Fitness barriers to spread of colistin-resistant *Klebsiella pneumoniae* overcome by establishing niche in patient population with elevated colistin use

Authors: Zena Lapp, Jennifer H Han, Divya Choudhary, Stuart Castaneda, Ali Pirani, Kevin Alby, Pam Tolomeo, Ellie JC Goldstein, Sean Muldoon, Ebbing Lautenbach, Evan S Snitkin

Author Information:

Name: Zena Lapp
Degree: BA
Affiliation: Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan
ORCID: https://orcid.org/0000-0003-4674-2176

Name: Jennifer Han
Degree: MD
Affiliation: GlaxoSmithKline
ORCID: https://orcid.org/0000-0003-4744-8067

Name: Divya Choudhary
Degree: BTech
Affiliation: Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK
ORCID: https://orcid.org/0000-0002-8401-9037

Name: Stuart Castaneda
Degree: MS
Affiliation: Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan
ORCID: https://orcid.org/0000-0002-7825-6888

Name: Ali Pirani
Degree: MS
Affiliation: Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan
ORCID: https://orcid.org/0000-0001-7810-0982

Name: Kevin Alby
Degree: PhD
Affiliation: University of North Carolina
ORCID: https://orcid.org/0000-0003-4637-8474

Name: Pam C Tolomeo
Degree: MPH, CCRP

Affiliation: Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Perelman School of Medicine
ORCID: https://orcid.org/0000-0001-5594-8884

Name: Ellie JC Goldstein
Degree: MD
Affiliation: R M Alden Research Laboratory; David Geffen School of Medicine, University of California, Los Angeles
ORCID: https://orcid.org/0000-0002-1044-6159

Name: Sean Muldoon
Degree: MD, MPH, FCCP
Affiliation: Kindred Healthcare
ORCID: NA

Name: Ebbing Lautenbach
Degree: MD
Affiliation: Department of Medicine (Infectious Diseases), Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania
ORCID: https://orcid.org/0000-0003-4772-2409

Name: Evan Snitkin
Degree: PhD
Affiliation: Department of Microbiology and Immunology, Department of Internal Medicine/Division of Infectious Diseases, University of Michigan, Ann Arbor, Michigan
Mailing address: 1520D MSRB I, 1150 W. Medical Center Dr., Ann Arbor, MI, 48109-5680
Email address: esnitkin@umich.edu
ORCID: https://orcid.org/0000-0001-8409-278X

Abstract

To combat antibiotic resistance, it is critical to improve our understanding of how new resistant strains emerge and spread. An antibiotic resistance threat of critical priority is the epidemic ST258 strain of carbapenem-resistant *Klebsiella pneumoniae* (CRKP). Here, we studied the spread of resistance among ST258 to an antibiotic of last resort, colistin, by tracking its evolution across 21 U.S. long-term acute care hospitals over the course of a year. Phylogenetic analysis suggested that a significant cost was associated with colistin resistance in most cases, as resistance emergence was common but resistance variants were rarely transmitted. The high cost of resistance was further supported by the observation that several of the resistance variants that were transmitted had acquired secondary variants that reverted the strain to colistin susceptibility. The exceptions to the general pattern of instability associated with colistin resistance were two large clusters of resistant strains in one ST258 clade II sublineage (clade IIB) present across 11 of the 12 sampled Southern California hospitals. Quantification of transmission fitness in the healthcare environment using time-scaled haplotypic density

indicated that while resistant isolates from other clades were less fit than their susceptible counterparts, clade IIB resistant isolates were more fit. Overlaying patient clinical data suggested that the increased fitness of colistin-resistant clade IIB isolates is in part driven by a lineage defining variant that increased clade IIB's association with patient subpopulations who were more likely to be treated with colistin. These results show that a favorable genetic background and sustained selective pressure led to the emergence and spread of a colistin-resistant ST258 sublineage across a regional healthcare network. More broadly, these findings highlight the utility of integrating pathogen genomic and corresponding clinical data from regional healthcare networks to detect and understand the origin and dissemination of antibiotic resistance threats.

Introduction

Multidrug resistant organisms (MDROs) pose a significant threat to public health due to uncontrolled transmission and dwindling treatment options.[1] Of greatest concern are epidemic MDRO lineages that have become adapted to healthcare settings and continually gain resistance due to selective antibiotic pressure.[2,3] As resistance evolution continues to outpace our ability to develop new therapies,[3] there is a critical need to improve our understanding of the forces driving the emergence and spread of resistance so that we can maintain the efficacy of last-line antibiotics.

When *de novo* evolution of antibiotic resistance occurs in an individual, there is often little to no subsequent transmission to others due to the fitness cost of maintaining resistance in the absence of antibiotic exposure.[4,5] However, in some cases this barrier is overcome, as evidenced by the proliferation of epidemic resistance clones.[5,6] This may happen when the resistance variant or gene has an inherently low fitness cost and is thus maintained in the population even in the absence of antibiotic exposure.[7] Alternatively, the fitness cost of a given resistance element may depend on the genetic background in which it emerges, leading to dissemination of these elements only in certain lineages. Indeed, experimental evidence has shown that the genetic background of a strain interacts with resistance determinants to influence the fitness of the resistant strain,[8,9] highlighting the importance of historical evolutionary events in determining the potential ability of a resistant strain to spread. While the preferential spread of resistance elements in specific strains lends support to the importance of genetic background, in most instances it is unclear what the underlying epistatic interactions are that lead to these different fitness effects.[5] Deciphering how the genetic background of a clinical strain influences transmissibility can provide insight into how fit strain/resistance combinations evolve and disseminate in the healthcare setting.

Carbapenem-resistant *Klebsiella pneumoniae* (CRKP) is a global threat as it is resistant to nearly all available antibiotics and is associated with high mortality rates.[1] One of the few remaining treatment options for CRKP is the antibiotic colistin.[10] Concerningly, the viability of colistin as a treatment option moving forward is threatened by the frequent emergence of colistin resistance during treatment with this drug.[11,12] However, despite the accessibility of these resistance variants to CRKP, their spread to other patients is rare, with most cases even

within a single healthcare setting stemming from parallel evolution.[12] Here, we document the emergence and spread of multiple colistin resistance variants in CRKP sequence type (ST) 258 across a regional healthcare network. Through integration of genomic and clinical data, we show how resistance spread was enhanced in genetic backgrounds that mitigate the fitness cost of resistance which, when combined with sustained selective pressure, enabled regional dissemination of colistin-resistant lineages.

Results

Most resistant isolates contain variants in known resistance genes.

Antimicrobial susceptibility testing revealed that 118/337 (35%) CRKP ST258 isolates were resistant to colistin, and that this resistance was found consistently across long-term acute care hospitals (LTACHs) over the yearlong study (**Fig S1**). To identify likely resistance variants, we first searched all isolates for *mcr*-containing mobile genetic elements and variants in canonical chromosomal genes known to confer resistance (pmrA, pmrB, phoP, phoQ, crrA, crrB, and mgrB). As expected due to the timeframe and locations in which the isolates were collected (July 2014 to August 2015 in the U.S.),[13] we did not find *mcr* genes in any of the isolates. However, the majority of colistin resistance we observed (103/118, 87%) could be explained by known and putative resistance variants in canonical resistance genes (**Fig S2**; **Table S1**; see methods for details on identification of resistance variants).

As not all resistance in our dataset could be explained by variants in known resistance genes, we next performed a genome-wide association study (GWAS) on the isolates with unknown resistance determinants to identify putative non-canonical resistance-conferring variants. We identified two additional putative resistance genes in this way: qseC and phosphotransferase system sugar transporter subunit IIB (**Fig S2**; **Table S1**). These variants explain 6/15 (40%) of the unknown colistin resistance in our dataset. Notably, qseC, which explained resistance in 4 isolates, has been shown to confer colistin resistance in an experimental evolution study using clinical isolates.[14] We were unable to identify putative colistin resistance-conferring variants for the other nine isolates. For all subsequent analyses, we defined resistance genes as the set of canonical and GWAS-identified resistance genes.

Epistatic interactions appear to influence resistance in isolates with more than one variant in resistance genes.

Next, we investigated the relationship between the number of variants in resistance genes and the level of colistin resistance. While the majority of isolates with one variant in a resistance gene were resistant, the majority of isolates with two or more variants in resistance genes were susceptible (**Fig 1A**). To better understand how the presence of multiple variants in resistance genes influenced the extent of resistance, we explored the relationship between the number of variants in resistance genes and minimum inhibitory concentration (MIC). We found that resistant isolates with two or more variants in resistance genes tended to have higher MICs on average than those with one or fewer (median MIC of 24 vs. 8, $p = 0.001$; **Fig 1B**). In contrast to resistant isolates where having two or more variants in resistance genes was associated with increased MIC, susceptible isolates with two or more variants in resistance genes were more

susceptible on average than those with one or fewer (median MIC of 1 vs. 0.5, p = 0.0002). Furthermore, in 37/45 (82%) cases where susceptible isolates harbored multiple variants in resistance genes, at least one of the variants had been previously shown to confer resistance. These findings suggest that, while variants in resistance genes often increase MIC as expected, there also exist epistatic interactions among variants in these genes that influence their impact on resistance phenotypes.

Colistin resistance exhibits patterns of *de novo* evolution, onward dissemination, and reversion to susceptibility.

To better understand the origin and fate of colistin resistance variants, we investigated the phylogenetic relationship among colistin-resistant and susceptible isolates (**Fig 2**; **Fig S3**). Visualization of resistance on the phylogeny revealed a striking dichotomy between a clade II sublineage (defined as clade IIB) and the other ST258 sublineages (clades I and IIA). Clade IIB contained large clusters of colistin-resistant isolates, while the other two sublineages largely exhibited sporadic parallel evolution of resistance. Therefore, while 28/118 (24%) isolates acquired colistin resistance from putative *de novo* resistance evolution events, two clonal expansions in clade IIB across 11/12 LTACHs in Southern California (10 in the Los Angeles area and 1 in San Diego; **Fig S4**) accounted for over half (69/118, 58%) of resistance in this isolate collection. These two clonal expansions were associated with a nonsense mutation in mgrB (Gln30*) and a missense mutation in phoQ (Thr244Asn), respectively. Notably, these same types of variants occurred in the sporadic colistin-resistant isolates (**Fig 2**; **Fig S5**), suggesting that these specific resistance variants were likely not inherently more fit. Within these clonal expansions, many of the isolates appeared to regain susceptibility to colistin via the accumulation of additional variants in resistance genes. These putative reversion mutations usually occurred within the same gene or in a downstream gene of the molecular pathway to resistance (**Fig 2**; **Fig S6**), and were sometimes followed by re-acquisition of resistance through a different mechanism (**Fig 2**; **Fig S5**).

Resistant strains in clade IIB are more fit than their susceptible non-revertant counterparts.

Next, we were interested in understanding the relative transmission fitness of susceptible, resistant, and revertant isolates in the different ST258 sublineages. In particular, we hypothesized that the resistance variants in clade IIB had a decreased associated fitness cost, and thus were able to disseminate more widely. To estimate transmission fitness in the context of the healthcare environment we applied an analytic approach to quantify the epidemic success of isolates in our patient population. In particular, we took advantage of the comprehensive nature of our sampling and used time-scaled haplotypic density (THD) to quantify the extent of relative spread of each genotype.[15] Applying the THD metric, we observed significant differences in fitness effects of resistance variants in clade IIB versus the other clades (**Fig 3**). In clades I and IIA we observed evidence of a significant fitness cost for resistance variants, as resistant clusters and susceptible revertants were less fit than susceptible non-revertants from those clades (R cluster: p = 0.01; S revertant: p = 0.0003). We did not observe a fitness cost for resistant singletons that did not transmit to others (p = 0.18), which may be because these strains very recently evolved resistance and are thus still closely related to their susceptible ancestor strains. In contrast, resistance variants in clade IIB were associated with a significant

fitness benefit, with resistant clusters and susceptible revertants from clade IIB being more fit than susceptible non-revertants from that clade (R cluster: p = 0.03; S revertant: p = 0.02). Taken together, these findings suggest that resistance variants in isolates from clade IIB conferred a fitness advantage, which may account for the independent emergence and spread of two different resistance alleles in this clade. In contrast, the accumulation of variants in resistance genes in the other clades appears to be associated with a transmission fitness cost, which is consistent with their limited clonal spread.

Clade IIB isolates have more colistin exposure compared to isolates from other clades.
Lastly, we were interested in investigating whether colistin use in study facilities influenced the emergence and spread of stable resistance in clade IIB. Supporting the role of colistin use in the spread of colistin resistance, we found a positive association between having a clade IIB isolate and both prior exposure to colistin (15/120, 12.5% in clade IIB vs. 8/217, 3.7% in other clades; Fisher's exact p = 0.003), as well as treatment with colistin (41/120, 34.2% in clade IIB vs. 39/217, 18.0% in other clades; Fisher's exact p = 0.001). We hypothesize that this association between colistin use and clade IIB is mediated by our previously reported observation that clade IIB is enriched in respiratory isolates[16] and the frequent use of colistin to treat respiratory tract infections during the study period (**Fig 4A**), due to limited available antibiotic options at that time. This hypothesis is supported by a slight enrichment in colistin treatment for clade IIB respiratory isolates compared to non-respiratory isolates (31/83, 37.3% for respiratory vs. 10/37, 27.0% for non-respiratory; p = 0.3), and a significant enrichment across all clades (53/183, 29% for respiratory vs. 27/154, 17.5%; p = 0.01). Furthermore, in addition to potentially influencing the spread of colistin resistance, the extensive use of colistin to treat infections in a setting of clonally disseminating colistin resistance led to 36/118 (30.5%) patients with colistin-resistant isolates being treated with this antibiotic, even though the treatment was likely ineffective (**Fig 4B**).

In putting our past and present observations together, we noted that a defining feature of clade IIB was disruption of the putative O-antigen glycosyltransferase *kfoC*.[16] Thus, we hypothesize that disruption of *kfoC* increased the affinity of clade IIB for the respiratory tract, increased exposure to colistin, and decreased the transmission fitness cost of colistin resistance, allowing colistin-resistant clade IIB strains to spread across a regional healthcare network (**Fig 5**).

Discussion

The means by which colistin resistance evolves and disseminates in the healthcare setting has important implications for antibiotic stewardship and infection prevention strategies. Here, we used genomic and clinical data to track the origin and fate of colistin resistance variants across a comprehensive regional sample of clinical CRKP ST258 isolates. Our analysis identified a sublineage whose genetic background appears to decrease the fitness cost of resistance, thus allowing colistin-resistant strains to spread among regional healthcare facilities.

Examination of colistin resistance variants within the comprehensive and longitudinal context of circulating CRKP strains allowed for inferences into their functional impact and clinical

significance. We discovered that resistant isolates with multiple variants in resistance genes have increased resistance, but susceptible isolates with multiple variants in resistance genes have decreased resistance. Additionally, we identified several instances where a known resistance variant was present in susceptible isolates, likely due to a suppressor variant in either the same resistance gene or a downstream resistance gene. This finding has two important implications. First, known resistance variants occurring in both resistant and susceptible isolates complicates the computational identification of novel resistance determinants using GWAS. Possible solutions include increasing sample size or removing isolates with known resistance variants. Second, the presence of a colistin resistance variant cannot necessarily be equated with *in vitro* phenotypic resistance. Clinically, this suggests that testing for resistance variants may not be a substitute for testing for phenotypic resistance, as harboring a resistance variant may not preclude the effectiveness of colistin treatment. In addition to being phenotypically susceptible, revertants that harbor resistance variants may in general have more difficulty becoming resistant once more, which may lend credence to treating these patients with colistin.[17]

Of particular genomic and clinical interest are two independent instances of emergence and spread of colistin resistance variants in clade IIB, in stark contrast to the sporadic emergence of resistance variants in the other clades. Notably, the types of resistance variants present in clade IIB also occurred in other clades, implying that the nature of the variant itself is not what allowed the clade IIB variants to spread. Instead, our analysis of fitness using THD suggests that these colistin resistance variants were able to spread in this clade because the genetic background of the clade reduced the fitness cost of the variants. We previously showed that a defining feature of clade IIB is that the majority of isolates contain a disruption in the lipopolysaccharide (LPS) O locus gene *kfoC*,[16] a putative glycosyltransferase. Given the key role that LPS plays in both colistin-mediated killing and colistin resistance,[11] it is possible that inactivation of *kfoC* facilitates the spread of colistin resistance by altering the outer membrane in a way that reduces the fitness cost of resistance-conferring LPS modifications. Furthermore, we hypothesize that *kfoC* may indirectly mitigate the fitness cost of colistin resistance by increasing the association between clade IIB and patient populations in which colistin use is more frequent. In particular, clade IIB isolates are significantly enriched in both association with the respiratory tract and colistin exposure. Thus, increased colistin use in clade IIB may have both set the stage for the emergence of multiple instances of stable resistance in this clade and provided a continued selective pressure to maintain resistance. These findings highlight the importance of continued real-time monitoring of resistance to improve antibiotic stewardship and decrease the likely ineffective use of colistin to treat colistin-resistant CRKP strains.

One strength of our study is that we have comprehensive sampling of clinical CRKP isolates from LTACHs over the course of a year. This not only allowed us to capture putative transmission events of colistin resistance, but also allowed us to estimate the fitness of different ST258 sublineages. Additionally, whole-genome sequencing of study isolates permitted us to interrogate not only variants in known resistance genes, but also investigate putative resistance-conferring variants in other genes using GWAS. Furthermore, our access to clinical information about colistin use allowed us to identify a relationship between this clinical variable and different

ST258 sublineages, providing us with insight into the selective pressures of colistin resistance evolution and spread in patients.

Our study also has several limitations. First, we do not have rectal surveillance cultures, which limits our ability to fully capture the population of colistin-resistant and susceptible CRKP in the LTACHs studied, and thus to fully define spread. While this could bias our clinical collection if certain strains are more likely to present clinically and be cultured at extraintestinal sites, the strains we have included are the most clinically relevant since they were identified in clinically derived samples. Also, we cannot be certain that our classification of isolates into different categories of colistin susceptibility and resistance is entirely accurate. For instance, putative *de novo* evolution of resistance in resistant singletons may not have occurred in the patient we sampled. However, our comprehensive sampling of clinical isolates from the facilities in the study provides some confidence that resistance in resistant singleton isolates did in fact evolve in that patient. Another limitation of our study is that we do not have information about colistin exposure prior to the patient entering the healthcare facility, and we do not have information about the dose or duration of colistin use. Even with this limitation, we were able to identify significant associations between colistin use and different ST258 sublineages.

In conclusion, we observed distinct dynamics of colistin resistance evolution in different CRKP ST258 sublineages that appear to be due to differences in the fitness cost of colistin resistance dependent on the genetic background of the strain. In particular, we identified an emerging ST258 sublineage that is more fit and more amenable to maintaining colistin resistance than other ST258 strains. This is of particular concern due to the already limited treatment options for CRKP, and therefore merits further surveillance to determine the extent of spread. Furthermore, our findings highlight the value of using high-resolution genomic analysis to inform antibiotic stewardship and the importance of surveillance and monitoring of MDROs for continued evolution and adaptation to the healthcare environment.

Methods

Study isolates and metadata

We used whole-genome sequences of clinical CRKP isolates from a prospective longitudinal study in 21 U.S. LTACHs over the course of a year (BioProject accession no. PRJNA415194).[18] Isolates and metadata were collected, and isolates were sequenced, as described in Han *et al*.[18] Prior patient use of colistin in the facility from which the isolate was taken (up to 30 days before the isolate collection date), as well as colistin use for empiric and definitive treatment, was extracted from the electronic health record. We only used information on colistin, not polymyxin B, for all analyses.

Antibiotic susceptibility testing

Custom Sensititre broth microdilution panels (ThermoFisher Scientific, Waltham, MA) were used to determine colistin minimum inhibitory concentrations (MIC). We used Clinical & Laboratory Standards Institute (CLSI) guidelines for interpretations with an MIC cutoff of ≥4 characterized as resistant.[19] CLSI guidelines classify isolates with an MIC ≤2 as being of intermediate

resistance and therefore are technically considered non-resistant using this interpretation; however, the European Committee on Antimicrobial Susceptibility Testing defines isolates with an MIC ≤2 as susceptible.[20] For clarity, we chose to use the term susceptible for isolates with an MIC ≤2.

Isolate selection

Multi-locus sequence types were called using ARIBA.[21] Over 90% of CRKP isolates collected over the course of the study belonged to ST258; therefore, we focus all of our analyses on this sequence type. We ordered the isolates by resistance status followed by collection date. For all analyses except GWAS, only the first patient ST258 isolate from this ordered list was used, thus prioritizing resistant isolates over susceptible isolates. Sample sizes were too small to glean insights into within-host colistin resistance evolution.

Single nucleotide variant calling, indel calling, and phylogenetic tree reconstruction

Variant calling was performed with a customized variant calling pipeline (https://github.com/Snitkin-Lab-Umich/variant_calling_pipeline/) as follows. The quality of sequencing reads was assessed with FastQC v0.11.9,[22] and Trimmomatic v0.39[23] was used for trimming adapter sequences and low-quality bases. Single nucleotide variants (SNVs) were identified by (i) mapping filtered reads to the ST258 KPNIH1 reference genome (BioProject accession no. PRJNA73191) using the Burrows-Wheeler short-read aligner (bwa v0.7.17),[24] (ii) discarding polymerase chain reaction duplicates with Picard v2.24.1,[25] and (iii) calling variants with SAMtools and bcftools v1.9.[26] Variants were filtered from raw results using VariantFiltration from GATK v4.1.9.0[27] (QUAL >100; MQ >50; >=10 reads supporting variant; and FQ < 0.025). Indels were called using the GATK HaplotypeCaller[28] with the following filters: root mean square quality (MQ) > 50.0, GATK QualbyDepth (QD) > 2.0, read depth (DP) > 9.0, and allele frequency (AF) > 0.9. In addition, a custom Python script was used to filter out (mask) variants in the whole-genome alignment that were: (i) SNVs <5 base pairs (bp) in proximity to indels, (ii) in a recombinant region identified by Gubbins v2.3.4,[29] in a phage region identified by the Phaster web tool[30] or (iii) they resided in tandem repeats of length greater than 20bp as determined using the exact-tandem program in MUMmer v3.23.[31] This whole-genome masked variant alignment was used to reconstruct a maximum likelihood phylogeny with IQ-TREE v1.6.12[32] using the general time reversible model GTR+G and ultrafast bootstrap with 1000 replicates (-bb 1000).[33]

Insertion calling

Large insertions relative to the reference genome were called using panISa v0.1.4.[34]

Variant preprocessing

We preprocessed variants to include multiallelic sites and used the major allele method for variant binarization, as described in Saund *et al*.[35] We used SnpEff to predict the functional impact of single nucleotide variants and indels (high, moderate, low, modifier).[36] Additionally, we considered all insertions in or upstream of genes as high impact, and those downstream of

genes as moderate impact. Only high and moderate impact variants were included in downstream analyses.

Identification of putative resistance variants

Known resistance genes

We consider the following genes known (canonical) resistance genes: mgrB, phoP, phoQ, crrA, crrB, pmrA, and pmrB.[37] We group variants in known resistance genes into the following categories, in order of decreasing confidence:

1. Known: experimentally confirmed resistance variants, including loss-of-function variants in mgrB.[37–40]
2. Known site: the variant occurs at a nucleotide site where there is an experimentally confirmed resistance variant, but that specific amino acid change has not been experimentally confirmed.[41–44]
3. Putative: nonsynonymous or disruptive variants in known resistance genes where >60% of the variants are present in resistant isolates.

We define the set of all of these as resistance variants. Significant differences between MIC of different isolates were determined using Wilcox tests.

Genome-wide association study

We performed a burden test using treeWAS v1.0,[45] a convergence-based GWAS method, as our sample size is relatively small and resistance is a very convergent phenotype. A burden test increases power to detect resistance genes, and convergence-based methods control for the structure of the phylogeny more than mixed model methods. For GWAS, we included only isolates with no mutations in known resistance genes as we were most interested in identifying novel resistance genes, and because the entire dataset contained a number of susceptible isolates with known resistance variants that would have confounded the analysis. We used pyseer v1.3.6[46] to calculate the number of unique patterns and determine a p-value cutoff ($p < 9.47e-5$). Putative resistance genes were considered ones that were identified as significant by treeWAS, had more than one convergence event on the phylogeny, and >60% of all the variants in the gene were found in resistant isolates. Variants within these genes were considered putative resistance variants using the same definition as for putative resistance variants in known resistance genes. Using these requirements, we included four putative resistance variants from two putative resistance genes based on the GWAS results.

Identification of putative suppressor variants

We define putative suppressor variants as those in resistance genes where >60% of isolates with the variant are susceptible and contain a resistance variant.

Determination of isolate resistance group

We assigned each isolate to one of four categories (**Fig S3**):

1. Resistant singletons: resistant isolates that do not cluster on the phylogeny, or that cluster but contain distinct resistance variants.

2. Resistant clusters: resistant isolates that cluster on the phylogeny and contain the same resistance variant. Additionally, if a cluster on the phylogeny has unknown resistance variants, we also defined it as a resistant cluster.
3. Susceptible revertants: susceptible isolates that contain a putative resistance-conferring variant.
4. Susceptible non-revertants: susceptible isolates that do not contain a putative resistance-conferring variant.

For the purposes of identifying resistant clusters that include susceptible revertants, we classified an isolate as "quasi-resistant" if it contained a known resistance variant and/or was resistant. Clusters of "quasi-resistant" isolates were identified using the get_clusters function in regentrans (https://github.com/Snitkin-Lab-Umich/regentrans) with a pureness of 1. Each isolate in these clusters was then denoted as a resistant cluster or susceptible revertant isolate depending on their corresponding phenotype.

Calculation of transmission fitness using time-scaled haplotypic density

Even though there is a clear fitness cost associated with colistin resistance in the healthcare setting as evidenced by our finding that the vast majority of resistant strains do not spread, previous experimental approaches to studying transmission fitness of colistin-resistant CRKP have found no fitness cost to resistance.[47,48] Therefore, we define transmission fitness in the healthcare setting as the epidemic success of the strain based on time-scaled haplotypic density (THD).[15] We calculated the THD for each isolate with the R package thd v1.0.1[15] using the KPNIH1 reference genome length (5,394,056 base pairs), a mutation rate of 1.03e-6,[49] the time scale parameter, and a timescale of one year. Pairwise single nucleotide variant distances were calculated with the dist.dna function in ape v5.4.1[50] (model = 'N', pairwise.deletion = TRUE). Significant differences between THD of different isolates were determined using Wilcox tests.

Calculation of colistin use significance

We used Fisher's exact tests to calculate the significance of the association between various groups of isolates and colistin use.

Data visualization

We performed all data visualization in R v4.0.2[51] using the following packages: tidyverse v1.3.0,[52] ggtree v2.2.4,[53,54] pheatmap v1.0.12,[55] ggplotify v0.0.5,[56] and cowplot v1.1.0.[57] Code and data corresponding to the manuscript figures can be found on GitHub (https://github.com/Snitkin-Lab-Umich/ltach-crkp-colistin-ms).

Acknowledgements

We gratefully acknowledge Kindred Healthcare for their assistance in collecting data and isolates used in this study. We also thank the patients and staff of the long-term acute-care hospitals for their gracious participation in this study, and Mary Hayden and Robert Weinstein for critical review of the manuscript.

Funding

This research was supported by a CDC Cooperative Agreement FOA no. CK16-004-Epicenters for the Prevention of Healthcare Associated Infections and the National Institutes of Health R01 AI139240-01 and 1R01 AI148259-01. Z.L. received support from the National Science Foundation Graduate Research Fellowship Program under grant no. DGE 1256260. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The funding bodies had no role in the design of the study or collection, analysis, and interpretation of data, or in writing the manuscript.

Conflicts of interest

J.H.H. is currently an employee of, and holds shares in, the GSK group of companies.

Author contributions

E.S.S., J.H.H., E.L., and Z.L. conceptualized the study and acquired funding to support the project. K.A. performed MIC testing. Z.L, A.P., S.C., and D.C. performed data curation and formal analysis. Z.L. performed investigation and visualization. E.S.S. and Z.L. developed methodology. E.S.S. supervised the project. Z.L. and E.S.S. wrote the original draft. All authors reviewed and edited the manuscript.

References

1. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases* (2017) doi:10.1016/S1473-3099(17)30753-3.
2. MacLean, R. C. & Millan, A. S. The evolution of antibiotic resistance. *Science* **365**, 1082–1083 (2019).
3. *Antimicrobial resistance: global report on surveillance*. (World Health Organization, 2014).
4. The cost of antibiotic resistance from a bacterial perspective. *Drug Resistance Updates* **3**, 237–245 (2000).
5. Baker, S., Thomson, N., Weill, F.-X. & Holt, K. E. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science* **360**, 733–738 (2018).
6. Logan, L. K. & Weinstein, R. A. The Epidemiology of Carbapenem-Resistant Enterobacteriaceae: The Impact and Evolution of a Global Menace. *J Infect Dis* **215**, S28–S36 (2017).
7. Melnyk, A. H., Wong, A. & Kassen, R. The fitness costs of antibiotic resistance mutations. *Evol Appl* **8**, 273–283 (2015).
8. Vogwill, T., Kojadinovic, M. & MacLean, R. C. Epistasis between antibiotic resistance mutations and genetic background shape the fitness effect of resistance across species of Pseudomonas. *Proc Biol Sci* **283**, (2016).
9. Vogwill, T., Kojadinovic, M., Furió, V. & MacLean, R. C. Testing the Role of Genetic

Background in Parallel Evolution Using the Comparative Experimental Evolution of Antibiotic Resistance. *Molecular Biology and Evolution* **31**, 3314–3323 (2014).
10. Doi, Y. Treatment Options for Carbapenem-resistant Gram-negative Bacterial Infections. *Clinical Infectious Diseases* **69**, S565–S575 (2019).
11. Poirel, L., Jayol, A. & Nordmann, P. Polymyxins: Antibacterial Activity, Susceptibility Testing, and Resistance Mechanisms Encoded by Plasmids or Chromosomes. *Clin. Microbiol. Rev.* **30**, 557–596 (2017).
12. Macesic, N. *et al.* Emergence of Polymyxin Resistance in Clinical Klebsiella pneumoniae Through Diverse Genetic Adaptations: A Genomic, Retrospective Cohort Study. *Clinical Infectious Diseases* **70**, 2084–2091 (2020).
13. Skov, R. L. & Monnet, D. L. Plasmid-mediated colistin resistance (mcr-1 gene): three months later, the story unfolds. *Eurosurveillance* **21**, 30155 (2016).
14. Pitt, M. E. *et al.* Octapeptin C4 and polymyxin resistance occur via distinct pathways in an epidemic XDR Klebsiella pneumoniae ST258 isolate. *J. Antimicrob. Chemother.* (2018) doi:10.1093/jac/dky458.
15. Rasigade, J.-P. *et al.* Strain-specific estimation of epidemic success provides insights into the transmission dynamics of tuberculosis. *Scientific Reports* **7**, 45326 (2017).
16. Lapp, Z. *et al.* Patient and Microbial Genomic Factors Associated with Carbapenem-Resistant Klebsiella pneumoniae Extraintestinal Colonization and Infection. *mSystems* **6**, (2021).
17. Snitkin, E. S. *et al.* Genomic insights into the fate of colistin resistance and Acinetobacter baumannii during patient treatment. *Genome Res.* **23**, 1155–1162 (2013).
18. Han, J. H. *et al.* Whole-Genome Sequencing To Identify Drivers of Carbapenem-Resistant Klebsiella pneumoniae Transmission within and between Regional Long-Term Acute-Care Hospitals. *Antimicrobial Agents and Chemotherapy* **63**, (2019).
19. Weinstein, M. P. M100-performance standards for antimicrobial susceptibility testing, 30th edition. http://em100.edaptivedocs.net/GetDoc.aspx?doc=CLSI%20M100%20ED30:2020&sbssok=CLSI%20M100%20ED30:2020%20TABLE%202A&format=HTML&hl=colistin%20klebsiella (2020).
20. EUCAST: Clinical breakpoints and dosing of antibiotics. https://www.eucast.org/clinical_breakpoints/.
21. Hunt, M. *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom* **3**, (2017).
22. Andrews, S. *s-andrews/FastQC*. (2021).
23. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
24. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
25. *broadinstitute/picard*. (Broad Institute, 2021).
26. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
27. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303 (2010).

28. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 201178 (2018) doi:10.1101/201178.
29. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43**, e15–e15 (2015).
30. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**, W16–W21 (2016).
31. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biology* **5**, R12 (2004).
32. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* **32**, 268–274 (2015).
33. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol Biol Evol* **30**, 1188–1195 (2013).
34. Treepong, P. *et al.* panISa: ab initio detection of insertion sequences in bacterial genomes from short read sequence data. *Bioinformatics* **34**, 3795–3800 (2018).
35. Saund, K., Lapp, Z., Thiede, S. N., Pirani, A. & Snitkin, E. S. prewas: data pre-processing for more informative bacterial GWAS. *Microbial Genomics,* **6**, e000368 (2020).
36. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6**, 80–92 (2012).
37. Berglund, B. Acquired Resistance to Colistin via Chromosomal And Plasmid-Mediated Mechanisms in Klebsiella pneumoniae. *Infectious Microbes & Diseases* **1**, 10–19 (2019).
38. Olaitan, A. O., Morand, S. & Rolain, J.-M. Mechanisms of polymyxin resistance: acquired and intrinsic resistance in bacteria. *Front. Microbiol.* **5**, (2014).
39. Jayol, A. *et al.* High-Level Resistance to Colistin Mediated by Various Mutations in the crrB Gene among Carbapenemase-Producing Klebsiella pneumoniae. *Antimicrob Agents Chemother* **61**, (2017).
40. Cheng, Y.-H., Lin, T.-L., Lin, Y.-T. & Wang, J.-T. Amino Acid Substitutions of CrrB Responsible for Resistance to Colistin through CrrC in Klebsiella pneumoniae. *Antimicrobial Agents and Chemotherapy* **60**, 3709–3716 (2016).
41. Matamouros, S., Hager, K. R. & Miller, S. I. HAMP Domain Rotation and Tilting Movements Associated with Signal Transduction in the PhoQ Sensor Kinase. *mBio* **6**, (2015).
42. Yadavalli, S. S. *et al.* Functional Determinants of a Small Protein Controlling a Broadly Conserved Bacterial Sensor Kinase. *Journal of Bacteriology* **202**, (2020).
43. Pitt, M. *et al.* Multifactorial Chromosomal Variants Regulate Polymyxin Resistance In Extensively Drug-Resistant Klebsiella pneumoniae. (2017) doi:10.1101/134684.
44. Olaitan, A. O. *et al.* Worldwide emergence of colistin resistance in Klebsiella pneumoniae from healthy humans and patients in Lao PDR, Thailand, Israel, Nigeria and France owing to inactivation of the PhoP/PhoQ regulator mgrB: an epidemiological and molecular study. *International Journal of Antimicrobial Agents* **44**, 500–507 (2014).
45. Collins, C. & Didelot, X. A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLOS Computational Biology* **14**, e1005958 (2018).
46. Lees, J. A., Galardini, M., Bentley, S. D., Weiser, J. N. & Corander, J. pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* **34**,

4310–4312 (2018).
47. Cannatelli, A., Santos-Lopez, A., Giani, T., Gonzalez-Zorn, B. & Rossolini, G. M. Polymyxin resistance caused by mgrB inactivation is not associated with significant biological cost in Klebsiella pneumoniae. *Antimicrob Agents Chemother* **59**, 2898–2900 (2015).
48. Kidd, T. J. *et al.* A Klebsiella pneumoniae antibiotic resistance mechanism that subdues host defences and promotes virulence. *EMBO Mol Med* **9**, 430–447 (2017).
49. Bowers, J. R. *et al.* Genomic Analysis of the Emergence and Rapid Global Dissemination of the Clonal Group 258 Klebsiella pneumoniae Pandemic. *PLoS ONE* **10**, e0133727 (2015).
50. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
51. R: The R Project for Statistical Computing. https://www.r-project.org/.
52. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).
53. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28–36 (2017).
54. Yu, G., Lam, T. T.-Y., Zhu, H. & Guan, Y. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol* **35**, 3041–3043 (2018).
55. Kolde, R. *pheatmap: Pretty Heatmaps*. (2019).
56. Yu, G. *ggplotify: Convert Plot to 'grob' or 'ggplot' Object*. (2020).
57. Wilke, C. O. *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'*. (2019).

Figures

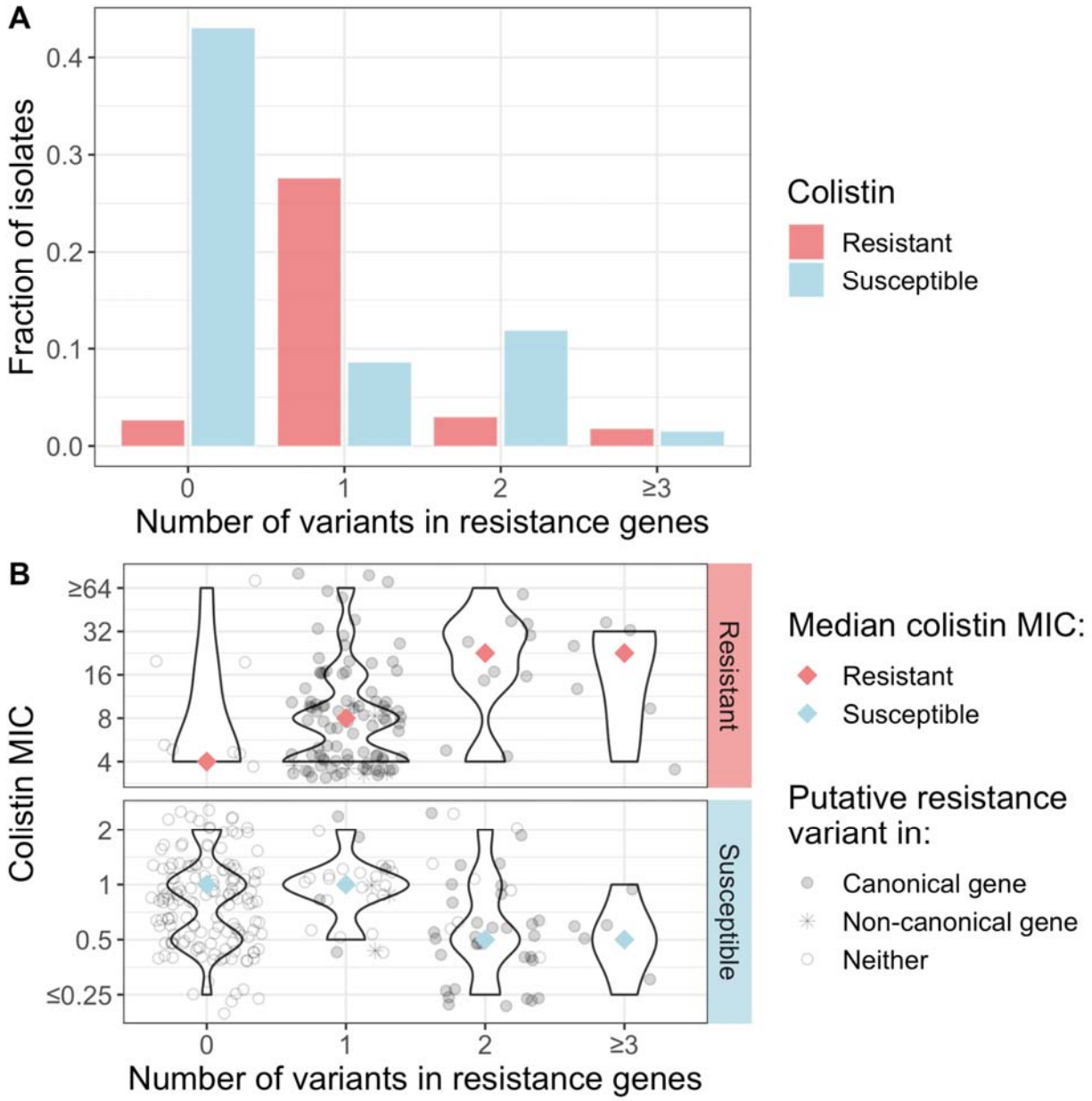

Figure 1: Epistatic interactions appear to influence the extent of colistin resistance in clinical CRKP isolates. (A) Most susceptible isolates have no variants in known resistance genes, while most resistant isolates have at least one variant in a known resistance gene. Many susceptible isolates also have two variants in known resistance genes, suggesting potential reversion events. (B) Colistin MIC split by total number of variants in resistance genes. Shapes indicate whether any of the variants are putative resistance variants (see methods for definition of resistance variants). On average, resistant isolates with two or more variants in resistance genes have a higher MIC than resistance isolates with one or fewer variants in resistance genes, and susceptible isolates with two or more variants in resistance genes have a lower MIC than susceptible isolates with one or fewer variants in resistance genes. Significance defined as Wilcox $p < 0.05$. MIC=minimum inhibitory concentration.

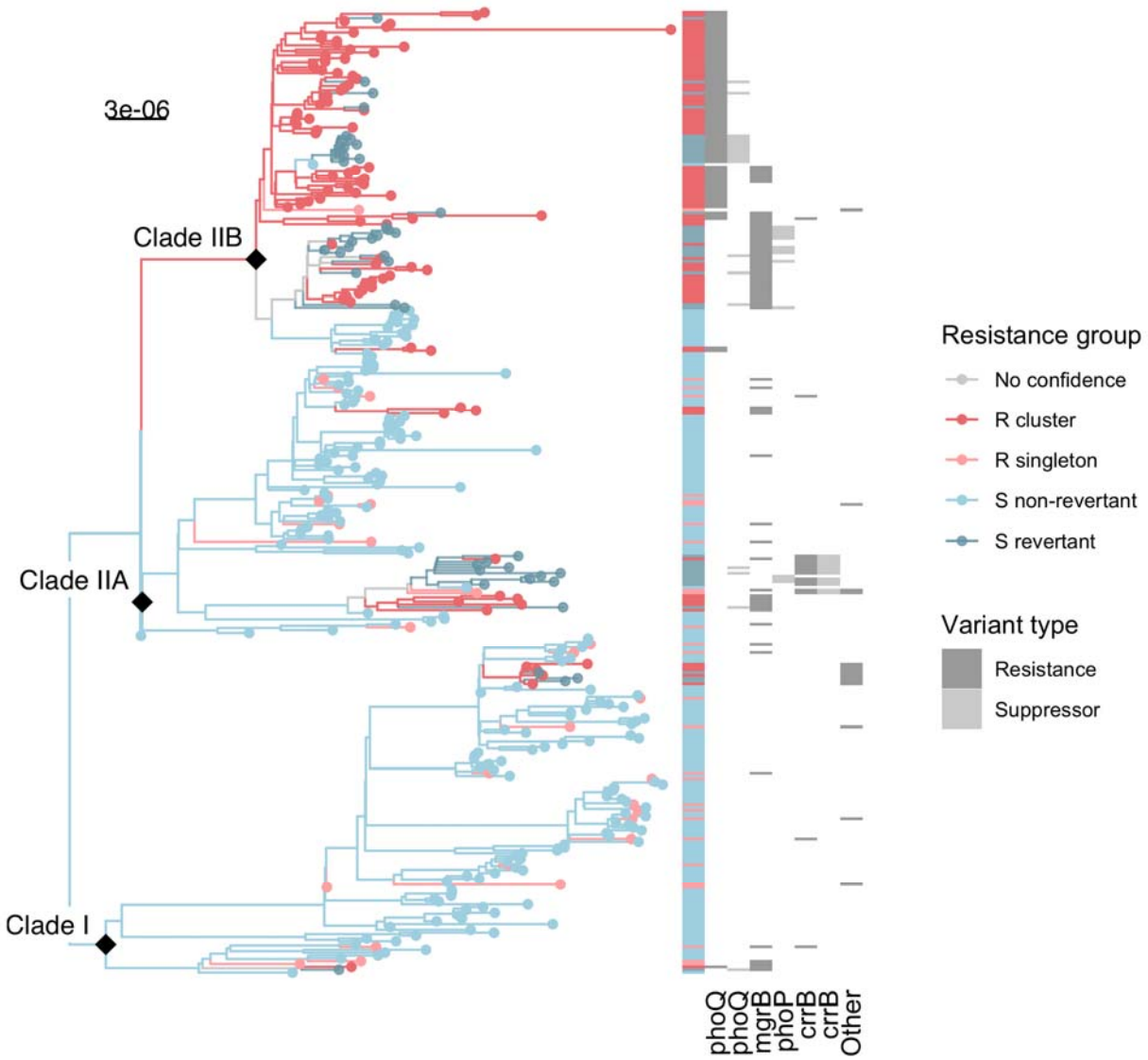

Figure 2: Distinct resistance evolution patterns exist in different CRKP ST258 sublineages. Clade IIB exhibits extensive dissemination of resistance variants, as well as several putative reversion events, while the other clades contain sporadic and less transmitted resistance variants. Each grey column of the heatmap is a gene, where all variants of a given type (resistance in dark grey, suppressor in light grey), are grouped for clarity. See Fig S5 for more details on individual variants. Scale bar is in substitutions per site per year. R=resistant; S=susceptible.

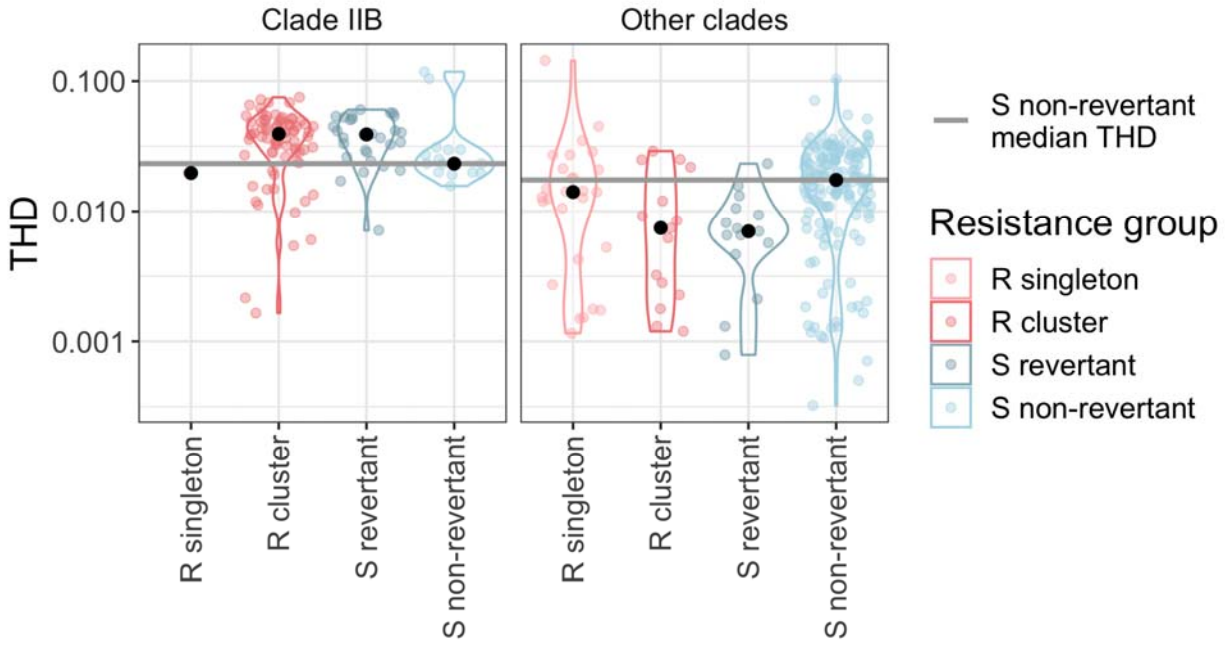

Figure 3: Colistin resistance does not impart a fitness cost in clade IIB strains. In clade IIB, resistant isolates are more fit than corresponding susceptible non-revertants. On the other hand, resistant isolates in other clades are less fit than susceptible non-revertants in those clades. Significance defined as Wilcox $p < 0.05$. R=resistant; S=susceptible; THD=time-scaled haplotypic density.

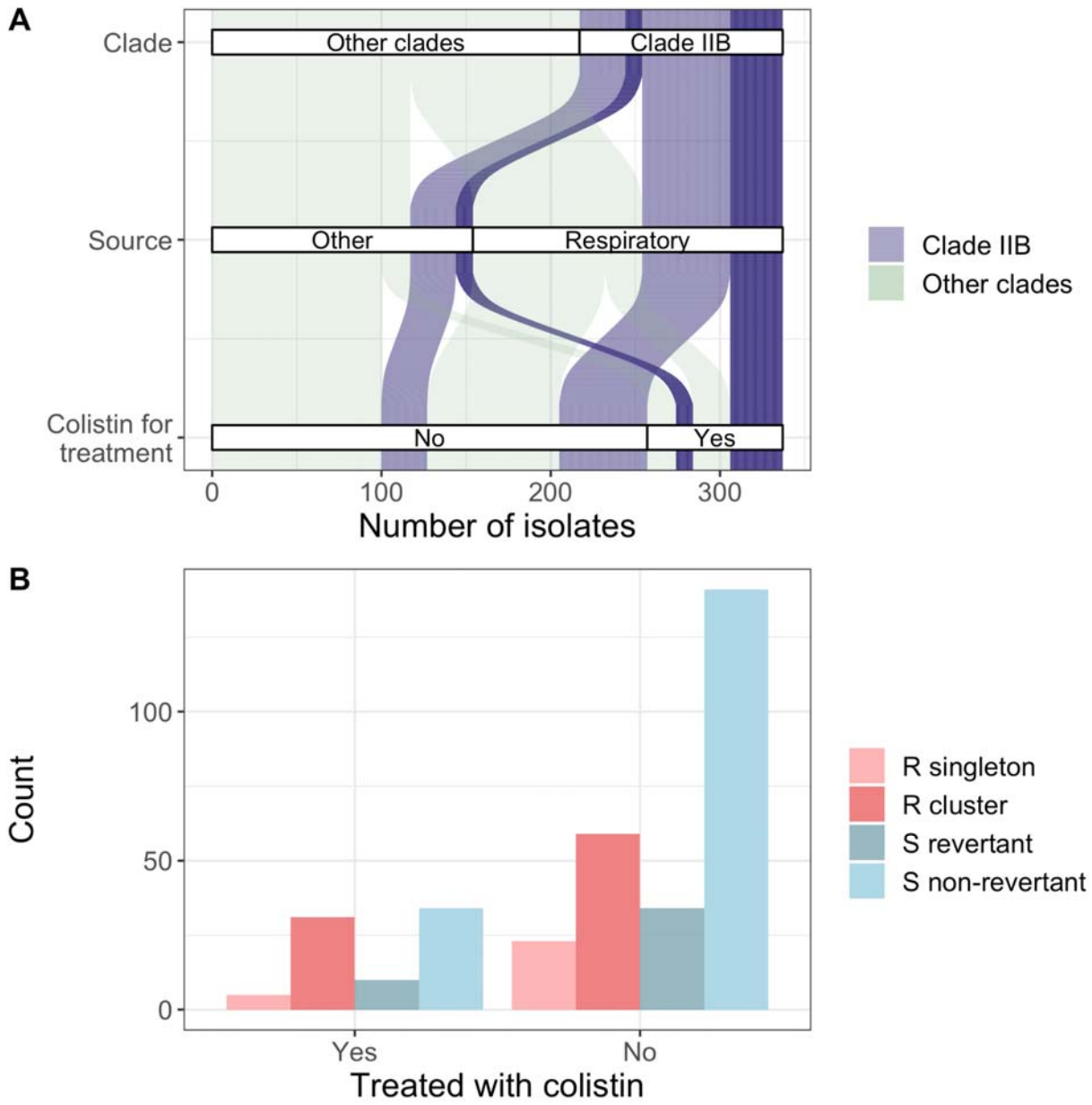

Figure 4: Patients with colistin-resistant isolates were treated with colistin. (A) The majority of clade IIB isolates are respiratory isolates, and many of the patients with those strains were treated with colistin (empiric and/or definitive). Differences in transparency are aesthetic to draw attention to clade IIB isolates treated with colistin. (B) Many colistin-resistant isolates were treated with colistin.

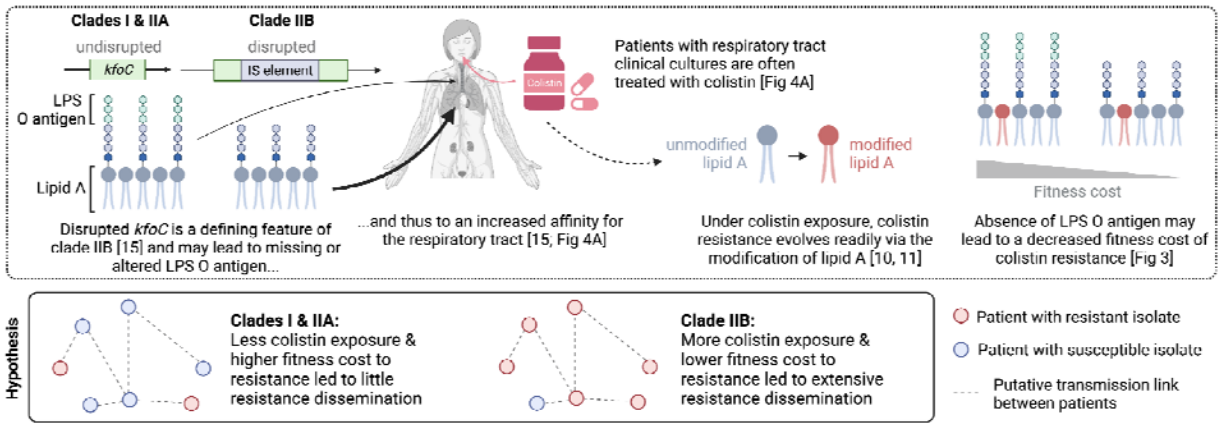

Figure 5: Disruption of the putative O-antigen glycosyltransferase *kfoC* may lead to decreased transmission fitness cost of colistin resistance. Created with BioRender.com

Supplementary figures

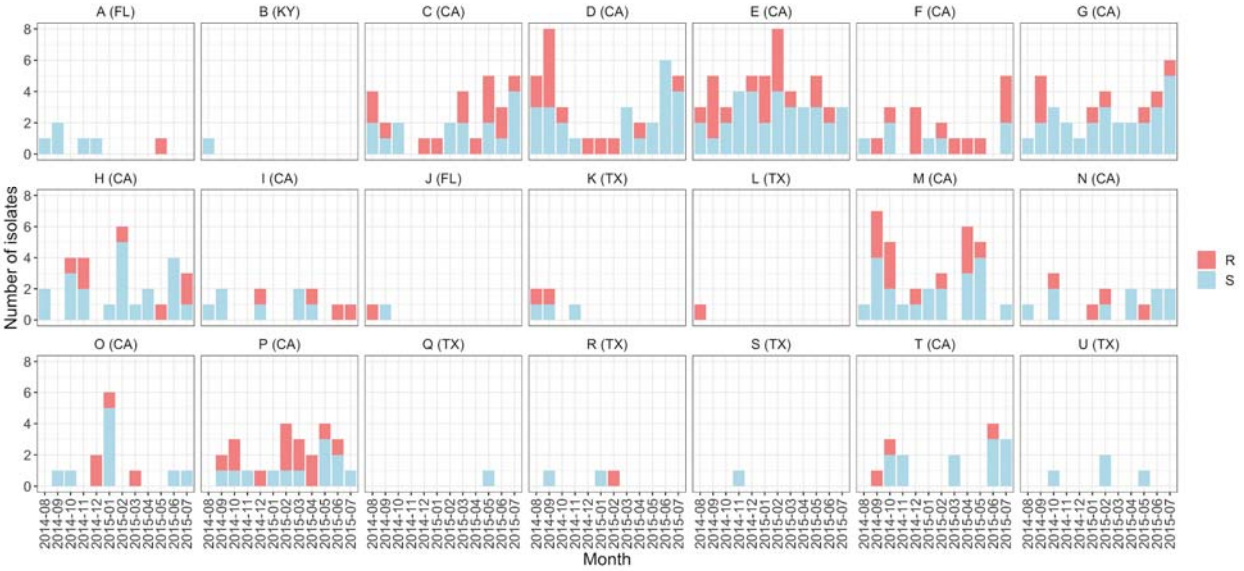

Figure S1: Colistin resistance occurs across time and geography. R=resistant, S=susceptible.

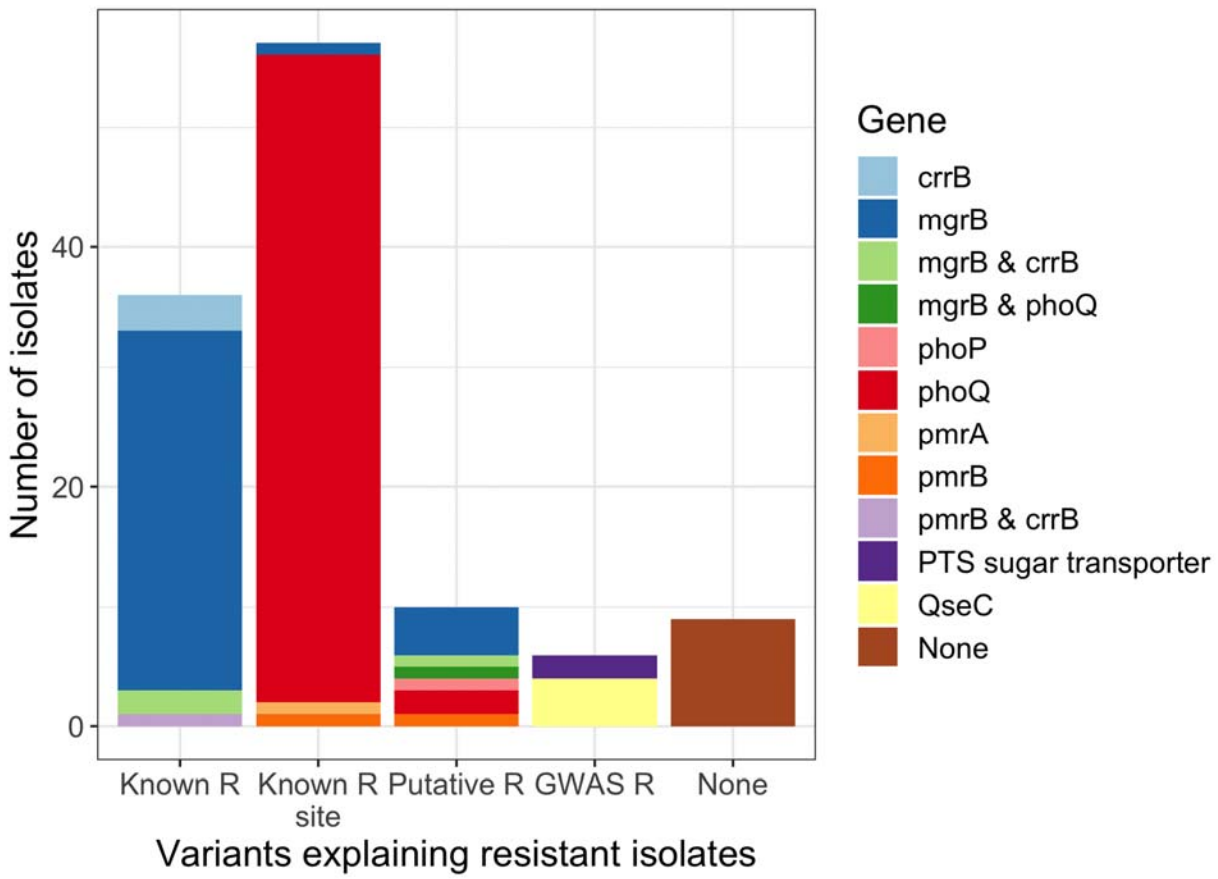

Figure S2: The majority of colistin resistance can be explained by variants in known resistance genes.

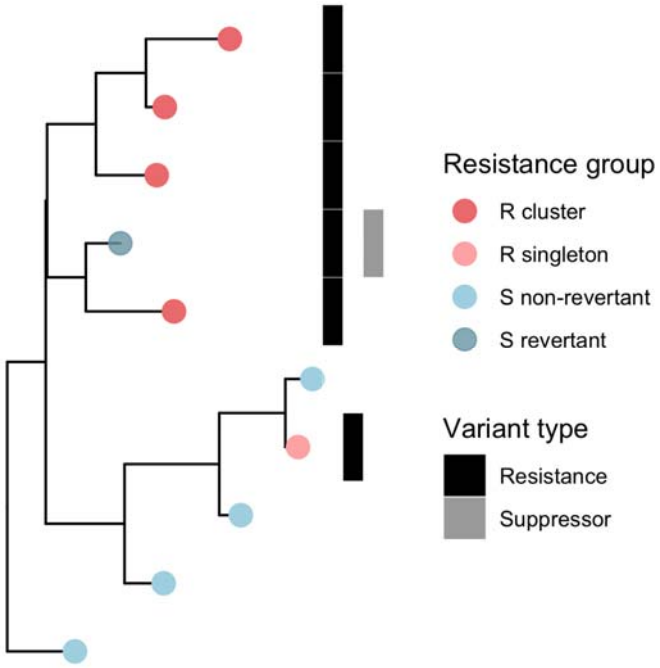

Figure S3: Toy example of how resistance groups are defined. Note that susceptible revertants are defined by the presence of a resistance variant, rather than a putative suppressor variant. See methods for more details. Heatmap columns are each variants found in a resistance gene. R=resistant, S=susceptible.

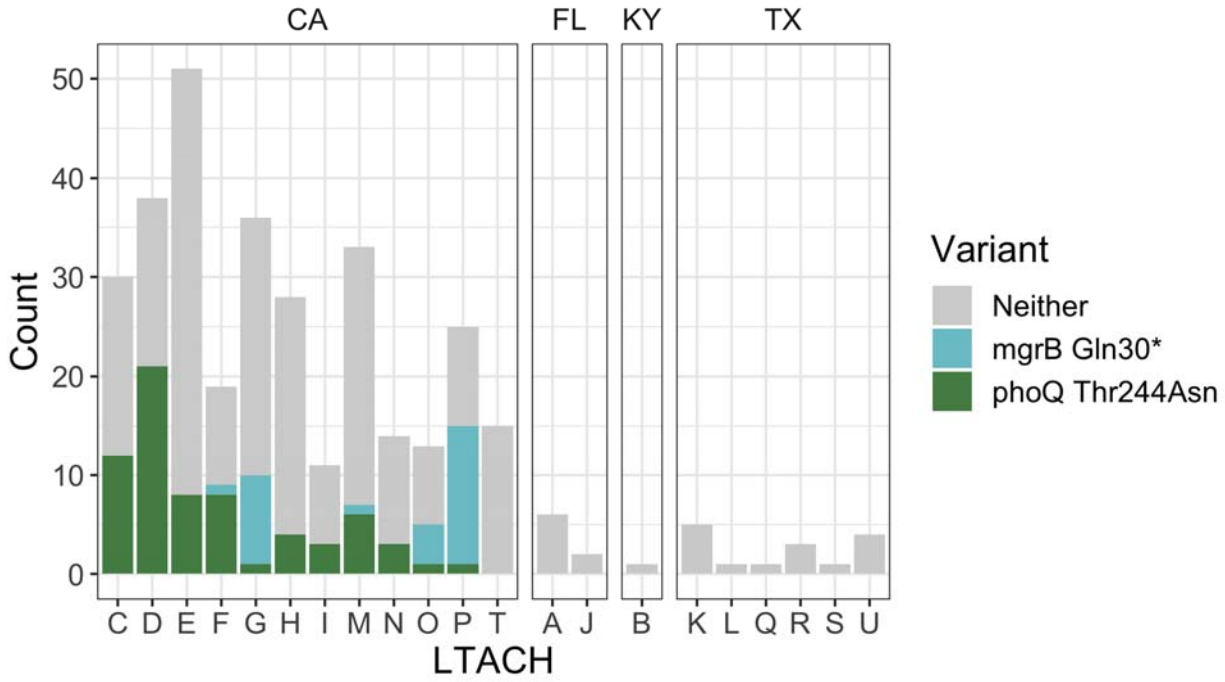

Figure S4: The two clonally expanded resistance variants occur across multiple LTACHs in California. 10/11 LTACHs in the Los Angeles area have the variant, as well as one LTACH in the San Diego area.

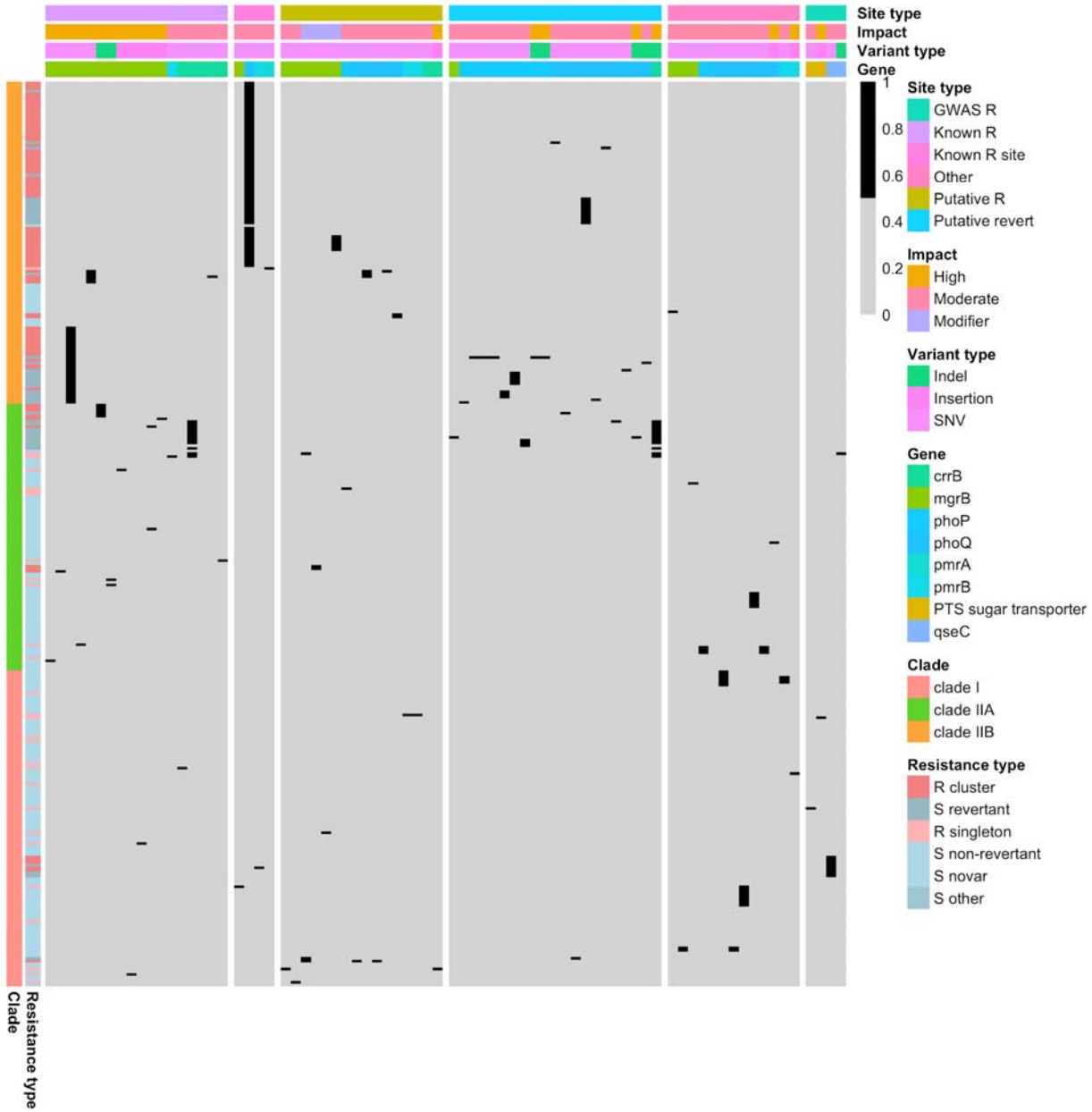

Figure S5: Putative resistance and suppressor variants in canonical and non-canonical resistance genes. Columns are variants and rows are isolates. All variants except the last panel are from canonical resistance genes. R=resistant, S=susceptible.

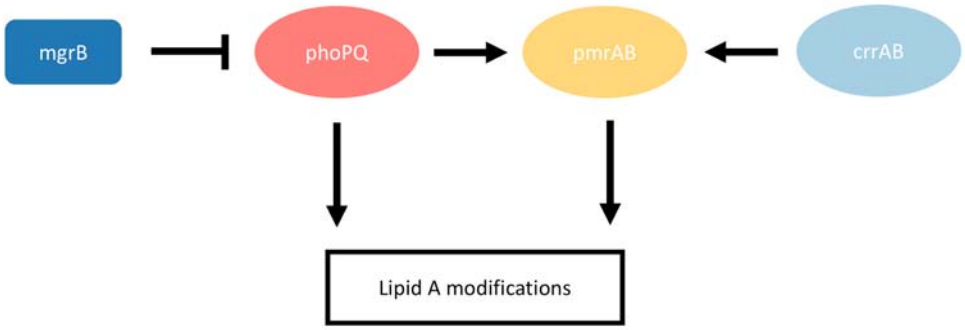

Figure S6: Schematic of the molecular pathway of canonical colistin resistance genes.

Table S1: Resistance variants

Gene	Variant type	Impact	Site type	Variant	Isolate count
mgrB	SNV	high	known	Trp47*	1
mgrB	SNV	high	known	Lys43*	1
mgrB	SNV	high	known	Gln30*	41
mgrB	SNV	high	known	Gln22*	1
mgrB	SNV	high	known	Lys3*	6
mgrB	Indel	high	known	Ile12fs	6
mgrB	Indel	high	known	Leu8fs	2
mgrB	Insertion	high	known	*NA*	1
mgrB	Insertion	high	known	*NA*	1
mgrB	Insertion	high	known	*NA*	1
mgrB	Insertion	high	known	*NA*	2
mgrB	Insertion	high	known	*NA*	1

pmrB	SNV	moderate	known	Leu82Arg	1
crrB	SNV	moderate	known	Gln10Leu	1
crrB	SNV	moderate	known	Leu94Met	16
crrB	SNV	moderate	known	Pro151Ser	2
crrB	SNV	moderate	known	Pro151Ser	1
crrB	SNV	moderate	known	Gly183Val	1
mgrB	SNV	moderate	known_site	Asp31Asn	1
phoQ	SNV	moderate	known_site	Thr244Asn	85
pmrB	SNV	moderate	known_site	Pro158Gln	1
pmrA	SNV	moderate	known_site	Gly53Ser	1
mgrB	SNV	moderate	putative_r	Ile41Asn	1
mgrB	SNV	moderate	putative_r	Asn25Thr	1
mgrB	SNV	modifier	putative_r		3
mgrB	SNV	modifier	putative_r		3

mgrB	SNV	modifier	putative_r		1
mgrB	SNV	modifier	putative_r		9
phoP	SNV	moderate	putative_r	Gly53Val	1
phoQ	SNV	moderate	putative_r	Leu308Gln	1
phoQ	SNV	moderate	putative_r	Leu257Gln	4
phoQ	SNV	moderate	putative_r	Thr248Pro	1
phoQ	SNV	moderate	putative_r	Leu105Gln	2
phoQ	SNV	moderate	putative_r	Leu62Pro	2
pmrB	SNV	moderate	putative_r	Leu262Pro	1
pmrB	SNV	moderate	putative_r	Ser203Pro	1
crrB	SNV	moderate	putative_r	Ala200Pro	1
crrB	Insertion	high	putative_r	*NA*	1
mgrB	SNV	moderate	putative_revert	Thr21Pro	1
phoP	SNV	moderate	putative_re	Gln202Pro	3

			vert		
phoP	SNV	moderate	putative_revert	Thr189Pro	1
phoP	SNV	moderate	putative_revert	His188Gln	1
phoP	SNV	moderate	putative_revert	His188Arg	1
phoP	SNV	moderate	putative_revert	Arg185Leu	3
phoP	SNV	moderate	putative_revert	Arg111His	11
phoP	SNV	moderate	putative_revert	Val18Ala	4
phoP	Indel	high	putative_revert	Thr189fs	1
phoP	Indel	high	putative_revert	His188fs	1
phoQ	SNV	moderate	putative_revert	Gln481Leu	1
phoQ	SNV	moderate	putative_revert	Ala451Thr	1

phoQ	SNV	moderate	putative_revert	Gly443Ser	1
phoQ	SNV	moderate	putative_revert	Arg433Leu	13
phoQ	SNV	moderate	putative_revert	Asp417Val	2
phoQ	SNV	moderate	putative_revert	Leu322Met	1
phoQ	SNV	moderate	putative_revert	Glu123Lys	2
phoQ	SNV	moderate	putative_revert	Ala36Thr	1
phoQ	Indel	high	putative_revert	Leu344fs	1
phoQ	Indel	moderate	putative_revert	Thr248_Leu254del	1
crrB	Indel	high	putative_revert	Ser338fs	16
mgrB	SNV	moderate	other	Phe44Cys	1
mgrB	SNV	moderate	other	Leu24Pro	3
mgrB	SNV	moderate	other	Leu19Gln	1

phoP	SNV	moderate	other	Ala80Gly	3
phoQ	SNV	moderate	other	Ile462Ser	1
phoQ	SNV	moderate	other	Leu365His	8
phoQ	SNV	moderate	other	Pro343Leu	3
phoQ	SNV	moderate	other	Thr332Pro	9
phoQ	SNV	moderate	other	Arg50Leu	7
phoQ	SNV	moderate	other	Gly33Ser	3
phoQ	Insertion	high	other	*NA*	1
pmrB	SNV	moderate	other	Pro95Gln	3
pmrB	Insertion	high	other	*NA*	1
PTS sugar transporter	SNV	moderate	gwas_r	Ser72Ile	1
PTS sugar transporter	Insertion	high	gwas_r	*NA*	1
qseC	SNV	moderate	gwas_r	Val180Ala	9
qseC	Indel	moderate	gwas_r	Arg12_Leu13	1

				del	