Genomic Epidemiology of Horizontal Plasmid Transfer Among Healthcare-Associated Bacterial Pathogens in a Tertiary Care Hospital

by

Daniel Evans

BS, Carnegie Mellon University, 2017

Submitted to the Graduate Faculty of the

Department of Infectious Diseases and Microbiology

Graduate School of Public Health

in partial fulfillment of the requirements for the degree of

Master of Science in Infectious Diseases and Microbiology

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

By

Daniel Richard Evans

It was defended on

April 5, 2021

and approved by

Daria Van Tyne PhD, Assistant Professor, Division of Infectious Diseases, University of Pittsburgh School of Medicine

Jeremy Martinson DPhil, Assistant Professor, Department of Infectious Diseases and Microbiology, University of Pittsburgh Graduate School of Public Health

Lee Harrison MD, Professor, Department of Epidemiology, University of Pittsburgh Graduate School of Public Health

Copyright © by Daniel Richard Evans

2021

Genomic Epidemiology of Horizontal Plasmid Transfer Among Healthcare-Associated Bacterial Pathogens in a Tertiary Care Hospital

Daniel Richard Evans, MS

University of Pittsburgh, 2021

Abstract

Healthcare-associated bacterial pathogens frequently carry plasmids that contribute to antibiotic resistance and virulence. The horizontal transfer of plasmids between pathogens within hospitals has been previously documented, but the epidemiology and clinical burden of nosocomial plasmid transfer remains poorly understood. Our primary objective was to systematically resolve plasmids from whole-genome sequences of nosocomial bacterial isolates, using thresholds of sequence similarity that were indicative of horizontal transfer. Our secondary objective was to identify potential routes and assess the clinical burden of horizontal plasmid transfer.

Whole-genome sequencing was performed on 3,074 nosocomial bacterial isolates from 2,322 hospitalizations of 1,960 patients, using the Illumina platform. Seventy-eight strains were also sequenced by long-read Oxford Nanopore technology, and hybrid genome assemblies were generated using Unicycler. Plasmids were resolved from Illumina-sequenced genomes by alignment using BLASTn. Single nucleotide polymorphisms (SNPs) were identified using Snippy. De-identified patient data associated with bacterial isolates – including length of hospital stay and Charlson comorbidity index – were collected using Theradoc.

Ninety-five percent of analyzed strains maintained at least 95% of the sequence content of reference plasmids, with SNPs occurring at rates of fewer than 1 per 5000bp of reference

plasmid sequence. Using these thresholds, we identified 41 plasmid lineages that were potentially horizontally transferred among non-clonal bacterial strains. Of these lineages, 28 (68.2%) were significantly associated with at least one medical procedure, room, or hospital ward. Hospitalizations involving the 41 plasmid lineages were significantly longer (+3 days; 95% CI +1 to +4; $p < 0.001$) than hospitalizations not involving those plasmids. Patients infected with strains carrying transferred plasmids had significantly greater overall comorbidity (Charlson comorbidity index +1, 95% CI +1 to +2; $p < 0.0001$) than patients whose infections did not involve plasmids in lineages.

Our findings show that the horizontal transfer of plasmids among bacterial isolates causing nosocomial infections is frequent, and that this phenomenon may impose an unappreciated clinical burden by exacerbating infections by isolates carrying these plasmids. Future directions will include more detailed analyses of comorbidity and mortality, as well as sequencing of additional samples to confirm or refute hypothesized routes of plasmid transfer.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CCI: Charlson comorbidity index

CDC: Centers for Disease Control and Prevention

CRE: carbapenem-resistant *Enterobacteriaceae*

EDS-HAT: Enhanced Detection System for Hospital-Associated Transmission

ESBL: extended-spectrum beta-lactamase

HAI: healthcare-associated infection

HAT: healthcare-associated transmission / hospital-acquired transmission

HGT: horizontal gene transfer

ICD: International Classification of Disease

IV: intravenous

KPC: *Klebsiella pneumoniae* carbapenemase

MDR: multidrug resistance (or resistant)

MGE: mobile genetic element

MLST: multi-locus sequence typing

MRI: magnetic resonance imaging

MRSA: methicillin-resistant *Staphylococcus aureus*

PDR: pan-drug resistance / resistant

PFGE: pulsed-field gel electrophoresis

PICC: peripherally inserted central catheter

SNP: single nucleotide polymorphism

ST: sequence type

TEG: thromboelastography

UTI: urinary tract infection

VRE: vancomycin-resistant *Enterococcus*

WGS: whole-genome sequencing

XDR: extensive drug resistance / extensively drug resistant

Preface

I would first like to thank all members of the team working on the Enhanced Detection System for Hospital-Associated Transmission (EDS-HAT) research study, whose groundbreaking efforts forged the bedrock upon which this study was based. This multidisciplinary team included members of the Harrison, Roberts, and Cooper labs at the University of Pittsburgh and the Auton Lab at Carnegie Mellon University – specifically, Dr. Lee Harrison, Dr. Jane Marsh, Dr. Mustapha Mustapha, Alexander Sundermann, Vatsala Srinivasa, Marissa Griffith, and Chinelo Ezeonwuka (Harrison lab), Dr. Vaughn Cooper, Dr. Daniel Snyder, and Dr. Alfonso Santos-Lopez (Cooper Lab), and Dr. Mark Roberts (Roberts lab). Additionally, the collection, generation, processing, and curation of data for this project is a culmination of countless hours of work by numerous physicians, nurses, infection preventionists, bioinformaticians, laboratory technicians, data managers, and clinical research coordinators. While many of these personnel may not be listed as authors, this project would not be possible without hard work from each and every one of them.

Second, I sincerely thank Dr. Daria Van Tyne for her continued supportive mentorship of my professional interests since long before the inception of this project. Since I joined her lab as a laboratory technician in 2018, she has given me numerous chances to explore my interests, try my strengths, and (most importantly) learn from my mistakes. The members of her lab – including Hayley Nordstrom, Dr. Gayatri Chilambi, and Shu-Ting Cho – have also steadfastly supported this project as it has developed. I would also like to thank Dr. Mustapha Mustapha for his subject-matter expertise in microbial genomics and hospital epidemiology, his wisdom, and his friendship. Marissa Griffith also deserves enormous credit for sharing her expertise in

programming, bioinformatics, and database management, and for her endless patience with my many questions and requests for support as this project evolved.

Dr. Robbie Mailliard and Dr. Jeremy Martinson, of the University of Pittsburgh Graduate School of Public Health, have both shown tremendous support for my career development and have offered much help with the complicated logistics of my part-time student schedule. Their appreciation for the unorthodox paths that lurk through academia was critical to my success in this program.

Further, I would like to thank my colleagues in the Infectious Disease Epidemiology division of the Allegheny County Health Department. They have graciously encouraged my academic growth despite the stress and time-intensity of the COVID-19 pandemic, and they have taken steps to blend some of my professional responsibilities with my academic interests. In particular, I would like to thank Dr. Kristen Mertz, Dr. LuAnn Brink, and Jennifer Fiddner.

My parents have been my fiercest cheerleaders and strongest supporters ever since I was born. Their dedication to active parenting and their endless support for all of my interests – from science to baseball and percussion and beyond – made me the man I am today. Words cannot express how grateful I am for their love and for their enthusiasm for my success.

Lastly, I want to thank my brilliant, talented, motivated, passionate, and beautiful wife, Marissa Evans. Every day since we first met in 2015, she has challenged me to be the best version of myself – academically, professionally, and personally. She has supported my work in every possible way, from advising me on the statistical methods I should use to analyze my data, to editing my papers and presentations, listening to my practice talks, sharing in my enthusiasm during my breakthroughs, and comforting me at my lowest moments. I wake up every morning feeling thankful for her love and excited for our happy future together.

1.0 Introduction

1.1 The clinical burden of healthcare-associated bacterial infections

Infections acquired in hospitals and other healthcare settings impose a serious public health burden every year, both in the United States and worldwide. The Centers for Disease Control and Prevention (CDC) estimates that the point prevalence of these nosocomial infections among hospitalized patients in the United States is greater than three percent (1). A 2015 report on the nationwide incidence and prevalence of healthcare-associated infections (HAIs) reported approximately 687,000 infections in American acute care hospitals, as well as 72,000 deaths of patients diagnosed with HAIs while admitted (1). Bacterial pathogens caused more than 90% of these infections, the most common of which were *Clostridioides difficile*, *Staphylococcus aureus*, *Escherichia coli*, *Enterococcus spp.*, *Enterobacter spp.*, *Pseudomonas aeruginosa*, and *Klebsiella spp.* (1). The direct healthcare costs imposed by HAIs in the United States have been estimated in the tens of billions of dollars every year (2), as these infections increase the lengths of hospital stays of infected patients and force the consumption of numerous additional medications, personal protective equipment, and clinical care (2, 3).

While the most recent iteration of the National and State HAI Progress Report showed significant declines in rates of many types of HAIs (4), nosocomial bacterial pathogens remain an enormous burden on American healthcare infrastructure. Additional resources and innovative strategies are needed to reduce both the rate and the severity of these infections, especially as hospital epidemiologists and infectious disease specialists are increasingly occupied with cases and outbreaks of COVID-19 (5).

1.2 Antibiotic resistance, mobile genetic elements, and plasmids

The growing global threat of antibiotic resistance is a major contributor to the public health burden imposed by healthcare-associated infections (HAIs). The 2019 CDC report on antibiotic resistance in the United States estimated that antibiotic-resistant microbial organisms cause approximately 2.8 million infections and nearly 35,000 deaths every year (6). Hundreds of thousands of these infections occur among hospitalized patients, including more than 320,000 infections by methicillin-resistant *Staphylococcus aureus* (MRSA) and 200,000 infections by bacterial species of the family *Enterobacteriaceae* that can produce extended-spectrum beta-lactamases (ESBLs) (7). HAIs caused by antibiotic-resistant bacterial pathogens complicate patients' regimens of care (8), worsen their prognoses (9), dramatically increase their risk of death (8), and substantially increase costs of medical care (2). These effects are particularly profound among infections by multidrug-resistant (MDR) bacterial pathogens, including strains that express ESBLs, carbapenemases, and/or multiple resistance genes (10).

Many genes that confer resistance to antibiotics are carried on mobile genetic elements (MGEs) – sequences of DNA that can move within and between different organisms' genomes independently of the reproduction of those organisms (11) (Figure 1). There are numerous types of MGEs – including plasmids, transposons, genomic islands, prophages, and integrons – that each have characteristic structures and mechanisms of movement within and between genomes (11). The four major mechanisms by which MGEs mobilize within or between bacterial genomes are (a) transformation, the direct uptake of DNA from the environment, (b) transduction, the addition and removal of DNA by prophages embedded in bacterial genomes that are expressed to form lytic bacteriophages, (c) conjugation, the direct shuttling of plasmids between bacterial cells temporarily conjoined by pili, and (d) vesiduction, whereby MGEs are packaged into

extracellular vesicles that fuse with plasma membranes of other bacterial cells (12). The movement of these MGEs is known as horizontal gene transfer (HGT) (Figure 1). In bacteria, many MGEs are composed partly or entirely of smaller MGEs that can be added to, rearranged in, or removed from larger MGEs within which they are held (12). In this regard, MGEs can be likened to Russian nesting dolls that can be rearranged and reshuffled within and between one another (13) (Figure 2).

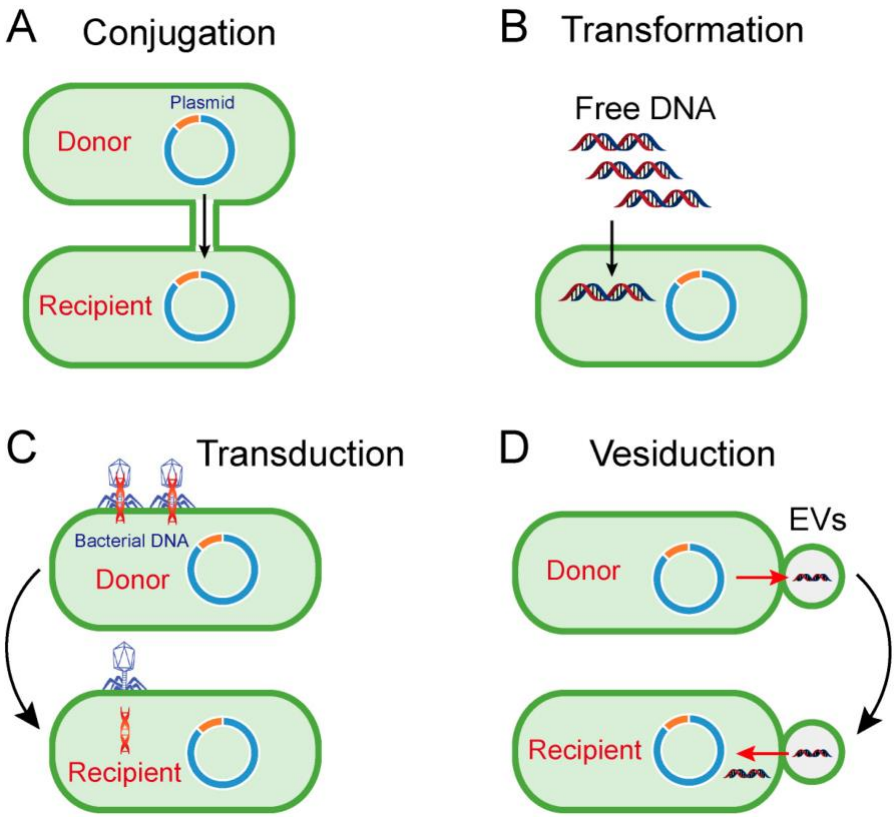

Figure 1: Major mechanisms of horizontal gene transfer in bacteria.
Plasmids can be exchanged directly between bacterial cells by conjugation (A), obtained by bacterial cells from the extracellular environment by transformation (C), or packaged and shuttled in extracellular vesicles by vesiduction (D). Figure originally presented in Liu et al, 2020 (14).

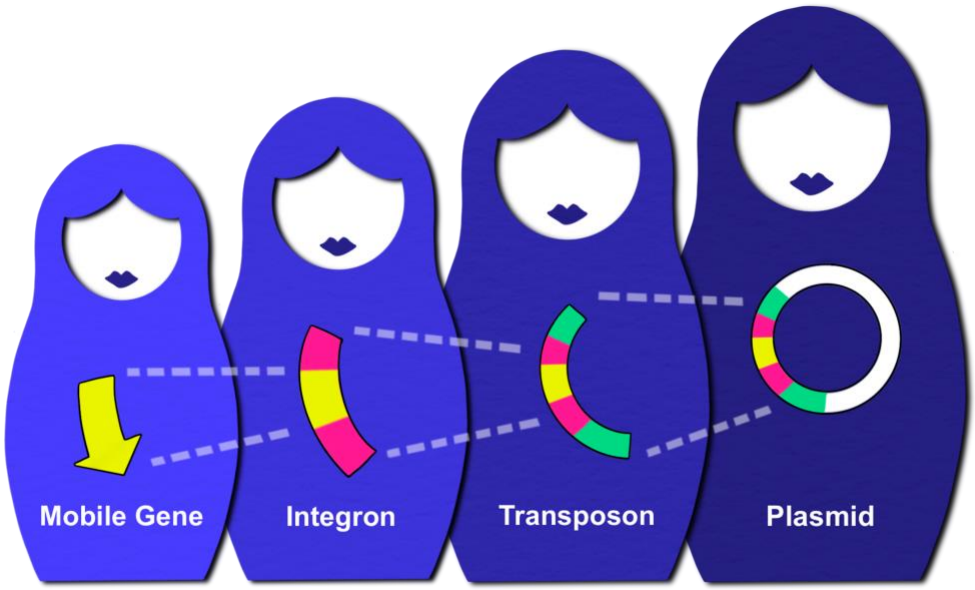

Figure 2: Russian nesting doll-like structures of mobile genetic elements (MGEs) in bacteria.
Plasmids can carry numerous smaller MGEs that themselves can carry smaller MGEs. Gene cargo for these nested MGEs frequently includes genes encoding antibiotic resistance, environmental persistence, and virulence. Figure created by and shared with the permission of Hayley Nordstrom of the Van Tyne Lab, University of Pittsburgh School of Medicine.

MGEs are key players in the spread of antibiotic resistance because HGT can spread beneficial evolutionary traits among bacterial populations much more quickly than classical Darwinian mutation and selection upon those populations (15, 16). With the added selective pressure of the application of enormous quantities of antibiotics both in hospitals, outpatient settings, and agriculture in recent decades (17), the rapid distribution of antibiotic resistance genes on MGEs has become increasingly advantageous for bacteria exposed to these drugs (15). Additionally, some larger MGEs have acquired genes that can confer resistance to other antimicrobial compounds – including metal cations like copper, arsenic, and silver – as well as metabolic operons and genes that improve the formation of biofilms (18). As a result, MGEs that encode antibiotic resistance can spread through and be maintained in bacterial populations that

are not frequently subjected to the selective pressure of antibiotics, as other genes present in those MGEs can provide beneficial adaptations to the bacterial hosts that carry them (17). Additionally, many of the open reading frames encoded on plasmids and transposons have not been fully characterized (19), indicating that these larger MGEs may confer numerous other functions that the scientific community does not yet fully understand.

With recent advances in the field of functional genomics of horizontal gene transfer, some MGEs have become notorious for carrying resistance determinants for different antibiotics (20). In their 3' conserved domains, class 1 integrons all carry the *qacΔE1* quaternary ammonium compound resistance gene and the *sul1* sulfonamide resistance genes (21). Class 2 integrons are highly variable, but typically encode the aminoglycoside resistance gene *aadA1* (21). The Tn4401 transposon is a highly conserved structure known for carrying the *blaKPC-2* carbapenemase gene (22). The ISEcp1 insertion sequence has been predominantly linked to the spread of *blaCTX-M* beta-lactamases (20, 23). Perhaps the most high-profile recently established link between a resistance gene and an MGE was the discovery of the polymyxin resistance gene *mcr-1* on a plasmid carried by an *Escherichia coli* strain from China (24). This discovery was particularly worrisome to the field, because it confirmed that the spread of resistance to this so-called "last-resort" antibiotic could spread between bacterial strains by HGT (25).

Of the numerous types of MGEs found in nature, plasmids are perhaps the most worrisome to the field as vectors of multidrug resistance (MDR) in healthcare settings. This is because plasmids – circularized extrachromosomal DNA that do not require integration into the bacterial chromosome to be functional (26) – can serve as a backbone structure upon which numerous smaller MGEs encoding antibiotic resistance genes can be scaffolded and horizontally transferred together (12). The presence of MGEs among bacterial strains in healthcare settings

not only increases rates of MDR, extensive drug resistance to numerous antibiotics (XDR), and occasionally pan-drug resistance to all available antibiotics (PDR) in nosocomial pathogens, but also facilitates the more rapid transfer of smaller MGEs between different plasmids and bacterial chromosomes (9, 11, 12, 15, 26). While some plasmid-borne resistance genotypes tend to localize geographically (27), other resistance-encoding plasmids have spread rapidly to healthcare settings across the world (28). This pandemic spread of MDR plasmids has greatly contributed to recent increases in cases of and deaths from ESBL-producing *Enterobacteriaceae*, CRE, and other resistant bacterial pathogens (29). Because of their enormous impact on HAIs, studying, tracking, and controlling the horizontal transfer of plasmids has increased in priority in the field over the past several years (30).

1.3 The rapidly evolving field of microbial genomic epidemiology in healthcare settings

Preventing and controlling the spread of healthcare-associated bacterial pathogens has become more innovative and effective in recent years (31–33). Much of this success can be attributed to advances in biotechnology that improved the accuracy and reliability of characterizing nosocomial bacterial strains (31). Wholly reliant barely a generation ago (and still somewhat reliant) on verifying the existence of outbreaks by techniques like pulsed-field gel electrophoresis (PFGE), biochemical typing, and DNA hybridization (34), the field has greatly benefited from recent advances in modern genomics (32, 35). The development of multi-locus sequence typing (MLST) in the late 1990s has allowed clinical microbiologists to differentiate subtypes of numerous bacterial species based on polymorphisms in the DNA sequences of a handful of housekeeping genes (36). MLST effectively balanced the precision of comparing and contrasting genetic polymorphisms between different strains with the cost and complexity of

sequencing technology at the time, and it has continued to be used to efficiently group bacterial strains in the modern genomics era (37, 38). The development and improvement of next-generation sequencing technologies – particularly Illumina's platform of "sequencing by synthesis" – then drove the rapid expansion of whole-genome sequencing (WGS) of bacterial genomes (35, 39, 40). The more recent development of so-called third-generation sequencing – namely Pacific Biosciences' single-molecule real-time (SMRT) (41) and Oxford Nanopore Technologies' nanopore sequencing platforms (42) – has strengthened the accessibility and extended the capabilities of WGS even further (35, 43, 44) (Figure 3). This technology allows for sequencing of long fragments or even complete molecules of DNA, without deliberately fragmenting them into very short sequences during library preparation.

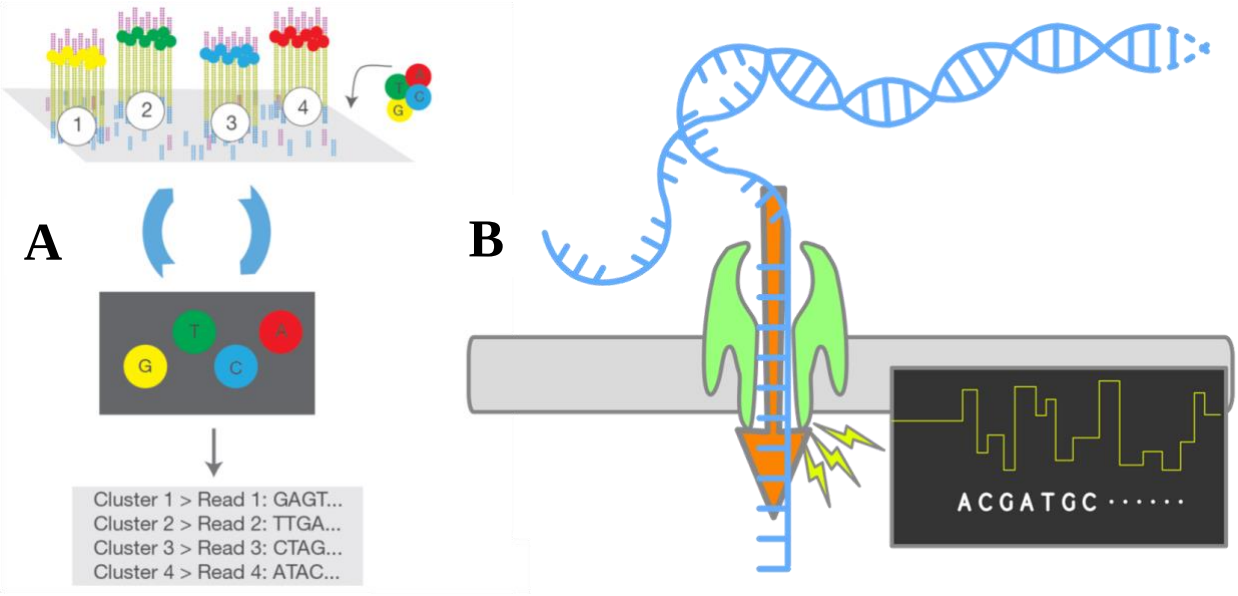

Figure 3: Short-read Illumina and long-read Oxford Nanopore genome sequencing technologies.

Illumina sequencing (A, left) applies the principle of "sequencing by synthesis" to simultaneously detect the presence of nucleotides on numerous short sequences of DNA in a prepared library. Oxford Nanopore sequencing (B, right) involves the detection of changes in electrical currents caused by nucleotides passing through nanopore structures embedded within a

synthetic membrane. Illumina figure originally presented by Illumina, Inc., San Diego CA (45). Oxford Nanopore figure from the Database Center for Life Sciences (DCLS), presented under the CC by 4.0 Creative Commons license.

The major advantage of WGS over prior techniques is that it permits high-resolution comparisons of single nucleotide polymorphisms (SNPs) in bacterial strains across their entire genomes (40). This makes WGS more accurate and high-resolution in the differentiation of strains than PFGE, which uses restriction digestion to generate banding patterns of DNA that is not sequenced and therefore not fully characterized (34, 40). It also makes WGS much more comprehensive and discriminatory than traditional MLST, as comparing thousands of bacterial gene sequences from WGS yields enough resolution to identify relationships within a single sequence type (40, 46). While not all hospitals have the capacity to perform WGS on-site, its utility, accuracy, and cost-effectiveness in identifying and controlling outbreaks have been repeatedly demonstrated in recent years (47–50).

The rapid expansion of WGS in clinical microbiology has created a new specialty within the diverse field of infection prevention – microbial genomic epidemiology. As its name implies, microbial genomic epidemiology (henceforth referred to as "genomic epidemiology") combines analyses of whole-genome sequences of bacterial isolates collected with reviews of geographic, chronological, and/or healthcare-related data linked to those patients to better understand how bacterial pathogens spread between patients or in the environment (32, 35, 50–52). The high specificity of clustering isolates confirmed by WGS to be identical or closely related has allowed genomic epidemiologists to track single bacterial clones throughout numerous locations in single healthcare settings (50, 52). It has also confirmed the global dissemination of emerging microbial pathogens and drug resistance genes (24, 28, 53). Furthermore, retrospective analyses of nosocomial WGS data have helped genomic epidemiologists identify bacterial outbreaks in

hospitals that had been previously undetected by more conventional infection prevention techniques (50, 54). Thanks to improvements in sequencing methods and equipment, as well as the development of new programs that simplify the analysis of WGS data (55, 56), the field of genomic epidemiology continues to expand and become more accessible worldwide.

One consequence of the rapid growth of genomic epidemiology is that it has produced enormous quantities of data that describe the contents of bacterial genomes. Identifying and tracking outbreaks of bacterial strains requires only the comparison of genomic data to one another, not necessarily the detailed characterization of the gene content of those strains (57–59). As a result, the work of genomic epidemiologists across the world has created veritable gold mines of raw data for microbiologists, computational biologists, and bioinformaticians to study. These data have helped the field better understand the functional genomics of virulence and metabolism in bacterial pathogens (60, 61), the distributions and functions of genes that encode resistance to antibiotics, antiseptics, and environmental stressors (20, 62, 63), and the structures and functions of numerous types of mobile genetic elements (MGEs) (20–23). Additionally, much of the WGS data that their studies have generated are linked to epidemiologic data on patient care, locations, and other factors (40, 50, 57). As a result, the field has paved the way for clinical microbiologists and infectious disease physicians to understand how genomic features influence diseases, environmental persistence, and transmission of bacterial pathogens, as well as horizontal gene transfer (35, 62, 64). In fact, the entirety of the study on horizontal gene transfer that is described in this document is based on supplementary analyses of bacterial WGS data that were generated as part of a cutting-edge genomic epidemiology study on the transmission of bacterial pathogens (57, 59).

1.4 Recent discoveries in the genomic epidemiology of plasmids

As discussed previously, plasmids contribute substantially to the antibiotic resistance, environmental persistence, and virulence of healthcare-associated bacterial pathogens (30, 40). Because of this, they have gained considerable attention from genomic epidemiologists and microbiologists in recent years. Improvements in sequencing technology have exponentially increased the ease of characterizing plasmids carried by bacterial strains (43, 65). As a result, tens of thousands of complete plasmid sequences are freely available for scientists all over the world to study (66). Furthermore, the increased application of WGS in healthcare settings has identified multidrug resistance plasmids that engage in horizontal transfer within the bacterial populations of those facilities (27, 43, 44). The combined availability of global plasmid characterization and increased WGS in localized hospitals has opened a subfield of microbial genomics dedicated to the epidemiology of plasmids (32, 35).

Plasmids present in hospital microflora have been shown to engage in two phenomena that impact patient care: mediating outbreaks of bacterial strains by conferring fitness advantages to them, and engaging in horizontal transfer among bacterial strains that are not clonally related to one another. A plasmid-mediated outbreak can occur when a bacterial strain present in a hospital acquires a plasmid whose gene cargo gives that bacterial strain distinct fitness advantages (e.g. resistance, environmental persistence, metabolic advantages, etc.) over other strains (26, 47, 67). The strain, in turn, is more likely to survive in the microbicide-rich hospital and/or to cause severe infections in patients, thereby increasing the probability that an outbreak of that strain will occur (26, 30). In addition to their involvement in these strain-specific outbreaks, plasmids can transfer to bacterial strains that are not clonal derivatives of their host strains by conjugation, transformation, or vesicle-mediated transfer (11, 30). These events of

HGT have been demonstrated both between bacterial strains of the same sequence type, as well as strains of different sequence types, species, and even genera (51, 59). These "secondary transmission networks" – to use one phrase of several that have been coined to describe this phenomenon – allow plasmids to move to new bacterial strains, some of which may end up causing nosocomial infections (68, 69). Importantly, these networks are very difficult to identify rapidly enough to intervene against horizontal transfer (70), and prior research has shown that they can continue undetected after outbreaks of bacterial strains that carry them are controlled (47). This is because conventional diagnostic techniques in clinical microbiology do not detect the presence or absence of plasmids within a strain (43, 70), and because the infection prevention infrastructure of healthcare settings seldom accommodate the detailed genomic analysis that is needed to resolve highly similar plasmids in co-circulating bacterial strains (32). The major focus of the study described in this document is on these secondary transmission networks of the horizontal transfer of plasmids in hospital settings.

When the genomic infrastructure to study plasmid-mediated outbreaks and secondary transmission networks has been available, genomic epidemiologists have made several key discoveries that have advanced the field. First, plasmids co-circulate with one another among nosocomial pathogens in transmission networks that cross the boundaries of bacterial outbreaks. Studies focused on individual species and sequence types (51), multiple species within the *Enterobacteriaceae* family (64), and multiple species of different families and orders have all successfully identified and followed the evolution of plasmids that are conserved within, introduced into, and lost from their bacterial populations of interest (43). Second, plasmids found in hospital environments can have highly conserved structures, highly dynamic structures, or both. Studies that contributed to this finding included Sheppard *et al*, which demonstrated that

nosocomial plasmids served as highly plastic platforms for the recombination of transposons that encoded *bla*$_{KPC}$-family carbapenemase genes (13), as well as David *et al,* which showed some beta-lactamase genes mobilized on highly conserved plasmids and others on numerous diverse plasmid backbones (12). Third, even with modern WGS techniques, establishing clear epidemiologic links for the HGT of plasmids has been inconsistent. Some prior studies have identified clear links between patients treated both in the same and in multiple healthcare facilities found to have carried the same plasmids (44, 67, 68, 71), whereas others were unable to establish links between larger numbers of patients infected by strains that carried the same plasmids (27). Other studies have identified possible locations where HGT of plasmids may take place in environmental reservoirs within the hospital – e.g. sinks, surfaces, and plumbing systems (72–74). Taken together, these findings demonstrate that while studying plasmids with contemporary techniques in genomics have revealed new insight into the wide range of their dynamics and contributions to antibiotic resistance, the epidemiologic factors that contribute to horizontal transfer in hospital settings remain poorly understood.

1.5 Key gaps in the field of the genomic epidemiology of plasmids in healthcare settings

The incomplete understanding of the epidemiology of plasmid transfer in hospital settings can be attributed to several gaps in the focuses and methodologies of recent studies. One of the most significant shortcomings is that most recent studies have focused either on plasmids found within sets of strains that were sequenced for outbreak investigations of bacterial transmission, or on plasmids that encoded resistance to specific classes of antibiotics. Far less attention has been paid to secondary plasmid transmission from more comprehensive, unbiased surveys of greater populations of nosocomial pathogens. As a result, the field of view from

which we have tried to understand the behavior of plasmids in clinical settings is severely limited. Greater understanding of their dynamics requires plasmid-focused analyses that are not biased by bacterial host species or resistance phenotype.

While the broader field is rich with discoveries linking plasmids to one another by high genetic similarity (43, 44, 65, 68, 72, 74), there remain no uniformly adopted standards by which they can be considered similar enough to be epidemiologically linked within a hospital environment. This starkly contrasts with the study of bacterial pathogens in clinical settings, where numerous studies have proposed or validated thresholds of genome identity to infer transmission (75–77). Because of this non-uniformity in the field, genomic inferences of potential plasmid transfer cannot be established with the same degree of confidence as can be made about the transmission of bacterial strains. Studies that move towards establishing such standards for plasmid epidemiology would therefore greatly advance the field.

More recent studies of greater numbers of strains from single hospital systems have begun to capture the influence of plasmid transfer on populations of different bacterial strains and species (51, 58), but the potential routes and risk factors of plasmid transfer that may be inferred from those studies' datasets remain largely unexplored. Additionally, while the field has identified both loose links between patients infected by strains that carry the same plasmids and potential sites or mechanisms by which horizontal transfer occurs in general (27, 44, 59, 78, 79), there have been few reports that simultaneously connected the genomics of plasmids and the precise routes by which they likely horizontally transferred (59). The result of this gap has been much speculation about potential routes of transfer within the hospital environment based on few confirmed or probable events with clear genomic and epidemiologic evidence. More studies that

provide these reliable connections are key to developing effective public health interventions against nosocomial plasmid transfer in hospitals.

1.6 Previous findings and purpose of the current study

Our recent work in the field grew out of our efforts to perform systematic surveys of mobile genetic elements (MGEs) within a single hospital setting. The goal of this prior work was to improve our understanding of the distribution and horizontal transfer of MGEs within a hospital, using methods that minimized our bias towards or against particular species or classes of antibiotic resistance (59). Using a large dataset of bacterial whole-genome sequences collected from healthcare-associated clinical isolates, we identified several plasmids that appeared to be maintained with high sequence preservation in populations of nosocomial bacterial pathogens. Some of these plasmids were present in strains of multiple sequence types, species, or genera. Additionally, we identified two events of likely horizontal transfer of plasmids between strains of different genera that had co-infected two individual patients. We also inferred the likely transfer of one of these plasmids to a third patient with a clear epidemiologic link (Figure 4). The results of this study indicated that applying the same unbiased approach specifically focused on plasmids (as opposed to MGEs of all types) may yield greater insight into how, when, and where HGT occurs within a hospital setting. To accomplish this, we acknowledged our need for more systematic methods both to resolve plasmids from our genomic data and to establish clear links between the patients infected with the isolates that carried those plasmids.

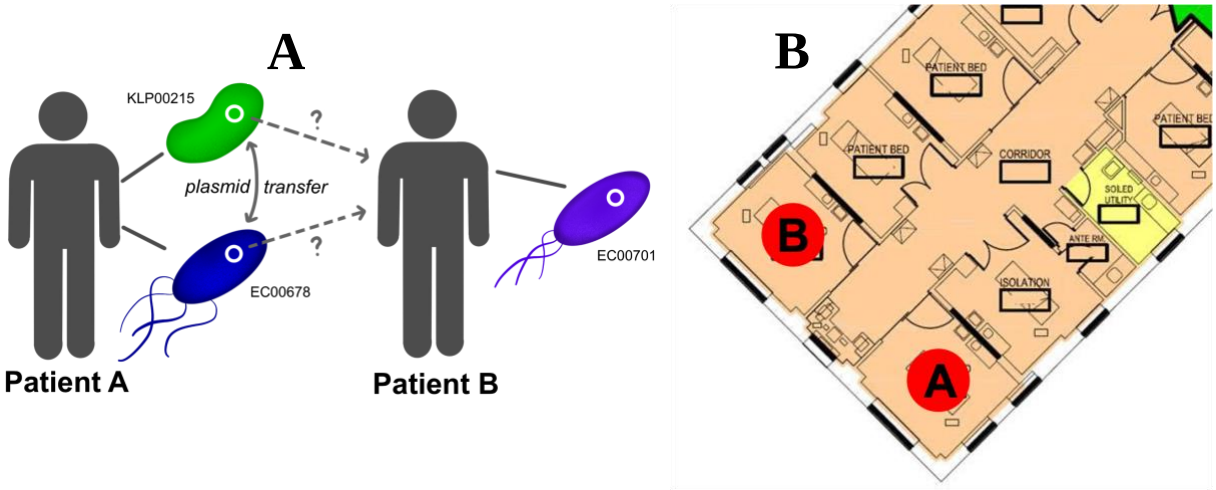

Figure 4: Cross-genus horizontal plasmid transfer among epidemiologically linked hospitalized patients.

(A, left) By performing hybrid assembly of both short-read Illumina and long-read Oxford Nanopore genome sequencing data, we resolved three nearly identical plasmids from clinical isolates that infected two patients. We concluded that two bacterial strains had likely exchanged a plasmid during co-infection of patient A, and then following an unknown intermediate step of bacterial transmission, the plasmid was horizontally transferred into a third strain that caused an infection in Patient B. Image originally presented in Evans et al, *eLife* 2020 (59). (B, right) Schematic of a hospital wing in which Patients A and B had stayed in adjacent rooms on the same unit for three days, after culture of isolates KLP00215 and EC00678 from Patient A but before culture of isolate EC00701 from Patient B.

The purpose of the study described in this document was to study the genomic epidemiology of horizontal plasmid transfer between patients in clinical settings, as well as to explore the clinical burden associated with this transfer. Our goal was to contribute to the field both by proposing new methods to systematically identify plasmids that may be engaged in horizontal transfer within a hospital setting and to study de-identified healthcare data for indications of how horizontal transfer both affected and linked infected patients. Our study helps fill the gaps in the field that were described in the previous section by: (1) using methods that focus primarily on plasmids from numerous bacterial species and with diverse gene cargo; (2) proposing new methods to identify and apply uniform standards of plasmid sequence similarity

for future studies; and (3) directly connecting genomic inference of horizontal transfer of plasmids with epidemiologic data that indicate where, when, and how transfer may have taken place within clinical settings.

2.0 Public Health Significance

Healthcare-associated infections (HAIs) impose a serious burden on healthcare infrastructure, both in the United States and worldwide. Each year in the United States, hundreds of thousands of Americans acquire bacterial infections while receiving medical care in hospitals, outpatient settings, and skilled nursing facilities (4). Plasmids that circulate among the strains that cause these infections carry genes that complicate patient care and increase severity of infection (69, 80). While the processes of horizontal gene transfer (HGT) that circulate these plasmids are known to take place in hospital settings (43, 59, 74), the clinical burden of healthcare-associated HGT and the routes by which it occurs remain poorly understood. Additionally, the field of plasmid epidemiology currently lacks consistently applied thresholds of sequence similarity by which recent horizontal transfer between strains can be reasonably inferred. Understanding both the burden and routes of this phenomenon could help infection prevention personnel prioritize and control horizontal gene transfer in hospitals, potentially limiting the severity of nosocomial bacterial infections.

This study contributes to public health by quantifying the clinical burden of, and describing potential routes and risk factors to, the horizontal transfer of plasmids in a hospital setting. Our findings describe the presence of an unmet clinical burden of the nosocomial horizontal transfer of plasmids between clinical bacterial isolates that cause HAIs, as well as potential routes and risk factors linked to this horizontal transfer. We also present novel strategies to study this phenomenon with systematic methods for screening genomic, epidemiologic, and clinical data linked to bacterial isolates. Our findings can help broaden the scope of infection prevention to control and prevent the exchange of plasmids between bacterial

strains by horizontal transfer, thereby mitigating the severity of nosocomial infections that benefit from the genetic cargo on those plasmids.

3.0 Hypotheses and Specific Aims

3.1 Hypotheses

Hypothesis 1: Plasmids engage in horizontal transfer in hospitals between bacterial strains that cause nosocomial infections. This phenomenon occurs often enough that it can be detected and characterized by systematic screening of whole-genome sequence data.

Hypothesis 2: Leveraging systematic genomic screening data to perform outbreak investigations of plasmids in hospitals will identify previously undetected routes of nosocomial plasmid transfer.

Hypothesis 3: Plasmids engaged in nosocomial horizontal gene transfer are associated with increased burden of disease in healthcare-associated bacterial infections caused by strains that carry those plasmids, compared to infections by strains that do not.

3.2 Specific Aims

Aim 1: To develop and apply novel methods to systematically infer the horizontal transfer of plasmids among bacterial strains that cause nosocomial infections.

> Aim 1a: To establish reasonable thresholds of plasmid sequence similarity that can be used to infer the occurrence of nosocomial plasmid transfer.
>
> Aim 1b: To systematically resolve and characterize lineages of plasmids engaged in horizontal transfer among clinical bacterial strains within a single hospital system.

Aim 2: To identify likely exposures and risk factors, and assess the clinical burden of plasmids engaged in nosocomial horizontal gene transfer.

Aim 2a: To identify common exposures associated with horizontal transfer of plasmids, as well as clinical factors that may predispose patients to greater risk of acquiring these plasmids.

Aim 2b: To compare measures of disease burden between bacterial infections involving plasmids engaged in nosocomial horizontal transfer with infections not involving those plasmids, using de-identified patient healthcare data.

4.0 Materials and Methods

4.1 Collection of clinical bacterial isolates and corresponding patient data

All bacterial isolates analyzed in this study were collected through the Enhanced Detection System for Hospital-Associated Transmission (EDS-HAT) project, an ongoing research initiative led by a team of clinicians, microbiologists, and epidemiologists at the University of Pittsburgh School of Medicine and the University of Pittsburgh Medical Center (50). Isolates were eligible for inclusion in this study with a positive clinical culture from a patient admitted to the university's flagship hospital system either at least 3 days after admission, and/or if the patient had any procedure or prior inpatient stay within 30 days of the collection of the isolate. All isolates of the genera *Acinetobacter*, *Burkholderia*, *Citrobacter*, *Proteus*, *Providencia*, *Serratia*, and *Stenotrophomonas* that were eligible for this study were included. Inclusion of strains from other genera were limited by expression of antibiotic resistance or toxin production, i.e. toxin-producing *Clostridioides difficile*, vancomycin-resistant *Enterococcus spp.* (VRE), extended-spectrum beta-lactamase (ESBL)-producing *Escherichia spp.* and *Klebsiella spp.*, and methicillin-resistant *Staphylococcus aureus.* These isolates were identified using TheraDoc software (Version 4.6, Premier, Inc, Charlotte NC). The University of Pittsburgh Institutional Review Board (IRB) has approved the EDS-HAT project, classifying it as being exempt from informed consent because direct contact with human subjects is not performed. The IRB also issued an exemption (STUDY20060252) to use data collected for the EDS-HAT project to study the horizontal transfer of MGEs in the hospital setting.

4.2 Short-read and long-read whole-genome sequencing of clinical bacterial isolates

Whole-genome sequencing (WGS) was performed using genomic DNA that was extracted from overnight cultures of single bacterial colonies, using the Qiagen DNeasy Tissue Kit (Qiagen, Germantown MD). Short-read WGS was performed using Illumina technology (Illumina, San Diego CA). Libraries for Illumina sequencing were constructed using the Illumina Nextera DNA Sample Prep Kit with 150bp paired-end reads. Libraries were sequenced on the Illumina NextSeq platform. Long-read sequencing and base-calling was performed using Oxford Nanopore technology (Oxford Nanopore Technologies, Oxford, United Kingdom). Libraries were constructed using a rapid multiplex barcoding kit (catalog number SQK-RBK004). Sequencing was performed using an Oxford Nanopore MinION device with R9.4.1 flow cells. Base-calling was performed using Albacore v2.3.3 or Guppy v2.3.1 (Oxford Nanopore Technologies, Oxford, United Kingdom).

4.3 Read processing and assembly of whole-genome sequencing data

Short-read Illumina sequencing data were processed prior to assembly. Trim Galore v0.6.1 was used to remove sequencing adaptors, low-quality bases, and poor-quality reads. Bacterial species were identified from processed Illumina reads by alignment to Kraken v1.0 and RefSeq databases (81, 82). Genomes of strains sequenced only with Illumina technology were assembled using SPAdes v3.11 (83). The quality of assembled genomes was then verified using QUAST (84). Assembled genomes were excluded if they failed to meet the following four quality control parameters: genome-wide read depth of at least 40X, cumulative length of assembled genome within 20% of the expected length for the assigned genus, fewer than 400 contigs in the assembled genome, and an N50 value of less than 50,000bp.

Long-read sequencing data were processed using default read quality parameters of Albacore v2.3.3 or Guppy v2.3.1 (Oxford Nanopore Technologies, Oxford, United Kingdom). Hybrid assembly was performed for genomes for which both short- and long-read sequencing data were available and whose short-read only assemblies passed the aforementioned parameters, using Unicycler v0.4.7 or v0.4.8-beta (85) (Figure 5).

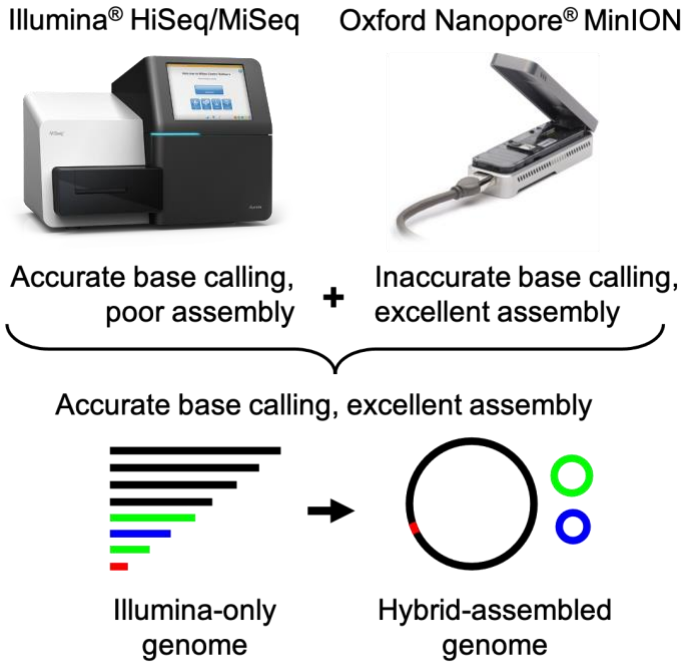

Figure 5: Conceptual framework and utility of hybrid genome assembly. Bacterial genomes sequenced only by Illumina technology have highly accurate nucleotide sequences, but assembly is typically fragmented. Genomes sequenced only by Oxford Nanopore technology have less accurately called sequences but can be assembled to much greater contiguity. Successfully using both technologies yields accurate, well assembled whole-genome sequences that clearly define the genomic context of mobile genetic elements.

4.4 Characterization, alignment, and phylogenetic analyses of whole-genome and plasmid sequences

Assembled genomes and plasmids were annotated with Prokka v1.13 or v1.14 (86). Multi-locus sequence types (STs) were assigned using *mlst* v2.16.1 with PubMLST typing

schemes (37, 38). Antibiotic resistance genes were identified by BLASTn alignments of assembled genomes to the ResFinder v4.1 database (87, 88). Plasmid replicons were identified by BLASTn alignments of assembled genomes to the PlasmidFinder v2.1 database (89). Other genomic features, e.g. metal resistance genes and virulence factors, were identified from annotations by Prokka (86) and by BLASTn alignments to the VFDB and VirulenceFinder databases (90, 91). Genomospecies were defined by grouping isolates with core genome average nucleotide identity (ANI) of at least 95% to one another and less than 95% ANI to any other isolate genome. Phylogenies of bacterial whole-genome sequences were constructed from core genome alignments generated by Roary v5.18.2 (92) and visualized using RAxML v8.0.26 with 1,000 bootstrap iterations (93). Phylogenies of plasmids were constructed from core genome alignments of reference plasmids to bacterial genomes using *snippy-core* v4.4.5 (94) and visualized using RAxML v8.0.26 with 1,000 bootstrap iterations (93). Annotated plasmid sequences were also aligned to one another using EasyFig v2.2.2 (95).

4.5 Analysis of plasmids from known outbreaks to establish parameters of sequence similarity

Plasmids were identified from hybrid-assembled genomes of isolates for which other clonal isolates existed within the dataset. Contigs were selected as plasmids if they met the following criteria: closed circular as determined during hybrid assembly (85), longer than 2kb but not longer than 300kb, and possessing at least one replicon as identified by PlasmidFinder (89). Single nucleotide polymorphisms (SNPs) relative to plasmid sequences were identified using *snippy-core* v4.4.5 (94). Preservation of reference plasmid gene content ("coverage") in

other bacterial strains was calculated by mapping plasmid sequences to contigs of bacterial genomes using BLASTn (96), using a sequence identity threshold for alignments of 95%.

4.6 Resolution and characterization of plasmid lineages using closed circular plasmids

Reference plasmid sequences used to resolve groups of strains carrying the same plasmid within the hospital setting ("lineages") were identified from 56 of 78 total hybrid-assembled whole-genome sequences of clinical bacterial isolates, using the aforementioned selection parameters. 3,074 whole-genome sequences from clinical bacterial isolates were then screened for the presence of plasmids by calculating preservation of plasmid gene content by BLASTn-based mapping (96) and identifying SNPs using *snippy-core* (94), as described earlier. A plasmid lineage was identified if a reference plasmid was resolved at sufficient nucleotide similarity and gene content preservation in at least three bacterial strains of the same ST, or in at least two strains of different STs, genomospecies, or genera. Lineages were de-duplicated by aligning and visualizing closed reference plasmid sequences of similar lengths and distribution among sequence types and genomospecies using EasyFig v2.2.2 (95).

4.7 Identification of potential routes of plasmid transfer within the hospital setting

Potential routes of horizontal transfer of plasmids of interest were identified by a case-control methodology, using systematic screening of patient electronic health record (EHR) data by a previously published automated data mining algorithm (57). For each plasmid lineage resolved by our genomics approach, hospital charge codes and admission records were systematically reviewed and assessed for statistically significant enrichment among patients infected by strains carrying plasmids in that lineage. Charge codes for identical procedures

performed on two or more patients, as well as admission records for patients who had the same rooms or units in their hospital records, were counted as potential shared exposures among infected patients if those events occurred at any time between 30 days prior to culture of the earliest isolate in the plasmid lineage and the date of culture of the latest isolate in the lineage. Admission records for roommates with overlapping stays were identified as shared exposures regardless of the length of cohabitation. The algorithm then applied the same parameters to identify the co-occurrence of the same charge codes or admission records in patients not infected by any of the plasmid-carrying bacterial strains. Potential routes of plasmid transfer identified by the algorithm were then manually filtered to exclude routes involving only infections by bacterial strains of the same sequence types, as well as any isolates whose patients were exposed to those routes after the dates of culture of isolates included in the plasmid lineages being examined.

Odds ratios were calculated from numbers of patients carrying or not carrying plasmids of interest and numbers of patients with or without potential exposures identified from charge codes or admission records. The Haldane-Anscombe correction was applied to all potential exposures for which one count of patients was equal to zero (97). P-values for each potential exposure were then calculated by a Z-score method based on standard errors of odds ratios that is described in Sheskin, 2004 (98). Multiple hypothesis corrections for statistical significance were performed by the Benjamini-Hochberg method (99), using false discovery rate thresholds equivalent to the inverses of the total numbers of procedures or geographic shared exposures.

4.8 Analyses of clinical data linked to patient hospitalization records

Clinical data on patient demographics, comorbidities, admission and discharge dates were collected and summarized by an honest broker. Charlson Comorbidity Index scores were

calculated using International Classification of Disease (ICD) diagnoses from inpatient and outpatient encounters, from the date one year prior to admission to the date of admission during which one or more isolates in the study were cultured (100). Length of hospital admission was calculated from recorded dates of admission and discharge, with hospitalizations with admission and discharge dates of the same date recorded as a length of 1 day. Hospitalizations with discharge and readmission to units in the same hospital on the same dates were combined and counted as single hospitalizations. Comparisons of clinical factors with continuous variables were performed using the Wilcoxon rank sum test using GraphPad Prism v9.0.0 software (GraphPad Software Inc., California, USA). Outliers within a group of continuous variables were defined as any points greater than 3 standard deviations from the mean, and were not included in calculations of significance.

5.0 Results

5.1 Establishment of plasmid sequence similarity thresholds to indicate potential horizontal transfer

While the epidemiology of plasmids among hospital bacterial isolates has previously been characterized (44, 44, 59), the field lacks consistent guidelines for determining thresholds of sequence similarity by which the occurrence of horizontal gene transfer can be hypothesized. We therefore concluded that to perform our study, we first needed to establish these thresholds for plasmid similarity. We reasoned that plasmids undergoing horizontal transfer between clinical bacterial strains would most likely maintain sequence identity to a degree that is similar to identity between plasmids carried by bacterial strains that were known to (1) have transmitted between patients or (2) been maintained within the same patient for a prolonged period of time. Thus, our first steps were to identify plasmids that were maintained among such bacterial strains and to measure their sequence similarity.

To collect these data, we studied four hybrid-assembled plasmid sequences that were present in three clusters of bacterial strains that showed strong genomic and epidemiologic evidence of nosocomial transmission (Table 1). Three of these plasmids were sequenced as part of outbreak investigations that have been previously published (50, 51). Two plasmids were from different sequence types of *Klebsiella pneumoniae* (ST258 and ST307), and two were present in the same cluster of *Enterococcus faecium* isolates (ST1471). We also studied three pairs and two triplets of previously published hybrid-assembled plasmids that had been resolved from isolates that infected the same patients or were reported as members of the same plasmid lineage within the hospital system (Table 1) (59). Additionally, we included 33 strains (23 *Klebsiella spp.*, 6 *Escherichia spp.*, and 4 *Acinetobacter spp.*) that (a) comprised pairs and triplets of isolates

collected from the same patients and (b) showed high nucleotide sequence coverage of any of 12 plasmids that had been previously sequenced and hybrid-assembled (58). Collectively, this dataset included 25 plasmids from 5 genera (4 Gram-negative and 1 Gram-positive) of bacterial pathogens. Elapsed time between dates of culture of plasmid-linked strains within this dataset ranged from 0 days to 427 days (mean 119.5 days, median 82 days).

Plasmid Name	Length (kb)	Replicons / Incompatibility	Source Isolate(s)	Reason for Inclusion
pKLP00149_2	165.2	IncFII (pBK30683)	*K. pneumoniae* ST258	Outbreak isolate described in Marsh *et al*, 2019
pKLP00218_2	164.7	IncFIB IncFII(K)	*K. pneumoniae* ST307	Clonal isolates identified from core genome phylogenetic analyses
pVRE32553_2	223.8	repUS15 repA(pNB2354p1)	*E. faecium ST1471*	Outbreak isolate described in Sundermann *et al*, 2019
pVRE32553_3	51.7	rep17 CDS29(pRUM)	*E. faecium ST1471*	Outbreak isolate described in Sundermann *et al*, 2019
pKLP00177_3 pKLP00203_3	170.8 170.1	IncFIB(K)	*K. pneumoniae* ST2712 *K. pneumoniae* ST25	Highly similar plasmids from a lineage described in Evans *et al*, 2020
pKLP00161_2 pKLO00017_2	236.5 226.8	IncFIB IncFII(K)	*K. pneumoniae* ST405 *K. oxytoca* ST207	Highly similar plasmids from a lineage described in Evans *et al*, 2020
pKLP00187_2 pCB00017_2	196.7 196.8	IncFIB IncFII(K)	*K. pneumoniae* ST231 *C. brakii* ST356	Nearly identical plasmids present in two isolates co-infecting the same patient, described in Evans *et al*, 2020
pKLP00215_4 pEC00678_3 pEC00701_3	113.6 113.6 113.6	IncFIB IncFII(K)	*K. pneumoniae* ST405 *E. coli* ST69 *E. coli* ST131	Nearly identical plasmids from infections of two patients with an epidemiologic link, described in Evans *et al*, 2020

Table 1: Reference sequences used to calculate plasmid sequence similarity.
The dataset included four closed plasmids from hybrid-assembled genomes of clinical isolates that caused outbreaks, two pairs of highly similar closed plasmids from genomes of different sequence types (STs) or species, and one pair and one triplet of highly similar closed plasmids with clear epidemiologic links of nosocomial horizontal transfer.

We used *snippy-core* to quantify normalized rates of single nucleotide polymorphisms (SNPs) in the core sequences shared by reference plasmids and the non-hybrid-assembled whole-genome sequences of bacterial strains to which they were epidemiologically linked ("core mutation rate"). We also used BLASTn to quantify percentages of reference plasmid sequence content that were preserved at high sequence identity among these isolates ("sequence coverage"). We observed that among groups of bacterial isolates linked to reference plasmids, 95% of isolates accrued fewer than 14 single nucleotide polymorphisms (SNPs) per 100kb of sequence length of the reference plasmid, i.e. 1 core SNP per 7.2kb of reference sequence

(Figure 6a). Additionally, 95% of isolates maintained at least 94.7% of the sequence content of their corresponding reference plasmid sequence (Figure 6b). Notably, 17 pairs of isolates whose plasmids fell within the 95th percentile thresholds of coverage and mutation rates were cultured more than 180 days apart. These results indicate that plasmids can remain remarkably well preserved in nosocomial pathogens, both within bacterial strains that are known to be transmitting between patients and within bacterial strains that exchange these plasmids by horizontal transfer.

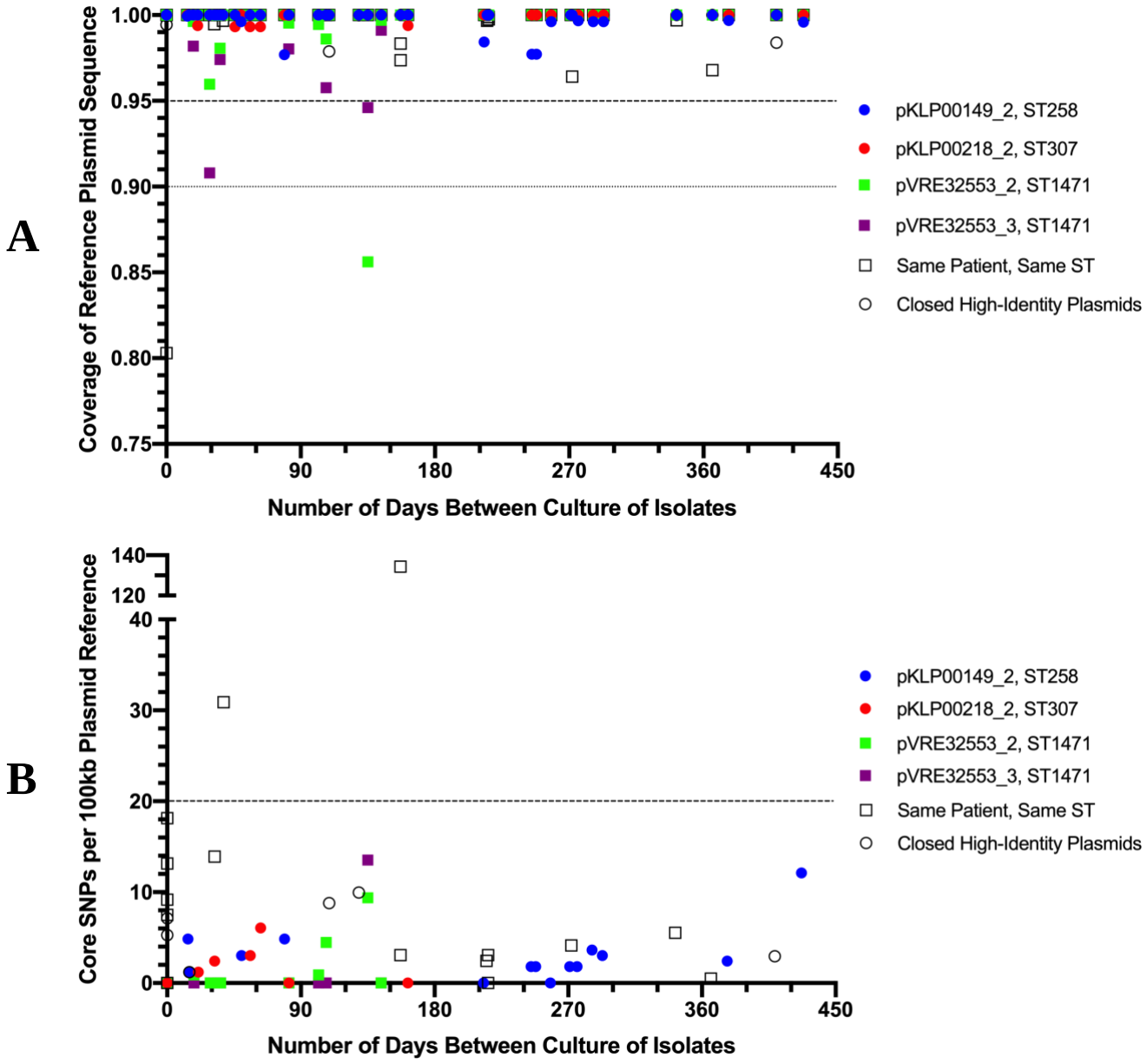

Figure 6: Preservation of gene content and nucleotide identity among plasmids involved in healthcare-associated transmission or horizontal transfer.

(A, top) Plots of loss of reference plasmid gene content, shown as proportions of lengths of reference plasmids, versus time between dates of culture of linked isolates. The thick dashed line shows the 5th percentile of plasmid coverage, indicating that 95% of linked strains shared at least 95% of the gene content of their shared plasmids. The thin dashed line shows the coverage threshold used for our prior study of plasmid transfer (58). (B, bottom). Plot of single nucleotide polymorphisms (SNPs) between reference plasmids and linked strains versus time between dates of culture of linked isolates. SNP values are normalized by the lengths of reference plasmids.

The thick dashed line shows the 95th percentile threshold of 20 SNPs per 100kb of plasmid length (1 per 5kb).

5.2 Application of plasmid similarity thresholds to systematically resolve lineages of plasmids in a large genomic dataset from a single hospital system

After calculating similarity thresholds based on plasmid sequences shared among bacterial strains that were likely transmitted between or preserved within infected hospitalized patients, we applied those thresholds to systematically resolve shared plasmids among all bacterial isolates collected from a single hospital system. To perform this study, we used 3,074 whole-genome sequences of clinical bacterial isolates that had been collected as part of the Enhanced Detection System for Healthcare-Associated Transmission (EDS-HAT) study, an ongoing research program that combines WGS data with de-identified patient care data to identify and predict routes of bacterial transmission in hospital settings (57). We used genomes of 3,074 bacterial strains that had been collected from 2,322 hospitalizations of 1,960 patients treated at a single tertiary hospital system over a two-year period from 2016 to 2018 (Tables 2a and 2b). Each bacterial strain met predetermined criteria for being classified as having been likely involved in nosocomial transmission (see Methods).

A

Genus of Plasmid Host	Isolates in Dataset	Isolates With Plasmid Lineage	Isolates Without Plasmid Lineage
Acinetobacter	82	5 (6.1%)	77 (93.9%)
Burkholderia	12	0 (0%)	12 (100%)
Citrobacter	126	21 (16.7%)	105 (83.3%)
Clostridioides	517	0 (0%)	517 (100%)
Escherichia	170	28 (16.5%)	142 (83.5%)
Enterococcus	250	208 (83.2%)	42 (16.8%)
Klebsiella	136	81 (59.6%)	55 (40.4%)
Proteus	151	0 (0%)	151 (100%)
Providencia	13	0 (0%)	13 (100%)
Pseudomonas	894	2 (0.2%)	892 (99.8%)
Serratia	180	2 (1.1%)	178 (98.9%)
Staphylococcus	420	2 (0.5%)	418 (99.5%)
Stenotrophomonas	123	9 (7.3%)	114 (92.7%)
Total	**3074**	**353**	**2721**

B

Source of Clinical Isolate	Isolates in Dataset	Isolates With Plasmid Lineage	Isolates Without Plasmid Lineage
Respiratory tract	1024	55 (5.4%)	969 (94.6%)
Tissue or wound	648	101 (15.6%)	547 (84.4%)
Urinary tract	627	151 (24.1%)	476 (75.9%)
Stool	517	0 (0%)	517 (100%)
Blood	258	46 (17.8%)	212 (82.2%)
Total	**3074**	**353**	**2721**

Table 2: Clinical bacterial isolates examined for the presence of plasmid lineages engaged in horizontal transfer.

(A, top) Distribution of isolates by genus. (B, bottom) Distribution of isolates by source of clinical isolate, a proxy for type of infection.

We had previously resolved 142 closed plasmids from 56 of the 78 bacterial strains in our dataset whose genomes were assembled using both short-read Illumina and long-read Oxford Nanopore sequencing data (Table 3). 12 of these plasmid sequences have previously been described in an earlier study of the diversity of mobile genetic elements in a hospital setting, 10 of which were shown to be members of plasmid lineages maintained by bacterial transmission or horizontal plasmid transfer (59). Some of the other sequences were confirmed or potential duplicates of one another or of these previously documented plasmids. We used BLASTn to query the 3,074 isolates' genomes against the 142 plasmid sequences and calculate sequence coverage of each plasmid within each genome. We then calculated core mutation rates among all genomes that had at least 95% sequence coverage of at least one of the plasmids.

Genus of Isolate Carrying Reference Plasmid Sequence	**Plasmid Sequences Used to Resolve Lineages (not de-deduplicated)**	**Plasmids in Lineages Screened for Epidemiologic Links (de-duplicated)**	**Plasmids in Lineages with Epidemiologic Links* (de-duplicated)**
Acinetobacter	1	1	1
Citrobacter	5	3	2
Clostridioides	1	0	0
Escherichia	48	4	1
Enterococcus	32	13	11
Klebsiella	45	16	12
Pseudomonas	4	1	0
Serratia	4	0	0
Staphylococcus	1	1	0
Stenotrophomonas	1	1	1
Total	**142**	**40**	**28**

Table 3: Genus distribution of reference plasmids used to identify plasmid lineages circulating in the hospital system by horizontal transfer.

From these isolates, we identified 40 groups of strains that met the following criteria: (1) they contained at least 3 strains in the group and/or included strains of different STs, genomospecies, or genera; (2) they shared the same plasmid with at least 95% sequence coverage and fewer than 20 SNPs per 100kb relative to the reference; and (3) if the reference plasmid was less than 10kb in length, it contained a known plasmid replicon or incompatibility group sequence (Table 3). We labeled these 40 groups of strains "plasmid lineages". Reference sequences for plasmid lineages ranged in length from 4.0kb to 242.3kb (mean 82.8kb, median 47.4kb) and carried 20 unique replicons and incompatibility groups. Collectively, the lineages that we identified contained 353 (11.4%) of the 3,074 isolates, from 15 of 96 genomospecies and 9 of 12 genera in the dataset. A total of 303 (13.0%) of the 2,322 hospitalizations, of 246 (12.6%) of the 1,960 patients in our dataset, involved infection by at least one strain that was included in at least one plasmid lineage. Within specific genera of bacterial pathogens, plasmids in lineages were most abundant in isolates of the genus *Enterococcus* (n = 203, 83.2% of isolates in genus), *Klebsiella* (n = 81, 59.6%), *Escherichia* (n = 28, 16.5%), and *Citrobacter* (n = 21, 16.7%) (Table 2) (Figure 7). Lineages of plasmids from *Enterococcus spp.* included by far the greatest numbers of isolates, with five lineages (pVRE32994_4, pVRE33085_3, pVRE33085_4, pVRE33562_4, and pVRE33562_5) containing 45 or more isolates each.

Twenty-six of the 40 reference plasmids for the lineages we resolved carried at least one known antibiotic resistance gene; these lineages included 249 (8.1%) of the strains from 214 (9.2%) hospitalizations of 194 (9.9%) patients in our dataset. Within these reference plasmids, the most abundantly present antibiotic resistance genes encoded aminoglycoside resistance (n = 21 lineages), ESBL enzymes (n = 14), sulfonamide resistance (n = 12), and macrolide resistance (n = 11). Five reference plasmids carried at least one carbapenem resistance gene, namely

blaKPC-2 (n = 4) and *blaKPC-3* (n = 1). Seventeen reference plasmids carried known resistance genes or operons for at least one metal; these lineages included 205 (6.7%) of the strains from 177 (7.6%) hospitalizations of 160 (8.2%) patients. The most abundant classes of metal resistance included resistance to copper (n = 9 lineages), silver (n = 6), arsenic (n = 6), and cadmium (n = 6). Fourteen plasmid lineages did not carry any known antibiotic or metal resistance genes; these lineages included 120 (3.9%) strains of 89 (5.4%) hospitalizations of 87 (3.7%) patients. Six of these lineages carried at least one virulence factor or other environmental persistence gene. The lineages pCB00073_2 and pKLP00155_6 carried cloacin operons (101); pVRE33085_3 and pVRE33562_6 carried the bacteriocins microcin C7 and hiracin JM79, respectively (102, 103); pACIN00156_2 carried a gene encoding resistance to formaldehyde; pSTEN00043_8 carried the non-hemolytic phospholipase *plcN* (104); and pVRE33085_3 also carried the bile salt hydrolase *bsh* (105).

Taken together, these results show that the plasmid lineages we resolved from our genomic dataset were highly diverse in length, incompatibility group, and gene content, and that they were present in many patients infected with nosocomial bacterial pathogens.

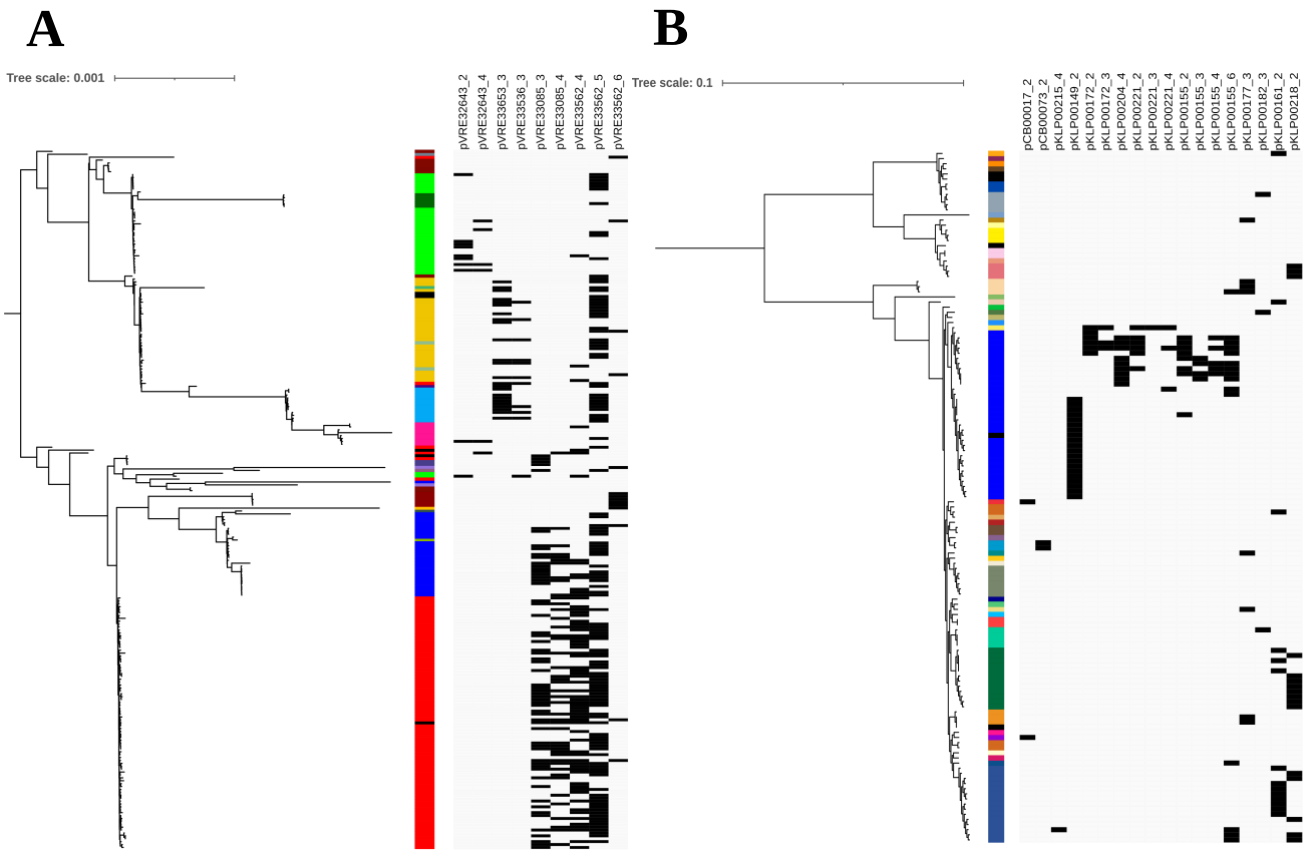

Figure 7: Phylogenetic profiling of plasmid lineages.

Presence or absence of plasmids in lineages among isolates of (A, top) *Enterococcus faecium* and (B, bottom) *Klebsiella spp.* The top clade of the phylogenetic tree of *Klebsiella spp.* contains isolates of *K. oxytoca*, *K. michiganensis*, and *K. grimontii*; the bottom clade contains isolates of *K. pneumoniae*. Phylogenies were constructed from bacterial core genome alignments using RAxML v8.0.26 with 100 bootstrap iterations and visualized using ITOL v5.7 (93, 106). Color strips identify sequence types (STs) of bacterial strains.

5.3 Improved accuracy and refined resolution of plasmid lineages over previously published methods

The 40 closed reference plasmids that we used in this study included 10 sequences that had been previously used to resolve lineages with a less stringent similarity threshold based solely on sequence coverage (59). We therefore compared the results of our newer, more stringent methods to the results of our prior study, with the hypothesis that our new method would more precisely identify and differentiate lineages than did our previously published methods.

Of the previously resolved plasmid lineages, the largest by number of strains was a small ColRNAI plasmid (pKLP00155_6, 9.5kb) whose primary cargo was a cloacin- and cloacin immunity-encoding gene cassette. When attempting to reconstruct this lineage with our new method, core mutation rates calculated relative to pKLP00155_6 plasmid indicated that this single lineage was a combination of two closely related plasmid lineages carried by different strains (Figure 8a). 13 of the strains in the original lineage showed consistently greater sequence identity to another closed reference plasmid (pCB00073_2, 9.3kb) in the original lineage than to the pKLP00155_6 reference plasmid. Constructing a maximum likelihood phylogeny of the core plasmid sequence shared by all strains in the lineage divided the strains in the lineage into two separate clades, one of which included pCB00073_2 and the other pKLP00155_6 (Figure 8b). This confirmed that two separate lineages had been combined under a single label, and that including the core mutation rate threshold improved our ability to accurately resolve plasmid lineages.

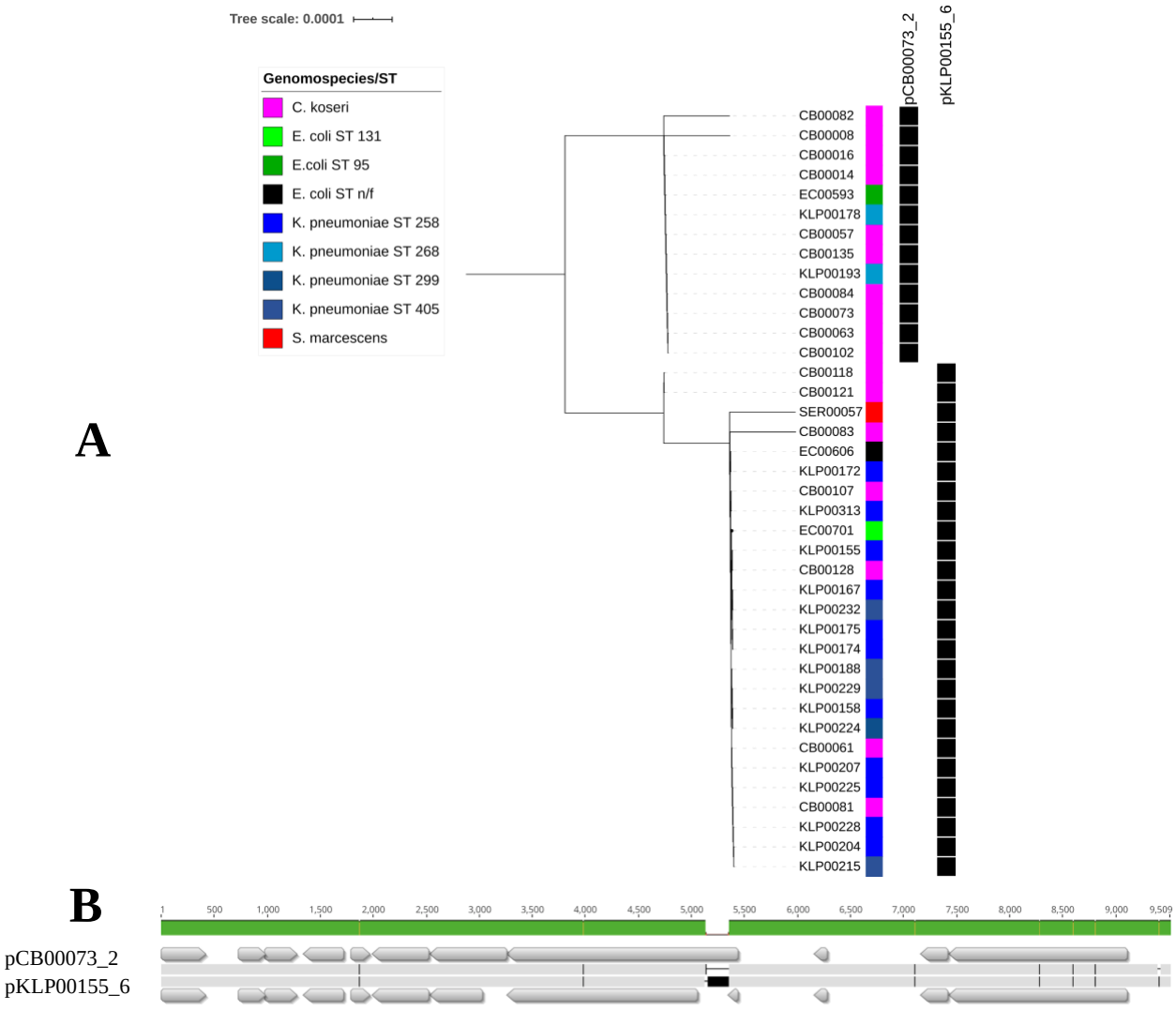

Figure 8: Separation of a previously defined plasmid lineage into two lineages using updated methods.

(A, top) Phylogeny of core plasmid sequences of the lineage pKLP00155_6, as defined in Evans *et al*, 2020 (58), showing sorting of plasmid sequences into two clades. The phylogenetic tree was constructed using RAxML v8.0.26 with 1000 bootstrap iterations and visualized using ITOL v5.7, using pKLP00155_6 as an internal reference (93, 106). (B, bottom) Alignment of plasmid sequences annotated using Prokka and visualized in Mauve (86, 107).

Our prior work also focused on two other plasmids, pKLP00161_2 and pKLP00218_2, that were substantially different in length, but pKLP00161_2 appeared to contain most of the sequence of pKLP00218_2 (Figure 9a). Using only sequence coverage data, we had previously sorted these plasmids and the strains that carried them into two separate lineages. When attempting to reconstruct these lineages with our updated methods, we found that all strains in the pKLP00161_2 lineage had core mutation rates relative to pKLP00218_2 that met the thresholds we had established, and vice versa. Additionally, the extra sequence that was present in pKLP00161_2 but absent in pKLP00218_2 included four additional antibiotic resistance genes: the ciprofloxacin resistance gene *aac(6')-Ib-cr*, the ESBL gene *blaOXA-1*, the quinolone resistance gene *qnrB1*, and the tetracycline resistance gene *tet(A)*. Given these new data, as well as the fact that the earliest-cultured strain in the pKLP00218_2 lineage was collected more than 6 months after the earliest-cultured strain in the pKLP00161_2 lineage, we hypothesized that the pKLP00218_2 lineage was a derivative of the pKLP00161_2 lineage that had evolved in the hospital by losing smaller mobile genetic elements – which, among other cargo of hypothetical open reading frames, carried multiple drug resistance genes – that were nested within pKLP00161_2. Constructing a maximum likelihood phylogeny of the core plasmid sequence shared by the isolates in both lineages validated this conclusion by grouping all members of the two lineages, save one, into a single clade (Figure 9b). These results further support the utility of calculating rates of point mutations as part of our pipeline to systematically resolve plasmid lineages.

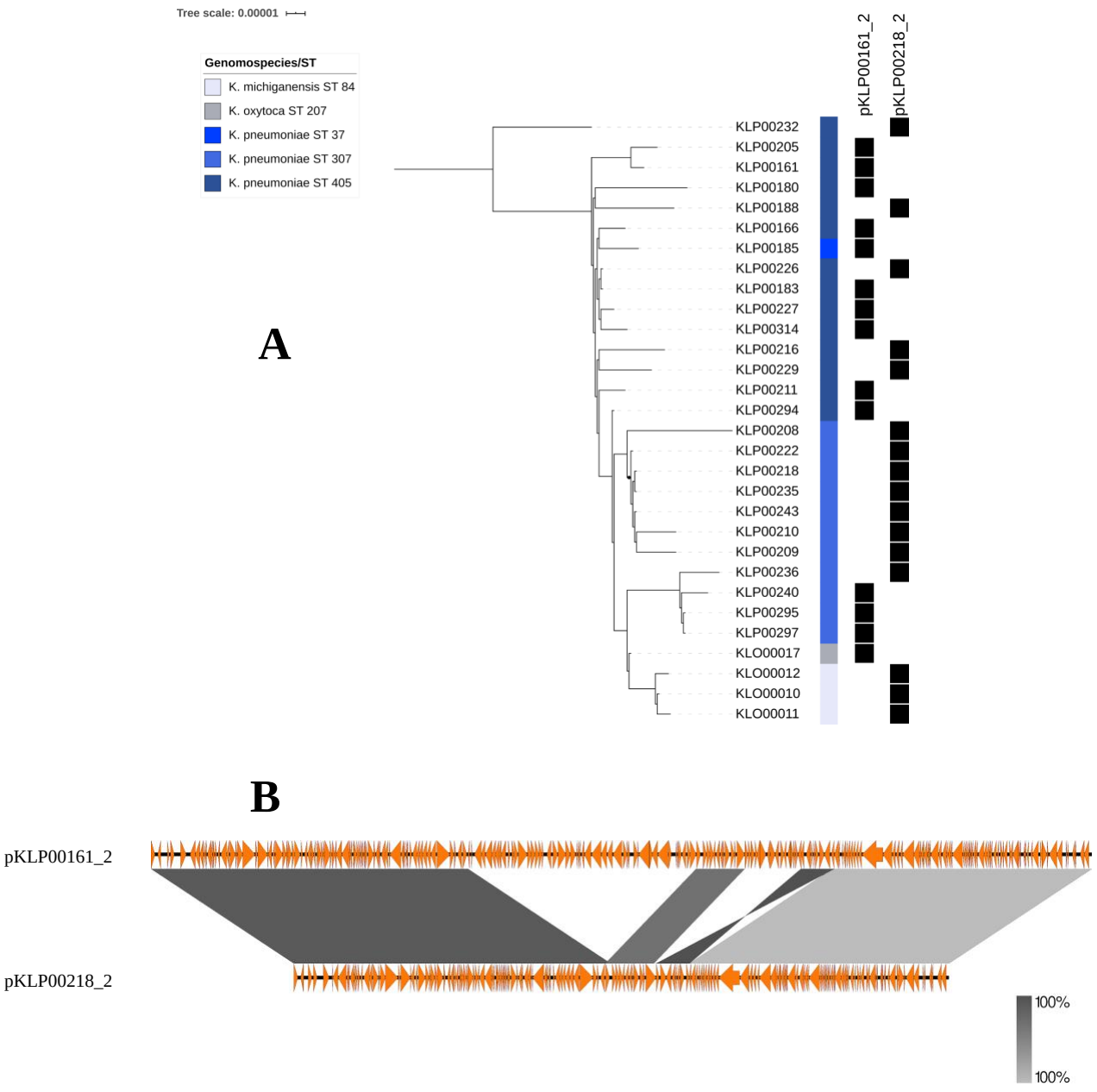

Figure 9: Recent common ancestry of two previously defined lineages identified by updated methods.

(A, top) Core sequence phylogeny of the plasmid lineages pKLP00161_2 and pKLP00218_2, as defined in Evans *et al*, 2020 (58), showing that common ancestry of plasmids was not delineated by lineage. The phylogenetic tree was constructed using RAxML v8.0.26 with 1000 bootstrap iterations and visualized using ITOL v5.7, using pKLP00218_2 as an internal reference (93, 106). (B, bottom) Alignment of plasmid sequences annotated using Prokka and visualized with EasyFig v2.2.2 (86, 95).

We performed similar comparisons between our prior and updated methods for the other seven plasmids that were references for lineages that we had previously characterized (59). We successfully reconstructed all seven lineages, each of which included strains from at least two different patients. Some of the bacterial strains that had been included within the plasmid lineages identified by our prior methods were excluded from the lineages resolved by our new strategy. This was either because the plasmids carried by those strains covered less than 95% of the sequence of their corresponding reference plasmids, or because they had sufficient coverage but had excess polymorphisms relative to the reference to exceed our cutoff.

Taken together, these results demonstrate that our updated methods substantially refined the resolution of plasmid lineages as compared to our previously published methods. This increased our confidence in our inference of potential horizontal gene transfer within the hospital setting.

5.4 Identification of potential routes of horizontal plasmid transfer by review of patient healthcare data

After resolving lineages of plasmids with high sequence similarity, our next goal was to search for epidemiologic evidence of potential routes and risk factors of transfer that aligned with our genomic data. To perform these investigations systematically and efficiently, we employed a previously published data mining algorithm that was designed to identify significantly enriched locations and procedures in the hospital system that may have served as routes of healthcare-associated bacterial transmission (57). Briefly, the algorithm reviewed patient data to identify charge codes and admission records that were present in the EHRs of two or more patients that were infected by strains that carried plasmids of the same lineage. Charge

codes served as detailed records of procedures performed on patients, and admission records identified rooms or units in which patients had stayed while admitted to the hospital system. As such, these records identified potential routes of bacterial transmission or plasmid transfer between patients. To minimize confounding by transmission of plasmid-carrying bacterial strains, potential routes identified by the algorithm were only considered for further analysis if patients implicated in each route were infected by strains of different sequence types (STs), species, or genera. We used this algorithm to investigate 41 plasmid lineages: the 40 lineages based on closed reference plasmids described previously (Table 3), plus an additional lineage that combined strains carrying the pKLP00161_2 and pKLP00218_2 plasmids (Figure 9).

Using this algorithm, we identified 133 geographic potential shared exposures (Table 4) and 81 procedural potential shared exposures (Table 5) among 28 (68.3%) of the 41 plasmid lineages we investigated. Among the geographic potential shared exposures were five sets of roommates – all ICU patients – who were infected with non-clonal bacterial strains that carried the same plasmid lineages (Table 4). Three plasmid lineages were present in isolates that infected these roommates: (1) the combined pKLP00161_2 / pKLP00218_2 lineage with IncFIB and IncFIIK incompatibility groups that carried multiple ESBLs and genes encoding resistance to aminoglycosides, sulfonamides, trimethoprim, arsenic, silver, and copper (n = 7 total patients, two groups of two roommates and one group of three); (2) the pSTEN00043_8 lineage, which carried the non-hemolytic phospholipase *plcN* but no metal or antibiotic resistance genes (n = 4 patients); and (3) the pVRE33536_2 lineage, which carried genes encoding resistance to aminoglycosides, cadmium, cobalt, and copper (n = 2 patients).

Patients infected with pathogens carrying lineage plasmids also had procedural shared exposures of 18 major types (Table 5), which each may have affected anywhere from three to 63

patients. The most abundant procedural shared exposures across plasmid lineages included insertion of a peripherally inserted central catheter (PICC) (n = 14 lineages), declotting or maintenance of faulty intravenous devices or PICCs (n = 8 lineages), insertion of vascular catheters of a type other than PICC (n = 7 lineages), venous ultrasounds (n = 7 lineages), and inpatient visits by physical therapy or occupational therapy teams (n = 7 lineages).

There was substantial epidemiologic evidence to support the hypothesis that the pKLP00161_2 and pKLP00218_2 lineages behaved as one lineage engaged in horizontal transfer between patients. Plasmids of the two separate lineages were identified in numerous shared exposures involving strains of different sequence types or genomospecies. These exposures included two sets of ICU roommates; six groups of patients with consecutive or nearly consecutive (within 90 days) admissions to the same rooms; seven groups of patients with concurrent, consecutive, or near-consecutive admissions on the same units; 15 patients who received peripherally inserted central catheters (PICCs); 20 patients who received ultrasounds; 14 patients who received respiratory bronchoscopies; and 12 patients who received respiratory nitric oxide.

The epidemiologic evidence we gathered from our investigations also bolstered the hypothesis that antibiotic resistance, metal resistance, and virulence are key drivers of horizontal transfer among nosocomial pathogens (30, 108). All of the 28 plasmid lineages that had at least one significant procedural or geographic shared exposure among strains of different STs, genomospecies, or genera carried at least one antibiotic resistance gene, metal resistance gene, or virulence factor whose characterized function was not involved in conjugative transfer. Furthermore, of the 14 plasmid lineages we resolved that did not carry any antibiotic or metal resistance genes, only six (pACIN00156_2, pCB00073_2, pKLP00155_6, pSTEN00043_8,

pVRE33085_3, and pVRE33562_6) were found to have at least one procedural or geographic shared exposure. These six lineages were the only ones of that group that encoded at least one previously characterized virulence factor.

Taken together, these results show that the horizontal transfer of plasmids within hospital settings is prevalent, affects large numbers of patients infected with nosocomial bacterial strains, and is likely driven in part by known selective advantages conferred by plasmids. Additionally, they show that investigations into horizontal plasmid transfer within hospitals can be aided by systematically combining genomic surveillance with well curated hospital epidemiologic data.

	Total	ICU	LTAC	Rehab	ED	Room	Mean No. of Pts.	Median No. of Pts.	Range
Roommates with patient(s) carrying the same plasmid lineage	5	5	0	0	0	0	2.6	2	2 to 5
Same room recently occupied by patient(s) carrying plasmid(s) of the same lineage	40	6	1	0	3	30	3.1	3	2 to 5
Same unit currently or previously housing patient(s) carrying plasmid(s) of the same lineage	88	38	10	1	6	33	8.8	8	2 to 37

Table 4: Epidemiology of plasmid lineages with statistically significantly enriched geographic potential shared exposures.

Potential exposures were included in these data only if they involved infections of two or more patients by strains of different sequence types (STs), genomospecies, or genera that were both in the same plasmid lineage.

Type of Procedure	Lineages with Significant Association	Mean No. of Patients	Median No. of Patients	Range
Insertion of peripherally inserted central catheter (PICC)	14	21.3	15.5	4 to 45
Declotting of / maintenance on IV or PICC	8	17.8	8	3 to 63
Insertion of non-PICC vascular catheter	7	17.9	3	2 to 58
Venous ultrasound	7	23.1	20	8 to 46
Occupational therapy or physical therapy inpatient visit	7	13.2	9	2 to 40
Stomal therapy, inpatient	6	20.7	7.5	5 to 51
Mechanical ventilation	5	15.1	9.5	8 to 32
Respiratory bronchoscopy	4	9	7.5	7 to 14
Blood transfusion / perfusion	4	5.5	5	4 to 8
Respiratory tracheostomy	3	32.7	31	18 to 49
Ultrasound	3	24.7	27	9 to 38
Speech therapy inpatient visit for evaluation/treatment of swallowing	3	7.3	8	5 to 9
Hemodialysis	3	5.3	6	3 to 7
Respiratory nitric oxide	2	9.5	9.5	7 to 12
Respiratory arterial puncture	2	9	9	3 to 15
Magnetic resonance imaging (MRI)	1	10	10	n/a
Thromboelastinography (TEG)	1	6	6	n/a
Insertion of urinary catheter	1	3	3	n/a

Table 5: Epidemiology of plasmid lineages with statistically significantly enriched procedural potential shared exposures.

Potential exposures were included in these data only if they involved infections of two or more patients by strains of different sequence types (STs), genomospecies, or genera that were both in the same plasmid lineage.

5.5 Excess clinical burden of secondary plasmid transmission networks on patients with nosocomial infections

While numerous previous studies have documented nosocomial plasmid transfer (26, 43, 58, 66, 73), and plasmids are known to increase the severity of bacterial infections (24), the clinical burden of nosocomial plasmid transfer on entire hospital systems has not been previously studied in a systematic manner. We therefore investigated whether the presence of plasmid lineages that we had resolved imposed an unmet clinical burden on patients infected with healthcare-associated bacterial pathogens. We reviewed de-identified healthcare data from 2,322 hospitalizations of 1,160 patients infected with pathogens collected by EDS-HAT, to identify associations between the carriage of plasmids by those pathogens and indicators of infection severity. We used admission and discharge records to calculate and analyze patients' lengths of hospital stay, as a gauge of severity of disease and need for continuous care (Figure 10) (109). We also calculated and analyzed Charlson comorbidity index (CCI) data for each patient, to assess their risk of mortality or severe disease state following infection (Figure 11) (110).

Hospitalizations involving infection by at least one strain carrying a plasmid in any lineage were significantly longer than hospitalizations involving infections only by strains that did not carry plasmids of any lineage (Hodges-Lehmann median of differences of +3 days; 95% CI +1 to +4 days; $p < 0.001$) (Figure 10a). Infection by multiple strains carrying at least one plasmid of any lineage was also linked to longer hospitalization; (+18 days; 95% CI +6 to +28 days; $p < 0.01$) (Figure 10b). Infections involving plasmids in any lineage that encoded genes for resistance to any antibiotic (+2 days; 95% CI +1 to +4 days; $p < 0.01$) or resistance to any metal (+3 days; 95% CI +1 to +5 days; $p < 0.01$) were also associated with significantly longer hospitalization than infections by strains not carrying any antibiotic or metal resistance plasmids

in resolved lineages (Figure 10c-d). Associations between length of hospitalization and presence of a plasmid in at least one nosocomial isolate were significant within respiratory tract infections (n = 710 hospitalizations; +10 days; 95% CI +3 to +15 days; $p < 0.01$) and tissue or wound infections (n = 552 hospitalizations; +5 days; 95% CI +2 to +8 days; $p < 0.001$) (Figure 10e-f). The association was trending yet not significant among urinary tract infections (n = 530 hospitalizations; +2 days; 95% CI 0 to +4 days; $p = 0.0671$) (Figure 10g). The association was trending yet not significant among bloodstream infections (n = 228 hospitalizations) prior to the removal of outliers ($p = 0.0940$), but significant following the removal of outliers (+5 days; 95% CI 0 to +13 days; $p = 0.0403$) (Figure 10h). Among hospitalizations involving infections by *Enterococcus spp.* (n = 228) or by members of the order *Enterobacterales* (n = 527) – the taxonomic groups carrying the greatest number of plasmids of any lineage – there were no significant links to length of hospitalization (0 days; 95% -4 to +5 days; $p = 0.8403$ and +2 days; 95% CI -1 to +5 days; $p = 0.2457$, respectively) (Figure 10i-j). Taken together, these results show that the presence of plasmids engaged in horizontal transfer in the hospital system we studied was linked to longer hospitalizations, and that this association was mediated by plasmid gene cargo and type of infection rather than by the taxonomic groups of pathogens exchanging plasmids.

Associations between the involvement of plasmids in nosocomial infections and CCIs of infected patients largely paralleled those found in length of hospitalization (Figure 11). Patients infected by bacterial strains carrying plasmids of any lineage (Hodges-Lehmann median of differences of +1; 95% CI +1 to +2; $p < 0.0001$), as well as strains carrying lineage plasmids encoding antibiotic resistance (+1; 95% CI +1 to +2; $p < 0.0001$) or metal resistance (+1; 95% CI +1 to +2; $p < 0.0001$), tended to have greater CCIs than those whose HAIs did not involve those

groups of plasmids (Figure 11a,c-d). Significantly greater CCIs were also associated with the carriage of lineage plasmids among patients with tissue or wound infections (+2; 95% CI +1 to +2; $p < 0.0001$) and bloodstream infections (+2; 95% CI 0 to +3; $p = 0.0057$) (Figure 11e,g). There was also no association between greater CCI and carriage of lineage plasmids among patients infected by *Enterococcus spp.* (0; 95% CI -1 to +1; $p = 0.777$) (Figure 11j). Contrary to analyses of length of hospitalization, greater CCI and carriage of plasmids in lineages were significantly associated among patients with urinary tract infections (+1; 95% CI +1 to +2; $p = 0.0002$) (Figure 11f) and among patients with infections by pathogens of the order *Enterobacterales* (+1; 95% CI 0 to +2; $p = 0.012$) (Figure 11i). Also contrary to analyses of length of hospitalization, there was no association between greater CCI and carriage of lineage plasmids among patients with respiratory HAIs (0; 95% CI -1 to +1; $p = 0.721$) (Figure 11e) or with co-infections by multiple bacterial strains carrying lineage plasmids (+1; 95% CI -1 to +3; $p = 0.195$) (Figure 11b), indicating that the link between plasmids and longer hospital stay in those groups of patients occurred without any overall differences in degree of comorbidity prior to hospitalization.

In summary, these results show clear links between nosocomial bacterial pathogens' carriage of plasmid lineages likely engaged in horizontal transfer in hospitals and the clinical burden imposed on patients by those pathogens. Many of these associations may have been moderated by patient comorbidity, but some evidence points to greater clinical burdens imposed by plasmids without moderation by comorbidity.

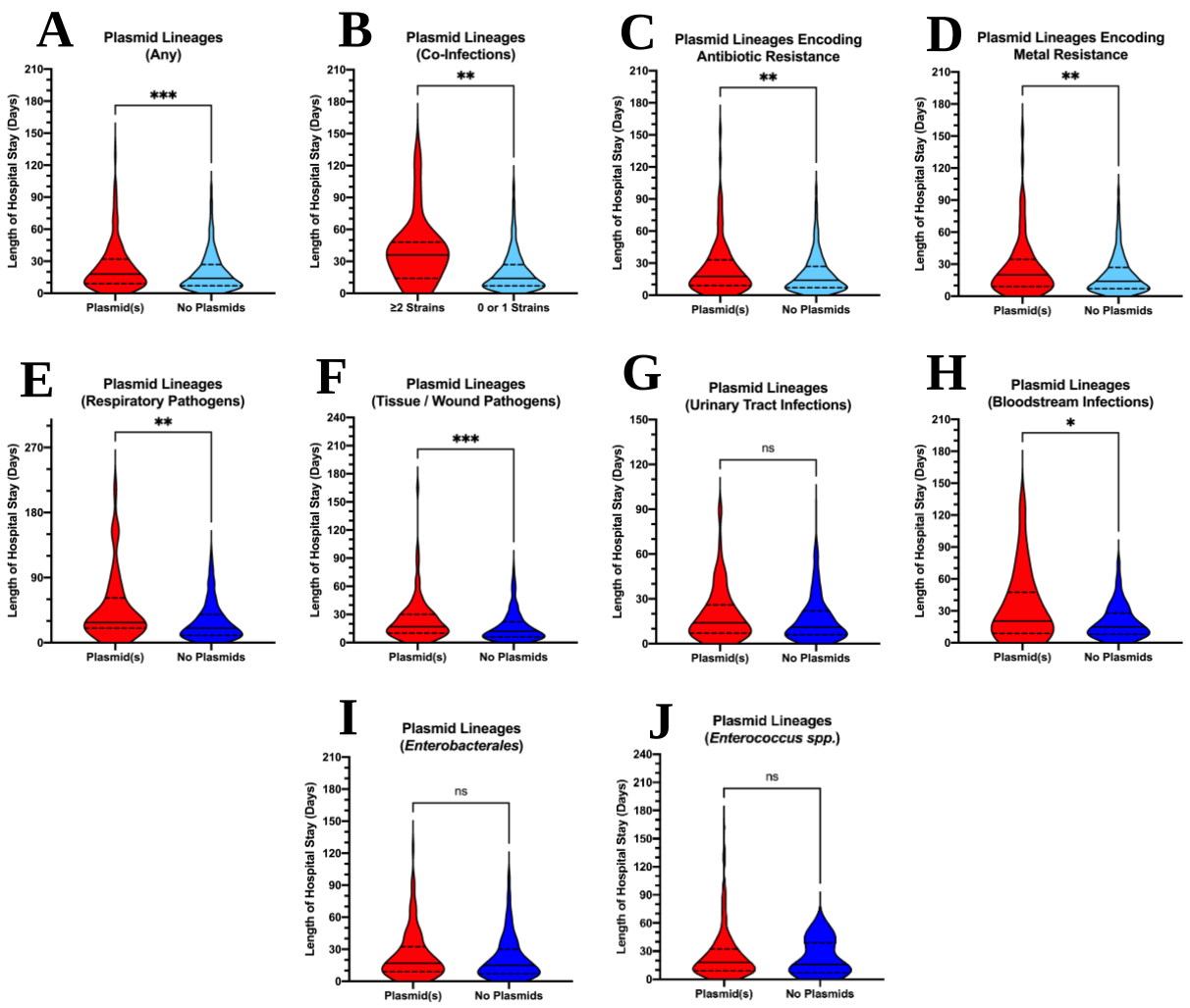

Figure 10: Carriage of plasmids engaged in nosocomial horizontal transfer is associated with increased length of hospitalization.

Data are shown for hospitalizations involving the following nosocomial isolates: (A) any strain carrying a plasmid of any lineage; (B) co-infection by multiple strains carrying plasmids of any lineage; any strain with a lineage plasmid encoding: (C) at least one known antibiotic resistance gene, or (D) at least one known gene; (E) respiratory infections, (F) tissue or wound infections, (G) urinary tract infections, metal resistance and (H) bloodstream infections involving a strain carrying a plasmid of any lineage; and infections by (I) *Enterobacterales* and (J) *Enterococcus spp.* strains carrying a plasmid of any lineage. Outliers from each compared group were removed prior to plotting and statistical comparison. Significance was assessed by p-values, which are shown as not significant (ns), $p < 0.05$ (*), $p < 0.01$ (**), or $p < 0.001$ (***).

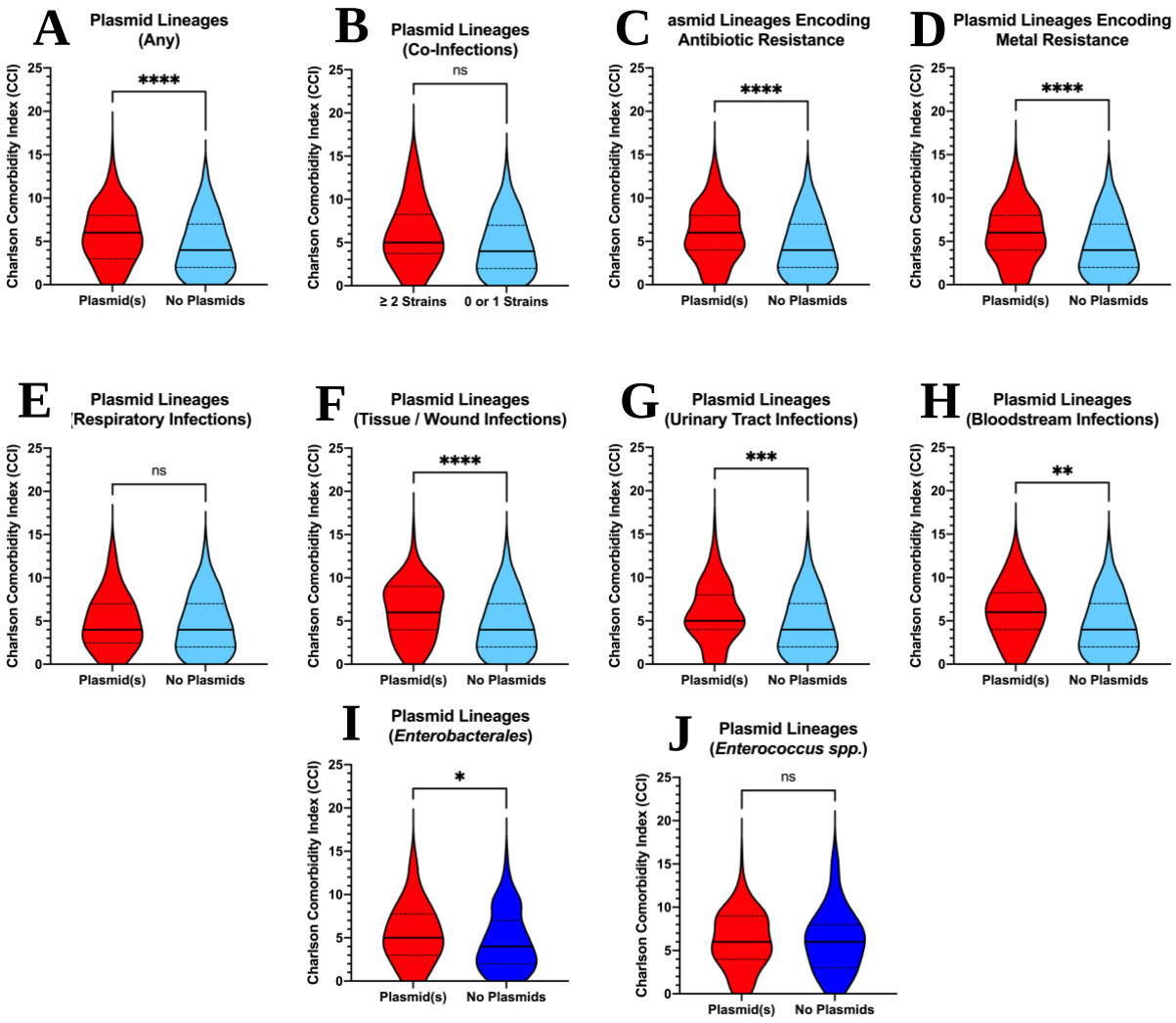

Figure 11: Carriage of plasmids engaged in nosocomial horizontal transfer is associated with greater Charlson comorbidity index.

Data are shown for Charlson comorbidity indices (CCIs) involving the following nosocomial isolates: (A) any strain carrying a plasmid of any lineage; (B) co-infection by multiple strains carrying plasmids of any lineage; any strain with a lineage plasmid encoding: (C) at least one known antibiotic resistance gene, or (D) at least one known metal resistance gene; (E) respiratory infections, (F) tissue or wound infections, (G) urinary tract infections, and (H) bloodstream infections involving a strain carrying a plasmid of any lineage; and infections by (I) *Enterobacterales* and (J) *Enterococcus spp.* strains carrying a plasmid of any lineage. Outliers from each compared group were removed prior to plotting and statistical comparison. Significance was assessed by p-values, which are shown as not significant (ns), $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), or $p < 0.0001$ (****).

6.0 Discussion

In this study, we thoroughly characterized the genomics, epidemiology, and clinical burden of the horizontal transfer of plasmids among nosocomial bacterial pathogens within a tertiary hospital setting. We used bacterial whole-genome sequences and closed plasmid sequences from known bacterial outbreaks and likely events of healthcare-associated horizontal plasmid transfer to identify uniform thresholds of sequence similarity to infer the occurrence of transfer. We applied these thresholds to systematically resolve plasmids likely engaged in horizontal transfer from a large genomic dataset of clinical bacterial isolates from a single hospital system. By repurposing an algorithm designed to identify routes of bacterial transmission from hospital admission and patient EHR data, we identified dozens of potential routes of horizontal transfer of plasmids among nosocomial strains of different sequence types (STs), genomospecies, and/or genera. We also found evidence that the carriage of plasmids likely engaged in horizontal transfer by bacterial isolates was associated with greater clinical burden among patients with healthcare-associated infections (HAIs).

Our study contributes to the fields of genomic epidemiology and hospital infection prevention in several ways. While the horizontal transfer of plasmids within healthcare settings has previously been described (26, 42, 43, 58, 73), the field has not previously proposed or established uniform thresholds of sequence similarity by which nosocomial plasmid transfer can be inferred. This study improves on prior studies (43, 58) by identifying and systematically applying thresholds of sequence content preservation and nucleotide identity that were drawn from evidence of horizontal or vertical inheritance of plasmids. Additionally, by using a large dataset of bacterial genomes collected for active surveillance of HAIs, we could draw conclusions about horizontal plasmid transfer regardless of the suspected occurrence of

outbreaks of bacterial pathogens. While many prior studies have described events of horizontal transfer within smaller groups of bacterial pathogens that are often implicated in outbreaks (26, 42, 43, 58, 73), our methods enabled us to study plasmid transfer at a much larger scope and more systematically than previous studies. Furthermore, while these prior studies with case-series designs have described individual cases of likely plasmid transfer, we systematically reviewed healthcare data associated with each genome in our dataset by using a data-mining algorithm. Because of this, we were able to quantify potential events and routes of horizontal plasmid transfer across an entire hospital system, providing more robust evidence about potential risk factors and interventions to prevent and control transfer.

To our knowledge, no previously published study has investigated in as much detail the clinical burden that horizontal plasmid transfer imposes on hospitalized patients with nosocomial infections. While the contribution of plasmids to the severity of infections has previously been described (6, 69, 108), and horizontal transfer is well known to occur in hospitals (43, 59, 69, 70, 74), our assessment of the scope, routes, and potential exacerbation of disease specifically linked to transfer of plasmids among infected patients contributes substantially to the field. As such, this study serves as a direct response to calls from the field to study this phenomenon in greater detail (30, 46, 59, 108).

Among our findings, two observations of particular interest were the association between carriage of plasmids in resolved lineages and increased length of hospitalization among 1) patients with respiratory tract infections, and 2) patients who were co-infected with multiple bacterial strains that carried plasmids in lineages. These groups were of interest because of the apparent lack of difference in overall comorbidity among patients that were included or excluded from these groups, which indicated the potential existence of a clinical burden of horizontally

transferring plasmids that was not linked to patient comorbidity. Importantly, our statistical approach prevented us from establishing or refuting causality among any of the significant links we identified. Additionally, the genomic dataset used for this project was generated for studies of bacterial transmission rather than plasmid transfer, so key signs or factors of transfer might not have been detected as a result. We also did not explore other markers of disease severity, namely mortality of infected patients or measures of phenotypic drug resistance. Nonetheless, the number of significant associations related to carriage of horizontally transferred plasmids among bacterial HAIs calls for further study, both of the epidemiologic factors of transfer and of interventions designed to reduce the likely clinical burden that plasmid transfer imposes on infected patients.

While we reported numerous potential shared exposures among patients that were likely affected by horizontal plasmid transfer, identifying exact routes by which transfer occurs in hospital settings remains challenging. Plasmids can be exchanged between bacterial strains by transformation and vesiduction, two mechanisms that do not require both the donor and recipient bacterial cells to be present in the same location at the same time (11, 30, 78). In addition to routes involving direct or indirect contact by patients, healthcare personnel, or other objects (48, 67, 72), it has been proposed that plasmids can also disperse through hospitals through airborne particulate matter (78). Additionally, bacterial strains can carry and exchange multiple plasmids that may have similar gene cargo, which likely led to some overlap or double-counting of potential exposures of different plasmids shared among the same isolates in our study. Furthermore, plasmids can remain nearly identical while being exchanged among multiple clinical bacterial isolates (58), indicating that plasmid transfer between intermediate bacterial strains can occur with minimal effect on the sequence of the plasmid. These potential

intermediate carriers of plasmids are difficult to identify without abundant sequencing of environmental isolates, which was not performed as part of this study. Studies of nosocomial plasmids that did include sequencing of environmental samples have identified potential intermediates of horizontal transfer (71), demonstrating the utility of these data. However, our ability to combine robustly characterized genomic data with detailed patient care records at a hospital-wide scale has provided valuable milestones to more precisely navigate this phenomenon.

The selective advantages that plasmids often confer to pathogenic bacteria – namely antibiotic resistance, environmental persistence, and virulence – are well documented (108, 111). The plasmid lineages that we resolved – particularly the lineages for which we could identify potential shared exposures among affected patients – carried genes for many of these functions. Intriguingly, of the fourteen lineages that did not encode any known antibiotic resistance or metal resistance genes, the six lineages for which we could identify potential shared exposures were the only ones that carried virulence factors with functions not directly involved in the horizontal transfer of plasmids. However, not all of the plasmid lineages that we resolved carried genes known to confer any of these properties to their bacterial hosts. Most plasmid lineages also carried genes with a variety of metabolic functions, as well as numerous hypothetical open reading frames. While characterizing plasmid sequences can vary depending on the bioinformatic tools used for gene annotation, the substantial proportion of plasmids that did not encode functions with known direct clinical effects indicates that antibiotic resistance and virulence are not the only functions that necessitate their preservation by bacterial strains. This indication of potentially unknown drivers of horizontal plasmid transfer concurs with previously published results that the gene content of plasmids – and therefore the selective

advantages that they confer to bacterial pathogens – remains poorly understood and should therefore be studied in greater detail (112).

Despite our attempts to perform this study with systematic methods in an unbiased manner, there were several limitations to our study design. First, the criteria for the collection of nosocomial bacterial strains excluded some well-known genera, e.g. *Enterobacter spp.*. Additionally, the collection of strains from some species and genera was limited by expression of antibiotic resistance, i.e. vancomycin-resistant *Enterococcus spp.* (VRE), extended-spectrum beta-lactamase (ESBL)-producing *Escherichia coli* and *Klebsiella spp.*, and methicillin-resistant *Staphylococcus aureus*. This introduced sampling bias among both our strains and the reference plasmids that were included in this study. Second, the diversity of reference plasmids used to resolve lineages was limited and likely biased, because the long-read sequencing and hybrid assembly that generated those reference plasmids was performed for previous studies that did not encompass the entire genomic dataset (50, 51, 59). As a result, there were likely a number of lineages that we failed to resolve because we lacked reference plasmids that were key to our methods to detect those lineages. Future studies in this field could correct this limitation by performing more active and systematic long-read sequencing and hybrid assembly of representative distributions of all nosocomial bacterial strains from a single hospital. Third, our analyses of length of hospital stay and patient comorbidity could have been confounded by other factors linked to patients' infections rather than the infectious isolates' plasmids – e.g. phenotypic antibiotic resistance, mutations or virulence factors present elsewhere in strains' genomes, or coinfection by other bacterial strains not collected through EDS-HAT. Lastly, significantly enriched procedure charge codes and admission records were identified using an algorithm that used parameters developed to screen for evidence of bacterial transmission, rather

than horizontal transfer; this retroactive approach was also not a full substitute for comprehensive outbreak investigations to confirm or refute potential routes of plasmid transfer.

In conclusion, we have shown that the horizontal transfer of plasmids among bacterial pathogens that cause healthcare-associated infections is highly prevalent and may impose an additional unmet clinical burden upon patients suffering from nosocomial infections. We demonstrated the utility of developing and applying systematic methods to resolve plasmids from genomic data, as well as to identify potential routes of, and risk factors for, horizontal transfer in hospital settings. The fields of genomic epidemiology and hospital infection prevention can build upon our findings to further investigate horizontal transfer in clinical settings, to hopefully develop effective interventions that control the dissemination of antibiotic resistance, virulence, and other selective advantages among nosocomial bacterial pathogens.

7.0 Future Directions

The results of this study can serve as the basis for several avenues of future research, both in microbial genomics and in hospital epidemiology. Further genomic analyses of bacterial whole-genome sequences could focus on the similarity of plasmid sequences and gene content, to better refine thresholds based on which horizontal transfer between bacterial strains can be inferred. Similar approaches or philosophies could be applied to more accurately infer the horizontal transfer of other classes of MGEs (e.g. transposons and prophages). Broader use of long-read sequencing methods – namely PacBio and Oxford Nanopore technology – would resolve more mobile genetic elements within the context in which they exist and exchange in bacterial genomes. Additional characterizations of the functions of genes encoded on plasmids would provide greater insight into the selective advantages that they confer to the bacterial strains that carry them.

To further describe and analyze the horizontal transfer of plasmids in hospital settings, additional genomic and epidemiologic methods could be applied in future studies. First, active surveillance of nosocomial pathogens by culturing clinical isolates could be refined to minimize sampling bias, namely by not excluding strains based on genera or phenotypic antibiotic resistance profiles. This surveillance could be bolstered by collection and analysis of bacterial strains that colonize, rather than infect, hospitalized patients, as these isolates may serve as reservoirs for horizontal transfer (113). More widespread, unbiased use of long-read sequencing of clinical isolates would likely identify more lineages of MGEs that are exchanged in clinical settings. Supplemental culture-based genome sequencing or non-culture-based metagenomic sequencing of environmental samples – either as a proactive measure to identify plasmids or as a reactive measure to confirm or refute potential routes of HGT – would provide great

supplemental value to the study of nosocomial HGT. Additionally, monitoring for enriched recurrence of phenotypic markers of MGEs (namely antibiotic resistance), in addition to recurrence of species and STs, among bacterial isolates from infected patients, may help detect outbreaks of nosocomial MGEs exchanged by horizontal transfer. Importantly, incorporating these techniques into real-time investigations of secondary plasmid transmission would yield key insights into horizontal transfer that this study was not able to capture.

To better understand the clinical burden imposed by horizontal plasmid transfer in hospitals, additional patient demographics such as age, sex, primary reasons for hospitalization, and diagnoses or known comorbidities could be examined in further detail. Analyses of more specific patient comorbidities – e.g. history of bacterial infection, immunodeficiency, etc. – may identify more specific mediators and moderators of the relationship between horizontally transferred plasmids and associated clinical burden. Additional measures of disease severity – including mortality, escalation of care, rates of readmission following discharge, and cost of care – would also provide value to these studies.

Bibliography

1. Magill SS, O'Leary E, Janelle SJ, Thompson DL, Dumyati G, Nadle J, Wilson LE, Kainer MA, Lynfield R, Greissman S, Ray SM, Beldavs Z, Gross C, Bamberg W, Sievers M, Concannon C, Buhr N, Warnke L, Maloney M, Ocampo V, Brooks J, Oyewumi T, Sharmin S, Richards K, Rainbow J, Samper M, Hancock EB, Leaptrot D, Scalise E, Badrun F, Phelps R, Edwards JR. 2018. Changes in Prevalence of Health Care–Associated Infections in U.S. Hospitals. N Engl J Med 379:1732–1744.

2. Scott RD. 2009. The Direct medical costs of healthcare-associated infections in U.S. hospitals and the benefits of prevention.

3. Haque M, Sartelli M, McKimm J, Abu Bakar MB. 2018. Health care-associated infections – an overview. Infect Drug Resist Volume 11:2321–2333.

4. Centers for Disease Control and Prevention (CDC). 2019. 2019 National and State Healthcare-Associated Infections (HAI) Progress Report. Atlanta, GA.

5. Stevens MP, Doll M, Pryor R, Godbout E, Cooper K, Bearman G. 2020. Impact of COVID-19 on traditional healthcare-associated infection prevention efforts. Infect Control Hosp Epidemiol 41:946–947.

6. CDC. 2019. Antibiotic Resistance Threats in the United States, 2019.

7. Kadri SS. 2020. Key Takeaways From the U.S. CDC's 2019 Antibiotic Resistance Threats Report for Frontline Providers. Crit Care Med 48:939–945.

8. Zilberberg MD, Shorr AF, Micek ST, Vazquez-Guillamet C, Kollef MH. 2014. Multi-drug resistance, inappropriate initial antibiotic therapy and mortality in Gram-negative severe sepsis and septic shock: a retrospective cohort study. Crit Care Lond Engl 18:596.

9. Struelens MJ. 1998. The epidemiology of antimicrobial resistance in hospital acquired infections: problems and possible solutions. BMJ 317:652–654.

10. Kadri SS, Adjemian J, Lai YL, Spaulding AB, Ricotta E, Prevots DR, Palmore TN, Rhee C, Klompas M, Dekker JP, Powers JH, Suffredini AF, Hooper DC, Fridkin S, Danner RL, National Institutes of Health Antimicrobial Resistance Outcomes Research Initiative (NIH–ARORI). 2018. 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 https://doi.org/10.1093/cid/ciy378.

11. Juhas M. 2015. Horizontal gene transfer in human pathogens. Crit Rev Microbiol 41:101–108.

12. David S, Cohen V, Reuter S, Sheppard AE, Giani T, Parkhill J, European Survey of Carbapenemase-Producing Enterobacteriaceae (EuSCAPE) Working Group, ESCMID

Study Group for Epidemiological Markers (ESGEM), Rossolini GM, Feil EJ, Grundmann H, Aanensen DM. 2020. Integrated chromosomal and plasmid sequence analyses reveal diverse modes of carbapenemase gene spread among Klebsiella pneumoniae. Proc Natl Acad Sci U S A 117:25043–25054.

13. Sheppard AE, Stoesser N, Wilson DJ, Sebra R, Kasarskis A, Anson LW, Giess A, Pankhurst LJ, Vaughan A, Grim CJ, Cox HL, Yeh AJ, Modernising Medical Microbiology (MMM) Informatics Group, Sifri CD, Walker AS, Peto TE, Crook DW, Mathers AJ. 2016. Nested Russian Doll-Like Genetic Mobility Drives Rapid Dissemination of the Carbapenem Resistance Gene blaKPC. Antimicrob Agents Chemother 60:3767–3778.

14. Liu Y, Tong Z, Shi J, Jia Y, Yang K, Wang Z. 2020. Correlation between Exogenous Compounds and the Horizontal Transfer of Plasmid-Borne Antibiotic Resistance Genes. Microorganisms 8:1211.

15. Andam CP, Fournier GP, Gogarten JP. 2011. Multilevel populations and the evolution of antibiotic resistance through horizontal gene transfer. FEMS Microbiol Rev 35:756–767.

16. Harrison E, Brockhurst MA. 2012. Plasmid-mediated horizontal gene transfer is a coevolutionary process. Trends Microbiol 20:262–267.

17. Hernando-Amado S, Coque TM, Baquero F, Martínez JL. 2019. Defining and combating antibiotic resistance from One Health and Global Health perspectives. Nat Microbiol 4:1432–1442.

18. Jutkina J, Marathe NP, Flach C-F, Larsson DGJ. 2018. Antibiotics and common antibacterial biocides stimulate horizontal transfer of resistance at low concentrations. Sci Total Environ 616–617:172–178.

19. Smalla K, Jechalke S, Top EM. 2015. Plasmid Detection, Characterization, and Ecology, p. 445–458. *In* Tolmasky, ME, Alonso, JC (eds.), Plasmids. ASM Press, Washington, DC, USA.

20. Stokes HW, Gillings MR. 2011. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. FEMS Microbiol Rev 35:790–819.

21. Deng Y, Bao X, Ji L, Chen L, Liu J, Miao J, Chen D, Bian H, Li Y, Yu G. 2015. Resistance integrons: class 1, 2 and 3 integrons. Ann Clin Microbiol Antimicrob 14:45.

22. Cuzon G, Naas T, Nordmann P. 2011. Functional Characterization of Tn *4401* , a Tn *3* -Based Transposon Involved in *bla* KPC Gene Mobilization. Antimicrob Agents Chemother 55:5370–5373.

23. Karim A, Poirel L, Nagarajan S, Nordmann P. 2001. Plasmid-mediated extended-spectrum beta-lactamase (CTX-M-3 like) from India and gene association with insertion sequence

ISEcp1. FEMS Microbiol Lett 201:237–241.

24. 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.

25. Jiang Y, Zhang Y, Lu J, Wang Q, Cui Y, Wang Y, Quan J, Zhao D, Du X, Liu H, Li X, Wu X, Hua X, Feng Y, Yu Y. 2020. Clinical relevance and plasmid dynamics of mcr-1-positive Escherichia coli in China: a multicentre case-control and molecular epidemiological study. Lancet Microbe 1:e24–e33.

26. Dolejska M, Papagiannitsis CC. 2018. Plasmid-mediated resistance is going wild. Plasmid 99:99–111.

27. Bosch T, Lutgens SPM, Hermans MHA, Wever PC, Schneeberger PM, Renders NHM, Leenders ACAP, Kluytmans JAJW, Schoffelen A, Notermans D, Witteveen S, Bathoorn E, Schouls LM. 2017. Outbreak of NDM-1-Producing Klebsiella pneumoniae in a Dutch Hospital, with Interspecies Transfer of the Resistance Plasmid and Unexpected Occurrence in Unrelated Health Care Centers. J Clin Microbiol 55:2380–2390.

28. Fu P, Tang Y, Li G, Yu L, Wang Y, Jiang X. 2019. Pandemic spread of bla among Klebsiella pneumoniae ST11 in China is associated with horizontal transfer mediated by IncFII-like plasmids. Int J Antimicrob Agents 54:117–124.

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

30. Lerminiaux NA, Cameron ADS. 2019. Horizontal transfer of antibiotic resistance genes in clinical environments. Can J Microbiol 65:34–44.

31. Ladner JT, Grubaugh ND, Pybus OG, Andersen KG. 2019. Precision epidemiology for infectious disease control. Nat Med 25:206–211.

32. Tang P, Croxen MA, Hasan MR, Hsiao WWL, Hoang LM. 2017. Infection control in the new age of genomic epidemiology. Am J Infect Control 45:170–179.

33. Ward DV, Hoss AG, Kolde R, van Aggelen HC, Loving J, Smith SA, Mack DA, Kathirvel R, Halperin JA, Buell DJ, Wong BE, Ashworth JL, Fortunato-Habib MM, Xu L, Barton BA, Lazar P, Carmona JJ, Mathew J, Salgo IS, Gross BD, Ellison RT. 2019. Integration of genomic and clinical data augments surveillance of healthcare-acquired infections. Infect Control Hosp Epidemiol 40:649–655.

34. Neoh H, Tan X-E, Sapri HF, Tan TL. 2019. Pulsed-field gel electrophoresis (PFGE): A

review of the “gold standard” for bacteria typing and current alternatives. Infect Genet Evol 74:103935.

35. Baker KS. 2020. Microbe hunting in the modern era: reflecting on a decade of microbial genomic epidemiology. Curr Biol CB 30:R1124–R1130.

36. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95:3140–3145.

37. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res 3:124.

38. Seemann T. MLST. Github.

39. Yohe S, Thyagarajan B. 2017. Review of Clinical Next-Generation Sequencing. Arch Pathol Lab Med 141:1544–1557.

40. Quainoo S, Coolen JPM, van Hijum SAFT, Huynen MA, Melchers WJG, van Schaik W, Wertheim HFL. 2017. Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. Clin Microbiol Rev 30:1015–1063.

41. Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics 13:278–289.

42. Jain M, Olsen HE, Paten B, Akeson M. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol 17:239.

43. Peter S, Bosio M, Gross C, Bezdan D, Gutierrez J, Oberhettinger P, Liese J, Vogel W, Dörfel D, Berger L, Marschal M, Willmann M, Gut I, Gut M, Autenrieth I, Ossowski S. 2020. Tracking of Antibiotic Resistance Transfer and Rapid Plasmid Evolution in a Hospital Setting by Nanopore Sequencing. mSphere 5:e00525-20, /msphere/5/4/mSphere525-20.atom.

44. Prussing C, Snavely EA, Singh N, Lapierre P, Lasek-Nesselquist E, Mitchell K, Haas W, Owsiak R, Nazarian E, Musser KA. 2020. Nanopore MinION Sequencing Reveals Possible Transfer of blaKPC–2 Plasmid Across Bacterial Species in Two Healthcare Facilities. Front Microbiol 11:2007.

45. 2015. An introduction to next-generation sequencing technology. Illumina.

46. Feng Y, Zou S, Chen H, Yu Y, Ruan Z. 2020. BacWGSTdb 2.0: a one-stop repository for bacterial whole-genome sequence typing and source tracking. Nucleic Acids Res https://doi.org/10.1093/nar/gkaa821.

47. Martin J, Phan HTT, Findlay J, Stoesser N, Pankhurst L, Navickaite I, De Maio N, Eyre DW, Toogood G, Orsi NM, Kirby A, Young N, Turton JF, Hill RLR, Hopkins KL, Woodford N, Peto TEA, Walker AS, Crook DW, Wilcox MH. 2017. Covert dissemination of carbapenemase-producing Klebsiella pneumoniae (KPC) in a successfully controlled outbreak: long- and short-read whole-genome sequencing demonstrate multiple genetic modes of transmission. J Antimicrob Chemother 72:3025–3034.

48. Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, Ip CLC, Golubchik T, Batty EM, Finney JM, Wyllie DH, Didelot X, Piazza P, Bowden R, Dingle KE, Harding RM, Crook DW, Wilcox MH, Peto TEA, Walker AS. 2013. Diverse sources of C. difficile infection identified on whole-genome sequencing. N Engl J Med 369:1195–1205.

49. Stadler T, Meinel D, Aguilar-Bultet L, Huisman JS, Schindler R, Egli A, Seth-Smith HMB, Eichenberger L, Brodmann P, Hübner P, Bagutti C, Tschudin-Sutter S. 2018. Transmission of ESBL-producing Enterobacteriaceae and their mobile genetic elements-identification of sources by whole genome sequencing: study protocol for an observational study in Switzerland. BMJ Open 8:e021823.

50. Sundermann AJ, Babiker A, Marsh JW, Shutt KA, Mustapha MM, Pasculle AW, Ezeonwuka C, Saul MI, Pacey MP, Van Tyne D, Ayres AM, Cooper VS, Snyder GM, Harrison LH. 2019. Outbreak of Vancomycin-resistant Enterococcus faecium in Interventional Radiology: Detection Through Whole Genome Sequencing-Based Surveillance. Clin Infect Dis Off Publ Infect Dis Soc Am https://doi.org/10.1093/cid/ciz666.

51. Marsh JW, Mustapha MM, Griffith MP, Evans DR, Ezeonwuka C, Pasculle AW, Shutt KA, Sundermann A, Ayres AM, Shields RK, Babiker A, Cooper VS, Van Tyne D, Harrison LH. 2019. Evolution of Outbreak-Causing Carbapenem-Resistant Klebsiella pneumoniae ST258 at a Tertiary Care Hospital over 8 Years. mBio 10.

52. Chng KR, Li C, Bertrand D, Ng AHQ, Kwah JS, Low HM, Tong C, Natrajan M, Zhang MH, Xu L, Ko KKK, Ho EXP, Av-Shalom TV, Teo JWP, Khor CC, MetaSUB Consortium, Chen SL, Mason CE, Ng OT, Marimuthu K, Ang B, Nagarajan N. 2020. Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment. Nat Med 26:941–951.

53. Singer AC, Kirchhelle C, Roberts AP. 2019. Reinventing the antimicrobial pipeline in response to the global crisis of antimicrobial-resistant infections. F1000Research 8:238.

54. Sundermann AJ, Chen J, Miller JK, Saul MI, Shutt KA, Griffith MP, Mustapha MM, Ezeonwuka C, Waggle K, Srinivasa V, Kumar P, Pasculle AW, Ayres AM, Snyder GM, Cooper VS, Van Tyne D, Marsh JW, Dubrawski AW, Harrison LH. 2020. Outbreak of *Pseudomonas aeruginosa* Infections from a Contaminated Gastroscope Detected by Whole Genome Sequencing Surveillance. Clin Infect Dis ciaa1887.

55. Giardine B. 2005. Galaxy: A platform for interactive large-scale genome analysis. Genome Res 15:1451–1455.

56. Wattam AR, Brettin T, Davis JJ, Gerdes S, Kenyon R, Machi D, Mao C, Olson R, Overbeek R, Pusch GD, Shukla MP, Stevens R, Vonstein V, Warren A, Xia F, Yoo H. 2018. Assembly, Annotation, and Comparative Genomics in PATRIC, the All Bacterial Bioinformatics Resource Center, p. 79–101. *In* Setubal, JC, Stoye, J, Stadler, PF (eds.), Comparative Genomics. Springer New York, New York, NY.

57. Sundermann AJ, Miller JK, Marsh JW, Saul MI, Shutt KA, Pacey M, Mustapha MM, Ayres A, Pasculle AW, Chen J, Snyder GM, Dubrawski AW, Harrison LH. 2019. Automated data mining of the electronic health record for investigation of healthcare-associated outbreaks. Infect Control Hosp Epidemiol 40:314–319.

58. Ellington MJ, Heinz E, Wailan AM, Dorman MJ, de Goffau M, Cain AK, Henson SP, Gleadall N, Boinett CJ, Dougan G, Brown NM, Woodford N, Parkhill J, Török ME, Peacock SJ, Thomson NR. 2019. Contrasting patterns of longitudinal population dynamics and antimicrobial resistance mechanisms in two priority bacterial pathogens over 7 years in a single center. Genome Biol 20:184.

59. Evans DR, Griffith MP, Sundermann AJ, Shutt KA, Saul MI, Mustapha MM, Marsh JW, Cooper VS, Harrison LH, Van Tyne D. 2020. Systematic detection of horizontal gene transfer across genera among multidrug-resistant bacteria in a single hospital. eLife 9.

60. Gao W, Howden BP, Stinear TP. 2018. Evolution of virulence in *Enterococcus faecium*, a hospital-adapted opportunistic pathogen. Curr Opin Microbiol 41:76–82.

61. Chilambi GS, Nordstrom HR, Evans DR, Ferrolino JA, Hayden RT, Marón GM, Vo AN, Gilmore MS, Wolf J, Rosch JW, Van Tyne D. 2020. Evolution of vancomycin-resistant *Enterococcus faecium* during colonization and infection in immunocompromised pediatric patients. Proc Natl Acad Sci 117:11703–11714.

62. Chandrangsu P, Rensing C, Helmann JD. 2017. Metal homeostasis and resistance in bacteria. Nat Rev Microbiol 15:338–350.

63. McDonnell G, Russell AD. 1999. Antiseptics and disinfectants: activity, action, and resistance. Clin Microbiol Rev 12:147–179.

64. Cerqueira GC, Earl AM, Ernst CM, Grad YH, Dekker JP, Feldgarden M, Chapman SB, Reis-Cunha JL, Shea TP, Young S, Zeng Q, Delaney ML, Kim D, Peterson EM, O'Brien TF, Ferraro MJ, Hooper DC, Huang SS, Kirby JE, Onderdonk AB, Birren BW, Hung DT, Cosimi LA, Wortman JR, Murphy CI, Hanage WP. 2017. Multi-institute analysis of carbapenem resistance reveals remarkable diversity, unexplained mechanisms, and limited clonal outbreaks. Proc Natl Acad Sci U S A 114:1135–1140.

65. George S, Pankhurst L, Hubbard A, Votintseva A, Stoesser N, Sheppard AE, Mathers A, Norris R, Navickaite I, Eaton C, Iqbal Z, Crook DW, Phan HTT. 2017. Resolving plasmid structures in Enterobacteriaceae using the MinION nanopore sequencer: assessment of MinION and MinION/Illumina hybrid data assembly approaches. Microb Genomics 3:e000118.

66. Brooks L, Kaze M, Sistrom M. 2019. A Curated, Comprehensive Database of Plasmid Sequences. Microbiol Resour Announc 8.

67. Heiden SE, Hübner N-O, Bohnert JA, Heidecke C-D, Kramer A, Balau V, Gierer W, Schaefer S, Eckmanns T, Gatermann S, Eger E, Guenther S, Becker K, Schaufler K. 2020. A Klebsiella pneumoniae ST307 outbreak clone from Germany demonstrates features of extensive drug resistance, hypermucoviscosity, and enhanced iron acquisition. Genome Med 12:113.

68. Schweizer C, Bischoff P, Bender J, Kola A, Gastmeier P, Hummel M, Klefisch F-R, Schoenrath F, Frühauf A, Pfeifer Y. 2019. Plasmid-Mediated Transmission of KPC-2 Carbapenemase in Enterobacteriaceae in Critically Ill Patients. Front Microbiol 10:276.

69. San Millan A. 2018. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. Trends Microbiol 26:978–985.

70. Müller V, Karami N, Nyberg LK, Pichler C, Torche Pedreschi PC, Quaderi S, Fritzsche J, Ambjörnsson T, Åhrén C, Westerlund F. 2016. Rapid Tracing of Resistance Plasmids in a Nosocomial Outbreak Using Optical DNA Mapping. ACS Infect Dis 2:322–328.

71. Seara N, Oteo J, Carrillo R, Pérez-Blanco V, Mingorance J, Gómez-Gil R, Herruzo R, Pérez-Vázquez M, Astray J, García-Rodríguez J, Ruiz-Velasco LM, Campos J, de Burgos C, Ruiz-Carrascoso G. 2015. Interhospital spread of NDM-7-producing Klebsiella pneumoniae belonging to ST437 in Spain. Int J Antimicrob Agents 46:169–173.

72. Weingarten RA, Johnson RC, Conlan S, Ramsburg AM, Dekker JP, Lau AF, Khil P, Odom RT, Deming C, Park M, Thomas PJ, NISC Comparative Sequencing Program, Henderson DK, Palmore TN, Segre JA, Frank KM. 2018. Genomic Analysis of Hospital Plumbing Reveals Diverse Reservoir of Bacterial Plasmids Conferring Carbapenem Resistance. mBio 9.

73. Schmithausen RM, Sib E, Exner M, Hack S, Rösing C, Ciorba P, Bierbaum G, Savin M, Bloomfield SF, Kaase M, Jacobshagen A, Gemein S, Gebel J, Engelhart S, Exner D. 2019. The washing machine as a reservoir for transmission of extended spectrum beta-lactamase (CTX-M-15)-producing Klebsiella oxytoca ST201 in newborns. Appl Environ Microbiol https://doi.org/10.1128/AEM.01435-19.

74. Tofteland S, Naseer U, Lislevand JH, Sundsfjord A, Samuelsen O. 2013. A long-term low-frequency hospital outbreak of KPC-producing Klebsiella pneumoniae involving Intergenus plasmid diffusion and a persisting environmental reservoir. PloS One 8:e59015.

75. Stimson J, Gardy J, Mathema B, Crudu V, Cohen T, Colijn C. 2019. Beyond the SNP Threshold: Identifying Outbreak Clusters Using Inferred Transmissions. Mol Biol Evol 36:587–603.

76. Zacharioudakis I, Ding D, Zervou F, Stachel A, Hochman S, Sterling S, Lighter J, Aguero-Rosenfeld M, Shopsin B, Phillips M. 2020. Use of Varying Single-Nucleotide Polymorphism Thresholds to Identify Strong Epidemiologic Links Among Patients with Methicillin-Resistant *Staphylococcus aureus* (MRSA). Infect Control Hosp Epidemiol 41:s423–s424.

77. Coll F, Raven KE, Knight GM, Blane B, Harrison EM, Leek D, Enoch DA, Brown NM, Parkhill J, Peacock SJ. 2020. Definition of a genetic relatedness cutoff to exclude recent transmission of meticillin-resistant Staphylococcus aureus: a genomic epidemiology analysis. Lancet Microbe 1:e328–e335.

78. Zhou Z-C, Shuai X-Y, Lin Z-J, Liu Y, Zhu L, Chen H. 2021. Prevalence of multi-resistant plasmids in hospital inhalable particulate matter (PM) and its impact on horizontal gene transfer. Environ Pollut 270:116296.

79. Arredondo-Alonso S, Top J, Corander J, Willems RJL, Schürch AC. 2021. Mode and dynamics of vanA-type vancomycin resistance dissemination in Dutch hospitals. Genome Med 13:9.

80. Arredondo-Alonso S, Top J, McNally A, Puranen S, Pesonen M, Pensar J, Marttinen P, Braat JC, Rogers MRC, van Schaik W, Kaski S, Willems RJL, Corander J, Schürch AC. 2020. Plasmids Shaped the Recent Emergence of the Major Nosocomial Pathogen *Enterococcus faecium*. mBio 11:e03284-19, /mbio/11/1/mBio.03284-19.atom.

81. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15:R46.

82. Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35:D61-65.

83. 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 J Comput Mol Cell Biol 19:455–477.

84. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinforma Oxf Engl 29:1072–1075.

85. 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.

86. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl 30:2068–2069.

87. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644.

88. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J Antimicrob Chemother 75:3491–3500.

89. Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58:3895–3903.

90. Joensen KG, Scheutz F, Lund O, Hasman H, Kaas RS, Nielsen EM, Aarestrup FM. 2014. Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic Escherichia coli. J Clin Microbiol 52:1501–1510.

91. Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. Nucleic Acids Res 47:D687–D692.

92. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693.

93. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinforma Oxf Engl 30:1312–1313.

94. Seemann, Torsten T. 2015. Snippy: rapid haploid variant calling and core SNP phylogeny. Available.

95. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinforma Oxf Engl 27:1009–1010.

96. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.

97. Ruxton GD, Neuhäuser M. 2013. Review of alternative approaches to calculation of a confidence interval for the odds ratio of a 2 × 2 contingency table. Methods Ecol Evol 4:9–13.

98. Sheskin, D.J. 2004. Handbook of parametric and nonparametric statistical procedures (3rd ed.), 3rd ed. Chapman & Hall/CRC.

99. Ferreira JA, Zwinderman AH. 2006. On the Benjamini–Hochberg method. Ann Stat 34:1827–1849.

100. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, Saunders LD, Beck CA, Feasby TE, Ghali WA. 2005. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 43:1130–1139.

101. van den Elzen PJ, Konings RN, Veltkamp E, Nijkamp HJ. 1980. Transcription of bacteriocinogenic plasmid CloDF13 in vivo and in vitro: structure of the cloacin immunity operon. J Bacteriol 144:579–591.

102. SÃ¡nchez J, Diep DB, Herranz C, Nes IF, Cintas LM, HernÃ¡ndez PE. 2007. Amino acid and nucleotide sequence, adjacent genes, and heterologous expression of hiracin JM79, a sec-dependent bacteriocin produced by *Enterococcus hirae* DCH5, isolated from Mallard ducks (*Anas platyrhynchos*). FEMS Microbiol Lett 270:227–236.

103. González-Pastor JE, San Millán JL, Castilla MA, Moreno F. 1995. Structure and organization of plasmid genes required to produce the translation inhibitor microcin C7. J Bacteriol 177:7131–7140.

104. Songer JG. 1997. Bacterial phospholipases and their role in virulence. Trends Microbiol 5:156–161.

105. Moser SA, Savage DC. 2001. Bile Salt Hydrolase Activity and Resistance to Toxicity of Conjugated Bile Salts Are Unrelated Properties in Lactobacilli. Appl Environ Microbiol 67:3476–3480.

106. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259.

107. Darling ACE. 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res 14:1394–1403.

108. Sun D, Jeannot K, Xiao Y, Knapp CW. 2019. Editorial: Horizontal Gene Transfer Mediated Bacterial Antibiotic Resistance. Front Microbiol 10:1933.

109. Vrijens F, Hulstaert F, Van de Sande S, Devriese S, Morales I, Parmentier Y. 2010. Hospital-acquired, laboratory-confirmed bloodstream infections: linking national surveillance data to clinical and financial hospital data to estimate increased length of stay and healthcare costs. J Hosp Infect 75:158–162.

110. McGregor JC, Kim PW, Perencevich EN, Bradham DD, Furuno JP, Kaye KS, Fink JC, Langenberg P, Roghmann M-C, Harris AD. 2005. Utility of the Chronic Disease Score and

Charlson Comorbidity Index as Comorbidity Measures for Use in Epidemiologic Studies of Antibiotic-resistant Organisms. Am J Epidemiol 161:483–493.

111. Carroll AC, Wong A. 2018. Plasmid persistence: costs, benefits, and the plasmid paradox. Can J Microbiol 64:293–304.

112. Acman M, van Dorp L, Santini JM, Balloux F. 2020. Large-scale network analysis captures biological features of bacterial plasmids. Nat Commun 11:2452.

113. McCarthy AJ, Loeffler A, Witney AA, Gould KA, Lloyd DH, Lindsay JA. 2014. Extensive Horizontal Gene Transfer during Staphylococcus aureus Co-colonization In Vivo. Genome Biol Evol 6:2697–2708.