Population structure, antibiotic resistance, and uropathogenicity of Klebsiella variicola

Robert F. Potter
Washington University School of Medicine in St. Louis

William Lainhart
Washington University School of Medicine in St. Louis

Joy Twentyman
Washington University School of Medicine in St. Louis

Meghan A. Wallace
Washington University School of Medicine in St. Louis

Bin Wang
Washington University School of Medicine in St. Louis

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation
https://digitalcommons.wustl.edu/open_access_pubs/7431
Population Structure, Antibiotic Resistance, and Uropathogenicity of *Klebsiella variicola*

Robert F. Potter, a William Lainhart, b, Joy Twentyman, c, Meghan A. Wallace, b, Bin Wang, b, Carey-Ann D. Burnham, b, c, d, David A. Rosen, c, d, Gautam Dantas, b, d, e

ABSTRACT *Klebsiella variicola* is a member of the *Klebsiella* genus and often misidentified as *Klebsiella pneumoniae* or *Klebsiella quasipneumoniae*. The importance of *K. pneumoniae* human infections has been known; however, a dearth of relative knowledge exists for *K. variicola*. Despite its growing clinical importance, comprehensive analyses of *K. variicola* population structure and mechanistic investigations of virulence factors and antibiotic resistance genes have not yet been performed. To address this, we utilized *in silico*, *in vitro*, and *in vivo* methods to study a cohort of *K. variicola* isolates and genomes. We found that the *K. variicola* population structure has two distant lineages composed of two and 143 genomes, respectively. Ten of 145 *K. variicola* genomes harbored carbapenem resistance genes, and 6/145 contained complete virulence operons. While the β-lactam *bla* _L_{EE},_ and quinolone _oqxAB_ antibiotic resistance genes were generally conserved within our institutional cohort, unexpectedly 11 isolates were nonresistant to the β-lactam ampicillin and only one isolate was nonsusceptible to the quinolone ciprofloxacin. *K. variicola* isolates have variation in ability to cause urinary tract infections in a newly developed murine model, but importantly a strain had statistically significant higher bladder CFU than the model uropathogenic *K. pneumoniae* strain TOP52. Type 1 pilus and genomic identification of altered _fim_ operon structure were associated with differences in bladder CFU for the tested strains. Nine newly reported types of pilus genes were discovered in the *K. variicola* pan-genome, including the first identified P-pilus in *Klebsiella* spp.

IMPORTANCE Infections caused by antibiotic-resistant bacterial pathogens are a growing public health threat. Understanding of pathogen relatedness and biology is imperative for tracking outbreaks and developing therapeutics. Here, we detail the phylogenetic structure of 145 *K. variicola* genomes from different continents. Our results have important clinical ramifications as high-risk antibiotic resistance genes are present in *K. variicola* genomes from a variety of geographic locations and as we demonstrate that *K. variicola* clinical isolates can establish higher bladder titers than *K. pneumoniae*. Differential presence of these pilus genes in *K. variicola* isolates may indicate adaption for specific environmental niches. Therefore, due to the potential of multidrug resistance and pathogenic efficacy, identification of *K. variicola* and *K. pneumoniae* to a species level should be performed to optimally improve patient outcomes during infection. This work provides a foundation for our improved understanding of *K. variicola* biology and pathogenesis.
**KEYWORDS** emerging pathogens, *Klebsiella*, antibiotic resistance, microbial genomics, urinary tract infection

*Klebsiella variicola* was initially believed to be a plant-associated, distant lineage of *Klebsiella pneumoniae*; however, it has subsequently been recovered from human clinical specimens (1). Despite increasing knowledge on the distinctness of *K. variicola*, *K. pneumoniae*, and *Klebsiella quasipneumoniae*, misidentification within the clinical microbiology lab commonly occurs (2, 3). This may have clinical implications, as one study demonstrated that *K. variicola*-infected patients have higher mortality than *K. pneumoniae*-infected patients (4). Furthermore, several virulence genes (VGs), including siderophores, allantoin utilization genes, and glycerate pathway genes, have been reported in select *K. variicola* strains (5, 6). *K. variicola* has been shown to contain a large pan-genome that is distinct from *K. quasipneumoniae* and *K. pneumoniae*, but the functional consequences of differential gene content have not been explored (2, 7).

In this study, we retrospectively analyzed a cohort of *Klebsiella* isolates collected from 2016 to 2017 at Washington University in St. Louis School of Medicine/Barnes-Jewish Hospital Clinical Microbiology Laboratory (WUSM) for possible *K. variicola* strains using matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) and yggE PCR-restriction fragment length polymorphism (RFLP) assays. We performed Illumina whole-genome sequencing (WGS) to compare *K. variicola* genomes from our institution with publicly available genomes in the first global evaluation of this species. We particularly focused on annotation of canonical *Klebsiella* VGs and antibiotic resistance genes (ARGs) and then assessed their functional consequences using *in vitro* assays and *in vivo* murine infections. Our results demonstrate that population structure, antibiotic resistance, and uropathogenicity of *K. variicola* are generally similar to *K. pneumoniae*, but variability among *K. variicola* genomes has important clinical implications with various strain efficacies in a murine model of urinary tract infection (UTI).

**RESULTS**

**Average nucleotide identity and MALDI-TOF MS can differentiate *K. variicola* from *K. pneumoniae***. We performed Illumina WGS on 113 isolates that are commonly misidentified as *K. pneumoniae* (*K. variicola* *n* = 56), *K. quasipneumoniae* *n* = 3, *K. pneumoniae* *n* = 53, and *Citrobacter freundii* *n* = 1). They were identified by Bruker Biotyper MALDI-TOF MS and yggE RFLP assays from a variety of adult infection sites (see Table S1 in the supplemental material). The isolates were retrieved from the Barnes-Jewish Hospital clinical microbiology laboratory (St. Louis, MO, USA) in 2016 to 2017. We used pyANI with the mummer method to calculate the pairwise average nucleotide identity (ANIm) between the isolates in our cohort and retrieved publicly available *Klebsiella* genomes (*n* = 90) (8, 9) (Table S1). The *C. freundii* isolate was originally classified as *K. pneumoniae* from the Vitek MS MALDI-TOF MS v2.3.3 but was later determined to be *Citrobacter freundii* by Bruker Biotyper MALDI-TOF MS. The yggE PCR-RFLP was indeterminate for this isolate. Confirmatory yggE PCR-RFLP had 94.6% (53/56) concordance with MALDI-TOF for prediction of *K. variicola* within our cohort (Fig. 1). While one genome was dropped from downstream analysis, the other 55 WUSM *K. variicola* genomes all had >95% ANIm with the reference genome of *K. variicola* At-22 (5). *K. variicola* HKUPOLA (GCA_001278905.1) had >95% ANIm with *K. quasipneumoniae* ATCC 700603 reference genome but not *K. variicola* At-22, indicating that it is likely a misannotated *K. quasipneumoniae* isolate and not a *K. variicola* isolate. The remainder of the NCBI *K. variicola* genomes clustered with *K. variicola* At-22 and the WUSM *K. variicola* cohort. One hundred percent (41/41) of the *K. pneumoniae* genomes from NCBI that were suspected to be *K. variicola* due to BLAST similarity had >95% ANIm with *K. variicola* At-22 but not *K. pneumoniae* HS11286 or *K. pneumoniae* CAV1042 (Fig. 1).

Hierarchical clustering of the pairwise ANIm values replicated previous phylogenetic analysis showing that *K. pneumoniae* and *K. quasipneumoniae* are more closely related to each other than to *K. variicola* (Fig. 1). Interestingly, the clustering pattern within *K.
variicola indicated that two isolates, KvMX2 (FLLH01.1) and YH43 (GCF_001548315.1), are more closely related to one another than to the remainder (143/145) of the K. variicola genomes. Given that K. quasipneumoniae can be differentiated into two subspecies based on ANI with the BLAST method (ANIb), we used the JSpecies ANIb program to specifically compare KvMX2 and YH43 with K. pneumoniae ATCC BAA-1705, K. quasipneumoniae ATCC 700603, and 3 other K. variicola genomes (10). KvMX2 and Yh43 have 98.02% ANIb with one another but an average of 96.67%, 96.65%, and 96.68% ANIb with WUSM_KV_53, WUSM_KV_15, and K. variicola At-22, respectively (Table S1). Consistent with our pyANI ANIm result, none of the K. variicola strains had >95% ANIb with K. pneumoniae ATCC BAA-1705 or K. quasipneumoniae ATCC 700603. These data suggest that MALDI-TOF MS or yggE PCR-RFLP may be effective means to differentiate K. variicola from K. pneumoniae in the absence of WGS.

K. variicola population structure has 2 lineages and 26 clusters in the second lineage. Core-genome alignment of the 1,262 genes at 90% identity shared by strains in all Klebsiella species and a Kluyvera georgiana outgroup shows that the K. variicola isolates are in a cluster with K. pneumoniae, K. quasipneumoniae, and the newly described K. quasivariicola (11) (Table S2; Fig. S1). Core-genome alignment of the 3,430 core genes at 95% nucleotide identity for the entire gene length by all 145 K. variicola genomes indicates that KvMX2 and Yh43 are distinctly related to the other 143 genomes (Fig. 2a; Table S2). These other genomes form a star-like phylogeny showing deep-branching clusters radiating from the center of the tree. FastGear, which uses hierBAPS to identify lineages and then searches for recombination between lineages, supported the differentiation of KvMX2 and Yh43 into a separate lineage from the other genomes and identified 6 instances of recombination between these two lineages (Table S3) (12, 13).

Phylogenomic network analysis and quantification of recombination from parSNP showed minimal recombination within the 143 K. variicola lineage 2 genomes, with approximately 1.62% of the K. variicola genome believed to be recombinant (Fig. S2a; Table S3) (14). The Nearest Neighbor network of the 3,496 genes shared by the lineage 2 genomes and a recombination-free phylogenetic tree of the 143 genomes from

![FIG 1 Pairwise average nucleotide identity cluster map of WUSM and NCBI Klebsiella. Hierarchical clustering and heat map of pairwise ANIm values among all isolates. The source of isolates (WUSM or NCBI) and initial species delineation (K. variicola, K. pneumoniae, or K. quasipneumoniae) are shown as colored bars adjacent to the heat map. The three major blocks are labeled by their final species determination.](http://mbio.asm.org/ on January 23, 2019 by guest)
FIG 2  Population structure of K. variicola genomes. (a) Approximate-maximum-likelihood tree of the total 145 K. variicola genomes and annotation of FastGear lineage identification. (b) Recombination-free parSNP tree of the closely related lineage 2 genomes with quantitative clustering from ClusterPicker.
parSNP showed many deep-branching clades with a star-like phylogeny (Fig. 2b; Table S2). This tree topology was similar with and without recombination, which suggests that K. variicola lineages emerged early from a single common ancestor into equally distant clades across different environments (Fig. S2b). Quantitative clustering of the 143 genomes in the second lineage with ClusterPicker showed that 56.6% (81/143) of genomes fall into 26 clusters, with 57.7% (15/26) of the clusters containing more than 2 genomes (Fig. 2b) (15). Only 46.2% (12/26) of clusters contain isolates from both WUSM_KV and NCBI. The largest clusters, 24 and 21, each contain 7 genomes. Cluster 21 contained WUSM_KV_10 and 6 genomes from an analysis of patient isolates at an intensive care unit in Seattle, WA (USA). Although they were in the same cluster, WUSM_KV_10 differed from these isolates at 1,882 sites across the 4,867 genes shared at 95% identity (Table S2 and Table S4).

To better understand the context of the 4 groups in lineage 2, we aligned the 2,932 genes shared among the 145 K. variicola genomes, Klebsiella (formerly Enterobacter) aerogenes KCTC 2190, K. quasipneumoniae ATCC 700603, and K. pneumoniae ATCC BAA-1705 at ≥90% identity to create a dendrogram (Fig. 2c; Table S2). This method preserved the conservation of the lineage 2 groups but showed a different order. The only discrepancy observed is that, in the lineage 2 phylogenetic tree, cluster 3 appeared to be in the A group; however, both 521_SSON and 524_SBOY are more similar to C group genomes in the dendrogram. This incongruence is consistent with cluster 3 radiating away from cluster 4 near the center of the phylogenetic tree (Fig. 2b).

Addition of metadata onto the dendrogram showed that the K. variicola cohort spans most geographic locations, with the notable exception of Africa and Oceania (Fig. 2c). The K. variicola genomes showed a remarkable level of source diversity, with representative isolates from animals (n = 4), fungi (n = 2), plants (n = 7), water (n = 3), and industrial waste (n = 6). However, as a testament to the pathogenic potential of K. variicola, 79.5% (114/145) of genomes came from sites associated with humans. Of the human-associated sites, 40.4% (46/114) came from urine and 19.2% (22/114) came from blood (Fig. 2c). We did not observe any apparent association with geography, habitat, or infection site for any of the K. variicola clades. Sixty-seven of 145 isolates had a sequence type (ST) identified using the K. pneumoniae multilocus sequence type scheme (Table S5). Consistent with the distance between lineages, 44 different STs were identified. ST1562 and ST641 had the highest number of isolates (n = 4). In summary, these data demonstrate that K. variicola has a diverse population structure and can be found in a variety of environmental and host niches.

Acquired ARGs and VGs are not restricted to any K. variicola cluster. We applied ResFinder to determine the burden of acquired ARGs among the K. variicola strains (Fig. 3a; Table S5) (16). β-Lactamase genes were the most abundant ARG in the K. variicola cohort (n = 26). As expected, bla LEN was almost universally conserved, as 837_KPNE was the only isolate without one identified. Ten different bla LEN alleles were found. bla LEN-16 was most common (51/145), followed by bla LEN-24 (40/145) and bla LEN-2 (31/145). Carbapenemases were rare, but bla KPC-2 (4/145), bla KPC-6 (1/145), bla NDM-1 (1/145), bla NDM-9 (3/145), and bla OXA-48 (1/145) were each identified across a total of 10/145 strains. bla CTX, bla SHV, bla TEM, and noncarbapenemase bla OXA genes were also identified, but we did not detect any class C β-lactamase genes or non-bla NDM class B β-lactamase genes. Aminoglycoside ARGs (n = 10), including members of the aac, aad, aph, and str families, comprised the second most abundant class. ARGs against folate synthesis inhibitors (n = 8), quinolones (n = 7), amphenicols (n = 4), tetracyclines (n = 2), macrolides/lincomesamides/streptogramins (n = 2), and fosfomycin (n = 1) were also found (Fig. 3a). In addition to the near-total conservation of bla LEN, the quinolone efflux
pump components oqxAB were found in almost all isolates (139/145). Across the 145 genomes, the median and mode number of ARGs were both 3. A 6.89% (10/145) proportion of genomes harbored ≥10 ARGs, including WUSM_KV_55 from our cohort.

We used the *K. pneumoniae* BIGSdb database (https://bigsdb.pasteur.fr/klebsiella/klebsiella.html) and BLASTN to identify canonical *Klebsiella* VGs in the *K. variicola* strains (Fig. 3b; Table S5). In contrast to ARGs, previously characterized *Klebsiella* VGs were found only sporadically in the *K. variicola* cohort. Interestingly, the *all* allantoin utilization operon and *arc, fdrA, gcl, glxKR, hyi,* and *ybbWY* genes were found in the distantly related YH43 genome as well as the closely related BIDMC90, k385, and WUSM_KV_03
the lowest number of acquired ARGs, as it lacked ARGs. Interestingly, it possesses ARGs found to harbor additional quinolone (84/145) but not resistance to aminoglycosides and sulfonamides, as WUSM_KV_55 contained the carbapenemase gene blaOXA-48. In addition to the core β-lactamase blaLEN-2, this isolate also contained a blaCTX-M-14 gene. Redundancy was again observed for the ARGs against aminoglycosides and sulfonamides, as WUSM_KV_55 contained aac(6’)-Ib-cr, aadA16, sul1, and sul2. Within our cohort, this isolate was the only isolate found to harbor additional quinolone (qnrB6), rifampin (arr-3), and amphenicol (floR) ARGs. Interestingly, it possesses oqxAB but not oqxA. Conversely, WUSM_KV_35 harbored the lowest number of acquired ARGs, as it lacked oqxAB but carried blaLEN-24.

We used Kirby-Bauer disk diffusion to quantify phenotypic resistance of the WUSM K. variicola strains to several clinically relevant antibiotics (Fig. 4b). Klebsiella species are generally considered intrinsically resistant to ampicillin due to a conserved β-lactamase gene. In our cohort, 3/55 isolates were unexpectedly susceptible to ampicillin while the rest were resistant. Despite phenotypic sensitivity to ampicillin, the genomes for WUSM_KV_25, WUSM_KV_34, and WUSM_KV_82 encode blaLEN-24, blaLEN-22, and blaLEN-10, respectively. These blaLEN alleles were also found in isolates intermediate and resistant to ampicillin. As expected, WUSM_KV_55 was the only isolate resistant to both meropenem and ceftazidime, presumably due to carriage of blaKPC-2. Additionally, it was the only isolate intermediate to ciprofloxacin. Four isolates were resistant to trimethoprim-sulfamethoxazole, but only WUSM_KV_50 and WUSM_KV_55 had identified ARGs that would explain this phenotype.

Review of a 2017 composite antibiogram from a microbiology laboratory serving 5 hospitals in the St. Louis region (Missouri, USA), based on first isolate per patient per year, revealed that, in general, K. pneumoniae (n = 1,522) had decreased susceptibility to all reported antimicrobials compared to K. variicola (n = 144), except for meropenem (99% susceptibility for both species). Most notably, K. pneumoniae exhibited decreased susceptibility, compared to K. variicola, with ampicillin-sulbactam (63% versus 93% susceptible), nitrofurantoin (66% versus 86% susceptible), and trimethoprim-sulfamethoxazole (80% versus 90% susceptible).

Changes in fim operon are associated with uropathogenicity in a murine UTI model. Given that 70% (39/56) of K. variicola strains from our cohort were isolated from the human urinary tract, we wanted to assess uropathogenicity in a diverse subset of these isolates. We transurethrally inoculated C3H/HeN mice with 10^7 CFU/ml of 5 individual K. variicola strains, or the model uropathogenic K. pneumoniae TOP52 strain, for comparison (Fig. 5a) (3, 18, 19). Similarly to previously published infections with K. pneumoniae TOP52, the K. variicola strains exhibited large variations in bacterial CFU recovered from the bladder at 24 h postinfection (hpi). Compared to TOP52, WUSM_KV_39 was the only isolate with a significantly increased bladder burden (P = 0.0094). Bacterial loads of WUSM_KV_10 and WUSM_KV_39 were both significantly higher than WUSM_KV_09 and WUSM_KV_14 (Fig. 5a). Despite this variability among
WUSM K. variicola strains have a low burden of ARGs and are generally susceptible to antibiotics. (a) Network diagram depicting each WUSM_KV isolate and ARG as nodes. ARGs are colored in accordance with predicted phenotypic resistance from ResFinder, and WUSM_KV genomes are colored by the burden of ARGs. (b) Scatter plots depicting Kirby-Bauer disk diffusion size (mm) from phenotypic susceptibility testing. Each plot represents an isolate, and the plots are colored according to CLSI interpretation. Those with atypical resistance are listed by name with putative ARGs.
FIG 5 Changes in fim operon are associated with outcomes in mouse UTI model. (a) CFU/bladder and CFU/kidney of *K. pneumoniae* TOP52 and WUSM_KV isolates 24 h after transurethral bladder inoculation of C3H/HeN mice. Short bars represent geometric means of each group, and dotted lines represent limits of detection. (b) *fimS* phase assay and quantification with respective bands indicating the “ON” and “OFF” position labeled. (c) Immunoblot for FimA and GroEL, with quantification shown below. (d) Easyfig illustration of genes in the *fim* operon and Jalview of the nucleotides and amino acids for the *fimB/fimE* intergenic region and *fimD* gene.
bladder CFU results, the results of kidney titer determinations at 24 hpi were not significantly different among strains by ANOVA (P = 0.1270). As observed in the bladder, however, WUSM_KV_10 and WUSM_KV_39 achieved significantly higher kidney CFU than WUSM_KV_14.

Given the variation in bladder burden, we wanted to assess if differences in uropathogenicity could be related to expression of type 1 pili, a key virulence factor for UTI encoded by the fim operon (19, 20). In K. pneumoniae and Escherichia coli, expression of type 1 pili is controlled by a region of invertible DNA (fimS site) (20, 21). Orientation of the fimS site in the “ON” position enables production of type1 pil and increased urovirulence. Under identical growth conditions, WUSM_KV_39 had a higher population with the fimS promoter region in the “ON” orientation than the other strains tested (Fig. 5b). Furthermore, consistent with its success in the bladder, WUSM_KV_39 was found to produce the greatest amount of FimA (the main structural component of type 1 pil), as measured by immunoblotting (Fig. 5c). WUSM_KV_03, WUSM_KV_09, and WUSM_KV_39 all produced significantly more FimA than K. pneumoniae TOP52. Interestingly, WUSM_KV_14 did not produce appreciable levels of FimA by this assay (Fig. 5c).

As we discovered significant variability in type 1 piliation, we specifically investigated changes in fim operon sequence between these isolates by viewing the Prokka coding sequence annotation in Easyfig and Jalview (Fig. 5d) (22, 23). We found that WUSM_KV_14 had a predicted truncated FimD usher sequence. A guanine-to-adenine single nucleotide polymorphism (SNP) in the fimD gene changed a predicted tryptophan residue into a premature stop codon, likely explaining the observed lack of production of type 1 pil. Additionally, in WUSM_KV_39, Prokka annotated a hypothetical protein in the intergenic region between fimB and fimE and included a gap replacing a thymine and a guanine-to-adenine SNP. The altered fimB/fimE intergenic region in WUSM_KV_39 may play a role in its increased expression of type 1 pil. Together, these data demonstrate that variation exists among K. variicola genomes that may account for differential urinary tract niche proclivity among isolates.

K. variicola contains both conserved and novel usher genes. The fim operon is one of the best-characterized chaperone-usher pathways (CUPs); given the observed importance of the fim operon in K. variicola uropathogenicity, we searched the pan-genome of our K. variicola cohort to identify the complete repertoire of CUP operons (24). Seventeen unique usher sequences at 95% identity were identified across the 55 WUSM K. variicola genomes, and an amino acid sequence alignment showed that they were distributed in 5 Nuccio and Baumler (25) clades (Fig. 6a; Table S6). From this analysis, we discovered 9 new usher genes previously undescribed in Klebsiella, which we name kva through kvi (Table S6). KviA and KveB usher sequences were found to cluster within the pi (∏) clade, making them the first description of a P-pilus apparatus in Klebsiella. The recently named ∏* subclade contained the greatest amount (7/17) of K. variicola usher sequences; 5 of these 7 were previously reported in K. pneumoniae, while KvcC and KvdB are first reported here.

FimD and the usher sequences for KpaC, KvaB, KpeC, and KpjC were present in all 55 WUSM K. variicola isolates (Fig. 6b). KvgC, KvhC, KviA, and KpcC were each found in only one isolate. KpgC, MrkC, KvbC, KpbC, KvcC, KveB, KvfC, and KvdB can be considered accessory usher sequences in this cohort, as they were absent in certain strains. The most notable pattern evident from the hierarchical clustering of the presence/absence for all usher genes in our K. variicola cohort is that isolates WUSM_KV_10 through WUSM_KV_21 all carry the KvdB sequence but not KpbC.

Eight of the 9 newly described usher sequences had highest BLASTP hits of ≥99% identity across the entire length of the gene against the nonredundant protein sequences database in April 2018, and all of them were previously annotated as being found in Enterobacteriaceae, Klebsiella, or K. variicola (Table S6). All of the usher genes except kvi were in operons that included a chaperone, at least one subunit, and a putative adhesin (Table S6). KvhC, the usher protein with the lowest BLASTP identity
value, had 76% identity to several genes from Enterobacter species (Fig. S3a). The contig with the *kvh* operon also contained several genes that had possible roles in prophage integration and transposase activity (Fig. S3b). Our results indicate that *K. variicola* strains harbor a diverse set of usher genes, which may augment *K. variicola* fitness across a variety of environmental niches, and these operons may be acquired from other Enterobacteriaceae.

**DISCUSSION**

A previous phylogenomic study used split-network analysis to demonstrate that the *K. variicola* phylogroup (formerly KPIII) is distinct from *K. pneumoniae* (KPI) and *K. quasipneumoniae* (KPII) (26). As an orthogonal method, we used ANI software, the gold standard for in silico species delineation, to recreate this differentiation of phylogroups as separate species (8). Historically, differentiation between *K. pneumoniae* and *K. variicola* has been difficult, as evidenced by misannotation of *K. variicola* as *K. pneu-
moniae in public genome sequence databases (Fig. 1). These misannotated K. variicola strains came from a variety of geographic regions and were not exclusive to any cluster. Within our sequenced cohort, differentiation of K. variicola from K. pneumoniae and K. quasipneumoniae using MALDI-TOF MS and yggE PCR-RFLP was supported by ANI. This indicates that yggE PCR-RFLP (3) would be a feasible alternative for clinical labs across the globe lacking access to MALDI-TOF MS or WGS. Additionally, hierarchical clustering of the ANI values and core-genome phylogeny demonstrated that 2 K. variicola genomes were distinctly separate from the other 143 in our cohort. ANIb values between these genomes and the other K. variicola genomes were ~96%, similar to what was observed for K. quasipneumoniae. The differences in ANIb values contributed to the delineation of K. quasipneumoniae into two subspecies, Klebsiella quasipneumoniae subsp. quasipneumoniae and Klebsiella quasipneumoniae subsp. similipneumoniae (27). Further phenotypic comparisons, including the sole carbon source utilization for differentiation of the K. quasipneumoniae subspecies, between KvMX2/ Yh43 and other K. variicola isolates is required to unequivocally qualify these as separate subspecies (27).

Numerous studies have shown that K. pneumoniae has a deep-branching phylogenetic structure with minimal recombination occurring within K. pneumoniae strains and between K. pneumoniae and K. variicola/K. quasipneumoniae (26, 28). Importantly, though, large-scale recombination events may be clinically relevant, as evidenced by research on the origin of the frequently carbapenem-resistant ST258 lineage (29, 30). Our results demonstrate that like K. pneumoniae, K. variicola shows minimal recombination within its genome, and its population structure is composed of numerous clades in a star-like phylogeny. A star-like population structure with deep-branching relationships between isolates (n = 29 and n = 28) was also found in two previously published K. variicola phylogenetic trees (2, 31).

Similarly to our work, a previous investigation did not identify any geographic distinction when genomes from within the United States were compared to those from outside the United States (2). The 6 genomes in cluster 21 with WUSM_KV_10 were from ICU patient samples in Seattle, WA, which provides the first evidence of clonal groups responsible for K. variicola infections in some settings (32). Although they were closely related compared against all K. variicola genomes, there were still 1,882 SNPs between WUSM_KV_10 and the other 6 genomes. Interestingly, clusters were not restricted to human infections, as cluster 24 contains 3 genomes from bovine mastitis (NL49, NL58, and NL58) and hospital isolates (VRCO0246, VRCO00242, VRCO00244, and VRCO00243) (https://www.ncbi.nlm.nih.gov/bioproject/361595) (33).

As expected for K. variicola, blaLEN β-lactamases were the most conserved ARGs. A previous report unexpectedly found a K. variicola isolate that harbored the blaOKR gene commonly found in K. quasipneumoniae; however, we did not identify such instances within our cohort (2). Although chromosomally carried in K. pneumoniae, fosA was identified in only 1/145 of the K. variicola genomes (34, 35). Additionally, as previously found in K. pneumoniae clinical isolate cohorts, we found oqxAB efflux pump genes widespread across K. variicola genomes (36–38). Although these genes may be ubiquitous in K. variicola, 0 of 55 isolates we tested had resistance to ciprofloxacin; the single example with intermediate susceptibility carried a qnrB6 gene. This is not atypical for Enterobacteriaceae possessing oqxAB, as one study found 100% prevalence of oqxAB in K. pneumoniae but no quinolone resistance (37). It is possible that for K. variicola, similarly to K. pneumoniae, high expression of oqxAB is essential for phenotypic resistance to quinolones (36). In K. pneumoniae, expansion of clonal groups is associated with carbapenemase carriage (i.e., ST258 and blaKPC); however, we did not observe any associations between carbapenemase genes and K. variicola clusters. Indeed, only 1.81% (1/55) of K. variicola strains within our institutional cohort had a carbapenemase gene and the regional resistance rate for meropenem between K. pneumoniae and K. variicola in 2017 was similar. blaNDM-positive K. variicola strains have been identified in clinical and environmental samples, but blaKPC-positive genomes came exclusively from clinical sources. KPN1481 (blaNDM-1) was annotated as a urine-derived isolate, but GJ1,
GJ2, and GJ3 (all blaNDM-9) were found in the Gwangju tributary in South Korea (2, 39). In contrast, WUSM_KV_55 (blaKPC-2) was isolated from bronchoalveolar lavage fluid, KP007 (blaKPC-2) from an intra-abdominal site, and 223/14 (blaKPC-4) from a laparotomy wound (40, 41). IncF plasmids, the most abundant replicon identified in the K. variicola cohort, are known carriers of antibiotic resistance genes, including blaCTX-M and blaOXA β-lactamases (42). Consistent with their widespread identification in K. variicola, IncF plasmids are frequently found in K. pneumoniae and E. coli (43, 44).

K. pneumoniae is the second leading cause of urinary tract infections (45). Given previous misclassification of K. variicola as K. pneumoniae and the high frequency at which K. variicola was isolated from the urinary tract, we were interested in comparing the uropathogenicity of our K. variicola isolates to the well-studied model K. pneumoniae TOP52 isolate (3, 18, 19). We identified strain-dependent virulence capacity, with UTIs from WUSM_KV_39 yielding statistically significant higher bladder CFU than K. pneumoniae TOP52. Quantification of metrics used to study uropathogenicity in E. coli and K. pneumoniae show increased fimS in the “ON” orientation and increased FimA production by WUSM_KV_39; these findings provide a plausible explanation for why WUSM_KV_39 performed better than K. pneumoniae TOP52 and all WUSM_KV isolates excluding WUSM_KV_10 (46). While we do not yet understand the role of the putative protein identified between recombinases fimB and fimE in WUSM_KV_39, one could postulate that this difference may affect fimbral expression. Additionally, the poorest performer in the urinary tract, WUSM_KV_14, encodes a mutation resulting in a truncated fimD usher sequence which likely explains its lack of FimA production. As with other bacterial pathogens, it is likely that specific virulence factors are required for K. variicola competency in distinct body niches (47, 48). Further work is therefore warranted to test if yersiniabactin and allantoin utilization promote lung and liver infections, respectively, in K. variicola as they do in K. pneumoniae (49–52).

K. variicola carries usher genes previously identified in K. pneumoniae and 9 novel ushers (53). Interestingly, KveB and KviA are the first report of π usher proteins in Klebsiella. The best-studied π operon, pap in E. coli, is a major contributor to pyelonephritis as the PapG adhesin can bind Gal-α-(1-4)-Gal exposed on human kidney cells (54). Other usher genes have been shown to be essential for biofilm formation, plant cell adhesion, and murine gut colonization, further demonstrating their role in niche differentiation (53). Clustering of the presence/absence of these ushers showed the absence of KpbC but presence of KvdB in 11 of the WUSM_KV genomes, a phenomenon similar to that observed for UshC and YraJ in E. coli (55). All 4 of these usher types were found in the γ clade, suggesting an exclusionary form of functional redundancy between usher genes (55). Usher genes and CUP operons are frequently exchanged horizontally between Enterobacteriaceae genera (55). Indeed, we have found that the KvhC usher protein has only 76% amino acid identity to any existing proteins in the nonredundant protein sequence database and that the kvh operon is situated next to multiple prophage- and transposase-associated genes.

In this investigation, we used phenotypic and genomic analyses to better understand the diversity of K. variicola genomes, both from our institution and across the globe (using publicly available NCBI genomes). Then, we assessed the functional consequences of ARGs and VGs toward antibiotic resistance and uropathogenicity. One limitation of our study is that our mouse infections and phenotypic analyses are performed with nonisogenic strains. If existing genetic modification systems in K. pneumoniae are shown to be useful for gene knockouts in K. variicola, further work can be performed to mechanistically validate our findings. An additional limitation is that ~30 genomes of K. variicola have been uploaded to NCBI since we initiated our comparison. These may further elucidate differences in population structure, although even with almost 300 genomes, one study indicates that K. pneumoniae diversity remains undersampled (26).

Our work represents the first large-scale genomic analysis of K. variicola across multiple institutions and the first use of a murine model to study K. variicola pathogenesis. We unequivocally show that whole-genome comparisons can separate K.
varicola from K. pneumoniae and offer convenient alternative methods for laboratories without access to WGS to differentiate these species. Importantly, we demonstrate that high-risk ARGs and VGs are present in K. varicola genomes from a variety of geographic locations. This may have clinical ramifications, as we demonstrate that some K. varicola clinical isolates can be superior uropathogens compared to K. pneumoniae. Similarly to E. coli and K. pneumoniae, the diversity of CUP operons in these isolates could complement additional acquired virulence genes and enable infection of specific niches. Therefore, it is imperative that K. varicola and K. pneumoniae continue to be differentiated in the clinical laboratory, so that we may apply data on differential gene repertoire, clinical behavior, and niche specificity to the goal of ultimately improving patient outcomes.

MATERIALS AND METHODS

Clinical Klebsiella collection. One hundred thirteen clinical Klebsiella species isolates recovered in the Barnes-Jewish Hospital microbiology laboratory (St. Louis, MI) from 2016 to 2017 were evaluated in this study. Of these, 56 were consecutively collected isolates identified by Bruker Biotype MALDI-TOF MS as K. varicola (research-use-only database v6). This identification was confirmed using a PCR-restriction fragment length polymorphism (RFLP) assay targeting the yggE gene (F: 5'-TGTTACTAAATCGCCTTTAGGG-3'; R: 5'-CACCGATCTGCAAAAACGTCTACT-3'; restriction enzyme: BciVI) that was designed to distinguish K. varicola from K. pneumoniae. A 94.6% proportion (53/56) of isolates were confirmed as K. varicola using the yggE PCR-RFLP assay.

The remaining 58 isolates were randomly selected from a banked collection of K. pneumoniae strains historically recovered from clinical specimens (29 from urine, 25 from blood, and 1 each from abdominal wound, tracheal aspirate, bronchial washing, and bile). Each of these isolates underwent Bruker MALDI-TOF MS and yggE PCR-RFLP to confirm their identification. Five percent (3/58) were confirmed as K. varicola using MALDI-TOF MS and the yggE PCR-RFLP assay.

Illumina whole-genome sequencing and publicly available Klebsiella genomes. Pure frozen stocks of the presumptive 113 Klebsiella isolates were plated on blood agar to isolate single colonies. Approximately 10 colonies were suspended using a sterile cotton swab into water, and total genomic DNA was extracted using the Bacteremia kit (Qiagen). An 0.5-ng amount of DNA was used as input for sequencing libraries using the Nextera kit (Illumina) (56). Libraries were pooled and sequenced on an Illumina NextSeq 2500 high-output system to obtain ~2.5 million 2 × 150-bp reads. Demultiplexed reads had Illumina adapters removed with Trimmomatic v.36 and decontaminated with DeconSeq v0.4.3 (57, 58). Draft genomes were assembled with SPAdes v3.11.0, and the scaffolds.fasta files were used as input for QUAST v 4.5 to measure the efficacy of assembly (see Table S1 in the supplemental material) (59, 60). All contigs of ≥500 bp in length were annotated for open reading frames with Prokka v1.12 (61).

To increase the number of genomes for downstream analysis, 50 K. varicola genomes were obtained from NCBI genomes (https://www.ncbi.nlm.nih.gov/genome/) in September 2017 (Table S1). Additionally, as it is possible that previously sequenced K. varicola may be incorrectly described as K. pneumoniae, we submitted the complete genome of the K. varicola reference strain AT-22 to NCBI BLASTN against the nonredundant nucleotide collection and the whole-genome shotgun sequence databases using default settings in September 2017. Using this method, we obtained 41 genomes of K. pneumoniae with the minimum observed query length of 38% at 99% identity (Table S1). Given that the cohort of genomes analyzed in our study includes isolates initially misannotated, we refer to them as either the NCBI genome or assembly (https://www.ncbi.nlm.nih.gov/assembly) accession key. Sequenced and acquired isolates were analyzed using a variety of computational programs (Text S1).

Antimicrobial susceptibility testing. K. varicola isolates underwent antimicrobial susceptibility testing per laboratory standard operating procedures using Kirby-Bauer disk diffusion on Mueller-Hinton agar (BD BBL Mueller-Hinton II agar), in accordance with Clinical and Laboratory Standards Institute (CLSI) standards. Disk diffusion results were interpreted using CLSI Enterobacteriaceae disk diffusion breakpoints (62). Briefly, 4 to 5 colonies from pure isolates were used to create a 0.5 McFarland suspension of the organism in sterile saline. A sterile, nontoxic cotton swab was dipped into the bacterial suspension, and a lawn of the organism was plated to Mueller-Hinton agar. Antimicrobial Kirby-Bauer disks were applied, and the plate was incubated at 35°C in room air for 16 to 24 h. The diameters of the zones of growth inhibition surrounding each antimicrobial disk were recorded in millimeters.

Mouse urinary tract infections. Bacterial strains from our K. varicola cohort and K. pneumoniae TOPS2 were used to inoculate 7- to 8-week-old female C3H/HeN mice (Envigo) by transurethral catheterization as previously described (18, 19, 63). The K. varicola strains were selected to encompass a range of genetically distinct isolates. WUSM_KV_03 and WUSM_KV_10 were specifically chosen as they contain the all and ybt operons, respectively. Static 20-ml cultures were started from freezer stocks, grown in Luria-Bertani (LB) broth at 37°C for 16 h, and centrifuged for 5 min at 8,000 × g, and the resultant pellet was resuspended in phosphate-buffered saline (PBS) and diluted to approximately 4 × 105 CFU/ml. Fifty milliliters of this suspension was used to infect each mouse with an inoculum of 2 × 107 CFU/ml. Inocula were verified by serial dilution and plating. At 24 hpi, bladders and kidneys were aseptically harvested, homogenized in sterile PBS via Bullet Blender (Next Advance) for 5 min, serially
diluted, and plated on LB agar. All animal procedures were approved by the Institutional Animal Care and Use Committee at Washington University School of Medicine.

Phase assays. To determine the orientation of the fimS phase switch in *Klebsiella*, a phase assay was adapted as previously described (20). An 817-bp fragment including *fimS* was PCR amplified using Taq polymerase (Invitrogen) and the primers 5′-GGGACAGATACGCGTTTGAT-3′ and 5′-GGCCTAACTGAACGGTTTGA-3′ and then digested with HinfI (New England Biolabs). Digestion products were separated by electrophoresis on a 1% agarose gel. A phase-ON switch yields products of 605 and 212 bp, and a phase-OFF switch yields products of 496 and 321 bp.

**FimA and GroEL immunoblots.** Acid-treated, whole-cell immunoblotting was performed as previously described using 1:2,000 rabbit anti-type 1 pilus and 1:500,000 rabbit anti-GroEL (Sigma-Aldrich) primary antibodies (64, 65). Amersham ECL horseradish peroxidase-linked donkey anti-rabbit IgG (GE Healthcare) secondary antibody (1:2,000) was applied, followed by application of Clarity enhanced chemiluminescence (ECL) substrate (Bio-Rad Laboratories). The membrane was developed and imaged using a ChemiDoc MP Imaging System (Bio-Rad Laboratories). Relative band intensities were quantified using Fiji (https://fiji.sc/) (66).

**Statistics.** CFU/bladder and CFU/kidney for both experimental replicates were used as input for ordinary one-way ANOVA to judge significance. Pairwise comparisons of CFU/bladder and CFU/kidney values were performed using the nonparametric Mann-Whitney U test. Similarly, normalized quantifications of relative FimA amounts (FimA/GroEL) and *fimS* in “ON” position (*fimS* “ON”/*fimS* “OFF”) were compared using the Mann-Whitney U test. All P values of <0.05 were considered significant, and all calculations were performed in GraphPad Prism v7.04.

**Accession number(s).** The genomes have all been deposited in NCBI under BioProject accession no. PRJNA473122.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.02481-18.

**TEXT S1, DOCX file, 0.04 MB.**

**FIG S1, TIF file, 1.5 MB.**

**FIG S2, TIF file, 3.4 MB.**

**FIG S3, TIF file, 1 MB.**

**TABLE S1, XLSX file, 0.1 MB.**

**TABLE S2, XLSX file, 0.6 MB.**

**TABLE S3, XLSX file, 0.02 MB.**

**TABLE S4, TXT file, 0.1 MB.**

**TABLE S5, XLSX file, 0.04 MB.**

**TABLE S6, XLSX file, 0.1 MB.**

**ACKNOWLEDGMENTS**

We thank members of the Dantas lab for insightful discussions of the results and conclusions. We thank Center for Genome Sciences & Systems Biology staff Brian Koebbe and Eric Martin for operation of the High-Throughput Computing Facility. We additionally thank David Hunstad for constructive feedback during manuscript authoring. We additionally thank Center for Genome Sciences & Systems Biology staff Jessica Hoisington-Lopez and MariaLynn Jaeger for performing the Illumina sequencing and demultiplexing.

This work is supported in part by awards to G.D. through the Edward Mallinckrodt, Jr. Foundation (Scholar Award) and from the National Institute of General Medical Sciences, the National Institute of Allergy and Infectious Diseases, and the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (NIH) under award numbers R01GM099538, R01AI123394, and R01HD092414, respectively. Experiments performed by J.T. and D.A.R. used funding from the NIH (award K08-AI127714) and the Children’s Discovery Institute of Washington University and St. Louis Children’s Hospital. R.F.P. was supported by an NIGMS training grant through award T32 GM007067 (PI: James Skeath) and the Monsanto Excellence Fund graduate fellowship.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.
REFERENCES


The document contains a list of scientific references, mostly related to antimicrobial resistance and its detection. The references are cited in the text, and the authors are mentioned when discussing their work. The document likely discusses the resistance mechanisms, detection methods, and the role of various genes and plasmids in bacterial resistance. It appears to be a review or a compilation of recent findings in the field of antimicrobial resistance.


The document also cites various other studies that likely discuss similar topics, such as the use of different detection methods, the role of specific genes like oqxAB, the prevalence of certain resistance mechanisms, and the importance of understanding these mechanisms for developing effective treatment strategies.

Overall, the document seems to be a comprehensive review or a compilation of recent research findings in the field of antimicrobial resistance, highlighting the importance of continuous monitoring and research in this area to combat the rapid evolution of bacterial resistance.