.809)	0.816 (0.793, 0.84)	0.027*	0.919
	Precision	0.738	0.799 (0.739, 0.859)	0.881 (0.848, 0.914)	0.037*	0.006*
	Recall	0.92	0.733 (0.635, 0.831)	0.763 (0.725, 0.801)	0.008*	0.006*
Non-CUIMC	AUROC	0.717	0.933 (0.884, 0.982)	0.933 (0.888, 0.979)	0.006*	0.002*
	bACC	0.717	0.753 (0.654, 0.853)	0.82 (0.76, 0.881)	0.415	0.006*
	Accuracy	0.699	0.873 (0.822, 0.925)	0.917 (0.894, 0.94)	0.006*	0.006*
	F1	0.471	0.59 (0.395, 0.785)	0.729 (0.648, 0.81)	0.185	0.002*
	Precision	0.345	0.711 (0.473, 0.949)	0.832 (0.759, 0.905)	0.018*	0.006*
	Recall	0.745	0.57 (0.362, 0.778)	0.669 (0.547, 0.791)	0.184	0.262
All	AUROC	0.791	0.894 (0.838, 0.95)	0.931 (0.915, 0.947)	0.006*	0.002*
	bACC	0.791	0.784 (0.73, 0.838)	0.801 (0.776, 0.827)	0.61	0.375
	Accuracy	0.761	0.827 (0.78, 0.874)	0.864 (0.84, 0.888)	0.019*	0.006*
	F1	0.694	0.702 (0.623, 0.781)	0.741 (0.702, 0.779)	0.76	0.02*
	Precision	0.577	0.8 (0.675, 0.926)	0.889 (0.846, 0.932)	0.011*	0.006*
	Recall	0.87	0.668 (0.549, 0.788)	0.638 (0.591, 0.686)	0.006*	0.002*

* denotes statistical significance with $P<0.05$. Abbreviations: AUROC – Area under receiver-operator curve; bACC – Balanced accuracy, CI – Confidence interval, CUIMC – Columbia University Irving Medical Center; GWAS – Genome-wide association study; Ref. – Reference.

777 **Table 2 – Comparison of different feature engineering approaches on performance of machine learning prediction of polymyxin**

778 **resistance**

Genomes used	Metric	Reference-based (95% CI)	Reference-based with GWAS (95% CI)	Reference - free (95% CI)	Reference-based with polymyxin exposure data (95% CI)	P value (GWAS vs ref. based)	P value (Ref. free vs ref. based)	P value (Poly. exposure vs ref. based)
CUIMC	AUROC	0.885 (0.849, 0.92)	0.893 (0.864, 0.922)	0.696 (0.564, 0.828)	0.923 (0.88, 0.965)	0.571	0.241	0.104
	bACC	0.789 (0.751, 0.827)	0.841 (0.82, 0.862)	0.64 (0.536, 0.743)	0.796 (0.714, 0.879)	0.026*	0.226	0.544
	Accuracy	0.796 (0.762, 0.83)	0.854 (0.83, 0.878)	0.649 (0.541, 0.758)	0.804 (0.716, 0.892)	0.009*	0.91	0.342
	F1	0.755 (0.701, 0.809)	0.816 (0.793, 0.84)	0.579 (0.454, 0.704)	0.768 (0.685, 0.85)	0.045*	0.734	0.733
	Precision	0.799 (0.739, 0.859)	0.881 (0.848, 0.914)	0.67 (0.506, 0.833)	0.866 (0.737, 0.996)	0.006*	0.023*	0.085
	Recall	0.733 (0.635, 0.831)	0.763 (0.725, 0.801)	0.56 (0.407, 0.714)	0.732 (0.607, 0.857)	1	0.011*	0.879
Non-CUIMC	AUROC	0.933 (0.884, 0.982)	0.933 (0.888, 0.979)	0.803 (0.692, 0.913)		0.85	0.677	

	bACC	0.753 (0.654, 0.853)	0.82 (0.76, 0.881)	0.5 (0.5, 0.5)		0.427	0.089	
	Accuracy	0.873 (0.822, 0.925)	0.917 (0.894, 0.94)	0.82 (0.811, 0.83)		0.185	0.005*	
	F1	0.59 (0.395, 0.785)	0.729 (0.648, 0.81)	0 (0, 0)		0.345	0.623	
	Precision	0.711 (0.473, 0.949)	0.832 (0.759, 0.905)	0 (0, 0)		0.703	0.569	
	Recall	0.57 (0.362, 0.778)	0.669 (0.547, 0.791)	0 (0, 0)		0.88	0*	
All	AUROC	0.894 (0.838, 0.95)	0.931 (0.915, 0.947)	0.692 (0.546, 0.838)		0.19	0.015*	
	bACC	0.784 (0.73, 0.838)	0.801 (0.776, 0.827)	0.5 (0.5, 0.5)		0.473	0.006*	
	Accuracy	0.827 (0.78, 0.874)	0.864 (0.84, 0.888)	0.688 (0.685, 0.691)		0.162	0.045*	
	F1	0.702 (0.623, 0.781)	0.741 (0.702, 0.779)	0 (0, 0)		0.384	0.003*	
	Precision	0.8 (0.675, 0.926)	0.889 (0.846, 0.932)	0 (0, 0)		0.363	0.472	
	Recall	0.668 (0.549, 0.788)	0.638 (0.591, 0.686)	0 (0, 0)		0.344	0.064	

779 * denotes statistical significance with $P<0.05$. Abbreviations: AUROC – Area under receiver-operator curve; bACC – Balanced accuracy, CI –

780 Confidence interval, CUIMC – Columbia University Irving Medical Center; GWAS – Genome-wide association study; Ref. – Reference.

781

Table 3 – Genes ranked by relative feature importance in machine learning model for polymyxin resistance prediction that incorporated all genomes and used Gradient-Boosted Trees Classifier

Annotation	Quantitative feature importance metric	Full name	Function and comments	Reference(s)
mgrB	0.317		Known determinant	
phoQ	0.0454		Known determinant	
pmrA	0.0343		Known determinant	
pmrB	0.0325		Known determinant	
lpdA	0.0284	Dihydrolipoyl dehydrogenase	Respiratory chain enzyme. Implicated in response to polymyxin B	(65)
ahpF	0.0250	Alkyl hydroperoxide reductase subunit F	Outer membrane protein conferring hydrogen peroxide resistance and implicated in stress response	(66)
grxD	0.0195	Grx4 family monothiol glutaredoxin	Iron regulation	(67)
envZ	0.0185		Sensing of osmotic signals, regulation of biofilm formation and capsule production. Implicated in response to polymyxin B.	(65, 68, 69)
pstB	0.0183	Phosphate import ATP-binding protein PstB	Capture and transport of periplasmic phosphate into cell	(70)
crrB	0.0160		Known determinant	
dmlR_5	0.0147		Unknown	
phoP	0.0143		Known determinant	
arnA	0.0136	UDP 4-deoxy-4-formamido-L-arabinose transferase	Part of arn operon that attaches arabinose to lipid A to confer polymyxin resistance	(33)
pepN	0.0136	aminopeptidase N	Cell wall protein, possible target of neutrophil elastase	(71)

pqiB_2	0.0133	Paraquat inducible protein B	Involved in transport pathways that contribute to membrane integrity	(72)
hypothetical protein	0.0113	H239_3063	Encodes putative RND-type efflux pump and newly discovered determinant of PR	(32)
gyrB	0.0104	DNA topoisomerase (ATP-hydrolyzing) subunit B	Implicated in fluoroquinolone resistance	(73)
hypothetical protein	0.00954		Unknown	(73)
cytR_1	0.00902		Unknown	
pgpB	0.00897	Phosphatidylglycero-phosphatase B	Involved in generating phospholipid for cell membrane	(74, 75)

784

Figures

Figure 1 – Summary of input datasets according to polymyxin resistance

Histograms showing the relative distribution of polymyxin resistance across all genomes, Columbia University Irving Medical Center (CUIMC) genomes only, non-CUIMC genomes only, then individual publicly-available datasets that formed non-CUIMC genomes. For further information regarding individual genomes please see Supplementary Table 1.

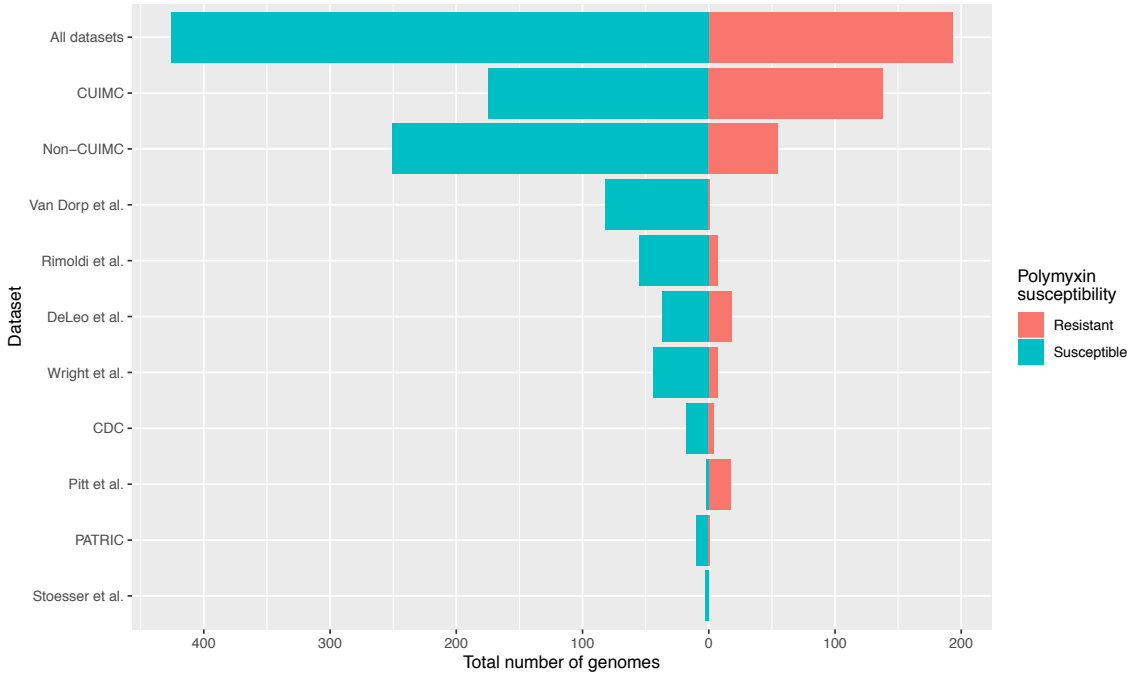


Figure 2 – Workflow for polymyxin genotype-susceptibility prediction pipeline

Publicly available g and genomes from Columbia University Irving Medical Center (CUIMC) were processed using either a reference-based or reference-free approach in order to generate a binary matrix. The binary matrix would represent either coding regions in the genome (reference-based approach) or individual k-mers (reference-free approach). For the reference-based approach, further feature engineering was performed by using a bacterial genome wide association study (GWAS) to prioritize genes and adding clinical polymyxin exposure data. Machine learning analyses were then performed using a bespoke pipeline implemented in Sci-kit Learn.

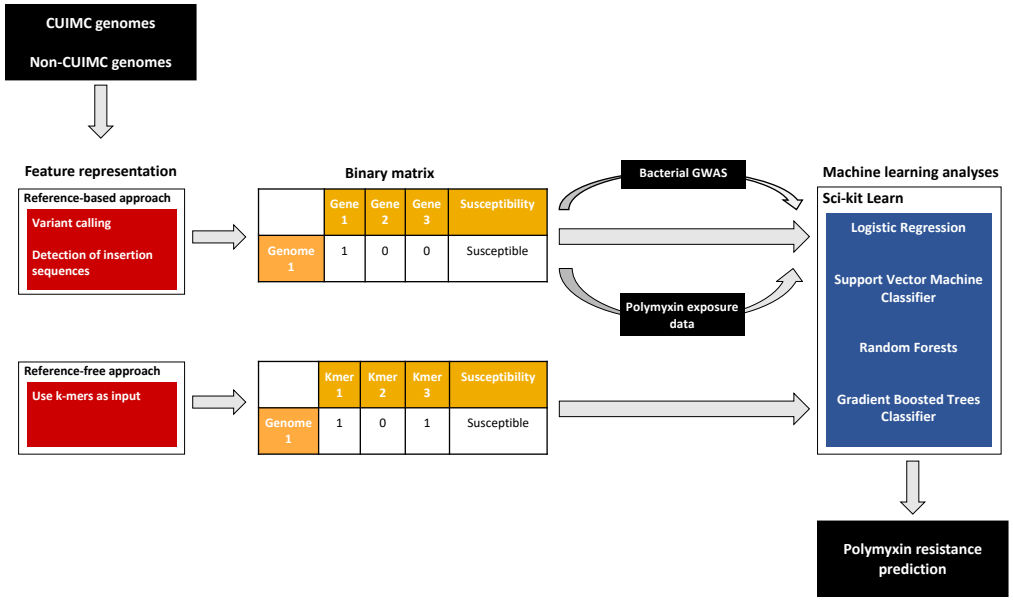
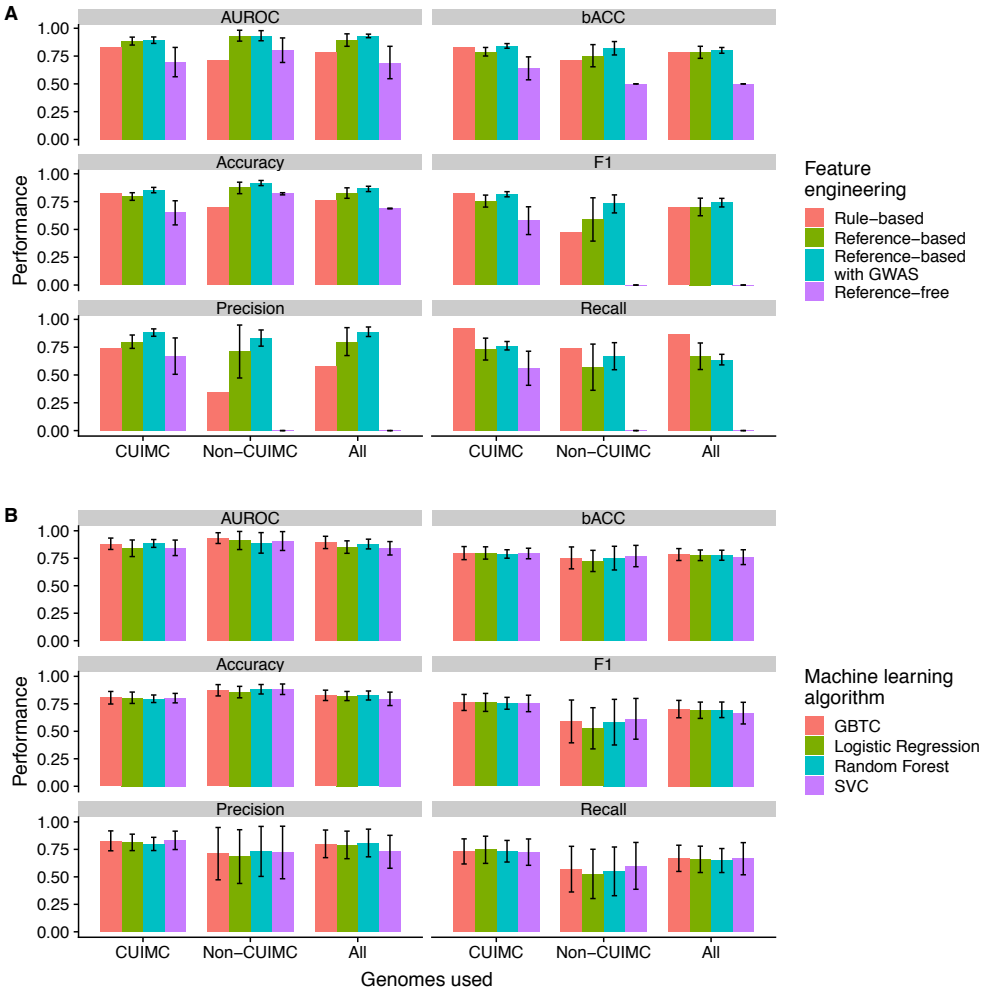


Figure 3 – Impact of feature engineering approach and machine learning algorithm on performance of machine learning models for polymyxin resistance prediction

Mean performance with 95% confidence intervals is shown across different performance metrics. Histograms showing how performance is impacted by A) feature engineering approach and B) choice of machine learning algorithm. Abbreviations: AUROC – Area under receiver-operator curve; bACC – Balanced accuracy; CUIMC – Columbia University Irving Medical Center; GBTC – Gradient Boosted Trees Classifier; GWAS – genome wide association study; SVC – Support Vector Machine Classifier.



Conclusion

In summary, this chapter was a showed that ML methods can be used to predict phenotypic PR in *K. pneumoniae* CG258 with high performance. This work contributes to the growing literature in AST genotype-phenotype prediction and highlights some of the issues the field faces.

Firstly, it shows that a 'one-size-fits-all' approach for different organisms and antimicrobial agents will be difficult to achieve due to the different biological underpinnings of AMR. This may be particularly true for highly complex polygenic forms of AMR, such as PR and vancomycin resistance in *S. aureus.* Secondly, genotype representation and feature selection rather than algorithm choice appear to be key to achieving high performance. From that perspective, use of GWAS as a filtering procedure may be applicable in other AST genotype-phenotype prediction settings. Finally, integration of genomic and clinical data represents an exciting new frontier for ML in healthcare and may play an important role in future efforts to improve the diagnosis, treatment and prevention of AMR.

Chapter 7:

Future directions

The implicit premise underlying this PhD is that in the foreseeable future genetic sequencing will be used as a ubiquitous part of the diagnosis and treatment of infectious diseases. This PhD therefore determined how we can use the novel tools afforded by advances in genetic sequencing and bioinformatics for studying multi-drug resistance in Gram negative bacteria, with a particular view towards impacting patient care. I noted several overarching themes.

The first was that WGS allowed me to identify an unexpected diversity in AMR, whether that is from the perspective of the MDROs or their mechanisms of resistance. This was illustrated by the findings on carbapenem resistance in the prospective study noted in Chapter 2. On a population level, I determined that CRE were far more diverse than previously thought. While the bulk of the literature on CRE epidemiology in the US has focused on the predominance of *K. pneumoniae* ST258 as the dominant clone [46], I noted not only emerging *K. pneumoniae* clones such as ST307 and ST392, but emerging carbapenem resistance in other organisms such as *E. coli* and *Enterobacter* spp. Some of the diversity detected was likely also due to the use of active surveillance, where we were able to note many subclinical colonization events. Furthermore, the use of WGS allowed us to identify putative mechanisms of carbapenem resistance and these appeared to correlate with risk of subsequent infection. This finding has important implications both for stratifying risk for individual patients, as well as better targeting

infection prevention interventions. My use of WGS to study PR similarly showed a remarkable complexity in the pathways to PR, with numerous variants and numerous genes implicated in individual isolates.

WGS also allowed me to appreciate that there is diversity in MDR-GNB on the patient level. This was shown by the different colonization trajectories in individual patients observed in Chapter 2. I was able to identify that patients are clearing colonization and having new colonization events much more frequently than previously thought. In addition, I noted the limitations of using phenotypic resistance categories to infer organism similarity: organisms in different phenotypic categories may be closely related, while organisms in the same phenotypic category may actually belong to different clades. With regards to PR, I showed that individual patients may have multiple pathways to PR suggesting that resistance is arising *de novo* rather than through clonal spread.

A second major theme was that the use of WGS to infer phylogenetic relationships will be crucial for understanding transmission of MDR-GNB in the healthcare setting. This was demonstrated at multiple points during this project. WGS was central to my detection of the earliest healthcare-associated cluster of *mcr-1*-carrying bacteria in the US, which arose out of work in Chapter 2 and became a subsequent study. In this instance, active surveillance allowed me to detect unusual late colonization events in the affected liver transplant recipients and WGS showed the presence of *mcr-1* and established the close phylogenetic relationship of the tested isolates, suggesting transmission. The use of WGS for phylogenetic analysis was also at the core of allowing me to determine that in most

patients, PR arose *de novo* rather than being due to healthcare-associated transmission. This finding will help shape future approaches at limiting the spread of PR, with a focus on antimicrobial stewardship as the primary intervention. Use of WGS for phylogenetic analysis also allowed me to track the regional emergence of *K. pneumoniae* ST307 as a novel multi-drug resistant clone, as described in Chapter 3.

A final theme was that new synergies between genomic data, clinical data and artificial intelligence approaches may radically alter the capabilities of the clinical microbiology laboratory. Many microbiological techniques in current practice have remained largely unchanged since the 19th century. The increasing availability of genetic sequencing is likely to alter this in several ways. On a basic level, as shown in Chapter 2, this will be through the increased resolution offered by WGS allowing more precise typing of organisms and molecular detection of mechanisms of resistance. With the advent of novel antimicrobial agents targeted at specific mechanisms of resistance, knowing the mechanism will become increasingly important. The next step after this basic usage will be the future foreshadowed by my work on using ML to predict PR in Chapter 6, where it is possible to develop increasingly accurate predictions through integrating complex genomic and clinical data in ways that go beyond what is offered by using simple rules such as the presence or absence of resistance determinants.

Despite much of the work in this thesis focusing on MDR-GNB isolates collected in the US, there are important implications of the findings for Australia. As noted in Chapter 1, CRE have become endemic in certain settings in Australia with *bla*$_{\text{IMP-4}}$ being the

dominant carbapenemase [178]. This is a metallo-beta-lactamase for which novel beta-lactam/beta-lactamase inhibitor combinations such as ceftazidime-avibactam are not active and polymyxins continue to be one of the only treatments. As the findings in Chapter 4 show, exposure to polymyxins is likely to result in PR and the underlying mechanisms are expected to be similar in *K. pneumoniae* regardless of the carbapenemase background. Furthermore, the recent bla_{KPC-2} outbreak in Victoria almost exclusively involved *K. pneumoniae* CG258 [189], thus making the work in this thesis immediately relevant as this is the same clonal background covered in Chapters 4 to 6 of the thesis. Finally, the discovery of *mcr-1* in Australian isolate collections makes the detection of *mcr-1* through active surveillance in Chapter 3 generalizable to the Australian setting and carry a particular relevance due to the high prevalence of *mcr-1* carriage in clinical isolates from neighbouring Asian countries [98, 146].

While these themes noted as part of my PhD have been very exciting developments, there has been ongoing innovation during this time. This has led to several novel therapeutics and technologies that will likely have a significant impact on the future diagnosis and treatment of AMR, which I will now focus on.

Novel therapeutics

Until recently, polymyxins were amongst the only treatments available for some MDR-GNB bacteria, in particular CRE, and thus PR was a focus of this PhD. Fortunately, the last decade has seen several novel agents developed and approved for use with activity against these organisms. These have largely included additions to existing classes

including novel beta-lactam/beta-lactamase inhibitor combinations (e.g. ceftazidime-avibactam), tetracyclines (e.g. eravacycline and omadacycline) and aminoglycosides (e.g. plazomicin). Cefiderocol is a possible exception: while being a cephalosporin, it has a novel siderophore mechanism of action that provides efficacy against a broad range of MDR-GNB pathogens [303].

The agents that have had the most significant clinical impact to date have been novel beta-lactam/beta-lactamase inhibitor combinations. The agents currently approved for clinical use in this class are ceftazidime-avibactam, meropenem-vaborbactam, imipenem-relebactam and ceftolozane-tazobactam. Apart from ceftolozane-tazobactam, other agents in this group have paired an existing beta-lactam agent with a novel beta-lactamase inhibitor that has good activity against serine beta-lactamases belonging to Ambler classes A and C and some activity against Ambler class D. Therefore the molecular mechanism underpinning carbapenem resistance has greater importance than for other beta-lactams and early detection of these mechanisms may assist in improving their use and avoiding delays typically associated with phenotypic susceptibility testing.

They have proven to be essential additions to the armamentarium for treating CRE infections, with possible mortality benefits over polymyxins [304]. However, resistance has already been documented with multiple mechanisms described and represents a new frontier of AMR with pan-drug resistance remaining an important threat [7]. Resistance to ceftazidime-avibactam appears to emerge in approximately 10% of treated patients [305]. Indeed, in Chapter 4 I noted that 4/11 patients with PR isolates who were treated with

ceftazidime-avibactam developed resistance. On the basis of those observations, our group has now begun preliminary work on ceftazidime-avibactam resistance [306].

We noted ceftazidime-avibactam resistance in several isolates from a single patient that had been included in the study in Chapter 4. In brief, the patient had KPC *K. pneumoniae* sepsis in the context of complicated pancreatitis requiring therapy with polymyxins, ceftazidime-avibactam and tigecycline. In the context of this treatment, we noted the emergence of two distinct phenotypes of ceftazidime-avibactam resistance. The first had high-level ceftazidime-avibactam resistance (MIC>256 mcg/ml) but meropenem susceptibility, conferred through a D179Y protein substitution in bla_{KPC-2}. The second phenotype had lower level ceftazidime-avibactam resistance (MIC 12 mcg/ml) but high-level meropenem resistance (MIC>128 mcg/ml). In this phenotype, use of long-read sequencing allowed us to identify amplification and transposition of wild-type bla_{KPC-2} into a novel plasmid, resulting in increased bla_{KPC-2} copy number. We also noted porin mutations not present in the first phenotype or the original 'wild type' isolates. Taken together, this work encapsulates all mechanisms of ceftazidime-avibactam resistance that have been identified to date [305, 307-317] and also shows the potential advantages of using long-read sequencing to dissect mechanisms of resistance.

Without extensive clinical use, reports of resistance to other classes of novel therapeutics have been limited. However, in vitro resistance has been noted in pre-clinical studies. Cefiderocol has the broadest spectrum of activity of the novel agents but resistance was noted in the SIDERO-WT-2014 and SIDERO-CR studies [318, 319], in particular amongst

bla$_{PER}$-producing *A. baumannii* and *bla*$_{NDM}$-producing Enterobacteriaceae. Interestingly, the resistance could be reversed with the addition of beta-lactamase inhibitors, suggesting that beta-lactamases were responsible for cefiderocol resistance [320].

The novel tetracyclines, omadacycline and eravacycline, have activity in the presence of the major mechanisms of tetracycline resistance including ribosomal protection (TetB and TetK) and efflux (TetO) [321, 322]. However they may be affected by less common mechanisms of tetracycline resistance such as enzymatic inactivation and target site mutations [322]. Similarly plazomicin, a novel aminoglycoside, can resist modification by typical aminoglycoside-modifying enzymes but may be affected by other aminoglycoside resistance mechanisms [321]. In particular, *bla*$_{NDM}$-producing Enterobacteriaceae often have co-production of a 16S ribosomal RNA methyltransferases that protect the aminoglycoside binding site of 16S rRNA and confer resistance to plazomicin and other aminoglycosides [323].

The clinical experience with novel beta-lactamase/beta-lactamase inhibitor combinations demonstrates that while novel therapeutic agents are promising, there are ongoing risks of resistance emerging as a function of evolutionary biology. This shows the limitations of traditional approaches to treating infectious diseases and has led to interest in novel paradigms. These include altering the microbiome in order to prevent or treat MDR-GNB colonization, with some evidence for use of faecal microbiota transplantation and further clinical trials currently underway [324-327]. Phage therapy has been another major area of interest, building on work done in the early 20th century prior to the widespread

availability of synthetic antimicrobial agents [328]. While resistance during use of phage therapy has been documented, there are other putative advantages of using phage therapy such as the potential to use them in conjunction with standard antimicrobials and the possibility that use of phage therapy may lead to reversion of resistance [329, 330]. There are now several case reports of use of phage therapy and clinical trials underway [328].

New sequencing approaches

In parallel to the development of these novel antimicrobials and interest in novel therapeutic approaches, we have seen the emergence of new sequencing technologies [331]. While long-read sequencing has been accessible for some time using single-molecule real-time technology from Pacific Biosciences, nanopore-based technologies such as those offered by Oxford Nanopore have become increasingly used in bacterial genomics and now constitute an indispensable tool. The putative advantages of long-read sequencing generally, and nanopore-based technologies more specifically, can be grouped into technical advantages and logistical advantages.

The technical advantages of long-read approaches lie in their ability to achieve reads of that can range from several kilobases to >1 megabase in length. This offers the ability to resolve repetitive regions and detect structural variants (e.g. gene loss and fusion events), significantly aiding genome assembly and allowing reconstruction of high quality *de novo* assembled bacterial genomes [332]. One drawback of long-read sequencing technologies is an increased error rate (5-10%) compared to short read (0.1%), therefore

there has been increasing interest in using hybrid assembly approaches that use both short- and long-read data such as that offered by Unicycler [332-334]. This approach was used to create reference genomes for variant calling in Chapters 3 – 6. From the perspective of AMR, these high quality assemblies have provided crucial information on MGEs such as plasmids, whose repetitive regions were significant barriers to assembly in the past. This ability to track plasmid dynamics is of vital importance to future work in genomic epidemiology and infection prevention and is the focus of recent publications [199, 224, 335].

The logistical advantages of long-read sequencing pertain most specifically to the Oxford MinION instrument. Firstly, the size of the instrument and the fact that it can connect as a USB device into a laptop computer has made it possible to consider using it in low-resource settings, with its use during the Ebola outbreak in West Africa in 2014 serving as an important proof-of-principle [336]. In addition to its small size, the MinION device has several other features that make it attractive for use in clinical settings. These include lower associated costs (both in terms of the sequencing device and consumables), the possibility of obtaining and interpreting results from the sequencer in real-time and also the ability to have runs with smaller numbers of samples than Illumina instruments [337, 338]. To date, attempts to utilize these technologies for AMR detection in clinical settings are in their infancy with only preliminary studies being conducted [339, 340], but ongoing decreases in cost and improvements in performance will likely result in greater use including in metagenomics applications [341-343].

Increased use of machine learning

Artificial intelligence (including ML) has been widely publicized as a technology with the potential to revolutionize healthcare [263] but there are few ML interventions that have reached clinical use. This is also true in the field of AMR. To date the most promising work has been done in AST genotype-phenotype prediction and this formed the basis for the work on predicting PR from genomic data in Chapter 6. However, the role of ML can be expanded further, giving the opportunity to personalize a variety of predictions about risk, diagnosis and treatment of AMR infections.

To date there have been several studies using ML approaches to predict AMR. In one instance, a recursive partitioning algorithm predicted bacteraemia with ESBL-producing organism with a positive predictive value of 91% and a negative predictive value of 92% [344]. In this study, LASSO regression was used to help select five features from a dataset of over 30 demographic and clinical variables to be used in the algorithm, which is an example of feature engineering similar to the work in Chapter 6. In another study, multivariable logistic regression models were compared to clinical decision trees in order to predict resistance to piperacillin-tazobactam, cefepime, and meropenem in patients with Gram negative bloodstream infection [345] and were able to correctly classify patients as low, medium and high-risk. Findings from both of these studies could therefore be used to tailor choice of empirical therapy for Gram negative bacteraemia and aid current antimicrobial stewardship approaches.

In addition to focusing on AMR specifically, several studies have tried to use ML methods to predict infection-related diagnoses more generally, whether this is in the setting of predicting infection in patients first presenting to the emergency department [346, 347] or predicting sepsis [348-352]. The use of clinical data in these studies presents several challenges. Firstly, there is the issue of which data should be used as features for prediction. Structured data such as vital signs, output from monitoring machines, drug administration data or laboratory test results form an obvious starting point and have been used in multiple studies [347, 349, 350, 352]. However, much information entered into electronic medical records is in the form of unstructured free text and will require natural language processing to be accessed and integrated into ML approaches. Horng et al. aimed to use free text in their prediction of infection and firstly used a simple 'bag of words' model then a topic model to represent data from the free text, with similar performance for both [346].

Secondly, reliable labels also present a challenge: positive cultures do not necessarily represent infection episodes, there are multiple criteria for defining sepsis, and ICD codes in electronic medical records are primarily used for billing purposes leading to possible inaccuracies [263]. Manual assessment of patient records therefore remains the gold standard and was used by Rawson et al. to define infective episodes in their prospective study [347]. However, this approach obviously does not scale well, and this presents a significant limitation for use of ML approaches in infectious diseases. Indeed, this challenge represents an active area of research in clinical bioinformatics more generally, (termed 'phenotyping') and has led to efforts such as the eMERGE (Electronic Medical

Records and Genomics) network that aims to create reliable phenotypes from electronic medical records that can be linked to genomic data [353].

Improving prediction of AMR infections through ML will play an important role in optimizing their treatment. Currently a ‘one-size-fits-all’ approach to AMR treatment is used that leads to poorer patient outcomes, waste of precious infection prevention resources and inappropriate antimicrobial use that perpetuates the cycle of AMR. Better prediction could guide when to test for an AMR organism, when to place in contact precautions while results are awaited, and what empirical antimicrobial therapy should be prescribed. Furthermore, ML may also help personalize treatment by better identifying subsets of patients at risk of treatment failure. Seymour et al. recently were able to use ML approaches to discover four novel sepsis phenotypes with distinct demographics, laboratory values, and patterns of organ dysfunction [354]. These phenotypes correlated with biomarkers and mortality, showing how clinical entities may be more complex than traditionally appreciated with important implications for treatment and outcomes. Li et al. used electronic medical record data to stratify patients with *Clostridium difficile* infection according to risk of developing complications, which may in turn allow clinicians to better target treatments [355]. While these studies did not focus on specifically on AMR infections, patient- and pathogen-related factors could similarly be integrated in ML approaches to impact AMR treatment and outcomes [356].

Final remarks

With these themes in mind, the major areas of focus for future research from my PhD include 1) expanding the use of AST genotype-phenotype prediction in the clinical setting through real-time use with other organisms and forms of AMR; 2) facilitating detection of healthcare-associated transmission of AMR by automating genomic phylogenetic analysis and integrating clinical data from electronic medical records, 3) assessing the clinical feasibility of ML for other forms of prediction related to AMR including rapidly identifying AMR and predicting treatment responses through use of genomics and ML. This is an exciting era with numerous technological advancements coalescing to potentially fundamentally change how we practice infectious diseases and microbiology.

I have been privileged to conduct research as part of this PhD that is very much at this crossroads. I hope the insights from my PhD and future work contribute to the next phase of implementation where these technologies become increasingly used in the clinical setting to prevent the spread of AMR and improve the diagnosis and treatment of AMR infections.

References

1. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. Available at: https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf. Accessed July 16, 2019.
2. World Health Organisation. Antimicrobial Resistance: Global Report on Surveillance Geneva, Switzerland: World Health Organisation, **2014**.
3. Peleg AY, Hooper DC. Hospital-acquired infections due to gram-negative bacteria. N Engl J Med **2010**; 362(19): 1804-13.
4. Nguyen M, Eschenauer GA, Bryan M, et al. Carbapenem-resistant *Klebsiella pneumoniae* bacteremia: factors correlated with clinical and microbiologic outcomes. Diagnostic microbiology and infectious disease **2010**; 67(2): 180-4.
5. Magiorakos AP, Srinivasan A, Carey RB, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect **2011**; 18(3): 268-81.
6. Logan LK, Weinstein RA. The Epidemiology of Carbapenem-Resistant Enterobacteriaceae: The Impact and Evolution of a Global Menace. J Infect Dis **2017**; 215(suppl_1): S28-S36.
7. Sherry N, Howden B. Emerging Gram negative resistance to last-line antimicrobial agents fosfomycin, colistin and ceftazidime-avibactam – epidemiology, laboratory detection and treatment implications. Expert Rev Anti Infect Ther **2018**; 16(4): 289-306.
8. Chen L, Todd R, Kiehlbauch J, Walters M, Kallen A. 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): 33.
9. Krapp F, Ozer EA, Qi C, Hauser AR. Case Report of an Extensively Drug-Resistant Klebsiella pneumoniae Infection With Genomic Characterization of the Strain and Review of Similar Cases in the United States. Open Forum Infect Dis **2018**; 5(5): ofy074.

10. Kwong JC, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. Pathology **2015**; 47(3): 199-210.

11. Pak TR, Kasarskis A. How next-generation sequencing and multiscale data analysis will transform infectious disease management. Clin Infect Dis **2015**; 61(11): 1695-702.

12. Macesic N, Polubriaginof F, Tatonetti NP. Machine learning: novel bioinformatics approaches for combating antimicrobial resistance. Curr Opin Infect Dis **2017**; 30(6): 511-7.

13. Infectious Diseases Society of America. The 10 x '20 Initiative: pursuing a global commitment to develop 10 new antibacterial drugs by 2020. Clin Infect Dis **2010**; 50(8): 1081-3.

14. Kaye KS, Bhowmick T, Metallidis S, et al. Effect of Meropenem-Vaborbactam vs Piperacillin-Tazobactam on Clinical Cure or Improvement and Microbial Eradication in Complicated Urinary Tract Infection: The TANGO I Randomized Clinical Trial. JAMA **2018**; 319(8): 788-99.

15. Lucasti C, Vasile L, Sandesc D, et al. Phase 2, Dose-Ranging Study of Relebactam with Imipenem-Cilastatin in Subjects with Complicated Intra-abdominal Infection. Antimicrob Agents Chemother **2016**; 60(10): 6234-43.

16. Torres A, Zhong N, Pachl J, et al. Ceftazidime-avibactam versus meropenem in nosocomial pneumonia, including ventilator-associated pneumonia (REPROVE): a randomised, double-blind, phase 3 non-inferiority trial. Lancet Infect Dis **2018**; 18(3): 285-95.

17. Solomkin J, Evans D, Slepavicius A, et al. Assessing the Efficacy and Safety of Eravacycline vs Ertapenem in Complicated Intra-abdominal Infections in the Investigating Gram-Negative Infections Treated With Eravacycline (IGNITE 1) Trial: A Randomized Clinical Trial. JAMA Surg **2017**; 152(3): 224-32.

18. Stets R, Popescu M, Gonong JR, et al. Omadacycline for Community-Acquired Bacterial Pneumonia. N Engl J Med **2019**; 380(6): 517-27.

19. Wagenlehner FME, Cloutier DJ, Komirenko AS, et al. Once-Daily Plazomicin for Complicated Urinary Tract Infections. N Engl J Med **2019**; 380(8): 729-40.

20. Pullman J, Gardovskis J, Farley B, et al. Efficacy and safety of delafloxacin compared with vancomycin plus aztreonam for acute bacterial skin and skin structure infections: a Phase 3, double-blind, randomized study. J Antimicrob Chemother **2017**; 72(12): 3471-80.

21. Portsmouth S, van Veenhuyzen D, Echols R, et al. Cefiderocol versus imipenem-cilastatin for the treatment of complicated urinary tract infections caused by Gram-negative uropathogens: a phase 2, randomised, double-blind, non-inferiority trial. Lancet Infect Dis **2018**; 18(12): 1319-28.

22. de Gouvea EF, Martins IS, Halpern M, et al. The influence of carbapenem resistance on mortality in solid organ transplant recipients with *Acinetobacter baumannii* infection. BMC Infect Dis **2012**; 12: 351.

23. Freire MP, Abdala E, Moura ML, et al. Risk factors and outcome of infections with Klebsiella pneumoniae carbapenemase-producing K. pneumoniae in kidney transplant recipients. Infection **2015**; 43(3): 315-23.

24. Freire MP, Oshiro IC, Pierrotti LC, et al. Carbapenem-Resistant Enterobacteriaceae Acquired Before Liver Transplantation: Impact on Recipient Outcomes. Transplantation **2017**; 101(4): 811-20.

25. Bartoletti M, Giannella M, Tedeschi S, Viale P. Multidrug-Resistant Bacterial Infections in Solid Organ Transplant Candidates and Recipients. Infect Dis Clin North Am **2018**; 32(3): 551-80.

26. Pouch SM, Satlin MJ. Carbapenem-resistant Enterobacteriaceae in special populations: Solid organ transplant recipients, stem cell transplant recipients, and patients with hematologic malignancies. Virulence **2017**; 8(4): 391-402.

27. Rex JH, Talbot GH, Goldberger MJ, et al. Progress in the Fight Against Multidrug-Resistant Bacteria 2005–2016: Modern Noninferiority Trial Designs Enable Antibiotic Development in Advance of Epidemic Bacterial Resistance. Clin Infect Dis **2017**; 65(1): 141-6.

28. Kadri SS, Adjemian J, Lai YL, et al. Difficult-to-Treat Resistance in Gram-negative Bacteremia at 173 US Hospitals: Retrospective Cohort Analysis of Prevalence, Predictors, and Outcome of Resistance to All First-line Agents. Clin Infect Dis **2018**; 67(12): 1803-14.

29. McDonnell A, Rex JH, Goossens H, Bonten M, Fowler Jr VG, Dane A. Efficient Delivery of Investigational Antibacterial Agents via Sustainable Clinical Trial Networks. Clin Infect Dis **2016**; 63(suppl 2): S57-S9.

30. Liu Y-Y, Wang Y, Walsh TR, et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. Lancet Infect Dis Vol. 16, **2016**:161-8.

31. Tamma PD, Goodman KE, Harris AD, et al. Comparing the Outcomes of Patients With Carbapenemase-Producing and Non-Carbapenemase-Producing Carbapenem-Resistant Enterobacteriaceae Bacteremia. Clin Infect Dis **2017**; 64(3): 257-64.

32. Centers for Disease Control and Prevention. Carbapenem-resistant Enterobacteriaceae in Healthcare Settings. Available at: https://www.cdc.gov/hai/organisms/cre/index.html. Accessed July 16, 2019.

33. Queenan AM, Bush K. Carbapenemases: the versatile beta-lactamases. Clin Microbiol Rev **2007**; 20(3): 440-58.

34. Gupta N, Limbago BM, Patel JB, Kallen AJ. Carbapenem-resistant Enterobacteriaceae: epidemiology and prevention. Clin Infect Dis **2011**; 53(1): 60-7.

35. Rhomberg PR, Jones RN. Summary trends for the Meropenem Yearly Susceptibility Test Information Collection Program: a 10-year experience in the United States (1999-2008). Diagnostic microbiology and infectious disease **2009**; 65(4): 414-26.

36. Hidron AI, Edwards JR, Patel J, et al. NHSN annual update: antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006-2007. Infect Control Hosp Epidemiol **2008**; 29(11): 996-1011.

37. Bradford PA, Bratu S, Urban C, et al. Emergence of carbapenem-resistant Klebsiella species possessing the class A carbapenem-hydrolyzing KPC-2 and inhibitor-resistant TEM-30 beta-lactamases in New York City. Clin Infect Dis **2004**; 39(1): 55-60.

38. Bratu S, Landman D, Alam M, Tolentino E, Quale J. Detection of KPC carbapenem-hydrolyzing enzymes in Enterobacter spp. from Brooklyn, New York. Antimicrob Agents Chemother **2005**; 49(2): 776-8.

39. Bratu S, Mooty M, Nichani S, et al. Emergence of KPC-possessing Klebsiella pneumoniae in Brooklyn, New York: epidemiology and recommendations for detection. Antimicrob Agents Chemother **2005**; 49(7): 3018-20.

40. Hall BG, Barlow M. Revised Ambler classification of {beta}-lactamases. J Antimicrob Chemother **2005**; 55(6): 1050-1.

41. Bush K, Jacoby GA. Updated functional classification of beta-lactamases. Antimicrob Agents Chemother **2010**; 54(3): 969-76.

42. Nordmann P, Cuzon G, Naas T. The real threat of Klebsiella pneumoniae carbapenemase-producing bacteria. Lancet Infect Dis **2009**; 9(4): 228-36.

43. Kukla R, Chudejova K, Papagiannitsis CC, et al. Characterization of KPC-Encoding Plasmids from Enterobacteriaceae Isolated in a Czech Hospital. Antimicrob Agents Chemother **2018**; 62(3): 2133-7.

44. Leavitt A, Carmeli Y, Chmelnitsky I, Goren MG, Ofek I, Navon-Venezia S. Molecular epidemiology, sequence types, and plasmid analyses of KPC-producing Klebsiella pneumoniae strains in Israel. Antimicrob Agents Chemother **2010**; 54(7): 3002-6.

45. Tzouvelekis LS, Markogiannakis A, Psichogiou M, Tassios PT, Daikos GL. Carbapenemases in Klebsiella pneumoniae and Other Enterobacteriaceae: an Evolving Crisis of Global Dimensions. Clin Microbiol Rev **2012**; 25(4): 682-707.

46. Munoz-Price LS, Poirel L, Bonomo RA, et al. Clinical epidemiology of the global expansion of Klebsiella pneumoniae carbapenemases. Lancet Infect Dis **2013**; 13(9): 785-96.

47. Annavajhala MK, Gomez-Simmonds A, Uhlemann AC. Multidrug-Resistant Enterobacter cloacae Complex Emerging as a Global, Diversifying Threat. Front Microbiol **2019**; 10: 44.

48. Britt NS, Ritchie DJ, Kollef MH, et al. Clinical epidemiology of carbapenem-resistant gram-negative sepsis among hospitalized patients: Shifting burden of disease? Am J Infect Control **2018**; 46(10): 1092-6.

49. Pitout JDD, Nordmann P, Poirel L. Carbapenemase-Producing Klebsiella pneumoniae, a Key Pathogen Set for Global Nosocomial Dominance. Antimicrob Agents Chemother **2015**; 59(10): 5873-84.

50. Deleo FR, Chen L, Porcella SF, et al. Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 Klebsiella pneumoniae. Proc Natl Acad Sci U S A **2014**; 111(13): 4988-93.

51. Adler A, Khabra E, Chmelnitsky I, et al. Development and validation of a multiplex PCR assay for identification of the epidemic ST-258/512 KPC-producing Klebsiella pneumoniae clone. Diagnostic microbiology and infectious disease **2014**; 78(1): 12-5.

52. Chen L, Mathema B, Pitout JD, DeLeo FR, Kreiswirth BN. Epidemic Klebsiella pneumoniae ST258 is a hybrid strain. mBio **2014**; 5(3): e01355-14.

53. Centers for Disease Control and Prevention. Tracking CRE. Available at: https://www.cdc.gov/hai/organisms/cre/trackingcre.html. Accessed July 16, 2019.

54. Leavitt A, Navon-Venezia S, Chmelnitsky I, Schwaber MJ, Carmeli Y. Emergence of KPC-2 and KPC-3 in carbapenem-resistant Klebsiella pneumoniae strains in an Israeli hospital. Antimicrob Agents Chemother **2007**; 51(8): 3026-9.

55. Giani T, Pini B, Arena F, et al. Epidemic diffusion of KPC carbapenemase-producing Klebsiella pneumoniae in Italy: results of the first countrywide survey, 15 May to 30 June 2011. Euro Surveill **2013**; 18(22): 2-10.

56. European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe. Available at: https://ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/antimicrobial-resistance-europe-2015.pdf. Accessed July 16, 2019.

57. Galani I, Karaiskos I, Karantani I, et al. Epidemiology and resistance phenotypes of carbapenemase-producing Klebsiella pneumoniae in Greece, 2014 to 2016. Euro Surveill **2018**; 23(31): 8023-12.

58. Villegas MV, Lolans K, Correa A, et al. First detection of the plasmid-mediated class A carbapenemase KPC-2 in clinical isolates of Klebsiella pneumoniae from South America. Antimicrob Agents Chemother **2006**; 50(8): 2880-2.

59. Maya JJ, Ruiz SJ, Blanco VM, et al. Current status of carbapenemases in Latin America. Expert Rev Anti Infect Ther **2014**; 11(7): 657-67.

60. Walther-Rasmussen J, Høiby N. OXA-type carbapenemases. J Antimicrob Chemother **2006**; 57(3): 373-83.

61. Poirel L, Heritier C, Tolun V, Nordmann P. Emergence of Oxacillinase-Mediated Resistance to Imipenem in Klebsiella pneumoniae. Antimicrob Agents Chemother **2003**; 48(1): 15-22.

62. Carrer A, Poirel L, Yilmaz M, et al. Spread of OXA-48-encoding plasmid in Turkey and beyond. Antimicrob Agents Chemother **2010**; 54(3): 1369-73.

63. Poirel L, Bonnin RA, Nordmann P. Genetic Features of the Widespread Plasmid Coding for the Carbapenemase OXA-48. Antimicrob Agents Chemother **2011**; 56(1): 559-62.

64. Poirel L, Naas T, Nordmann P. Diversity, epidemiology, and genetics of class D beta-lactamases. Antimicrob Agents Chemother **2010**; 54(1): 24-38.

65. Lyman M, Walters M, Lonsway D, Rasheed K, Limbago B, Kallen A. Notes from the Field: Carbapenem-resistant Enterobacteriaceae Producing OXA-48-like Carbapenemases--United States, 2010-2015. MMWR Morb Mortal Wkly Rep **2015**; 64(47): 1315-6.

66. Girmenia C, Serrao A, Canichella M. Epidemiology of Carbapenem Resistant Klebsiella pneumoniae Infections in Mediterranean Countries. Mediterr J Hematol Infect Dis **2016**; 8(1): e2016032.

67. Ito H, Arakawa Y, Ohsuka S, Wacharotayankun R, Kato N, Ohta M. Plasmid-mediated dissemination of the metallo-beta-lactamase gene blaIMP among clinically isolated strains of Serratia marcescens. Antimicrob Agents Chemother **1995**; 39(4): 824-9.

68. Yan JJ. Metallo-beta-lactamase-producing Enterobacteriaceae isolates in a university hospital in Taiwan: prevalence of IMP-8 in Enterobacter cloacae and first identification of VIM-2 in Citrobacter freundii. J Antimicrob Chemother **2002**; 50(4): 503-11.

69. Yong D, Toleman MA, Giske CG, 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): 5046-54.

70. Kumarasamy KK, Toleman MA, Walsh TR, et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect Dis **2010**; 10(9): 597-602.

71. Karlowsky JA, Lob SH, Kazmierczak KM, et al. In Vitro Activity of Imipenem against Carbapenemase-Positive Enterobacteriaceae Isolates Collected by the SMART Global Surveillance Program from 2008 to 2014. J Clin Microbiol **2017**; 55(6): 1638-49.

72. Eichenberger EM, Thaden JT. Epidemiology and Mechanisms of Resistance of Extensively Drug Resistant Gram-Negative Bacteria. Antibiotics (Basel) **2019**; 8(2): 37-21.

73. Walsh TR, Weeks J, Livermore DM, Toleman MA. Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study. Lancet Infect Dis **2011**; 11(5): 355-62.

74. Wu W, Feng Y, Tang G, Qiao F, McNally A, Zong Z. NDM Metallo-β-Lactamases and Their Bacterial Producers in Health Care Settings. Clin Microbiol Rev **2019**; 32(2): 588-45.

75. Peleg AY, Franklin C, Bell JM, Spelman DW. Dissemination of the metallo-beta-lactamase gene blaIMP-4 among gram-negative pathogens in a clinical setting in Australia. Clin Infect Dis **2005**; 41(11): 1549-56.

76. Wei Z, Yu T, Qi Y, et al. Coexistence of plasmid-mediated KPC-2 and IMP-4 carbapenemases in isolates of Klebsiella pneumoniae from China. J Antimicrob Chemother **2011**; 66(11): 2670-1.

77. Jacoby GA. AmpC beta-lactamases. Clin Microbiol Rev **2009**; 22(1): 161-82, Table of Contents.

78. Goodman KE, Simner PJ, Tamma PD, Milstone AM. Infection control implications of heterogeneous resistance mechanisms in carbapenem-resistant Enterobacteriaceae (CRE). Expert Rev Anti Infect Ther **2016**; 14(1): 95-108.

79. Satlin MJ, Chen L, Patel G, et al. Multicenter Clinical and Molecular Epidemiological Analysis of Bacteremia Due to Carbapenem-Resistant Enterobacteriaceae (CRE) in the CRE Epicenter of the United States. Antimicrob Agents Chemother **2017**; 61(4): 1-13.

80. Li X-Z, Plésiat P, Nikaido H. The Challenge of Efflux-Mediated Antibiotic Resistance in Gram-Negative Bacteria. Clin Microbiol Rev **2015**; 28(2): 337-418.

81. Wozniak A, Villagra NA, Undabarrena A, et al. Porin alterations present in non-carbapenemase-producing Enterobacteriaceae with high and intermediate levels of carbapenem resistance in Chile. Journal of Medical Microbiology **2012**; 61(Pt_9): 1270-9.

82. Poirel L, Jayol A, Nordmann P. Polymyxins: Antibacterial Activity, Susceptibility Testing, and Resistance Mechanisms Encoded by Plasmids or Chromosomes. Clin Microbiol Rev **2017**; 30(2): 557-96.

83. Aquilini E, Merino S, Knirel YA, Regue M, Tomas JM. Functional identification of Proteus mirabilis eptC gene encoding a core lipopolysaccharide phosphoethanolamine transferase. Int J Mol Sci **2014**; 15(4): 6689-702.

84. Jiang SS, Liu MC, Teng LJ, Wang WB, Hsueh PR, Liaw SJ. Proteus mirabilis pmrI, an RppA-regulated gene necessary for polymyxin B resistance, biofilm formation, and urothelial cell invasion. Antimicrob Agents Chemother **2010**; 54(4): 1564-71.

85. Lin QY, Tsai YL, Liu MC, Lin WC, Hsueh PR, Liaw SJ. Serratia marcescens arn, a PhoP-regulated locus necessary for polymyxin B resistance. Antimicrob Agents Chemother **2014**; 58(9): 5181-90.

86. Gales AC, Jones RN, Sader HS. Contemporary activity of colistin and polymyxin B against a worldwide collection of Gram-negative pathogens: results from the SENTRY Antimicrobial Surveillance Program (2006-09). J Antimicrob Chemother **2011**; 66(9): 2070-4.

87. Bratu S, Tolaney P, Karumudi U, et al. Carbapenemase-producing Klebsiella pneumoniae in Brooklyn, NY: molecular epidemiology and in vitro activity of polymyxin B and other agents. J Antimicrob Chemother **2005**; 56(1): 128-32.

88. Rojas LJ, Salim M, Cober E, et al. Colistin Resistance in Carbapenem-Resistant Klebsiella pneumoniae: Laboratory Detection and Impact on Mortality. Clin Infect Dis **2017**; 64(6): 711-8.

89. Sutherland CA, Nicolau DP. Susceptibility Profile of Ceftolozane/Tazobactam and Other Parenteral Antimicrobials Against Escherichia coli, Klebsiella pneumoniae, and Pseudomonas aeruginosa From US Hospitals. Clin Ther **2015**; 37(7): 1564-71.

90. Tijet N, Sheth PM, Lastovetska O, Chung C, Patel SN, Melano RG. Molecular characterization of Klebsiella pneumoniae carbapenemase (KPC)-producing Enterobacteriaceae in Ontario, Canada, 2008-2011. PLoS One **2014**; 9(12): e116421.

91. Monaco M, Giani T, Raffone M, et al. Colistin resistance superimposed to endemic carbapenem-resistant Klebsiella pneumoniae: a rapidly evolving problem in Italy, November 2013 to April 2014. Euro Surveill **2014**; 19(42).

92. Meletis G, Oustas E, Botziori C, Kakasi E, Koteli A. Containment of carbapenem resistance rates of Klebsiella pneumoniae and Acinetobacter baumannii in a Greek hospital with a concomitant increase in colistin, gentamicin and tigecycline resistance. New Microbiol **2015**; 38(3): 417-21.

93. Zagorianou A, Sianou E, Iosifidis E, et al. Microbiological and molecular characteristics of carbapenemase-producing Klebsiella pneumoniae endemic in a tertiary Greek hospital during 2004-2010. Euro Surveill **2012**; 17(7).

94. Chiu SK, Wu TL, Chuang YC, et al. National surveillance study on carbapenem non-susceptible Klebsiella pneumoniae in Taiwan: the emergence and rapid dissemination of KPC-2 carbapenemase. PLoS One **2013**; 8(7): e69428.

95. Tan TY, Ng SY. The in-vitro activity of colistin in gram-negative bacteria. Singapore Med J **2006**; 47(7): 621-4.

96. Suh JY, Son JS, Chung DR, Peck KR, Ko KS, Song JH. Nonclonal emergence of colistin-resistant Klebsiella pneumoniae isolates from blood samples in South Korea. Antimicrob Agents Chemother **2010**; 54(1): 560-2.

97. Jayol A, Poirel L, Dortet L, Nordmann P. National survey of colistin resistance among carbapenemase-producing Enterobacteriaceae and outbreak caused by colistin-resistant OXA-48-producing Klebsiella pneumoniae, France, 2014. Euro Surveill **2016**; 21(37).

98. Ellem JA, Ginn AN, Chen SC, Ferguson J, Partridge SR, Iredell JR. Locally Acquired mcr-1 in Escherichia coli, Australia, 2011 and 2013. Emerg Infect Dis **2017**; 23(7): 1160-3.

99. Giani T, Arena F, Vaggelli G, et al. Large Nosocomial Outbreak of Colistin-Resistant, Carbapenemase-Producing Klebsiella pneumoniae Traced to Clonal Expansion of an mgrB Deletion Mutant. J Clin Microbiol **2015**; 53(10): 3341-4.

100. Mammina C, Bonura C, Di Bernardo F, et al. Ongoing spread of colistin-resistant Klebsiella pneumoniae in different wards of an acute general hospital, Italy, June to December 2011. Euro Surveill **2012**; 17(33).

101. Mezzatesta ML, Gona F, Caio C, et al. Outbreak of KPC-3-producing, and colistin-resistant, Klebsiella pneumoniae infections in two Sicilian hospitals. Clin Microbiol Infect **2011**; 17(9): 1444-7.

102. Antoniadou A, Kontopidou F, Poulakou G, et al. Colistin-resistant isolates of Klebsiella pneumoniae emerging in intensive care unit patients: first report of a multiclonal cluster. J Antimicrob Chemother **2007**; 59(4): 786-90.

103. Kontopoulou K, Protonotariou E, Vasilakos K, et al. Hospital outbreak caused by Klebsiella pneumoniae producing KPC-2 beta-lactamase resistant to colistin. J Hosp Infect **2010**; 76(1): 70-3.

104. Weterings V, Zhou K, Rossen JW, et al. An outbreak of colistin-resistant Klebsiella pneumoniae carbapenemase-producing Klebsiella pneumoniae in the Netherlands (July to December 2013), with inter-institutional spread. Eur J Clin Microbiol Infect Dis **2015**; 34(8): 1647-55.

105. Garza-Ramos U, Barrios H, Reyna-Flores F, et al. Characteristics of KPC-2-producing Klebsiella pneumoniae (ST258) clinical isolates from outbreaks in 2 Mexican medical centers. Diagnostic microbiology and infectious disease **2014**; 79(4): 483-5.

106. Marchaim D, Chopra T, Pogue JM, et al. Outbreak of colistin-resistant, carbapenem-resistant Klebsiella pneumoniae in metropolitan Detroit, Michigan. Antimicrob Agents Chemother **2011**; 55(2): 593-9.

107. Bogdanovich T, Adams-Haduch JM, Tian GB, et al. Colistin-resistant, Klebsiella pneumoniae carbapenemase (KPC)-producing Klebsiella pneumoniae belonging to the international epidemic clone ST258. Clin Infect Dis **2011**; 53(4): 373-6.

108. Cannatelli A, D'Andrea MM, Giani T, 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): 5521-6.

109. Cannatelli A, Giani T, D'Andrea MM, et al. MgrB inactivation is a common mechanism of colistin resistance in KPC-producing Klebsiella pneumoniae of clinical origin. Antimicrob Agents Chemother **2014**; 58(10): 5696-703.

110. Lippa AM, Goulian M. Feedback inhibition in the PhoQ/PhoP signaling system by a membrane peptide. PLoS Genet **2009**; 5(12): e1000788.

111. Olaitan AO, Diene SM, Kempf M, 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): 500-7.

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

113. Pitt ME, Elliott AG, Cao MD, et al. Multifactorial chromosomal variants regulate polymyxin resistance in extensively drug-resistant Klebsiella pneumoniae. Microb Genom **2018**; 4(3).

114. Aires CA, Pereira PS, Asensi MD, Carvalho-Assef AP. mgrB Mutations Mediating Polymyxin B Resistance in Klebsiella pneumoniae Isolates from Rectal Surveillance Swabs in Brazil. Antimicrob Agents Chemother **2016**; 60(11): 6969-72.

115. Berglund B, Hoang NTB, Tarnberg M, et al. Insertion sequence transpositions and point mutations in mgrB causing colistin resistance in a clinical strain of carbapenem-resistant Klebsiella pneumoniae from Vietnam. Int J Antimicrob Agents **2018**; 51(5): 789-93.

116. Giordano C, Barnini S, Tsioutis C, et al. Expansion of KPC-producing Klebsiella pneumoniae with various mgrB mutations giving rise to colistin resistance: the role of ISL3 on plasmids. Int J Antimicrob Agents **2018**; 51(2): 260-5.

117. Wright MS, Suzuki Y, Jones MB, 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): 536-43.

118. Cheng YH, Lin TL, Lin YT, Wang JT. Amino Acid Substitutions of CrrB Responsible for Resistance to Colistin through CrrC in Klebsiella pneumoniae. Antimicrob Agents Chemother **2016**; 60(6): 3709-16.

119. Cheng YH, Lin TL, Lin YT, Wang JT. A putative RND-type efflux pump, H239_3064, contributes to colistin resistance through CrrB in Klebsiella pneumoniae. J Antimicrob Chemother **2018**; 73(6): 1509-16.

120. Groisman EA. The pleiotropic two-component regulatory system PhoP-PhoQ. J Bacteriol **2001**; 183(6): 1835-42.

121. Park SY, Groisman EA. Signal-specific temporal response by the Salmonella PhoP/PhoQ regulatory system. Mol Microbiol **2014**; 91(1): 135-44.

122. Gunn JS. The Salmonella PmrAB regulon: lipopolysaccharide modifications, antimicrobial peptide resistance and more. Trends Microbiol **2008**; 16(6): 284-90.

123. Cannatelli A, Di Pilato V, Giani T, et al. In vivo evolution to colistin resistance by PmrB sensor kinase mutation in KPC-producing Klebsiella pneumoniae is associated with low-dosage colistin treatment. Antimicrob Agents Chemother **2014**; 58(8): 4399-403.

124. Cheng YH, Lin TL, Pan YJ, Wang YP, Lin YT, Wang JT. Colistin resistance mechanisms in Klebsiella pneumoniae strains from Taiwan. Antimicrob Agents Chemother **2015**; 59(5): 2909-13.

125. Jayol A, Poirel L, Brink A, Villegas MV, Yilmaz M, Nordmann P. Resistance to colistin associated with a single amino acid change in protein PmrB among Klebsiella pneumoniae isolates of worldwide origin. Antimicrob Agents Chemother **2014**; 58(8): 4762-6.

126. Olaitan AO, Dia NM, Gautret P, et al. Acquisition of extended-spectrum cephalosporin- and colistin-resistant Salmonella enterica subsp. enterica serotype Newport by pilgrims during Hajj. Int J Antimicrob Agents **2015**; 45(6): 600-4.

127. Sun S, Negrea A, Rhen M, Andersson DI. Genetic analysis of colistin resistance in Salmonella enterica serovar Typhimurium. Antimicrob Agents Chemother **2009**; 53(6): 2298-305.

128. De Majumdar S, Yu J, Fookes M, et al. Elucidation of the RamA regulon in Klebsiella pneumoniae reveals a role in LPS regulation. PLoS Pathog **2015**; 11(1): e1004627.

129. Telke AA, Olaitan AO, Morand S, Rolain JM. soxRS induces colistin hetero-resistance in Enterobacter asburiae and Enterobacter cloacae by regulating the acrAB-tolC efflux pump. J Antimicrob Chemother **2017**; 72(10): 2715-21.

130. Moffatt JH, Harper M, Harrison P, et al. Colistin resistance in Acinetobacter baumannii is mediated by complete loss of lipopolysaccharide production. Antimicrob Agents Chemother **2010**; 54(12): 4971-7.

131. Llobet E, Martinez-Moliner V, Moranta D, et al. Deciphering tissue-induced Klebsiella pneumoniae lipid A structure. Proc Natl Acad Sci U S A **2015**; 112(46): E6369-78.

132. Yan A, Guan Z, Raetz CR. An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in Escherichia coli. J Biol Chem **2007**; 282(49): 36077-89.

133. Kadar B, Kocsis B, Toth A, et al. Colistin resistance associated with outer membrane protein change in Klebsiella pneumoniae and Enterobacter asburiae. Acta Microbiol Immunol Hung **2017**; 64(2): 217-27.

134. Llobet E, March C, Gimenez P, Bengoechea JA. Klebsiella pneumoniae OmpA confers resistance to antimicrobial peptides. Antimicrob Agents Chemother **2009**; 53(1): 298-302.

135. Snesrud E, Ong AC, Corey B, et al. Analysis of Serial Isolates of mcr-1-Positive Escherichia coli Reveals a Highly Active ISApl1 Transposon. Antimicrob Agents Chemother **2017**; 61(5).

136. Snesrud E, He S, Chandler M, et al. A Model for Transposition of the Colistin Resistance Gene mcr-1 by ISApl1. Antimicrob Agents Chemother **2016**; 60(11): 6973-6.

137. Kieffer N, Nordmann P, Poirel L. Moraxella Species as Potential Sources of MCR-Like Polymyxin Resistance Determinants. Antimicrob Agents Chemother **2017**; 61(6).

138. AbuOun M, Stubberfield EJ, Duggett NA, et al. mcr-1 and mcr-2 variant genes identified in Moraxella species isolated from pigs in Great Britain from 2014 to 2015. J Antimicrob Chemother **2017**; 72(10): 2745-9.

139. Nang SC, Li J, Velkov T. The rise and spread of mcr plasmid-mediated polymyxin resistance. Crit Rev Microbiol **2019**; 45(2): 131-61.

140. Matamoros S, van Hattem JM, Arcilla MS, et al. Global phylogenetic analysis of Escherichia coli and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction. Sci Rep **2017**; 7(1): 15364.

141. Berrazeg M, Hadjadj L, Ayad A, Drissi M, Rolain JM. First Detected Human Case in Algeria of mcr-1 Plasmid-Mediated Colistin Resistance in a 2011 Escherichia coli Isolate. Antimicrob Agents Chemother **2016**; 60(11): 6996-7.

142. Doumith M, Godbole G, Ashton P, et al. Detection of the plasmid-mediated mcr-1 gene conferring colistin resistance in human and food isolates of Salmonella enterica and Escherichia coli in England and Wales. J Antimicrob Chemother **2016**; 71(8): 2300-5.

143. Aires CAM, da Conceicao-Neto OC, Tavares EOTR, et al. Emergence of the Plasmid-Mediated mcr-1 Gene in Clinical KPC-2-Producing Klebsiella pneumoniae Sequence Type 392 in Brazil. Antimicrob Agents Chemother **2017**; 61(7).

144. Kluytmans-van den Bergh MF, Huizinga P, Bonten MJ, et al. Presence of mcr-1-positive Enterobacteriaceae in retail chicken meat but not in humans in the Netherlands since 2009. Euro Surveill **2016**; 21(9): 30149.

145. Paveenkittiporn W, Kerdsin A, Chokngam S, Bunthi C, Sangkitporn S, Gregory CJ. Emergence of plasmid-mediated colistin resistance and New Delhi metallo-beta-lactamase genes in extensively drug-resistant Escherichia coli isolated from a patient in Thailand. Diagnostic microbiology and infectious disease **2017**; 87(2): 157-9.

146. Zhong LL, Phan HTT, Shen C, et al. High Rates of Human Fecal Carriage of mcr-1-Positive Multidrug-Resistant Enterobacteriaceae Emerge in China in Association With Successful Plasmid Families. Clin Infect Dis **2018**; 66(5): 676-85.

147. Liu Y, Liu JH. Monitoring Colistin Resistance in Food Animals, An Urgent Threat. Expert Rev Anti Infect Ther **2018**; 16(6): 443-6.

148. Dall C. Reports: India set to ban use of colistin in livestock. Available at: http://www.cidrap.umn.edu/news-perspective/2018/12/reports-india-set-ban-use-colistin-livestock. Accessed July 16, 2019.

149. Carroll LM, Gaballa A, Guldimann C, Sullivan G, Henderson LO, Wiedmann M. Identification of Novel Mobilized Colistin Resistance Gene mcr-9 in a Multidrug-Resistant, Colistin-Susceptible Salmonella enterica Serotype Typhimurium Isolate. mBio **2019**; 10(3).

150. Partridge SR, Di Pilato V, Doi Y, et al. Proposal for assignment of allele numbers for mobile colistin resistance (mcr) genes. J Antimicrob Chemother **2018**; 73(10): 2625-30.

151. Chopra T, Rivard C, Awali RA, et al. Epidemiology of Carbapenem-Resistant Enterobacteriaceae at a Long-term Acute Care Hospital. Open Forum Infect Dis **2018**; 5(10): ofy224.

152. Lin MY, Lyles-Banks RD, Lolans K, et al. The Importance of Long-term Acute Care Hospitals in the Regional Epidemiology of Klebsiella pneumoniae Carbapenemase-Producing Enterobacteriaceae. Clin Infect Dis **2013**; 57(9): 1246-52.

153. Prabaker K, Lin MY, McNally M, et al. Transfer from high-acuity long-term care facilities is associated with carriage of Klebsiella pneumoniae carbapenemase-producing Enterobacteriaceae: a multihospital study. Infect Control Hosp Epidemiol **2012**; 33(12): 1193-9.

154. Ray MJ, Lin MY, Weinstein RA, Trick WE. Spread of Carbapenem-Resistant EnterobacteriaceaeAmong Illinois Healthcare Facilities: The Role of Patient Sharing. Clin Infect Dis **2016**; 63(7): 889-93.

155. Kassis-Chikhani N, Saliba F, Carbonne A, et al. Extended measures for controlling an outbreak of VIM-1 producing imipenem-resistant Klebsiella pneumoniae in a liver transplant centre in France, 2003-2004. Euro Surveill **2010**; 15(46).

156. Patel G, Perez F, Bonomo RA. Carbapenem-resistant Enterobacteriaceae and Acinetobacter baumannii: assessing their impact on organ transplantation. Curr Opin Organ Transplant **2010**; 15(6): 676-82.

157. Singh N, Wagener MM, Obman A, Cacciarelli TV, de Vera ME, Gayowski T. Bacteremias in liver transplant recipients: shift toward gram-negative bacteria as predominant pathogens. Liver Transpl **2004**; 10(7): 844-9.

158. Pagano L, Caira M, Trecarichi EM, et al. Carbapenemase-producing Klebsiella pneumoniaeand Hematologic Malignancies. Emerg Infect Dis **2014**; 20(7): 1235-6.

159. Satlin MJ, Cohen N, Ma KC, et al. Bacteremia due to carbapenem-resistant Enterobacteriaceae in neutropenic patients with hematologic malignancies. Journal of Infection **2016**; 73(4): 336-45.

160. Bar-Yoseph H, Hussein K, Braun E, Paul M. Natural history and decolonization strategies for ESBL/carbapenem-resistant Enterobacteriaceae carriage: systematic review and meta-analysis. J Antimicrob Chemother **2016**; 71(10): 2729-39.

161. Ciobotaro P, Flaks-Manov N, Oved M, et al. Predictors of Persistent Carbapenem-Resistant Enterobacteriaceae Carriage upon Readmission and Score Development. Infect Control Hosp Epidemiol **2016**; 37(2): 188-96.

162. Kim YK, Song SA, Lee JN, et al. Clinical factors predicting persistent carriage of Klebsiella pneumoniae carbapenemase-producing carbapenem-resistant Enterobacteriaceae among patients with known carriage. J Hosp Infect **2018**; 99(4): 405-12.

163. Kontopidou F, Plachouras D, Papadomichelakis E, et al. Colonization and infection by colistin-resistant Gram-negative bacteria in a cohort of critically ill patients. Clin Microbiol Infect **2011**; 17(11): E9-E11.

164. Falcone M, Russo A, Iacovelli A, et al. Predictors of outcome in ICU patients with septic shock caused by Klebsiella pneumoniae carbapenemase-producing K. pneumoniae. Clin Microbiol Infect **2016**; 22(5): 444-50.

165. van Duin D, Doi Y. Outbreak of Colistin-Resistant, Carbapenemase-Producing Klebsiella pneumoniae: Are We at the End of the Road? J Clin Microbiol **2015**; 53(10): 3116-7.

166. Giacobbe DR, Del Bono V, Trecarichi EM, et al. Risk factors for bloodstream infections due to colistin-resistant KPC-producing Klebsiella pneumoniae: results from a multicenter case-control-control study. Clin Microbiol Infect **2015**; 21(12): 1106 e1-8.

167. Giamarellou H. Epidemiology of infections caused by polymyxin-resistant pathogens. Int J Antimicrob Agents **2016**; 48(6): 614-21.

168. Richter SE, Miller L, Uslan DZ, et al. Risk Factors for Colistin Resistance among Gram-Negative Rods and Klebsiella pneumoniae Isolates. J Clin Microbiol **2018**; 56(9).

169. Rossi F, Girardello R, Cury AP, Di Gioia TS, Almeida JN, Jr., Duarte AJ. Emergence of colistin resistance in the largest university hospital complex of Sao Paulo, Brazil, over five years. Braz J Infect Dis **2017**; 21(1): 98-101.

170. Arjun R, Gopalakrishnan R, Nambi PS, Kumar DS, Madhumitha R, Ramasubramanian V. A Study of 24 Patients with Colistin-Resistant Gram-negative Isolates in a Tertiary Care Hospital in South India. Indian J Crit Care Med **2017**; 21(5): 317-21.

171. Parisi SG, Bartolini A, Santacatterina E, et al. Prevalence of Klebsiella pneumoniae strains producing carbapenemases and increase of resistance to colistin in an Italian teaching hospital from January 2012 To December 2014. BMC Infect Dis **2015**; 15: 244.

172. Toth A, Damjanova I, Puskas E, et al. Emergence of a colistin-resistant KPC-2-producing Klebsiella pneumoniae ST258 clone in Hungary. Eur J Clin Microbiol Infect Dis **2010**; 29(7): 765-9.

173. Poirel L, Nordmann P. Emerging plasmid-encoded colistin resistance: the animal world as the culprit? J Antimicrob Chemother **2016**; 71(8): 2326-7.

174. Gu DX, Huang YL, Ma JH, et al. Detection of Colistin Resistance Gene mcr-1 in Hypervirulent Klebsiella pneumoniae and Escherichia coli Isolates from an Infant with Diarrhea in China. Antimicrob Agents Chemother **2016**; 60(8): 5099-100.

175. Olaitan AO, Chabou S, Okdah L, Morand S, Rolain JM. Dissemination of the mcr-1 colistin resistance gene. Lancet Infect Dis **2016**; 16(2): 147.

176. Prim N, Turbau M, Rivera A, Rodriguez-Navarro J, Coll P, Mirelis B. Prevalence of colistin resistance in clinical isolates of Enterobacteriaceae: A four-year cross-sectional study. J Infect **2017**; 75(6): 493-8.

177. Shenoy ES, Pierce VM, Walters MS, et al. Transmission of Mobile Colistin Resistance (mcr-1) by Duodenoscope. Clin Infect Dis **2019**; 68(8): 1327-34.

178. Harris P, Paterson D, Rogers B. Facing the challenge of multidrug-resistant gram-negative bacilli in Australia. Med J Aust **2015**; 202(5): 243-7.

179. Coombs GW BJ, Daley DA, Collignon P, Gottlieb T, Iredell J, Kotsanas D, Nimmo GR, Robson J, Turnidge JD, on behalf of the Australian Group on Antimicrobial Resistance. Australian Group on Antimicrobial Resistance: Sepsis Outcomes Programs 2016 report. Australian Commission on Safety and Quality in Health Care, ACSQHC, Sydney. Available at: https://www.safetyandquality.gov.au/wp4content/uploads/2018/03/AGAR-Sepsis-Outcome-Program-2016-Report-February-2018.pdf. Accessed May 15, 2020.

180. Brett JA, Johnson SA, Cameron DRM, et al. Carbapenemase-producing Enterobacteriaceae in Australian hospitals: outcome of point-prevalence screening in high-risk wards. J Hosp Infect **2019**; 101(2): 163-6.

181. Espedido BA, Partridge SR, Iredell JR. bla(IMP-4) in different genetic contexts in Enterobacteriaceae isolates from Australia. Antimicrob Agents Chemother **2008**; 52(8): 2984-7.

182. Kizny Gordon A, Phan HTT, Lipworth SI, et al. Genomic dynamics of species and mobile genetic elements in a prolonged blaIMP-4-associated carbapenemase outbreak in an Australian hospital. J Antimicrob Chemother **2020**; 75(4): 873-82.

183. Peleg AY, Franklin C, Walters LJ, Bell JM, Spelman DW. OXA-58 and IMP-4 carbapenem-hydrolyzing beta-lactamases in an Acinetobacter junii blood culture isolate from Australia. Antimicrob Agents Chemother **2006**; 50(1): 399-400.

184. Roberts LW, Catchpoole E, Jennison AV, et al. Genomic analysis of carbapenemase-producing Enterobacteriaceae in Queensland reveals widespread transmission of bla IMP-4 on an IncHI2 plasmid. Microb Genom **2020**; 6(1).

185. Roberts LW, Harris PNA, Forde BM, et al. Integrating multiple genomic technologies to investigate an outbreak of carbapenemase-producing Enterobacter hormaechei. Nat Commun **2020**; 11(1): 466.

186. Chua KY, Grayson ML, Burgess AN, Lee JY, Howden BP. The growing burden of multidrug-resistant infections among returned Australian travellers. Med J Aust **2014**; 200(2): 116-8.

187. Sherry NL, Lane CR, Kwong JC, et al. Genomics for Molecular Epidemiology and Detecting Transmission of Carbapenemase-Producing Enterobacterales in Victoria, Australia, 2012 to 2016. J Clin Microbiol **2019**; 57(9).

188. Chang LW, Buising KL, Jeremiah CJ, et al. Managing a nosocomial outbreak of carbapenem-resistant Klebsiella pneumoniae: an early Australian hospital experience. Intern Med J **2015**; 45(10): 1037-43.

189. Kwong JC, Lane CR, Romanes F, et al. Translating genomics into practice for real-time surveillance and response to carbapenemase-producing Enterobacteriaceae: evidence from a complex multi-institutional KPC outbreak. PeerJ **2018**; 6: e4210.

190. Department of Health and Human Services Victoria. Victorian guideline on carbapenemase-producing Enterobacteriaceae for health services (version 2.1). Department of Communicable Diseases Prevention and Control, Victorian Govenment, Melbourne., **2018**.

191. Baker S, Thomson N, Weill FX, Holt KE. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. Science **2018**; 360(6390): 733-8.

192. Didelot X, Walker AS, Peto TE, Crook DW, Wilson DJ. Within-host evolution of bacterial pathogens. Nat Rev Microbiol **2016**; 14(3): 150-62.

193. Hawken SE, Snitkin ES. Genomic epidemiology of multidrug-resistant Gram-negative organisms. Ann N Y Acad Sci **2019**; 1435(1): 39-56.

194. Arias CA, Panesso D, McGrath DM, et al. Genetic basis for in vivo daptomycin resistance in enterococci. N Engl J Med **2011**; 365(10): 892-900.

195. Pak TR, Altman DR, Attie O, et al. Whole-Genome Sequencing Identifies Emergence of a Quinolone Resistance Mutation in a Case of Stenotrophomonas maltophilia Bacteremia. Antimicrob Agents Chemother **2015**; 59(11): 7117-20.

196. Peleg AY, Miyakis S, Ward DV, et al. Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of Staphylococcus aureus. PLoS One **2012**; 7(1): e28316.

197. Mathers AJ, Peirano G, Pitout JDD. The Role of Epidemic Resistance Plasmids and International High-Risk Clones in the Spread of Multidrug-Resistant Enterobacteriaceae. Clin Microbiol Rev **2015**; 28(3): 565-91.

198. Sidjabat HE, Silveira FP, Potoski BA, et al. Interspecies Spread of Klebsiella pneumoniaeCarbapenemase Gene in a Single Patient. Clin Infect Dis **2009**; 49(11): 1736-8.

199. Conlan S, Park M, Deming C, et al. Plasmid Dynamics in KPC-Positive Klebsiella pneumoniae during Long-Term Patient Colonization. mBio **2016**; 7(3): 263-9.

200. Mathers AJ, Cox HL, Kitchel B, et al. Molecular Dissection of an Outbreak of Carbapenem-Resistant Enterobacteriaceae Reveals Intergenus KPC Carbapenemase Transmission through a Promiscuous Plasmid. mBio **2011**; 2(6): 425-7.

201. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with Antimicrobial Resistance. Clin Microbiol Rev **2018**; 31(4): 1079-61.

202. Snesrud E, He S, Chandler M, et al. A Model for Transposition of the Colistin Resistance Gene mcr-1by IS Apl1. Antimicrob Agents Chemother **2016**; 60(11): 6973-6.

203. Snesrud E, Ong AC, Corey B, et al. Analysis of Serial Isolates of mcr-1-Positive Escherichia coli Reveals a Highly Active IS Apl1Transposon. Antimicrob Agents Chemother **2017**; 61(5): 391-10.

204. Peacock SJ, Parkhill J, Brown NM. Changing the paradigm for hospital outbreak detection by leading with genomic surveillance of nosocomial pathogens. Microbiology **2018**; 164(10): 1213-9.

205. Koser CU, Holden MT, Ellington MJ, et al. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med **2012**; 366(24): 2267-75.

206. Harris SR, Cartwright EJ, Torok ME, et al. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study. Lancet Infect Dis **2013**; 13(2): 130-6.

207. Snitkin ES, Zelazny AM, Thomas PJ, et al. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. Sci Transl Med **2012**; 4(148): 148ra16.

208. Gorrie CL, Mirceta M, Wick RR, et al. Antimicrobial-Resistant Klebsiella pneumoniae Carriage and Infection in Specialized Geriatric Care Wards Linked to Acquisition in the Referring Hospital. Clin Infect Dis **2018**; 67(2): 161-70.

209. Jiang Y, Wei Z, Wang Y, Hua X, Feng Y, Yu Y. Tracking a hospital outbreak of KPC-producing ST11 Klebsiella pneumoniae with whole genome sequencing. Clin Microbiol Infect **2015**; 21(11): 1001-7.

210. Johnson RC, Deming C, Conlan S, et al. Investigation of a Cluster of Sphingomonas koreensis Infections. N Engl J Med **2018**; 379(26): 2529-39.

211. Kanamori H, Parobek CM, Juliano JJ, et al. A Prolonged Outbreak of KPC-3-Producing Enterobacter cloacae and Klebsiella pneumoniae Driven by Multiple Mechanisms of Resistance Transmission at a Large Academic Burn Center. Antimicrob Agents Chemother **2017**; 61(2): AAC.01516-16-12.

212. Marsh JW, Krauland MG, Nelson JS, et al. Genomic Epidemiology of an Endoscope-Associated Outbreak of Klebsiella pneumoniae Carbapenemase (KPC)-Producing K. pneumoniae. PLoS One **2015**; 10(12): e0144310-18.

213. Ruppé E, Olearo F, Pires D, et al. Clonal or not clonal? Investigating hospital outbreaks of KPC-producing Klebsiella pneumoniae with whole-genome sequencing. Clin Microbiol Infect **2017**; 23(7): 470-5.

214. Halachev MR, Chan JZ, Constantinidou CI, et al. Genomic epidemiology of a protracted hospital outbreak caused by multidrug-resistant Acinetobacter baumannii in Birmingham, England. Genome Med **2014**; 6(11): 70.

215. Mathers AJ, Stoesser N, Sheppard AE, et al. Klebsiella pneumoniae Carbapenemase (KPC)-Producing K. pneumoniae at a Single Institution: Insights into Endemicity from Whole-Genome Sequencing. Antimicrob Agents Chemother **2015**; 59(3): 1656-63.

216. Stoesser N, Sheppard AE, Shakya M, et al. Dynamics of MDR Enterobacter cloacae outbreaks in a neonatal unit in Nepal: insights using wider sampling frames and next-generation sequencing. J Antimicrob Chemother **2015**; 70(4): 1008-15.

217. Davis RJ, Jensen SO, Van Hal S, et al. Whole Genome Sequencing in Real-Time Investigation and Management of a Pseudomonas aeruginosa Outbreak on a Neonatal Intensive Care Unit. Infect Control Hosp Epidemiol **2015**; 36(9): 1058-64.

218. Willmann M, Bezdan D, Zapata L, et al. Analysis of a long-term outbreak of XDR Pseudomonas aeruginosa: a molecular epidemiological study. J Antimicrob Chemother **2015**; 70(5): 1322-30.

219. Decraene V, Phan HTT, George R, et al. A Large, Refractory Nosocomial Outbreak of Klebsiella pneumoniae Carbapenemase-Producing Escherichia coli Demonstrates Carbapenemase Gene Outbreaks Involving Sink Sites Require Novel Approaches to Infection Control. Antimicrob Agents Chemother **2018**; 62(12).

220. Ito K, Honda H, Yoshida M, et al. A metallo-beta-lactamase producing Enterobacteriaceae outbreak from a contaminated tea dispenser at a children's hospital in Japan. Infect Control Hosp Epidemiol **2018**; 40(2): 1-4.

221. Kizny Gordon AE, Mathers AJ, Cheong EYL, et al. The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections-A Systematic Review of the Literature. Clin Infect Dis **2017**; 64(10): 1435-44.

222. Mathers AJ, Vegesana K, German Mesner I, et al. Intensive Care Unit Wastewater Interventions to Prevent Transmission of Multi-species Klebsiella pneumoniae Carbapenemase (KPC) Producing Organisms. Clin Infect Dis **2018**.

223. Pecora ND, Li N, Allard M, et al. Genomically Informed Surveillance for Carbapenem-Resistant Enterobacteriaceae in a Health Care System. mBio **2015**; 6(4): 1170-11.

224. Weingarten RA, Johnson RC, Conlan S, et al. Genomic Analysis of Hospital Plumbing Reveals Diverse Reservoir of Bacterial Plasmids Conferring Carbapenem Resistance. mBio **2018**; 9(1): 1198-16.

225. Stoesser N, Sheppard AE, Moore CE, et al. Extensive Within-Host Diversity in Fecally Carried Extended-Spectrum-Beta-Lactamase-Producing Escherichia coli Isolates: Implications for Transmission Analyses. J Clin Microbiol **2015**; 53(7): 2122-31.

226. Mulvey MR, Haraoui LP, Longtin Y. Multiple Variants of Klebsiella pneumoniae Producing Carbapenemase in One Patient. N Engl J Med **2016**; 375(24): 2408-10.

227. Worby CJ, Lipsitch M, Hanage WP. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. PLoS Comp Biol **2014**; 10(3): e1003549.

228. Read TD, Massey RC. Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. Genome Med **2014**; 6(11): 1063-11.

229. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet **2001**; 17(9): 502-10.

230. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. Science **2005**; 308(5720): 385-9.

231. Power RA, Parkhill J, de Oliveira T. Microbial genome-wide association studies: lessons from human GWAS. Nat Rev Genet **2017**; 18(1): 41-50.

232. Wang K, Li M, Hadley D, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res **2007**; 17(11): 1665-74.

233. Saund K. Hogwash algorithms. Available at: https://github.com/katiesaund/hogwash/wiki/B)-hogwash-algorithms. Accessed July 16, 2019.

234. Chen PE, Shapiro BJ. The advent of genome-wide association studies for bacteria. Curr Opin Microbiol **2015**; 25: 17-24.

235. Wattam AR, Davis JJ, Assaf R, et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res **2017**; 45(D1): D535-D42.

236. Gupta SK, Padmanabhan BR, Diene SM, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother **2014**; 58(1): 212-20.

237. Zankari E, Hasman H, Cosentino S, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother **2012**; 67(11): 2640-4.

238. Jia B, Raphenya AR, Alcock B, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res **2017**; 45(D1): D566-D73.

239. Didelot X, Maiden MC. Impact of recombination on bacterial evolution. Trends Microbiol **2010**; 18(7): 315-22.

240. Noble WS. How does multiple testing correction work? Nat Biotechnol **2009**; 27(12): 1135-7.

241. Selle K, Barrangou R. Harnessing CRISPR-Cas systems for bacterial genome editing. Trends Microbiol **2015**; 23(4): 225-32.

242. Sheppard SK, Didelot X, Méric G, 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(29): 11923-7.

243. Collins C, Didelot X. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. PLoS Comp Biol **2018**; 14(2): e1005958.

244. Earle SG, Wu CH, Charlesworth J, et al. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. Nat Microbiol **2016**; 1(5): 16041.

245. Jaillard M, Lima L, Tournoud M, et al. A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. PLoS Genet **2018**; 14(11): e1007758.

246. Farhat MR, Shapiro BJ, Kieser KJ, et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. Nat Genet **2013**; 45(10): 1183-9.

247. Alam MT, Petit RA, 3rd, Crispell EK, et al. Dissecting vancomycin-intermediate resistance in staphylococcus aureus using genome-wide association. Genome Biol Evol **2014**; 6(5): 1174-85.

248. Laabei M, Recker M, Rudkin JK, et al. Predicting the virulence of MRSA from its genome sequence. Genome Res **2014**; 24(5): 839-49.

249. Chewapreecha C, Marttinen P, Croucher NJ, et al. Comprehensive identification of single nucleotide polymorphisms associated with beta-lactam resistance within pneumococcal mosaic genes. PLoS Genet **2014**; 10(8): e1004547.

250. Salipante SJ, Roach DJ, Kitzman JO, et al. Large-scale genomic sequencing of extraintestinal pathogenic Escherichia coli strains. Genome Res **2015**; 25(1): 119-28.

251. Holt KE, Wertheim H, Zadoks RN, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proc Natl Acad Sci U S A **2015**; 112(27): E3574-81.

252. Weinert LA, Chaudhuri RR, Wang J, et al. Genomic signatures of human and animal disease in the zoonotic pathogen Streptococcus suis. Nat Commun **2015**; 6: 6740.

253. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol **2016**; 17(1): 238.

254. Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. Bioinformatics **2018**; 34(24): 4310-2.

255. Lees JA, Vehkala M, Valimaki N, et al. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. Nat Commun **2016**; 7: 12797.

256. Hasman H, Saputra D, Sicheritz-Ponten T, et al. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. J Clin Microbiol **2014**; 52(1): 139-46.

257. Koser CU, Fraser LJ, Ioannou A, et al. Rapid single-colony whole-genome sequencing of bacterial pathogens. J Antimicrob Chemother **2014**; 69(5): 1275-81.

258. Bradley P, Gordon NC, Walker TM, et al. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. Nat Commun **2015**; 6: 10063.

259. Gordon NC, Price JR, Cole K, et al. Prediction of Staphylococcus aureus Antimicrobial Resistance by Whole-Genome Sequencing. J Clin Microbiol **2014**; 52(4): 1182-91.

260. Walker TM, Kohl TA, Omar SV, et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis **2015**; 15(10): 1193-202.

261. Samuel AL. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development **1959**; 3(3): 210-29.

262. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning : data mining, inference, and prediction. 2nd ed. New York, NY: Springer, **2009**.

263. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. N Engl J Med **2019**; 380(14): 1347-58.

264. Morgan XC, Huttenhower C. Chapter 12: Human microbiome analysis. PLoS Comp Biol **2012**; 8(12): e1002808.

265. Ho DSW, Schierding W, Wake M, Saffery R, O'Sullivan J. Machine Learning SNP Based Prediction for Precision Medicine. Front Genet **2019**; 10: 267.

266. Okser S, Pahikkala T, Aittokallio T. Genetic variants and their interactions in disease risk prediction - machine learning and network perspectives. BioData Min **2013**; 6(1): 5.

267. Wei Z, Wang K, Qu H-Q, et al. From Disease Association to Risk Assessment: An Optimistic View from Genome-Wide Association Studies on Type 1 Diabetes. PLoS Genet **2009**; 5(10): e1000678-11.

268. Wei Z, Wang W, Bradfield J, et al. Large sample size, wide variant spectrum, and advanced machine-learning technique boost risk prediction for inflammatory bowel disease. Am J Hum Genet **2013**; 92(6): 1008-12.

269. Brownlee J. An Introduction to Feature Selection. Available at: https://machinelearningmastery.com/an-introduction-to-feature-selection/. Accessed July 16, 2019.

270. Brownlee J. A Tour of The Most Popular Machine Learning Algorithms. Available at: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/. Accessed July 16, 2019.

271. Chen ML, Doddi A, Royer J, et al. Deep learning predicts tuberculosis drug resistance status from genome sequencing data. bioRxiv **2018**: 1-29.

272. Davis JJ, Boisvert S, Brettin T, et al. Antimicrobial Resistance Prediction in PATRIC and RAST. Sci Rep **2016**; 6: 27930.

273. Drouin A, Giguere S, Deraspe M, et al. Predictive computational phenotyping and biomarker discovery using reference-free genome comparisons. BMC Genomics **2016**; 17(1): 754.

274. Drouin A, Letarte G, Raymond F, Marchand M, Corbeil J, Laviolette F. Interpretable genotype-to-phenotype classifiers with performance guarantees. Sci Rep **2019**; 9(1): 4071.

275. Eyre DW, De Silva D, Cole K, et al. WGS to predict antibiotic MICs for Neisseria gonorrhoeae. J Antimicrob Chemother **2017**; 72(7): 1937-47.

276. Her H-L, Wu Y-W. A pan-genome-based machine learning approach for predicting antimicrobial resistance activities of the Escherichia coli strains. Bioinformatics **2018**; 34(13): i89-i95.

277. Hicks AL, Wheeler N, Sanchez-Buso L, Rakeman JL, Harris SR, Grad YH. Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. PLoS Comp Biol **2019**; 15(9): e1007349.

278. Kavvas ES, Catoiu E, Mih N, et al. Machine learning and structural analysis of Mycobacterium tuberculosis pan-genome identifies genetic signatures of antibiotic resistance. Nat Commun **2018**; 9(1): 4306.

279. Kouchaki S, Yang Y, Walker TM, et al. Application of machine learning techniques to tuberculosis drug resistance analysis. Bioinformatics **2018**; 35(13): 2276-82.

280. Li Y, Metcalf BJ, Chochua S, et al. Validation of β-lactam minimum inhibitory concentration predictions for pneumococcal isolates with newly encountered penicillin binding protein (PBP) sequences. BMC Genomics **2017**: 1-10.

281. Long SW, Olsen RJ, Eagar TN, 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).

282. MacFadden DR, Melano RG, Coburn B, Tijet N, Hanage WP, Daneman N. Comparing Patient Risk Factor-, Sequence Type-, and Resistance Locus Identification-Based Approaches for Predicting Antibiotic Resistance in Escherichia coli Bloodstream Infections. J Clin Microbiol **2019**; 57(6): h2346-9.

283. Mahé P. Predicting bacterial resistance from whole-genome sequences using k-mers and stability selection. BMC Bioinformatics **2018**; 19(1): 1-11.

284. Mahé P, El Azami M, Barlas P, Tournoud M. A large scale evaluation of TBProfiler and Mykrobe for antibiotic resistance prediction in Mycobacterium tuberculosis. PeerJ **2019**; 7(1): e6857-21.

285. Moradigaravand D, Palm M, Farewell A, Mustonen V, Warringer J, Parts L. Prediction of antibiotic resistance in Escherichia coli from large-scale pan-genome data. PLoS Comp Biol **2018**; 14(12): e1006258-17.

286. Nguyen M, Brettin T, Long SW, et al. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. Sci Rep **2017**: 1-11.

287. Nguyen M, Long SW, McDermott PF, et al. Using Machine Learning To Predict Antimicrobial MICs and Associated Genomic Features for Nontyphoidal Salmonella. J Clin Microbiol **2019**; 57(2): 495-15.

288. Pesesky MW, Hussain T, Wallace M, et al. Evaluation of Machine Learning and Rules-Based Approaches for Predicting Antimicrobial Resistance Profiles in Gram-negative Bacilli from Whole Genome Sequence Data. Front Microbiol **2016**; 7: 414-17.

289. Rishishwar L, Petit RA, Kraft CS, Jordan IK. Genome Sequence-Based Discriminator for Vancomycin-Intermediate Staphylococcus aureus. J Bacteriol **2014**; 196(5): 940-8.

290. Yang Y, Niehaus KE, Walker TM, et al. Machine learning for classifying tuberculosis drug-resistance from DNA sequencing data. Bioinformatics **2018**; 34(10): 1666-71.

291. Farhat MR, Sultana R, Iartchouk O, et al. Genetic Determinants of Drug Resistance in Mycobacterium tuberculosis and Their Diagnostic Value. Am J Respir Crit Care Med **2016**; 194(5): 621-30.

292. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. arXiv **2012**; cs.LG.

293. Kuhn M. Building Predictive Models in R Using the caret Package. Journal of Statistical Software **2008**; 28(5): 1-26.

294. Katuwal GJC, R. Machine Learning Model Interpretability for Precision Medicine. arXiv **2016**.

295. Lipton Z. The Mythos of Model Interpretability. arXiv **2016**.

296. Kalpoe JS, Sonnenberg E, Factor SH, et al. Mortality associated with carbapenem-resistant *Klebsiella pneumoniae i*nfections in liver transplant recipients. Liver Transpl **2012**; 18(4): 468-74.

297. Pereira MR, Scully BF, Pouch SM, et al. Risk factors and outcomes of carbapenem-resistant *Klebsiella pneumoniae* infections in liver transplant recipients. Liver Transpl **2015**.

298. Pereira MR, Uhlemann AC. The evolving threat of carbapenem-resistant infections after liver transplantation: The case of Acinetobacter baumannii. Liver Transpl **2016**; 22(5): 579-81.

299. Lubbert C, Hau HM, Rodloff A, et al. [Clinical impact of infections with carbapenem-resistant enterobacteriaceae in liver transplant recipients]. Z Gastroenterol **2015**; 53(11): 1276-87.

300. Macesic N, Green D, Wang Z, et al. Detection of mcr-1-Carrying Escherichia coli Causing Bloodstream Infection in a New York City Hospital: Avian Origins, Human Concerns? Open Forum Infect Dis **2017**; 4(3): ofx115.

301. Wyres KL, Hawkey J, Hetland MAK, et al. Emergence and rapid global dissemination of CTX-M-15-associated Klebsiella pneumoniae strain ST307. J Antimicrob Chemother **2019**; 74(3): 577-81.

302. Hindler JA, Humphries RM. Colistin MIC variability by method for contemporary clinical isolates of multidrug-resistant Gram-negative bacilli. J Clin Microbiol **2013**; 51(6): 1678-84.

303. Bonomo RA. Cefiderocol: A Novel Siderophore Cephalosporin Defeating Carbapenem-resistant Pathogens. Clin Infect Dis **2019**; 69(Supplement_7): S519-S20.

304. van Duin D, Lok JJ, Earley M, et al. Colistin Versus Ceftazidime-Avibactam in the Treatment of Infections Due to Carbapenem-Resistant Enterobacteriaceae. Clin Infect Dis **2018**; 66(2): 163-71.

305. Shields RK, Nguyen MH, Chen L, Press EG, Kreiswirth BN, Clancy CJ. Pneumonia and Renal Replacement Therapy Are Risk Factors for Ceftazidime-Avibactam Treatment Failures and Resistance among Patients with Carbapenem-Resistant EnterobacteriaceaeInfections. Antimicrob Agents Chemother **2018**; 62(5): e02497-17-12.

306. Giddins MJ, Macesic N, Annavajhala MK, et al. Successive Emergence of Ceftazidime-Avibactam Resistance through Distinct Genomic Adaptations in blaKPC-2-Harboring Klebsiella pneumoniae Sequence Type 307 Isolates. Antimicrob Agents Chemother **2018**; 62(3).

307. Lahiri SD, Johnstone MR, Ross PL, McLaughlin RE, Olivier NB, Alm RA. Avibactam and class C beta-lactamases: mechanism of inhibition, conservation of the binding pocket, and implications for resistance. Antimicrob Agents Chemother **2014**; 58(10): 5704-13.

308. Livermore DM, Warner M, Jamrozy D, et al. In vitro selection of ceftazidime-avibactam resistance in Enterobacteriaceae with KPC-3 carbapenemase. Antimicrob Agents Chemother **2015**; 59(9): 5324-30.

309. Papp-Wallace KM, Winkler ML, Taracila MA, Bonomo RA. Variants of β-lactamase KPC-2 that are resistant to inhibition by avibactam. Antimicrob Agents Chemother **2015**; 59(7): 3710-7.

310. Shields RK, Potoski BA, Haidar G, et al. Clinical Outcomes, Drug Toxicity, and Emergence of Ceftazidime-Avibactam Resistance Among Patients Treated for Carbapenem-Resistant Enterobacteriaceae Infections: Table 1. Clin Infect Dis **2016**: ciw636-4.

311. Shields RK, Nguyen MH, Press EG, Chen L, Kreiswirth BN, Clancy CJ. In Vitro Selection of Meropenem Resistance among Ceftazidime-Avibactam-Resistant, Meropenem-Susceptible Klebsiella pneumoniae Isolates with Variant KPC-3 Carbapenemases. Antimicrob Agents Chemother **2017**; 61(5).

312. Shields RK, Nguyen MH, Press EG, Chen L, Kreiswirth BN, Clancy CJ. Emergence of Ceftazidime-Avibactam Resistance and Restoration of Carbapenem Susceptibility in Klebsiella pneumoniae Carbapenemase-Producing K pneumoniae: A Case Report and Review of Literature. Open Forum Infect Dis **2017**; 4(3): ofx101.

313. Winkler ML, Papp-Wallace KM, Taracila MA, Bonomo RA. Avibactam and inhibitor-resistant SHV beta-lactamases. Antimicrob Agents Chemother **2015**; 59(7): 3700-9.

314. Castanheira M, Mendes RE, Sader HS. Low Frequency of Ceftazidime-Avibactam Resistance among Enterobacteriaceae Isolates Carrying blaKPC Collected in U.S. Hospitals from 2012 to 2015. Antimicrob Agents Chemother **2017**; 61(3): 1-9.

315. Drawz SM, Bonomo RA. Three decades of beta-lactamase inhibitors. Clin Microbiol Rev **2010**; 23(1): 160-201.

316. Humphries RM, Hemarajata P. Resistance to Ceftazidime-Avibactam in Klebsiella pneumoniae Due to Porin Mutations and the Increased Expression of KPC-3. Antimicrob Agents Chemother **2017**; 61(6).

317. Sun D, Rubio-Aparicio D, Nelson K, Dudley MN, Lomovskaya O. Meropenem-Vaborbactam Resistance Selection, Resistance Prevention, and Molecular Mechanisms in Mutants of KPC-Producing Klebsiella pneumoniae. Antimicrob Agents Chemother **2017**; 61(12): 1-14.

318. Hackel MA, Tsuji M, Yamano Y, Echols R, Karlowsky JA, Sahm DF. In Vitro Activity of the Siderophore Cephalosporin, Cefiderocol, against a Recent Collection of Clinically Relevant Gram-Negative Bacilli from North America and Europe, Including Carbapenem-Nonsusceptible Isolates (SIDERO-WT-2014 Study). Antimicrob Agents Chemother **2017**; 61(9).

319. Hackel MA, Tsuji M, Yamano Y, Echols R, Karlowsky JA, Sahm DF. In Vitro Activity of the Siderophore Cephalosporin, Cefiderocol, against Carbapenem-Nonsusceptible and Multidrug-Resistant Isolates of Gram-Negative Bacilli Collected Worldwide in 2014 to 2016. Antimicrob Agents Chemother **2018**; 62(2).

320. Yamano Y. In Vitro Activity of Cefiderocol Against a Broad Range of Clinically Important Gram-negative Bacteria. Clin Infect Dis **2019**; 69(Supplement_7): S544-S51.

321. Doi Y. Treatment Options for Carbapenem-resistant Gram-negative Bacterial Infections. Clin Infect Dis **2019**; 69(Supplement_7): S565-S75.

322. Watkins RR, Deresinski S. Omadacycline: A Novel Tetracycline Derivative With Oral and Intravenous Formulations. Clin Infect Dis **2019**; 69(5): 890-6.

323. Doi Y, Wachino JI, Arakawa Y. Aminoglycoside Resistance: The Emergence of Acquired 16S Ribosomal RNA Methyltransferases. Infect Dis Clin North Am **2016**; 30(2): 523-37.

324. Aira A, Feher C, Rubio E, Soriano A. The Intestinal Microbiota as a Reservoir and a Therapeutic Target to Fight Multi-Drug-Resistant Bacteria: A Narrative Review of the Literature. Infect Dis Ther **2019**; 8(4): 469-82.

325. ClinicalTrials.gov. Autologous Fecal Microbiota Transplantation (Auto-FMT) for Prophylaxis of Clostridium Difficile Infection in Recipients of Allogeneic Hematopoietic Stem Cell Transplantation. Available at: https://clinicaltrials.gov/ct2/show/NCT02269150. Accessed January 6, 2020.

326. Taur Y, Coyte K, Schluter J, et al. Reconstitution of the gut microbiota of antibiotic-treated patients by autologous fecal microbiota transplant. Sci Transl Med **2018**; 10(460).

327. Woodworth MH, Hayden MK, Young VB, Kwon JH. The Role of Fecal Microbiota Transplantation in Reducing Intestinal Colonization With Antibiotic-Resistant Organisms: The Current Landscape and Future Directions. Open Forum Infect Dis **2019**; 6(7).

328. Kortright KE, Chan BK, Koff JL, Turner PE. Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. Cell Host Microbe **2019**; 25(2): 219-32.

329. Chan BK, Sistrom M, Wertz JE, Kortright KE, Narayan D, Turner PE. Phage selection restores antibiotic sensitivity in MDR Pseudomonas aeruginosa. Sci Rep **2016**; 6: 26717.

330. Oechslin F, Piccardi P, Mancini S, et al. Synergistic Interaction Between Phage Therapy and Antibiotics Clears Pseudomonas Aeruginosa Infection in Endocarditis and Reduces Virulence. J Infect Dis **2017**; 215(5): 703-12.

331. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. Hum Mol Genet **2018**; 27(R2): R234-R41.

332. De Maio N, Shaw LP, Hubbard A, et al. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. Microb Genom **2019**; 5(9).

333. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC, McCombie WR. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res **2015**; 25(11): 1750-6.

334. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comp Biol **2017**; 13(6): e1005595.

335. Conlan S, Lau AF, Deming C, et al. Plasmid Dissemination and Selection of a Multidrug-Resistant Klebsiella pneumoniae Strain during Transplant-Associated Antibiotic Therapy. mBio **2019**; 10(5).

336. Quick J, Loman NJ, Duraffour S, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature **2016**; 530(7589): 228-32.

337. Cao MD, Ganesamoorthy D, Elliott AG, Zhang H, Cooper MA, Coin LJ. Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION(TM) sequencing. Gigascience **2016**; 5(1): 32.

338. Edwards HS, Krishnakumar R, Sinha A, Bird SW, Patel KD, Bartsch MS. Real-Time Selective Sequencing with RUBRIC: Read Until with Basecall and Reference-Informed Criteria. Sci Rep **2019**; 9(1): 11475.

339. Schmidt K, Mwaigwisya S, Crossman LC, et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. J Antimicrob Chemother **2017**; 72(1): 104-14.

340. Tamma PD, Fan Y, Bergman Y, et al. Applying Rapid Whole-Genome Sequencing To Predict Phenotypic Antimicrobial Susceptibility Testing Results among Carbapenem-Resistant Klebsiella pneumoniae Clinical Isolates. Antimicrob Agents Chemother **2019**; 63(1).

341. Bertrand D, Shaw J, Kalathiyappan M, et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. Nat Biotechnol **2019**; 37(8): 937-44.

342. Leggett RM, Alcon-Giner C, Heavens D, et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. Nat Microbiol **2019**.

343. Charalampous T, Kay GL, Richardson H, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol **2019**; 37(7): 783-92.

344. Goodman KE, Lessler J, Cosgrove SE, et al. A Clinical Decision Tree to Predict Whether a Bacteremic Patient Is Infected With an Extended-Spectrum β-Lactamase–Producing Organism. Clin Infect Dis **2016**; 63(7): 896-903.

345. Vazquez-Guillamet MC, Vazquez R, Micek ST, Kollef MH. Predicting resistance to piperacillin-tazobactam, cefepime and meropenem in septic patients with bloodstream infection due to Gram-negative bacteria. Clin Infect Dis **2017**; 65(10): 1607-1614.

346. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. PLoS One **2017**; 12(4): e0174708.

347. Rawson TM, Hernandez B, Moore LSP, et al. Supervised machine learning for the prediction of infection on admission to hospital: a prospective observational cohort study. J Antimicrob Chemother **2019**; 74(4): 1108-15.

348. Ghosh S, Li J, Cao L, Ramamohanarao K. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. J Biomed Inform **2017**; 66: 19-31.

349. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. Sci Transl Med **2015**; 7(299): 299ra122.

350. Mani S, Ozdas A, Aliferis C, et al. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. J Am Med Inform Assoc **2014**; 21(2): 326-36.

351. Nemati S, Holder A, Razmi F, Stanley MD, Clifford GD, Buchman TG. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. Crit Care Med **2018**; 46(4): 547-53.

352. Taneja I, Reddy B, Damhorst G, et al. Combining Biomarkers with EMR Data to Identify Patients in Different Phases of Sepsis. Sci Rep **2017**; 7(1): 10800.

353. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med **2013**; 15(10): 761-71.

354. Seymour CW, Kennedy JN, Wang S, et al. Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. JAMA **2019**; 321(20): 2003-17.

355. Li BY, Oh J, Young VB, Rao K, Wiens J. Using Machine Learning and the Electronic Health Record to Predict Complicated Clostridium difficile Infection. Open Forum Infect Dis **2019**; 6(5): ofz186.

356. Falagas ME, Lourida P, Poulikakos P, Rafailidis PI, Tansarli GS. Antibiotic treatment of infections due to carbapenem-resistant Enterobacteriaceae: systematic evaluation of the available evidence. Antimicrob Agents Chemother **2014**; 58(2): 654-63.