The Journal of Infectious Diseases
MAJOR ARTICLE

Phylogenetic Context of Antibiotic Resistance Provides Insights into the Dynamics of Resistance Emergence and Spread

Kyle J. Gontjes,[1] Aryan Singh,[1] Sarah E. Sansom,[2] James D. Boyko,[3] Stephen A. Smith,[3] Ebbing Lautenbach,[4,5,6] and Evan Snitkin[1,7]

[1]Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan, USA; [2]Department of Internal Medicine, Division of Infectious Diseases, Rush University Medical Center, Chicago, Illinois, USA; [3]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, USA; [4]Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA; [5]Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA; [6]Department of Medicine, Division of Infectious Diseases, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA; and [7]Department of Internal Medicine, Division of Infectious Diseases, University of Michigan Medical School, Ann Arbor, Michigan, USA

Background. To ameliorate the antibiotic resistance crisis, the drivers of resistance emergence and resistance spread must be better understood.

Methods. Whole-genome sequencing and susceptibility testing were performed on clinical carbapenem-resistant *Klebsiella pneumoniae* isolates collected from August 2014 to July 2015 across 12 long-term acute care hospitals. Ancestral state reconstruction partitioned patients with resistant strains into those that likely acquired resistance via de novo evolution or cross-transmission. Logistic regression was used to evaluate the associations between patient characteristics/exposures and these 2 pathways: resistance due to predicted within-host emergence of resistance and resistance due to predicted cross-transmission. This framework is available in the user-friendly R package, *phyloAMR* (https://github.com/kylegontjes/phyloAMR).

Results. Phylogenetic analysis of 386 epidemic lineage carbapenem-resistant *K. pneumoniae* sequence type 258 isolates revealed differences in the relative contribution of de novo evolution and cross-transmission to the burden of resistance to 5 antibiotics. Clade-specific variations in rates of resistance emergence and their frequency and magnitude of spread were detected for each antibiotic. Phylogenetically informed regression modeling identified distinct clinical risk factors associated with each pathway. Exposure to the cognate antibiotic was an independent risk factor for resistance emergence (trimethoprim-sulfamethoxazole, colistin, and novel beta-lactam/beta-lactamase inhibitors) and resistance spread (trimethoprim-sulfamethoxazole, amikacin, and colistin). In addition to antibiotic exposures, comorbidities (eg, stage IV + decubitus ulcers) and indwelling medical devices (eg, gastrostomy tubes) were detected as unique risk factors for resistance spread.

Conclusions. Phylogenetic contextualization generated insights and hypotheses into how bacterial genetic background, patient characteristics, and clinical practices influence the emergence and spread of antibiotic resistance.

Keywords. *Klebsiella pneumoniae*; antibiotic resistance; whole-genome sequencing; phylogenetics; clinical drivers.

Antibiotic resistance is a significant public health challenge [1]. The development of infections with antibiotic-resistant pathogens significantly reduces treatment options and clinical cure rates. Of greatest concern are multidrug-resistant lineages that have evolved resistance to multiple classes of antibiotics and disseminated worldwide [2].

The increasing problem of antibiotic resistance demands urgent clarity on the precise drivers of resistance proliferation. To this end, previous studies have sought to identify risk factors associated with harboring a pathogen that exhibits resistance [3–5]. This approach has a significant limitation—it overlooks the distinct routes to the acquisition of resistance, notably de novo evolution (ie, the emergence of a unique resistant strain spontaneously in an individual via chromosomal mutation or the acquisition of mobile genetic elements) and cross-transmission of circulating resistant lineages (ie, the acquisition of a resistant strain that is epidemiologically linked to other individuals) [6–8]. These pathways call for distinct intervention strategies. For example, judicious antimicrobial stewardship may be useful to reduce selection for de novo resistance evolution,

Received 04 June 2025; accepted 10 September 2025; published online 15 September 2025
Correspondence: Evan Snitkin, PhD, Department of Microbiology and Immunology, University of Michigan Medical School, 1520D Medical Sciences Research Building I, Ann Arbor, MI 48109-5680 (esnitkin@med.umich.edu).

The Journal of Infectious Diseases® 2025;232:e992–1002

https://doi.org/10.1093/infdis/jiaf478

whereas measures to minimize cross-transmission include patient cohorting, hand hygiene compliance, and contact isolation [9–12]. A deeper understanding of these pathways to resistance acquisition will facilitate the application of tailored prevention strategies and sharpen our comprehension of how bacterial genetic background, patient characteristics, and medical practices drive the emergence and spread of antibiotic resistance.

Here, we developed an approach that uses phylogenetic context to partition antibiotic-resistant isolates into those putatively derived from de novo evolution versus cross-transmission. We then applied this method to understand drivers of these 2 pathways to resistance in the epidemic carbapenem-resistant *Klebsiella pneumoniae* (CRKP) sequence type 258 (ST258) lineage, using a comprehensive collection of clinical isolates and patient metadata collected over 1 year from a network of long-term acute care hospitals (LTACHs). The emergence and spread of 5 critical antibiotic resistance phenotypes were identified, with the CRKP ST258 sublineage, antibiotic exposures, and patient clinical characteristics being differentially associated with the emergence and spread of each phenotype.

METHODS

Sample Collection

Clinical CRKP isolates were obtained from 1 August 2014, to 25 July 2015, at 12 southern California LTACHs [13]. Bacterial isolates from blood, respiratory, urine, and wound cultures identified as *K. pneumoniae* were tested for phenotypic carbapenem resistance using the 2015 Centers for Disease Control and Prevention criteria [13].

Clinical Metadata Collection

Clinical metadata were extracted from electronic medical records, as described previously [13]. Patient demographics, medical comorbidities, and indwelling medical devices were recorded at specimen collection, along with antibiotic exposures up to 30 days prior to specimen collection. Antibiotic exposure data were available only during a patient's stay in the LTACH.

Antibiotic Susceptibility Testing

Antibiotic susceptibility testing was performed for 2 aminoglycosides, colistin, trimethoprim–sulfamethoxazole (TMP–SMX), and 2 recently approved β-lactam/β-lactamase inhibitor (BL/BLI) combinations that were not clinically available at the time of specimen collection. Susceptibility profiles for TMP–SMX and gentamicin were extracted from regional microbiology laboratory records. Susceptibility to amikacin, colistin, imipenem–relebactam, and meropenem–vaborbactam was determined using broth microdilution (SENSITITRE, Thermo Scientific) [14]. Antibiotic resistance was defined as a minimum inhibitory concentration within the intermediate or resistant category by the 2021 Clinical and Laboratory Standards Institute interpretative criteria for all antibiotics except TMP–SMX [14], for which resistance was defined as a minimum inhibitory concentration ≥16/80 μg/mL (Supplementary Table 1). Resistance to BL/BLI agents was defined as resistance to imipenem–relebactam and/or meropenem–vaborbactam.

Whole-Genome Sequencing

Whole-genome sequencing of bacterial isolates was performed, as described previously [13]. The maximum-likelihood phylogeny was reconstructed from Gubbins-recombination filtered polymorphic sites using a custom, in-house Snakemake pipeline (https://github.com/Snitkin-Lab-Umich/phylokit) [15–24]. Details on whole-genome sequencing, variant calling, and phylogenetic tree reconstruction are present in Supplementary Methods.

Phylogenetic Analysis of Antibiotic Resistance Using Joint Ancestral State Reconstruction

We developed the open-source R package, *phyloAMR*, to study the emergence and spread of antibiotic resistance (https://github.com/kylegontjes/phyloAMR).

PhyloAMR's core function, *asr*, leveraged ancestral state reconstruction to characterize the evolution of resistance across the phylogeny. The model with the lowest sample size–corrected Akaike information criterion (AICc) was used for joint ancestral state reconstruction using *corHMM* [25]. Edges on the phylogeny were evaluated to determine episodes where the trait continued (ie, susceptible→susceptible or resistant→resistant), was gained (ie, susceptible→resistant), or was lost (ie, resistant→susceptible).

PhyloAMR's phylogenetic tree traversal algorithm, *asr_cluster_detection*, inferred the evolutionary history of antibiotic resistance. This algorithm classifies trait-containing isolates as phylogenetic singletons (ie, evidence of de novo evolution of a trait) or members of a phylogenetic cluster of the trait (ie, evidence that the trait was inherited from a common ancestor of circulating trait-containing lineage). Resistant isolates with gain events inferred at the tip were classified as phylogenetic singletons. However, these isolates were eligible for classification as members of a phylogenetic cluster if a reversion event was detected at its parental node. Resistant isolates were classified as members of a phylogenetic cluster if their ancestral gain event was shared with at least one additional resistant isolate. Resistant isolates that did not share an ancestral gain event were classified as phylogenetic singletons. Phylogenetic clusters where all isolates belonged to one patient were reclassified as redundant phylogenetic singletons.

To describe the evolutionary history of antibiotic resistance, the transitional data and phylogenetic clustering of each phenotype were characterized. Specifically, the frequency of resistance gain, loss, and continuation events was determined using *phyoAMR*'s *asr_transition_analysis* function. Descriptive statistics for phylogenetic clustering were generated using *phyloAMR*'s *asr_cluster_analysis* function (see Supplementary Materials).

Testing for Heterogeneous Rates of Antibiotic Resistance Across Clades

To further evaluate the influence of genetic background on the evolution of antibiotic resistance, we tested whether differences in rates of resistance evolution exist across the 2 clades of ST258. To achieve this, we utilized *phytool*'s *fitmultiMK* function, which implements a modified Markov model that allows for heterogeneous rates of discrete character evolution across user-specified regions on the phylogeny [26, 27]. For each phenotype, a *fitmultiMK* model was constructed with a single regime, indicative of uniform rates of resistance evolution across the phylogeny. Next, a *fitmultiMK* model was constructed with 2 regimes, permitting each clade to have distinct rates of resistance evolution. The fit of the single- and multi-regime models was compared using the likelihood ratio test. Finally, hidden-rate modeling using *corHMM*, a data-driven approach, was performed to validate the presence of 2 distinct rate classes. Support for the existence of 2 rate classes was determined by comparing the AICc of the models. For these analyses, we fit 2 model structures: equal rates and all rates differerent.

Identification of Risk Factors for Resistance Emergence and Spread

Logistic regression modeling was performed to evaluate the association between genetic background, patient demographics, and clinical variables with the primary resistance outcomes of crude resistance and our 2 phylogenetically informed classifications of resistance. For each phenotype, patients with resistant isolates were partitioned as acquiring resistance via putative in vivo emergence (ie, phylogenetic singletons) or the acquisition of a strain belonging to a resistant lineage (ie, members of a phylogenetic cluster), presumably due to a cross-transmission event. For patients with more than one isolate, their first isolate was selected unless they contributed a later resistant isolate, in which case the resistant isolate was chosen. Patients contributing only susceptible isolates served as the reference group.

First, logistic regression was performed to evaluate the association between these outcomes and 3 explanatory variables: prior exposure to the resistance phenotype's cognate antibiotic, exposure to dysbiotic antibiotics, and having an isolate belonging to clade I of ST258. Exposure to dysbiotic agents—drugs that increase the risk for disruption of the gut microbiota and *Clostridioides difficile* infection—was defined as having a history of exposure to at least one of the following antibiotics: cefepime, ceftaroline, ceftazidime, ceftriaxone, ciprofloxacin, ertapenem, metronidazole, levofloxacin, meropenem, imipenem, and piperacillin/tazobactam [28, 29]. Antibiotic exposure history was binarized as the receipt of ≥1 day of therapy in the past 30 days. Regression models evaluating the association between our resistance outcomes and antimicrobial exposures were adjusted for patient age and sex.

Next, we performed data-driven, multivariable regression modeling to identify shared and unique risk factors for our resistance outcomes [30]. Eligible variables included patient demographics, clinical comorbidities, indwelling medical devices, and antibiotic exposure histories (Supplementary Table 2). First, unadjusted logistic regression was performed. Next, all variables with an unadjusted *P*-value of <.20 were included in a logistic regression model. An iterative process removed variables if their *P*-value was >.10. After compiling this model, all initially ineligible variables were iteratively included in the model and retained if their *P*-value was <.10 .

Data Analysis and Visualization

All data analysis and visualization, unless otherwise stated, were performed using R version 4.5.0 [31]. The *ggplot2* and *ggnewscale* packages were used to generate common figures [32, 33]. Phylogenetic visualizations were generated using *ape*, *phytools*, and *ggtree* [26, 34, 35]. Descriptive tables were created using *tableone* [36]. Forest plots were generated using *forestplot* [37]. Multipanel figures were constructed using *cowplot* [38]. Code for this project is available at https://github.com/kylegontjes/phylogenetic-resistance-ms/.

RESULTS

Study Population

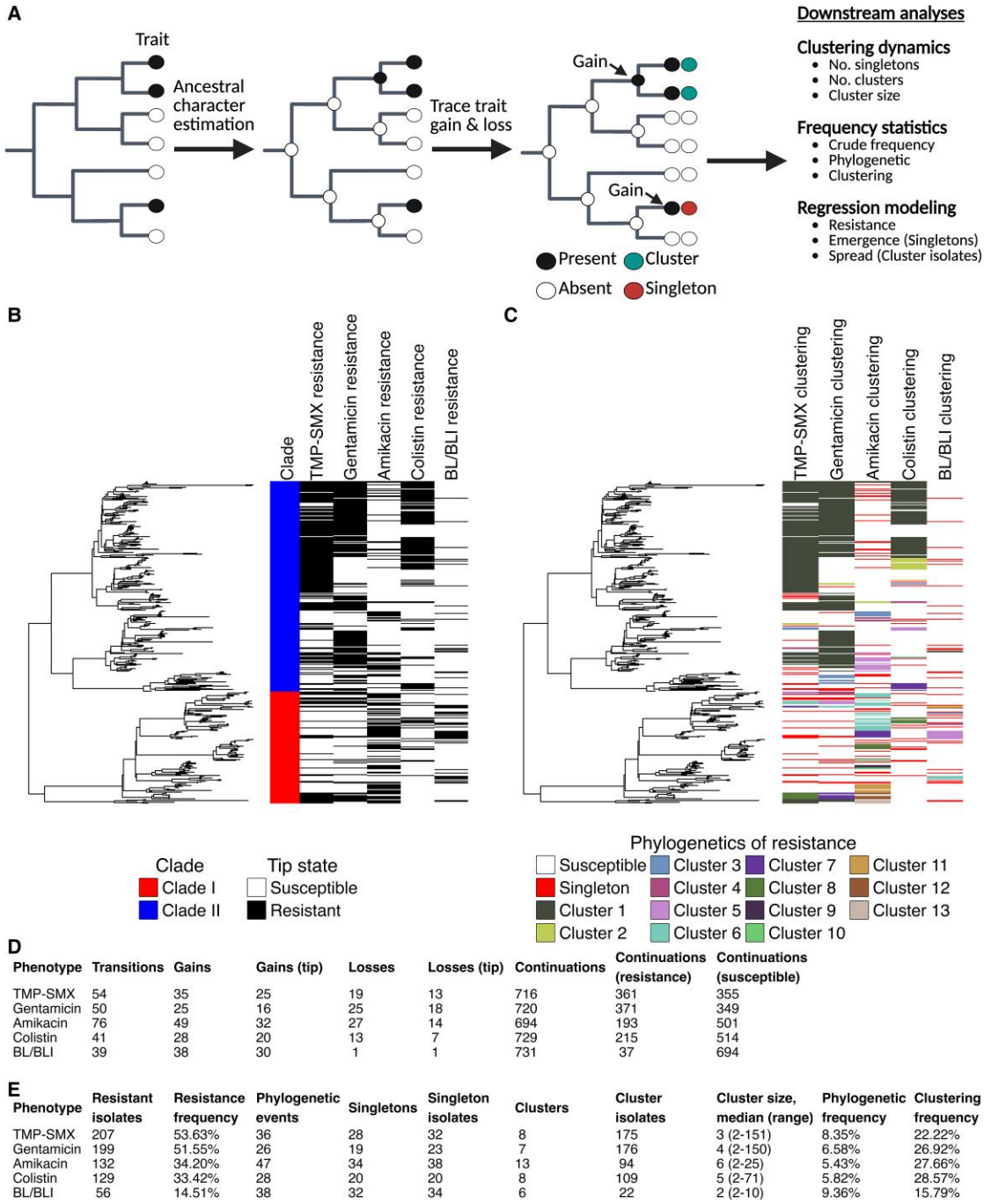
A total of 386 clinical CRKP ST258 isolates were collected from 312 patients across 12 California LTACHs over nearly 12 months (Supplementary Figure 1*A* and 1*B*). Fifty-five patients (17.6%) contributed more than one isolate during the study period (Supplementary Figure 1*C*). On first isolate collection, the median age was 72.7 years (interquartile range [IQR], 18.2 years), and 149 (47.8%) were female. Most patients had at least one indwelling medical device (87.8%), multiple medical comorbidities (64.4%), and received antibiotics in the 30 days preceding isolate collection (72.8%). The median LTACH length of stay before clinical culture was 21 days (IQR, 37 days). Cohort characteristics are provided in Supplementary Table 3.

Antibiotic Susceptibility Profiles

On application of clinical breakpoints, resistance was observed in 207 (53.6%) to TMP-SMX, 199 (51.6%) to gentamicin, 132 (34.2%) to amikacin, 129 (33.4%) to colistin, and 56 (14.5%) to novel BL/BLI combinations (meropenem–vaborbactam, 39 [10.1%]; imipenem–relebactam, 36 [9.3%]) (Supplementary Figure 1*D*). Minimum-inhibitory concentration distributions are reported in Supplementary Figure 2.

Whole-Genome Sequencing Revealed Variable Phylogenetic Clustering of Resistance

We implemented a phylogenetic algorithm to characterize the emergence and spread of antibiotic resistance in this densely sampled population (Figure 1*A*). Overlaying resistance on the phylogeny revealed numerous independent episodes of resistance emergence and spread for each antibiotic (Figure 1*B*). Quantification using our ancestral state reconstruction-



D

Phenotype	Transitions	Gains	Gains (tip)	Losses	Losses (tip)	Continuations	Continuations (resistance)	Continuations (susceptible)
TMP-SMX	54	35	25	19	13	716	361	355
Gentamicin	50	25	16	25	18	720	371	349
Amikacin	76	49	32	27	14	694	193	501
Colistin	41	28	20	13	7	729	215	514
BL/BLI	39	38	30	1	1	731	37	694

E

Phenotype	Resistant isolates	Resistance frequency	Phylogenetic events	Singletons	Singleton isolates	Clusters	Cluster isolates	Cluster size, median (range)	Phylogenetic frequency	Clustering frequency
TMP-SMX	207	53.63%	36	28	32	8	175	3 (2-151)	8.35%	22.22%
Gentamicin	199	51.55%	26	19	23	7	176	4 (2-150)	6.58%	26.92%
Amikacin	132	34.20%	47	34	38	13	94	6 (2-25)	5.43%	27.66%
Colistin	129	33.42%	28	20	20	8	109	5 (2-71)	5.82%	28.57%
BL/BLI	56	14.51%	38	32	34	6	22	2 (2-10)	9.36%	15.79%

Figure 1. Phylogenetic characterization of resistance revealed differential patterns of resistance emergence and spread for 5 antibiotic phenotypes among clinical carbapenem-resistant *Klebsiella pneumoniae* sequence type 258 isolates. *A*, Ancestral character estimation was performed to characterize the phylogenetic clustering of antibiotic resistance. Phenotypes were binarized into susceptible and resistant using clinical breakpoints. Resistance was inferred at internal nodes using *corHMM*'s joint ancestral reconstruction algorithm. The phylogenetic tree was traversed from parent to child node to determine episodes of trait continuation (ie, susceptible→susceptible and resistant→resistant), gain events (eg, susceptible→resistant), and loss events (ie, resistant→susceptible). Resistant isolates were inferred as phylogenetic singletons if the resistant gain event was inferred at the tip with no history of susceptible reversion at their parental node or the resistant isolate did not share its gain event with another isolate. Resistant isolates were classified as members of a phylogenetic cluster if their gain event was shared with at least one additional resistant isolate. Lineages where all resistant isolates belonged to one patient were reclassified as redundant resistant singletons. This schematic was generated using BioRender. *B*, Resistance overlaid across the *K. pneumoniae* sequence type 258 phylogeny. *C*, Phylogenetic clustering of resistance, as inferred from the ancestral state reconstruction algorithm, was overlaid across the phylogenetic tree. *D*, Phylogenetic transition and (*E*) clustering statistics for each antibiotic. The frequency of phylogenetic occurrence accounts for the number of episodes a trait occurs across the phylogenetic tree (ie, clusters + singleton events) relative to the total number of possible events (ie, clusters + singleton events + isolates without the trait). The frequency of clustering characterizes the proportion of phylogenetic events that are phylogenetic clusters. Abbreviations: BL/BLI, beta-lactam/beta-lactamase inhibitor; NS, resistance; TMP-SMX, trimethoprim-sulfamethoxazole.

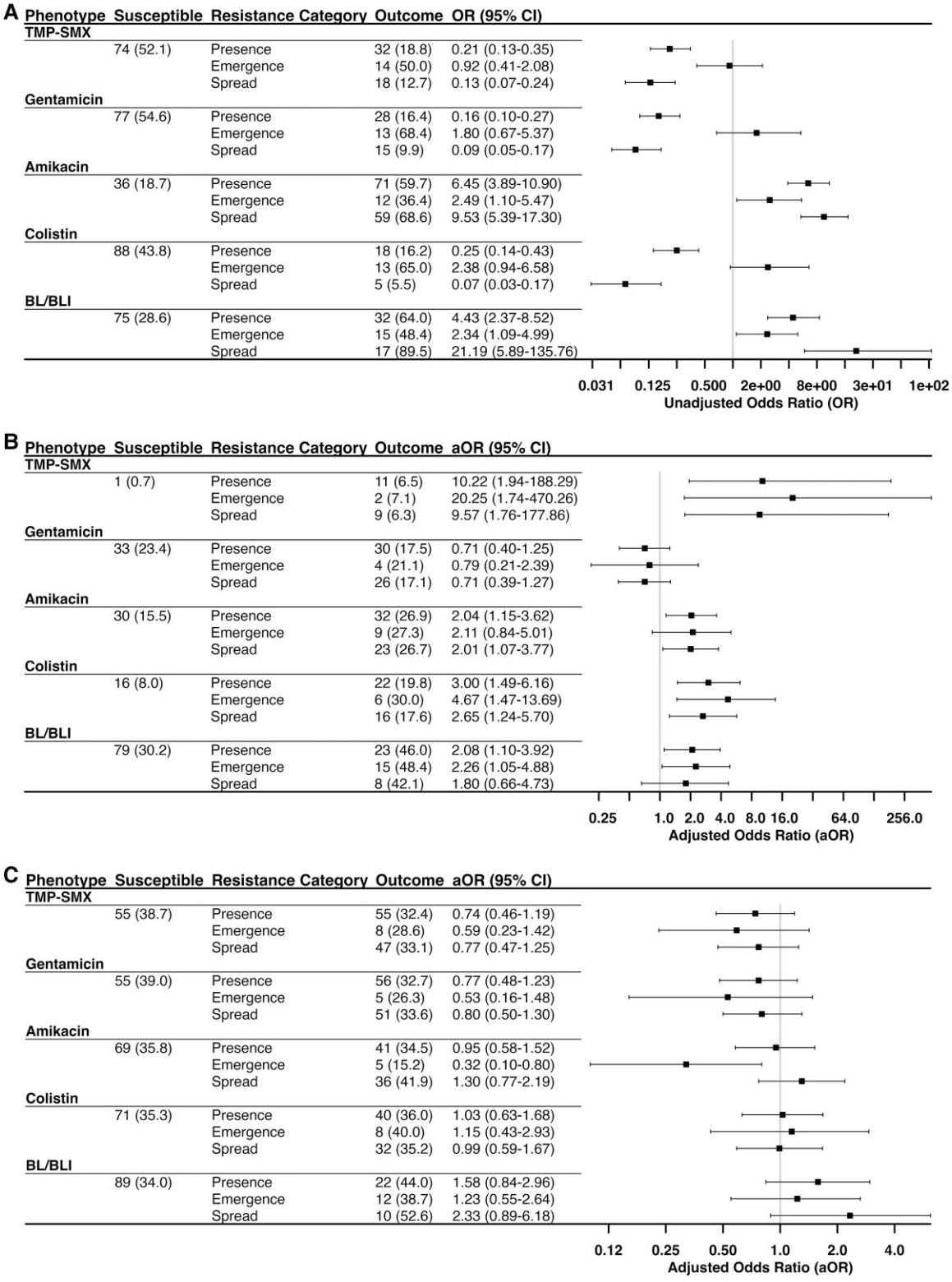


Figure 2. Influence of genetic background and antibiotic exposures on the emergence and spread of resistant lineages. Logistic regression was used to analyze the association between our resistance outcomes and (*A*) having an isolate belonging to clade I of the sequence type 258 phylogeny, (*B*) exposure to cognate antibiotics, and (*C*) exposure to dysbiotic antibiotics. For patients with more than one isolate, the first resistant isolate was retained. Each outcome (resistant, emergence, and spread) was compared with susceptible isolates. Exposure to an outcome's cognate antibiotic was defined as follows: TMP-SMX for TMP-SMX, aminoglycoside exposure for gentamicin and amikacin, polymyxin exposure for colistin, and carbapenem exposure for resistance to BL/BLI agents. Exposure to a dysbiotic agent was defined as a history of exposure to at least one of the following antibiotics: cefepime, ceftaroline, ceftazidime, ceftriaxone, ciprofloxacin, ertapenem, metronidazole, levofloxacin, meropenem, imipenem, and piperacillin/tazobactam [30, 31]. Models for antibiotic exposures were adjusted for patient sex and age at culture collection. Abbreviations: BL/BLI, beta-lactam/beta-lactamase inhibitor; OR, odds ratio; TMP-SMX, trimethoprim-sulfamethoxazole.

informed algorithm revealed differences in the number, size, and spatiotemporal overlap of resistant clusters for each antibiotic (Figure 1*C*–*E*; Supplementary Figures 3 and 4), suggesting differential propensities for resistance to emerge and spread. Indeed, ancestral state reconstruction produced evolutionary models with evolutionary rates that varied in magnitude (Supplementary Table 4), further supporting differences in emergence rates across antibiotics. The ancestral states and phylogenetic clustering of each phenotype were overlaid on the phylogeny in Supplementary Figure 5.

In addition to differences across antibiotics, we also observed differences in phylogenetic clustering between the 2 major ST258 clades (Figure 2*A*; Supplementary Table 5). Testing for heterogeneity in evolutionary rates supported a role for the genetic background of ST258 in shaping the emergence and spread of antibacterial resistance, with differences between the 2 clades in the rates of resistance evolution and reversion to susceptibility observed for gentamicin, amikacin, colistin, and BL/BLI agents (Supplementary Table 6). Data-driven modeling of rate heterogeneity using *corHMM* also supported the existence of distinct evolutionary rates of resistance across this phylogeny (Supplementary Table 7). Lastly, we employed logistic regression to evaluate the association of ST258 clade with the emergence and spread of resistance. On unadjusted regression, clade I was positively associated with the following outcomes: amikacin resistance emergence (odds ratio [OR], 2.49; 95% CI, 1.10–5.47), amikacin resistance spread (OR, 9.53; 95% CI, 5.39–17.30), BL/BLI resistance emergence (OR, 2.34; 95% CI, 1.09–4.99) and BL/BLI resistance spread (OR, 21.19; 95% CI, 5.89–135.76). Conversely, clade II was positively associated with the following outcomes: TMP-SMX resistance spread (OR, 7.50; 95% CI, 4.22–13.91), gentamicin resistance spread (OR, 10.99; 95% CI, 6.01–21.24), and colistin resistance spread (OR, 13.39; 95% CI, 5.72–39.27).

Phylogenetically Informed Risk Factor Analysis Revealed Shared and Unique Risk Factors Associated With the Emergence and Spread of Antibiotic Resistance

Having identified cases of resistance emergence and spread, we next explored whether differential risk factors exist across these 2 groups: phylogenetic singletons (ie, resistance emergence) and clusters (ie, spread of circulating resistant lineages). Initially focusing on antibiotic use, we postulated that exposure to cognate antibiotics would be preferentially associated with selection for resistance emergence. After adjusting for age and sex, cognate antibiotic exposure was positively associated with resistance emergence for TMP-SMX (adjusted OR [aOR], 20.25; 95% CI, 1.74–470.26), colistin (aOR, 4.67; 95% CI, 1.47–13.69), and BL/BLI agents (aOR, 2.26; 95% CI, 1.05–4.88) (Figure 2*B*). Interestingly, cognate antibiotic exposure was also positively associated with resistance spread for TMP–SMX (aOR, 9.57, 95% CI, 1.76–177.86), colistin (aOR, 2.65, 95% CI, 1.24–5.70), and amikacin (aOR, 2.01, 95% CI, 1.07–3.77). As microbiome-disrupting antibiotics can increase susceptibility to colonization with multidrug-resistant organisms, we tested whether these antibiotics were associated with the spread of antibiotic-resistant lineages. No statistically significant, positive association was detected (Figure 2*C*).

To more broadly identify clinical and patient factors that drive the emergence and spread of antibiotic resistance, we performed data-driven, multivariable regression modeling (Table 1; Supplementary Table 8). In addition, models were constructed that did not consider phylogenetic context (ie, phenotypic resistance) to understand how phylogenetic contextualization might provide nuance beyond the standard approach. Inspection of multivariable models revealed risk factors with positive associations for the emergence and spread of resistance (Figure 3). Cognate antibiotic exposures were detected in both resistance emergence and spread models. Trimethoprim–sulfamethoxazole exposure was associated with the emergence and spread of resistance to TMP–SMX. Tobramycin and gentamicin, 2 aminoglycoside antibiotics, were associated with the emergence and spread of amikacin resistance. Exposure to polymyxin antibiotics was associated with colistin emergence and spread. Increased age was positively associated with the emergence and spread of resistance to BL/BLI agents.

Independent risk factors in resistant models tended to segregate into emergence or spread models, supporting the hypothesis that the standard approach imprecisely merges 2 distinct populations (Figure 3). Several variables were identified as independent risk factors for the emergence of resistance, but not its spread. Antibiotic exposures were often associated with the emergence of resistance. Tigecycline was associated with emergence of resistance to TMP–SMX, colistin, and BL/BLI agents. Exposure to carbapenems was associated with the emergence of resistance to BL/BLI agents but not spread. Additional risk factors that were unique to resistance emergence were detected, notably associations between congestive heart failure and the emergence of gentamicin resistance and malignancy with the emergence of amikacin resistance.

While models for resistance spread included exposure to antibiotics, additional proxies for complexity of care, notably patient comorbidities and indwelling medical devices, were uniquely associated with resistance spread. The presence of stage IV+ decubitus ulcers was associated with the spread of resistance to TMP–SMX and gentamicin. Chronic respiratory comorbidities, specifically the presence of chronic obstructive pulmonary disorder or chronic bronchitis, were associated with the spread of amikacin and BL/BLI resistance. This comorbidity was also associated with the emergence of colistin resistance. Indwelling medical devices were also often associated with the spread of resistant lineages, with presence of a gastrostomy and tracheostomy associated with the spread of TMP–SMX resistance, while presence of an indwelling urinary

Table 1. Multivariable Logistic Regression Modeling Identified Distinct Statistically Significant Risk Factors for Resistance Emergence and Spread

Phenotype TMP-SMX	Multivariable Regression Models								
	Present ($n = 170$)	OR (95% CI)	*P* Value	Emergence ($n = 28$)	OR (95% CI)	*P* Value	Spread ($n = 142$)	OR (95% CI)	*P* Value
	Cephalosporin exposure	0.40 (.22–.72)	.0022	TMP-SMX exposure	119.13 (5.95–5891.04)	.0041	Acute kidney injury	0.41 (.23–.70)	.0014
	TMP-SMX exposure	24.97 (4.02–504.50)	.0044	Female sex	0.35 (.13–.86)	.0284	Cephalosporin exposure	0.40 (.22–.73)	.0031
	Acute kidney injury	0.49 (.29–.81)	.006	Tigecycline exposure	3.66 (1.10–11.75)	.0292	TMP-SMX exposure	18.93 (3.13–368.12)	.0077
	Gastrostomy tube	1.86 (1.11–3.18)	.0203	…	…		Fluoroquinolone exposure	0.42 (.20–.82)	.0132
	Decubitus ulcer stage IV/V	2.08 (1.12–3.96)	.0227	…	…		Decubitus ulcer stage IV/V	2.10 (1.11–4.06)	.0247
	Tigecycline exposure	2.34 (1.12–5.11)	.0278	…	…		Gastrostomy tube	1.89 (1.09–3.33)	.0256
	Gentamicin exposure	0.17 (.02–.76)	.0346	…	…		Gentamicin exposure	0.09 (.00–.54)	.0286
	Amikacin exposure	0.46 (.21–.98)	.0465	…	…		…	…	
Gentamicin	Present ($n = 171$)	OR (95% CI)	*P* Value	Emergence ($n = 19$)	OR (95% CI)	*P* Value	Spread ($n = 152$)	OR (95% CI)	*P* Value
	Decubitus ulcer stage IV/V	2.23 (1.22–4.20)	.0105	Congestive heart failure	3.61 (1.21–10.75)	.0199	Decubitus ulcer stage IV/V	2.57 (1.37–4.96)	.0038
	Cephalosporin exposure	0.49 (.28–.84)	.0108	…	…		Chronic kidney disease	0.47 (.27–.82)	.0082
	Chronic kidney disease	0.52 (.31–.88)	.0152	…	…		TMP-SMX exposure	10.85 (1.96–203.63)	.0259
	TMP-SMX exposure	12.95 (2.22–249.77)	.0193	…	…		Cephalosporin exposure	0.56 (.32–.98)	.0438
Amikacin	Present ($n = 119$)	OR (95% CI)	*P* Value	Emergence ($n = 33$)	OR (95% CI)	*P* Value	Spread ($n = 86$)	OR (95% CI)	*P* Value
	Tobramycin exposure	2.82 (1.14–7.40)	.0279	Tobramycin exposure	3.79 (1.03–12.65)	.0334	Cephalosporin exposure	1.88 (1.06–3.33)	.0297
	Gentamicin exposure	4.44 (1.24–20.80)	.0318	Gentamicin exposure	5.61 (.94–33.04)	.0474	Gentamicin exposure	4.61 (1.08–23.30)	.0424
Colistin	Present ($n = 111$)	OR (95% CI)	*P* Value	Emergence ($n = 20$)	OR (95% CI)	*P* Value	Spread ($n = 91$)	OR (95% CI)	*P* Value
	Polymyxin exposure	3.85 (1.85–8.28)	4e−04	Indwelling urinary catheter	0.15 (.04–.49)	.0039	Polymyxin exposure	3.19 (1.45–7.11)	.0039
	Ventilator-dependent respiratory failure	0.54 (.30–.95)	.0353	Polymyxin exposure	7.87 (1.92–34.78)	.0042	Underweight or malnourished	0.47 (.24–.87)	.0208
	Underweight or malnourished	0.53 (.29–.95)	.0361	COPD or chronic bronchitis	4.60 (1.54–14.26)	.0064	…	…	
	…	…		Tigecycline exposure	3.95 (1.11–13.41)	.0277	…	…	
BL/BLI	Present ($n = 50$)	OR (95% CI)	*P* Value	Emergence ($n = 31$)	OR (95% CI)	*P* Value	Spread ($n = 19$)	OR (95% CI)	*P* Value
	Age, 1 y unit	1.04 (1.01–1.06)	.0036	Age, 1 y unit	1.03 (1.01–1.06)	.025	Cephalosporin exposure	5.74 (2.03–17.14)	.0011
	Cephalosporin exposure	2.21 (1.13–4.27)	.0186	Carbapenem exposure	2.41 (1.08–5.37)	.0299	COPD or chronic bronchitis	4.16 (1.38–12.54)	.0102
	Carbapenem exposure	1.94 (1.02–3.69)	.0424	…	…		Age, 1 y unit	1.04 (1.00–1.09)	.0383
	Underweight or malnourished	0.42 (.16–.95)	.0498	…	…		Indwelling urinary catheter	3.10 (1.07–10.49)	.0478

Data-driven regression modeling was performed to identify risk factors for antibiotic resistance in our long-term acute care hospital population. For each phenotype, patients with resistant isolates were partitioned as acquiring resistance via (1) putative in vivo emergence (ie, phylogenetic singletons) or (2) the acquisition of a strain belonging to a resistant lineage (ie, members of a phylogenetic cluster), presumably due to a cross-transmission event. For patients with more than one isolate, their first isolate was selected unless they contributed a later resistant isolate, in which case the resistant isolate was chosen. Patients contributing only susceptible isolates served as the reference group for each analysis. Alongside risk factor modeling for resistance emergence and spread, analysis was also performed to identify risk factors for harboring a resistant isolate. This table includes all model components with a *P*-value <.05. Full multivariable regression modeling results are found in Supplementary Table 8.

Abbreviations: BL/BLI, beta-lactam/beta-lactamase inhibitor; COPD, chronic obstructive pulmonary disease; OR, odds ratio; TMP-SMX, trimethoprim-sulfamethoxazole.

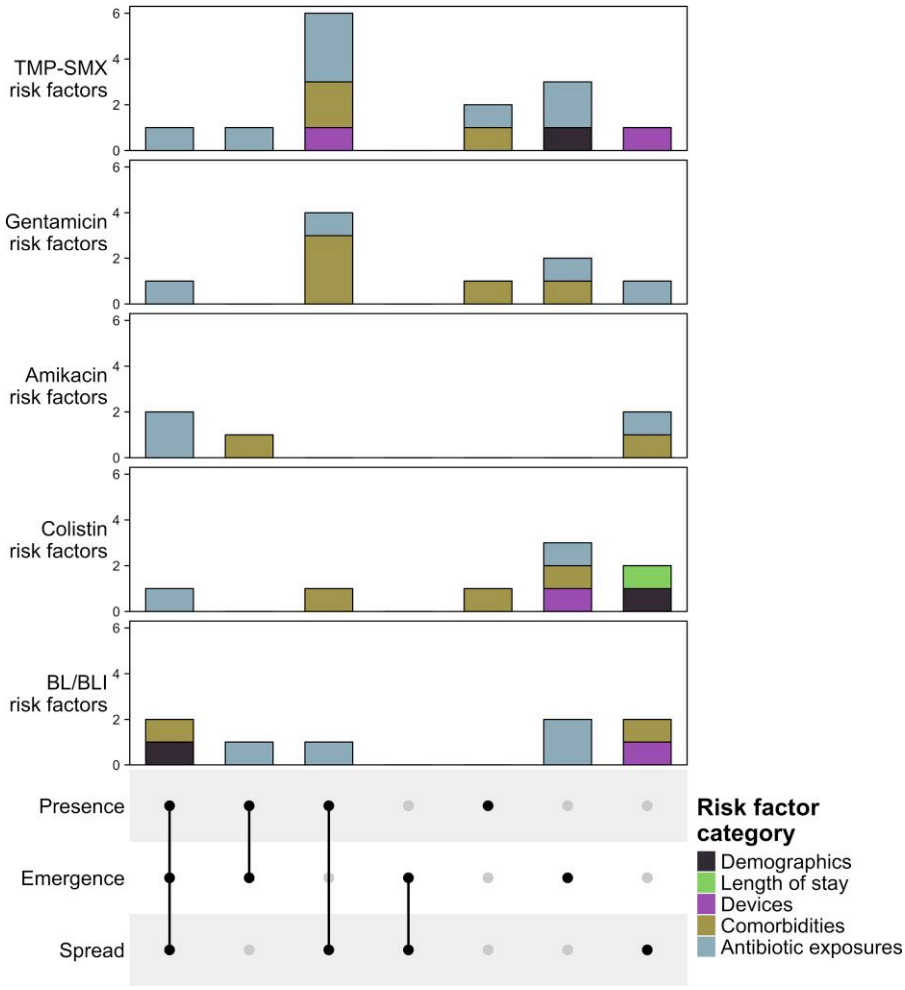


Figure 3. Shared and unique risk factors for the emergence and spread of antibiotic resistance. Logistic regression identified risk factors for harboring a resistant strain, harboring a resistant strain inferred to be due to de novo evolution, and harboring a resistant strain inferred to be acquired due to cross-transmission. The sharing of model components across these 3 regression analyses is visualized for resistance to trimethoprim-sulfamethoxazole, gentamicin, amikacin, colistin, and beta-lactam/beta-lactamase inhibitor agents. Represented as an UpSet plot, the heatmap indicates the possible membership patterns for an identified risk factor. For each phenotype, the bar plot indicates the number of variables in each intersection, while the fill colors indicate the variable category. Full multivariable regression models are reported in Supplementary Table 8. Abbreviations: BL/BLI, beta-lactam/beta-lactamase inhibitor; OR, odds ratio; TMP-SMX, trimethoprim-sulfamethoxazole.

catheter was associated with the spread of BL/BLI resistance. Highlighting its potentially complex role in resistance dynamics, cephalosporin exposure was positively associated with spread of amikacin and BL/BLI resistance but negatively associated with spread of TMP–SMX resistance, alongside the emergence and spread of resistance to gentamicin. Collectively, these analyses suggest that the traditional modeling approaches may mask important differences in the patient characteristics and clinical practices that drive the emergence and spread of antibiotic resistance.

DISCUSSION

To test the hypothesis that more refined insights into the drivers of antibiotic resistance could be attained by considering the distinct pathways to resistance acquisition, we developed a phylogenetically informed approach, compiled as the open-source R package *phyloAMR*. This easy-to-use tool enables the partitioning of patients from a densely sampled cohort into those putatively acquiring resistant strains via de novo evolution or cross-transmission events. Applying *phyloAMR* to a regional collection of CRKP ST258 isolates revealed differences in the relative roles of emergence and spread for 5 critical antibiotics. We also identified differential impacts of ST258 clade, antibiotic exposures, and clinical characteristics on the emergence and spread of each resistance phenotype.

Modeling the acquisition of antibiotic resistance as a binary state of phenotypic resistance (ie, resistant or susceptible) oversimplifies and masks complexity. Our phylogenetic characterization revealed considerable differences in the contribution of de novo evolution and cross-transmission to resistance

proliferation in a regionally disseminating bacterial lineage, even among antibiotics with comparable frequencies of resistance. Differences in the dynamics of resistance evolution may reflect differences in magnitudes of selection, the fitness and transmissibility of resistant strains, or the contribution of within-host dynamics [39]. These findings highlight the potential for whole-genome sequencing to enhance traditional epidemiological investigations into precise microbial and host-associated factors that influence the evolution and spread of resistance. Furthermore, phylogenetic contextualization can be leveraged to inform the development of innovative antimicrobial stewardship and infection prevention strategies to account for the dynamics of resistance evolution and spread to prolong the long-term efficacy of an antibiotic.

Our analysis revealed striking differences in the rates of resistance emergence and spread across clades of epidemic lineage CRKP ST258. Prior in vitro experimentation and epidemiological studies have observed that resistance-conferring genotypes can impose varying fitness costs across distinct genetic backgrounds and that genetic background influences the propensity for mutational resistance [40–43]. Genetic background's influence on cross-transmission, as observed with colistin resistance, may also influence the dynamics of resistance evolution [44]. Our analysis also identified clade-specific differences in the dynamics of resistance to meropenem–vaborbactam and imipenem–relebactam, 2 recently approved BL/BLI agents that were not yet clinically available in this population. Investigations of phylogenetic clustering, like those presented in this manuscript, could inform surveillance efforts by identifying genetic backgrounds with a high propensity for resistance development or spread across healthcare systems, as well as distinguishing microbial risk factors for resistance emergence and spread.

Integrating phylogenetic context into risk factor modeling supports the emerging hypothesis that patient comorbidities and medical practices exert distinct selective pressures for the evolution [45] and cross-transmission of antibiotic-resistant lineages [46]. Like prior studies [3–5], we observed that cognate antibiotic exposure was often an independent risk factor for acquisition of antibiotic resistance, potentially through selective pressure. Our modeling additionally suggests that other noncognate antibiotic exposures can also differentially drive the evolution of resistance. While we observed that antibiotic exposures were predominantly identified as risk factors when modeling resistance emergence, distinct antibiotic exposures and markers of clinical severity, notably comorbidities and indwelling medical devices, also contribute to the spread of resistant lineages. Therefore, we postulate that antimicrobial stewardship interventions could be tailored to the specific dynamics for each antibiotic-resistant lineage while also considering the underlying patient population. Leveraging the rich phylogenetic contextualization of antibiotic resistance provides novel insights for development of innovative strategies to detect and prevent the emergence and spread of antibiotic-resistant organisms.

Our study has several limitations. First, our collection only included clinical CRKP isolates, but not those from asymptomatic active surveillance, which could have facilitated a more comprehensive view of resistance dynamics. This limitation is anticipated to be mitigated due to the LTACH setting, in which dense culture sampling is frequently performed for clinical evaluation [47–49]. Second, we analyzed only select clinical factors that may select for resistance or increase the susceptibility for acquiring a resistant strain. Future investigations should evaluate additional medical comorbidities and exposure to nonantibiotic medications. Third, our analysis was performed on a complex, high-acuity LTACH population with a high frequency of antibiotic exposures and use of indwelling medical devices, which may limit the generalizability of our risk factor analyses. Nonetheless, identifying the drivers of antibiotic resistance in this setting is critical due to their high colonization burden and the role these facilities play in the regional transmission of antibiotic-resistant organisms [48–50]. Fourth, our ability to detect precise and robust statistical associations between patient characteristics and our phylogenetically informed resistance outcomes, notably the emergence of resistance, was limited by a small sample size. Finally, our study did not seek to directly resolve the evolutionary predictions of resistance with genotypic data. Prior work in this population of resistance to colistin and BL/BLI combinations suggests concordance between our evolutionary predictions and the chromosomal and plasmid-mediated mechanisms of resistance [7, 8], underscoring the feasibility of our phylogenetic framework to generate inferences regarding the evolution of antibiotic resistance.

Conclusions

Whole-genome sequencing and phylogenetic characterization of antibiotic resistance in a densely sampled LTACH cohort improved our understanding of how antibiotic resistance emerges and spreads across a healthcare network. Applying a phylogenetically informed approach can generate testable hypotheses regarding the pathways to acquisition of antibiotic resistance and inform the development of targeted interventions to reduce the global burden of resistance and extend the long-term efficacy of antibiotics. Future studies, including additional clinical populations (eg, acute care hospitals, skilled nursing facilities) and microbial species, should investigate whether phylogenetics can consistently reveal trends concerning how bacteria develop resistance and how resistant bacteria spread to new hosts.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the

authors. Questions or messages regarding errors should be addressed to the author.

Notes

Data availability. Code, results, and select deidentified data can be found at https://github.com/kylegontjes/phylogenetic-resistance-ms.

Financial support. E. L., E. S., and K. J. G. were supported by National Institute of Allergy and Infectious Diseases grant 5R01AI148259. J. D. B., K. J. G., S. A. S., and E. S. were supported by the National Institute of Allergy and Infectious Diseases U19AI181767. S. A. S. was also supported by the United States National Science Foundation IntBio 2217116. K. J. G. reports training fellowships from the National Human Genome Research Institute (T32-HG000040) and the National Institute of Allergy and Infectious Diseases (F31-AI186288).

Potential conflict of interest. The authors report no conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Murray CJ, Ikuta KS, Sharara F, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet **2022**; 399:629–55.
2. Bowers JR, Kitchel B, Driebe EM, et al. Genomic analysis of the emergence and rapid global dissemination of the clonal group 258 *Klebsiella pneumoniae* pandemic. PLoS One **2015**; 10:e0133727.
3. Fish DN, Piscitelli SC, Danziger LH. Development of resistance during antimicrobial therapy: a review of antibiotic classes and patient characteristics in 173 studies. Pharmacotherapy **1995**; 15:279–91.
4. Harris AD, Samore MH, Lipsitch M, Kaye KS, Perencevich E, Carmeli Y. Control-group selection importance in studies of antimicrobial resistance: examples applied to *Pseudomonas aeruginosa*, *Enterococci*, and *Escherichia coli*. Clin Infect Dis **2002**; 34:1558–63.
5. Kaye KS, Harris AD, Samore M, Carmeli Y. The case–control study design: addressing the limitations of risk factor studies for antimicrobial resistance. Infect Control Hosp Epidemiol **2005**; 26:346–51.
6. Baquero F, Martínez JL, Lanza VF, et al. Evolutionary pathways and trajectories in antibiotic resistance. Clin Microbiol Rev **2021**; 34:e00050-19.
7. Lapp Z, Han JH, Choudhary D, et al. Fitness barriers to spread of colistin-resistant *Klebsiella* pneumoniae overcome by establishing niche in patient population with elevated colistin use. medRxiv:2021.06.11.21258758 [preprint]. https://doi.org/10.1101/2021.06.11.21258758. **2021**.
8. Gontjes KJ, Pirani A, Badri Narayanan D, et al. Identifying drivers of β-lactam/β-lactamase inhibitor resistance emergence and spread before their clinical deployment. medRxiv:2025.07.29.25331838 [preprint]. https://doi.org/10.1101/2025.07.29.25331838. **2025**.
9. Cantón R, Morosini M-I. Emergence and spread of antibiotic resistance following exposure to antibiotics. FEMS Microbiol Rev **2011**; 35:977–91.
10. Stracy M, Snitser O, Yelin I, et al. Minimizing treatment-induced emergence of antibiotic resistance in bacterial infections. Science **2022**; 375:889–94.
11. Baquero F, Lanza VF, Cantón R, Coque TM. Public health evolutionary biology of antimicrobial resistance: priorities for intervention. Evol Appl **2015**; 8:223–39.
12. Baym M, Stone LK, Kishony R. Multidrug evolutionary strategies to reverse antibiotic resistance. Science **2016**; 351:aad3292.
13. Han JH, Lapp Z, Bushman F, et al. Whole-genome sequencing to identify drivers of carbapenem-resistant *Klebsiella pneumoniae* transmission within and between regional long-term acute-care hospitals. Antimicrob Agents Chemother **2019**; 63:e01622-19.
14. Zhang HL, Gontjes KJ, Han JH, et al. Characterization of resistance to newer antimicrobials among carbapenem-resistant *Klebsiella pneumoniae* in the post–acute-care setting. Infect Control Hosp Epidemiol **2023**; 44:1159–62.
15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **2014**; 30:2114–20.
16. Andrews S. FastQC: a quality control tool for high throughput sequence data. **2010**.
17. Mölder F, Jablonski KP, Letcher B, et al. Sustainable data analysis with Snakemake. F1000Research **2021**; 10:33.
18. Croucher NJ, Page AJ, Connor TR, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic acids research **2015**; 43:e15-e.
19. Wishart DS, Han S, Saha S, et al. PHASTEST: faster than PHASTER, better than PHAST. Nucleic Acids Res **2023**; 51:W443–W50.
20. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol **2018**; 14:e1005944.
21. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol **2009**; 26:1641–50.
22. Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol **2020**; 37:1530–4.
23. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **2014**; 30:1312–3.

24. Kalyaanamoorthy S, Minh BQ, Wong TK, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods **2017**; 14:587–9.
25. Boyko JD, Beaulieu JM. Generalized hidden Markov models for phylogenetic comparative datasets. Methods Ecol Evol **2021**; 12:468–78.
26. Revell LJ. Phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). PeerJ **2024**; 12:e16505.
27. Revell LJ, Schliep KP, Mahler DL, Ingram T. Testing for heterogeneous rates of discrete character evolution on phylogenies. J Evol Biol **2024**; 37:1591–602.
28. Baggs J, Jernigan JA, Halpin AL, Epstein L, Hatfield KM, McDonald LC. Risk of subsequent sepsis within 90 days after a hospital stay by type of antibiotic exposure. Clin Infect Dis **2018**; 66:1004–12.
29. Gontjes KJ, Gibson KE, Lansing BJ, et al. Association of exposure to high-risk antibiotics in acute care hospitals with multidrug-resistant organism burden in nursing homes. JAMA Netw Open **2022**; 5:e2144959-e.
30. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. Source Code Biol Med **2008**; 3:1–8.
31. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, **2025**. https://www.r-project.org/.
32. Wickham H, Averick M, Bryan J, et al. Welcome to the Tidyverse. J Open Source Softw **2019**; 4:1686.
33. Campitelli E. ggnewscale: multiple fill and colour scales in'-ggplot2'. R Package Version 04. **2020**.
34. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics **2004**; 20:289–90.
35. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol **2017**; 8:28–36.
36. Yoshida K, Bohn J, Yoshida MK. Package 'tableone'. Vienna, Austria: R Foundation for Statistical Computing. **2020**.
37. Gordon M, Lumley T, Gordon MM. Package 'forestplot'. Advanced forest plot using 'grid' graphics. Vienna: The Comprehensive R Archive Network. **2019**.
38. Wilke CO, Wickham H, Wilke MCO. Package 'cowplot'. Streamlined plot theme and plot annotations for 'ggplot2. **2019**.
39. Day T, Huijben S, Read AF. Is selection relevant in the evolutionary emergence of drug resistance? Trends Microbiol **2015**; 23:126–33.
40. Holt KE, Wertheim H, Zadoks RN, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proc Natl Acad Sci **2015**; 112:E3574–81.
41. Carrilero L, Dunn SJ, Moran RA, McNally A, Brockhurst MA. Evolutionary responses to acquiring a multidrug resistance plasmid are dominated by metabolic functions across diverse *Escherichia coli* lineages. MSystems **2023**; 8:e00713-22.
42. Buckner MM, Saw HT, Osagie RN, et al. Clinically relevant plasmid-host interactions indicate that transcriptional and not genomic modifications ameliorate fitness costs of *Klebsiella pneumoniae* carbapenemase-carrying plasmids. MBio **2018**; 9:e02303-17.
43. Dunn S, Carrilero L, Brockhurst M, McNally A. Limited and strain-specific transcriptional and growth responses to acquisition of a multidrug resistance plasmid in genetically diverse *Escherichia coli* lineages. Msystems **2021**; 6:e00083-21.
44. Bray AS, Smith RD, Hudson AW, et al. MgrB-dependent colistin resistance in *Klebsiella pneumoniae* is associated with an increase in host-to-host transmission. MBio **2022**; 13:e03595-21.
45. Santos-Lopez A, Marshall CW, Haas AL, Turner C, Rasero J, Cooper VS. The roles of history, chance, and natural selection in the evolution of antibiotic resistance. Elife **2021**; 10:e70676.
46. Donskey CJ. The role of the intestinal tract as a reservoir and source for transmission of nosocomial pathogens. Clin Infect Dis **2004**; 39:219–26.
47. Munoz-Price LS. Long-term acute care hospitals. Clin Infect Dis **2009**; 49:438–43.
48. Han JH, Goldstein EJ, Wise J, Bilker WB, Tolomeo P, Lautenbach E. Epidemiology of carbapenem-resistant *Klebsiella pneumoniae* in a network of long-term acute care hospitals. Clin Infect Dis **2017**; 64:839–44.
49. Lin MY, Lyles-Banks RD, Lolans K, et al. The importance of long-term acute care hospitals in the regional epidemiology of *Klebsiella pneumoniae* carbapenemase—producing Enterobacteriaceae. Clin Infect Dis **2013**; 57:1246–52.
50. Toth DJ, Khader K, Slayton RB, et al. The potential for interventions in a long-term acute care hospital to reduce transmission of carbapenem-resistant Enterobacteriaceae in affiliated healthcare facilities. Clin Infect Dis **2017**; 65:581–7.