Differentiation of hypervirulent and classical *Klebsiella pneumoniae* with acquired drug resistance

Thomas A. Russo[1,2,3,4#], Cassandra L. Alvarado[1,2], Connor J. Davies[1,2], Zachary J. Drayer[2], Ulrike Carlino-MacDonald[1,2], Alan Hutson[5], Ting L. Luo[6], Melissa J. Martin[6], Brendan W. Corey[6], Kara A. Moser[7], J. Kamile Rasheed[7], Alison L. Halpin[7], Patrick T. McGann[6], and Francois Lebreton[6]

[1]Veterans Administration Western New York Healthcare System; [2]Department of Medicine; [3]Department of Microbiology and Immunology; [4]The Witebsky Center for Microbial Pathogenesis; University at Buffalo, State University of New York, Buffalo, New York; [5]Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, New York; [6]Multidrug-Resistant Organism Repository and Surveillance Network (MRSN), Walter Reed Army Institute of Research, Silver Spring, Maryland; [7] Division of Healthcare Quality Promotion, U.S. Centers for Disease Control and Prevention, Atlanta, Georgia.

Running title: Differentiating hypervirulent from classical *K. pneumoniae*

Correspondent footnote:

Thomas A. Russo. Jacobs School of Medicine and Biomedical Sciences, 955 Main St, Room 5212, Buffalo, NY 14203 and The Veterans Administration Western New York Healthcare System, 3495 Bailey Ave, Buffalo, NY 14215.

Email: trusso@buffalo.edu

Abstract word count: 250, Importance word count: 150, Text word count: 5703

Abstract:

Distinguishing hypervirulent (hvKp) from classical *Klebsiella pneumoniae* (cKp) strains is important for clinical care, surveillance, and research. Some combination of *iucA, iroB, peg-344, rmpA,* and *rmpA2* are most commonly used, but it is unclear what combination of genotypic or phenotypic markers (e.g. siderophore concentration, mucoviscosity) most accurately predicts the hypervirulent phenotype. Further, acquisition of antimicrobial resistance may affect virulence and confound identification. Therefore, 49 *K. pneumoniae* strains that possessed some combination of *iucA, iroB, peg-344, rmpA,* and *rmpA2* and had acquired resistance were assembled and categorized as hypervirulent hvKp (hvKp) (N=16) or cKp (N=33) via a murine infection model. Biomarker number, siderophore production, mucoviscosity, virulence plasmid's Mash/Jaccard distances to the canonical pLVPK, and Kleborate virulence score were measured and evaluated to accurately differentiate these pathotypes. Both stepwise logistic regression and a CART model were used to determine which variable was most predictive of the strain cohorts. The biomarker count alone was the strongest predictor for both analyses. For logistic regression the area under the curve for biomarker count was 0.962 (P = 0.004). The CART model generated the classification rule that a biomarker count = 5 would classify the strain as hvKP, resulting in a sensitivity for predicting hvKP of 94% (15/16), a specificity of 94% (31/33), and an overall accuracy of 94% (46/49). Although a count of ≥ 4 was 100% (16/16) sensitive for predicting hvKP, the specificity and accuracy decreased to 76% (25/33) and 84% (41/49) respectively. These findings can be used to inform the identification of hvKp.

Importance:

Hypervirulent *Klebsiella pneumoniae* (hvKp) is a concerning pathogen that can cause life-threatening infections in otherwise healthy individuals. Importantly, although strains of hvKp have been acquiring antimicrobial resistance, the effect on virulence is unclear. Therefore, it is of critical importance to determine whether a given antimicrobial resistant *K. pneumoniae* isolate is hypervirulent. This report determined which combination of genotypic and phenotypic markers could most accurately identify hvKp strains with acquired resistance. Both logistic regression and a machine-learning prediction model demonstrated that biomarker count alone was the strongest predictor. The presence of all 5 of the biomarkers *iucA, iroB, peg-344, rmpA,* and *rmpA2* was most accurate (94%); the presence of ≥ 4 of these biomarkers was most sensitive (100%). Accurately identifying hvKp is vital for surveillance and research, and the availability of biomarker data could alert the clinician that hvKp is a consideration, which in turn would assist in optimizing patient care.

Introduction:

Hypervirulent *Klebsiella pneumoniae* (hvKp) is capable of causing a variety of severe infectious syndromes in healthy hosts from the community as well as in hospitalized hosts (1, 2). By contrast, classical *K. pneumoniae* (cKp) primarily causes healthcare-associated infections (1). The ability to accurately identify hvKp has important clinical implications, such as identifying occult abscesses to enable source control and monitoring for endophthalmitis (3). Presently, clinical microbiology laboratories are unable to differentiate between these two pathotypes.

As a first step to be able to identify hvKp, our group previously designed a study in which two strain cohorts were assembled. Clinical criteria were used to assemble an hvKp-rich strain cohort; random blood isolates from geographical regions predicted to have a low prevalence of hvKp were used to create the cKp-rich strain cohort. Subsequent assessment of these strains identified 5 genotypic biomarkers that achieved a diagnostic accuracy of ≥0.95 for differentiating the hvKp-rich and cKp-rich strain cohorts, namely *iucA, iroB, peg-344, rmpA,* and *rmpA2.* This epidemiological analysis was experimentally validated in a murine sepsis model (4).

Subsequently, other investigators have used the presence of some of these markers to define a strain as being hypervirulent (5-8); often by the presence of some combination of *iucA* and *rmpA/A2* (9-12). However, it is important to note that *iucA, iroB, peg-344, rmpA,* and *rmpA2* biomarkers are linked on the canonical hvKp virulence plasmid pLVKP (13), which has been shown to be critical for the expression of the hvKp hypervirulent phenotype (14-16). The absence of some of these markers may reflect the presence of an incomplete virulence plasmid, which in turn could affect the hypervirulent phenotype. In our study (4) of the strains defined as hvKp (LD_{50} <10^7 CFU) by virulence in the murine sepsis model, 80/85 (94%) strains possessed 5/5 of these biomarkers, 4/85 (5%) possessed 4/5 biomarkers (*rmpA2* was absent) and only 1/85 strains possessed 3/5 of these markers (*rmpA2*, *iucA* were absent) (4). Of note, of the

strains that did not possess any of these 5 markers only 1/82 (1.2%) had an LD_{50} of <1 x 10^7 CFU (the LD_{50} of that strain was 3.0 x 10^5 CFU) (4). Further, the strains assessed in this initial study were identified by clinical syndrome irrespective of their antimicrobial susceptibility profile. As a result, these data did not adequately address the impact of possessing fewer than five of these markers or the effect of acquired antimicrobial resistance on the hypervirulent phenotype.

To address this knowledge gap, two collections of *K. pneumoniae* strains were assembled that possessed acquired antimicrobial resistance and the presence of the biomarkers *iucA, iroB, peg-344, rmpA,* and *rmpA2*. Forty-nine strains were identified that possessed 1-5 of these genes, their virulence was assessed in a CD1 murine subcutaneous challenge model, and these strains were categorized as hvKp or cKp (17). Next, we hypothesized that some combination of biomarker number, total siderophore production, mucoviscosity, Kleborate virulence score (18), and Mash/Jaccard distances (19, 20) to the canonical pLVPK could accurately determine the pathotype of these *K. pneumoniae* strains. Analysis by both stepwise logistic regression and a machine-learning based prediction model using a classification and regression tree (CART) demonstrated that the biomarker count alone proved to be the strongest predictor and was highly accurate. These data are of value for informing the development of a diagnostic test for the clinical microbiology laboratory as well as guiding surveillance, translational and clinical research, and patient care.

Methods and Materials:

Bacterial strains. Forty-nine strains used in this study and selected genotypic and phenotypic characteristics are listed in Supplementary Table 1. Fifteen strains were obtained from the U.S. Centers for Disease Control and Prevention and 34 were obtained from the Walter Reed Army Institute of Research (WRAIR). All strains were stored at -70°C or -80°C prior to use.

CD1 mouse subcutaneous (SQ) challenge infection models.

Animal studies were reviewed and approved by the Veterans Administration Institutional Animal Care Committee and the University at Buffalo-SUNY and were carried out in strict accordance with the recommendations in the guidelines delineated in the "NIH Guide for the Care and Use of Laboratory Animals"(revised 1985) and the "Ethics of Animal Experimentation Statement" (Canadian Council on Animal Care, July 1980) as monitored by the Institutional Animal Care and Use Committee (IACUC). All efforts were made to minimize suffering. Veterinary care for the animals was supplied by the staff of the Veterans Administration Animal Facility under the direction of a fully licensed veterinarian. Studies were performed as described (4). The model is clinically relevant since hypervirulent *K. pneumoniae* (hvKp) may enter through micro-breaks in the skin, similar to *S. aureus*. Further, published data has established its ability to discriminate between cKp and hvKp strains (4). Lastly, we have established that this model produces similar results as intra-peritoneal challenge (17). The various challenge inocula and number of mice used for each strain are delineated in Supplementary Table S1. The animals were closely monitored for 14 days after challenge for the development of the study endpoints, survival or severe illness (*in extremis* state)/death, which was recorded as a dichotomous variable. Signs that were monitored and which resulted in immediate euthanasia using methods consistent with the recommendations of the American Veterinary Medical Association Guidelines included hunched posture, ruffled fur, labored breathing, reluctance to move, photophobia, and dehydration.

Quantitative siderophore assay.

Strains were grown individually overnight at 37°C in iron-chelated M9 minimal media containing casamino acids (c-M9-CA) (21) and culture supernatants were assessed using the chromeazurol S dye assay as described (22). A minimum of 3 biological assays with 3 technical repeats were performed and the results were reported as the mean ± the SD.

Mucoviscosity assay.

Since the hypermucoviscosity phenotype is more strongly associated with the hypervirulent phenotype compared to capsule (23, 24) this variable was chosen for study. Strains grown overnight in lysogeny broth (LB) were used to inoculate 10 mL of either LB or c-M9-CA plus added trace elements (5 µg/mL $CaCl_2$, 1 µg/mL $CoCl_2$, 20 µg/mL $MgCl_2$, 10 µg/mL $MnCl_2$) (c-M9-CA-te) to a starting OD_{600} of approximately 0.2. Strains were grown for 24 hours, and the assay was performed as described (17). A minimum of 3 biological assays were performed and the results were reported as the mean ± the SD.

Growth/survival in human ascites fluid

Growth/survival experiments in human ascites were performed as previously described (25). The procedures for obtaining human ascites fluid were reviewed and approved by the Western New York Veterans Administration Institutional Review Board; informed consent for ascites fluid was waived because it was collected from de-identified patients who were undergoing therapeutic paracentesis for symptoms due to abdominal distension. These individuals were not being treated with antimicrobials and were not infected with human immunodeficiency virus, hepatitis B virus, or hepatitis C virus. The ascites fluid was cultured to confirm sterility, divided into aliquots, and stored at -80°C. For these experiments ascites obtained from a single individual was used.

Sequencing and Assembly.

All strains were sequenced from Illumina MiSeq or NextSeq platforms. CDC strains were assembled with Spades (Phoenix v1.) WRAIR strains were assembled with de novo assembly using Newbler (v2.7) with minimum thresholds for contig size and coverage set at 200 bp and 49.5×, respectively. Long-read sequencing was completed on a subset of WRAIR isolates using

with PacBio (Pacific Biosciences: Menlo Park, CA. USA) or MinION (Oxford Nanopore Technologies: Oxford, United Kingdom) platforms. Assembly was performed using the Trycycler tool (26). After filtering the bottom 5% of reads by quality score, reads were subsampled, with replacement, 12 times. Four assemblies were generated using Flye, Miniasm, and Raven using the 12 independent subset of reads. The consensus assembly was constructed with the assistance of Trycycler and was finally polished with short reads from Illumina MiSeq or NextSeq platforms.

Identification of acquired antimicrobial resistance and virulence genes.

Study strains possessed acquired antimicrobial resistance and some combination of the biomarkers *iucA, iroB, peg-344, rmpA,* and *rmpA2.* Kleborate (v2.3.2) was used to screen each isolate for any relevant virulence and antimicrobial resistance determinants (18). An agglomerative virulence score was determined by the presence or absence of the five genes: *iroB, iucA, peg-344*, and plasmid-borne *rmpA* and *rmpA2*. Possible loss-of-function mutations in the virulence biomarkers were predicted based on the Kleborate results as well as BLASTN and BLASTP analysis (Supplementary Table S1). All biomarkers except for *peg-344* are within the interrogative scope of Kleborate. The metabolite transporter *peg-344* was included in the agglomerative scoring due to its high association with fully virulent *K. pneumoniae* isolates (4). To calculate the prevalence of *iroB, iucA, peg-344*, *rmpA* and *rmpA2* in MDR *K. pneumoniae,* an unbiased collection of 3,123 clinical isolates (27) from global origins was queried using BLASTN and Kleborate (Supplementary Table S2). Strains also underwent antimicrobial susceptibility testing using reference broth microdilution with panels that were prepared in-house according to the Clinical and Laboratory Standards Institute (CLSI) guidelines (28) and stored at -70°C until use. Results of testing were interpreted according to guidance provided in CLSI document M100 (29), with the exception of tigecycline, for which criteria established by the U.S. Food and Drug

Administration were applied. *E. coli* ATCC 25922, *P. aeruginosa* ATCC 27853, and *K. pneumoniae* ATCC 700603 were used for quality control.

Mash and Jaccard distance indices.

To determine genetic relatedness of each isolates' virulence plasmid to the canonical pLVPK virulence plasmid (13), Mash distance (19) and Jaccard similarity indices (20) were calculated between closed plasmid sequences containing virulence biomarkers and pLVPK for the 29 closed genomes (hvKp N=9, cKp N=20). Mash distance calculation was performed with the dependency-free binary provided freely by the developers (https://github.com/marbl/mash). The Jaccard similarity is defined by the proportion of genes shared by two comparator plasmids over the set of all observed genes in the two plasmids. Hence, the Jaccard distance is the complement of the similarity value and represents the proportion of all observed genes that are not shared in the two comparator plasmids. The Jaccard metric was calculated from Roary output (30), which provided a gene presence or absence table from a set of annotated input genome assemblies. The Vegan package in R was used to calculate Jaccard distance between each pairwise comparison of virulence plasmids to pLVPK.

Comparative genomics

Two strains (AKp14 and AKp35) carried all five biomarkers and did no exhibit an hypervirulent phenotype in the murine model. Gene content analysis was performed using Roary (30) to investigate possible genetic factors involved in the loss of virulence. For strain AKp14 (ST23), we compared its genome to those of all ST23 confirmed hvkp in this study (AKp4, AKp10, AKp25, AKp27 and RCA1) (Supplementary Table S3). For strain AKp35 (ST412), there was no other ST412 with available virulence data to compare. Instead, its genome was compared to all the 16 strains (various ST) with confirmed hvKp phenotype in this study (Supplementary Table S3).

Statistical Analyses.

The LD_{50} was estimated using a logistic regression model with the factors for strain and inoculum (CFU/mL). The estimated LD_{50} was given as $(\hat{s}_i + \hat{\beta}_0)/\hat{\beta}_1$ where $\hat{s}_i$ is the strain effect, $\hat{\beta}_0$ is the intercept and $\hat{\beta}_1$ is the regression parameter associated with inoculum. A model fit statistical analysis was performed (c-stat). C was equal to 0.879 (C = 1.0 is a perfect fitting model), which supports that the model was a good fit. The k-means cluster analysis was employed to examine whether distinct groupings could be formed based on the LD_{50} values, using two, three, or more clusters. The analysis revealed two well-defined clusters that were categorized as cKP (LD_{50} of > 1 x 10^7 CFU) and hvKP (LD_{50} of ≤ 1 x 10^7 CFU).

Mucoviscosity when grown in LB or c-M9-CA-te, biomarker numbers (*iucA*, *iroB*, *peg-344*, *rmpA*, *rmpA2*), Mash and Jaccard distances, and Kleborate scores for the hvKp and cKp strain cohorts were initially compared using a Mann-Whitney test. Siderophore production for the two cohorts was compared utilizing an unpaired t test (Prism version 9.5.1).

Exact logistic regression was used to model the association of class with each marker of virulence and to estimate the area under the ROC curve (AUC) in both a univariate and multivariate fashion. The AUC ranges from 0.5 to 1 and is a measure of the percent agreement between the observed and predicted values. A value of AUC=0.5 implies no predictive utility. A value of 1.0 implies 100% accuracy in terms of prediction (SAS).

A machine-learning based prediction model using a classification and regression tree (CART) was used to examine the prediction of the strain cohorts cKP (LD_{50} of > 1 x 10^7 CFU) and hvKP (LD_{50} of ≤ 1 x 10^7 CFU) as a function of count (sum of positive *iucA*, *iroB*, *peg-344*, *rmpA* and

rmpA2 markers), quantitative siderophore production, mucoviscosity when grown in LB or c-M9-CA-te, Mash Distance, Jaccard Distance, and Kleborate Score. The method of cost-complexity pruning was utilized to determine the final tree fit (SAS).

Results:

Strain cohorts

Forty-nine *K. pneumoniae* strains were collected from 8 countries (n= 30 from USA; n= 4 from Germany; n= 3 from Ukraine; n= 3 from Middle East; n= 2 from Thailand; n= 5 from Afghanistan; n= 1 from Uganda; n= 1 from Turkey). Twenty-eight sequence types that possessed some combination of the biomarkers *iucA, iroB, peg-344, rmpA,* and *rmpA2* and antimicrobial resistance other than the ubiquitous ampicillin resistance were chosen for study (Supplementary Table S1). These included the ST23, ST11, ST29, and ST268 hvKp lineages, but also the "high-risk" emerging multi-drug resistant (MDR) lineages ST395, ST307, and ST15**.** Of these 49 strains 34 (69%) possessed an extended-spectrum ß-lactamase (ESBL) and 20 (41%) possessed a carbapenemase; the mean number of resistance genes was 7.51 (range 1-14) (Supplementary Table S1). Of these 49 strains, 17 (35%) possessed all 5 biomarkers, 7 (14%) possessed 4 biomarkers, 19 (39%) possessed 2 biomarkers, and 6 (12%) possessed 1 biomarker (Supplementary Table S1). To place this in perspective, the natural prevalence of biomarkers in an unbiased collection of 3,123 MDR isolates (27) from global origins were assessed. The most represented biomarkers were *peg-344* (8% of isolates), *iucA* (5%) *rmpA2* and *rmpA* (each at 4%) and *iroB* (1%). But only 0.9% (n= 28), 2.7% (n= 84), 1% (n= 30) and 4% (n= 127) of isolates carried 5, 4, 3 or ≤ 2 biomarkers respectively (Supplementary Table S2).

Next, these 49 strains underwent pathotype characterization as defined by virulence in the outbred CD1 SQ challenge model. hvKp were defined by an LD_{50} of ≤1 x 10^7 CFU and cKp was defined by an LD_{50} of > 1 x 10^7 CFU (17). Sixteen (33%) strains were categorized as hvKp

(mean log LD_{50} 4.06 ± 1.91) and 33 (67%) as cKp (mean log LD_{50} 8.91 ± 0.31) (Figure 1A, Table 1).

Phenotypic assessment of the hvKp and cKp strain cohorts

Previous studies have demonstrated that quantitative siderophore production and mucoviscosity assessment held promise for differentiating hvKp from cKp strains (4, 17). Therefore, we evaluated these phenotypes for all 49 strains in the hvKp and cKp strain cohorts.

Quantitative siderophore production was similar for the hvKp (135.8 ± 55.5 µg/mL) and cKp (139.8 ± 84.9 µg/mL) strain cohorts (Figure 1B, Table 1). When grown in LB medium, the hvKp strain cohort (0.194 ± 0.171) was significantly more mucoviscous than the cKp strain cohort (0.1002 ± 0.1097) (P = 0.0113) (Figure 1C, Table 1). When grown in c-M9-CA-te medium, the hvKp (0.366 ± 0.254) was significantly more mucoviscous than the cKp strain cohort (0.1269 ± 0.1893) (P < 0.0001) (Figure 1D, Table 1). These data demonstrate that quantitative siderophore production was unable to distinguish the hvKp and cKp strain cohorts, but quantitative mucoviscosity could.

Genotypic Assessment of the hvKp and cKp strain cohorts

Previous studies demonstrated that the presence of *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2*, which are linked on the hvKp-specific virulence plasmid possessed a >95% accuracy for differentiating the hvKp-rich and cKp-rich strain cohorts (4). Therefore, all 49 strains in the hvKp and cKp strain cohorts were assessed for these genes and their Kleborate virulence score was calculated.

The hvKp strain cohort has a significantly higher marker count (some combination of *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2*) (4.94 ± 0.25) compared to the cKp strain cohort (2.36 ± 1.17) (P <

0.0001) (Figure 2A, Table 1). Of note, all 25 strains that possessed only one or two markers were categorized as cKp based on the murine infection model (Figure 3A, Supplementary Table S1) (31). Of these, 19/25 (76%) possessed *iucA* and *rmpA2* and 5/25 (20%) possessed *iucA* alone. In addition, the Kleborate virulence score was assessed, which is defined as the genes for aerobactin (3 points), yersiniabactin (1 point), and colibactin (1 point). The average Kleborate score was significantly different for the hvKp (4.3 ± 0.70) compared to the cKp (3.5 ± 0.87) (P= 0.001) strain cohort (Figure 2D, Table 1). Individually, within the 16 hvKp strains, 14 (87%) received a Kleborate score of five or four, and the remaining two scored as three on the virulence scale (Figure 3B, Supplementary Table S1). However, a virulence score of five or four also represented 50% of cKp isolates (17/33) (Figure 3B, Supplementary Table S1).

Besides the presence of a biomarker gene, possible loss-of-function mutations were also characterized (Supplementary Table S1). Compared to reference allele, 88% (37/42) of the isolates with *rmpA2* carried a predicted truncated allele. This was observed at similar frequencies in cKp (88% 23/26) and hvKp (87% 14/16) isolates with identical late frameshifts (E96fs*101, E96fs*115 and N118fs*127) being the most represented. Similarly, predicted truncations in *rmpA* were observed in 27% (6/22) of isolates carrying this biomarker despite all being categorized as hvKp phenotypically. For other biomarkers, an early stop codon (K7*) was identified in *iucA* in 5 cKP strains and a single strain carried a late frameshift (T291fs*298) in *peg-344* but was still categorized as hvKp based on the mouse survival data.

Beyond individual biomarkers, it was hypothesized that carriage of an hvKp virulence plasmid would also be predictive of the hvKp pathotype as measured by Mash/Jaccard distances compared to the canonical virulence plasmid pLVKP. Mash/Jaccard distances for pLVLP-like virulence plasmids from the 29 closed genomes (hvKp N=9, cKp N=20) were measured. The

Mash distance was significantly different for the hvKp (0.0086 ± 0.001) compared to cKp (0.04705 ± 0.042) strain cohort (P = 0.0002) (Figure 2B, Table 1). Likewise, the Jaccard distance was significantly different for the hvKp (0.33 ± 0.16) compared to the cKp (0.6718 ± 0.2320) (P = 0.0001) (Figure 2C, Table 1). Each pLVPK-like plasmid can be compared to the canonical pLVPK sequence by its Mash and Jaccard distance. The two distances are highly correlated, but do not scale linearly (Figure 4). Figure 4 show scatterplots constructed using these two distances whereby each point (n=29) represents a strain's pLVPK-like plasmid. As Mash and Jaccard distances approach 0, (i.e., the strain's pLVPK-like sequence and shared gene content is more similar to pLVPK), the strain is more likely to belong to a hvKp pathotype (Figure 4, top left panel, Pathotype). Viewing from different perspectives, we observe that pLVPK-like plasmids that are more similar to pLVPK, are also more likely to carry all five biomarkers within that plasmid (Figure 4 top right panel, # of Biomarkers). Within the same scope, we can extend generalizations at the genome level from the sequences of their respective pLVPK-like plasmids. In general, strains that carry a pLVPK-like that is highly similar to pLVPK are more likely to score ≥3 in Kleborate virulence (Figure 4 bottom left panel, Kleborate Virulence Score) and are also less likely to carry ESBLs/carbapenemases (Figure 4 bottom right panel, AMR) elsewhere in the genome.

<u>Logistic regression and CART prediction models</u>

Biomarker number, total siderophore production, mucoviscosity, Kleborate virulence score, and Mash/Jaccard distances to the canonical pLVPK were analyzed by both stepwise logistic regression and a CART model to determine which factor or combination of factors were most accurate for predicting the hvKp and cKp strain cohorts. The strongest predictor in both models was the presence of all five of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* (Table 2). In the logistic regression model the AUC for biomarker count was 0.962 (P = 0.004). In that model the Jaccard distance was the next most predictive with an AUC of 0.919 (P < 0.001).

Likewise, the CART model predicted biomarker count alone as most accurate. Using a cut-point of possession of all 5 biomarkers resulted in a sensitivity for predicting hvKP of 94% (15/16), a specificity of 94% (31/33), and an overall accuracy of 94% (46/49). A count of ≥ 4 resulted in a 100% (16/16) sensitivity for predicting hvKp, but specificity was decreased to 76% (25/33) and accuracy decreased to 84% (41/49). The Jaccard distance, Mash distance, mucoviscosity (c-M9-CA-te), Kleborate score, mucoviscosity (LB), and siderophore concentration demonstrated accuracies of 90%, 86%, 80%, 78%, 71%, and 49% respectively (Table 2).

Assessment of cKp strains that possessed all five of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2*.

Although imperfect, the presence of all five of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* was 94% accurate for predicting a strain that exhibited the hvKp phenotype. However, two strains, AKp14 and AKp35, both of which possessed all 5 biomarkers, exhibited a cKp phenotype as defined by the murine infection model (Supplementry Table S1). Further, both strains possessed a nearly complete hvKp-specific virulence plasmid based on Mash/Jaccard distances (Figure 2B and 2C) with the exception of an approximately 15 kb region of pLVPK missing in AKp35 as well as 18 other cKp isolates (Figure 5). Besides the pLVPK plasmid, the entire gene content of AKp14 (ST23) was compared to those of all ST23 with a confirmed hvKp phenotype in this study (AKp4, AKp10, AKp25, AKp27 and RCA1) (Supplementary Table S3). This revealed that 4,862 orthologous genes were shared by all ST23 isolates but that AKp14 uniquely carried 264 genes (due to a large insertion in the chromosome, the carriage of a prophage, and the acquisition of 3 plasmids of 45.4, 35 and 6.5 kb), and uniquely lacked 3 genes annotated as a putative ArsR family transcriptional regulator, DdpA dipeptide-binding protein, and alkanesulfonate ABC transporter permease. For strain AKp35 (ST412), no other ST412 isolate with available virulence data was available for comparison. Instead, the gene content of AKp35 was compared to all 16 strains (various ST) with confirmed hvKp phenotype in

this study. While 4,203 orthologous genes were shared by all isolates, 98 genes were uniquely found in AKp35 and, importantly, 22 genes were uniquely missing in this isolate (Supplementary Table S3). Thirteen of these 22 genes were missing because of a large chromosomal deletion and belonged to a functional cluster coding for the cytochrome c maturation system (*ccmA-H*). Finally, to gain further insight into this phenotypic and genotypic virulence disparity a growth/survival analysis was performed for all 49 strains from this study in human ascites ex vivo. This medium was chosen because it mimics the in vivo growth environment and possesses complement and likely antimicrobial peptide mediated antimicrobial activity. Interestingly, by 24 hours 48/49 strains were able to grow/survive in ascites. But AKp14 was unable to do so (Supplementary Figure 1). These data support that AKp14 has a biofitness deficiency that is independent of the possession of biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* and a nearly complete hvKp-specific virulence plasmid.

Discussion:

The goal of this report was to test the hypothesis that some combination of pre-selected genotypic and phenotypic factors could differentiate hvKp and cKp strains that have acquired varying degrees of antimicrobial resistance. The effect of resistance on virulence is unresolved and may depend on the strain, the number and nature of resistance genes, and whether resistance genes on mobile elements are linked with virulence genes (32, 33). However, in some cases antimicrobial resistance may decrease virulence (34-37). Further, and more importantly, there are numerous examples of MDR- and XDR-cKp strains acquiring only a portion of the hvKp-specific virulence plasmid with the resultant pathotype designation of hvKp in the absence of their virulence potential assessed in valid infection models. Therefore, taken together, we felt it was important to assess a variety of variables that held the potential for accurately identifying the hypervirulent phenotype in this strain population. Data from this study support that the presence of all five biomarkers (*iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2*) was

most accurate in both the logistic regression model (AUC 0.962) and the CART model (94% accuracy) (Table 2). In support of this concept, Jaccard and Mash distances in comparison with the canonical pLVPK were the second-most accurate factors for predicting the hvKp strain cohort with an AUC of 0.919 and 0.90 respectively in the logistic regression model and accuracies of 90% and 86% in the CART model using the defined cut-points (Table 2). However, it is important to note that all the strains studied in this report had acquired antimicrobial resistance beyond the ubiquitous ampicillin resistance in *K. pneumoniae*. This comparison may not be valid for strains that have not acquired additional antimicrobial resistance. These data are consistent with our earlier study in which 94% of the strains in the hvKp-rich cohort possessed all 5 biomarkers, which likely reflects the presence of a complete hvKp virulence plasmid. Notably, these strains were identified by clinical syndrome irrespective of their antimicrobial susceptibility profile (4). Interestingly, only 0.9% of a 3,123 MDR *K. pneumoniae* global strain set (27) carried all 5 biomarkers, which we find here to be most predictive of hypervirulence.

The reasonable accuracy of differentiating hvKp from cKp strains using only the five biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* was particularly encouraging. Assays for these markers could be readily developed and validated for use by clinical microbiology laboratories. It is important to note that *iutA* and *iroE* can be present on the chromosome and should not be used to differentiate hvKp from cKp (1). Phenotypic assays, such as those assessed in this study, are more challenging to implement outside of the research setting. Likewise, a whole genome genetic analysis could be equally or more insightful, but presently is best suited for research purposes and is not yet ready for use in the clinical microbiology laboratory.

The findings from this report hold promise for clinical utility. The ability to differentiate hvKp from cKp is needed for optimal clinical management. For example, hvKp infection requires vigilance

for endophthalmitis and assessment for occult abscesses that could require source control. Further, the sites of infection due to hvKp could dictate modifications of the antimicrobial regimen to optimize tissue concentrations (e.g. prostate, central nervous system (CNS)) and may affect the duration of therapy. This is even more important for antimicrobial resistant hvKp (e.g. CR-hvKp) due to more limited treatment options. Although the classical clinical syndrome of an otherwise healthy individual from the community presenting with hepatic abscess, multiple sites of infection, or unusual sites of infection for cKp (e.g. endophthalmitis, CNS infection, or necrotizing fasciitis) is highly suggestive of hvKp, an increasing number of non-classical presentations, and healthcare-associated infections are being reported (38-40). This dictates the need for an accurate clinical microbiology laboratory-based test to identify hvKp. Although some laboratories have used the string test as a phenotype to identify hvKp, it lacks optimal sensitivity or specificity (4, 41, 42). The genotypic markers described in this study could form the basis of such a test, but these data demonstrate the need for all 5 biomarkers to be included (43). Although the presence of all 5 biomarkers was most accurate, using a cut-point of 4 markers versus 5 increases the sensitivity to 100% (versus 94%), albeit this comes at the cost of decreased specificity (76% versus 94%) and accuracy (84% versus 92%). Nonetheless for the clinical perspective, over-diagnosing hvKp infection could be considered more pragmatic since the consequences of missing such an infection may be significant (e.g., loss of vision).

Putative loss-of-function mutations were observed in some biomarkers and, in particular, *rmpA* and *rmpA2* due to frameshift-causing mutations in poly(G) tracks, as previously described (44, 45). Whether these presumed truncated alleles retain functionality remains to be established but, in this study, their occurrence was identical in hvKp and cKp populations. Likewise, while it cannot be excluded that putative loss-of-function mutations seen in *iuc3* plays a role in the low virulence of these strains, the remaining 19 strains with wild type *iuc* and a biomarker count ≤ 2 also exhibited low virulence in the murine model. However similar to the concept of which

marker number cut-point should be utilized, even if a strain has a mutation in one of these genes that results in loss-of-function and perhaps decreased virulence, it is safer to overcall a strain as being hvKp.

Although the accuracy of predicting a strain to be either hvKp or cKp strain was 94% based on the possession of all 5 biomarkers, it was imperfect. This is due to the complexity of pathogenesis, which is a multifactorial process. The possession of defined virulence factors requisite for the hypervirulent phenotype does not necessarily reflect its overall pathogenic potential. It may be missing undefined virulence traits that were not measured, produce less surface polysaccharides which may facilitate the acquisition of resistance genes (46), or have decreased biofitness due to a variety of other factors such as antimicrobial resistance, increased sensitivity to antimicrobial factors, or other occult metabolic changes. This proved to be the case for two of the strains in our study (AKp14 and AKp35). While it remains to be functionally established, it is possible that AKp35 lost some of its virulence capabilities because of the deletion of the cytochrome c maturation system (*ccmA-H*), a chromosomal cluster ubiquitous in Gram-negative bacteria and previously linked to growth in low iron, intracellular infection, and virulence of *L. pneumophila* (47). By contrast, no major chromosomal deletions were observed in AKp14. Instead, this ST23 isolate uniquely acquired mobile elements (including 3 plasmids) accruing to a total of 264 (+5%) additional auxiliary genes. While genome reduction is a known trait associated with bacterial pathogens (48) it remains to be established whether the accumulation of plasmids in AKp14 could result in decreased biofitness. Reinforcing this hypothesis, this strain demonstrated significantly decreased growth/survival in human ascites compared to the other 48 strains assessed in this report (Supplementary Figure S1). It is important to note that the ability to grow/survive in ascites does not equate to a hypervirulent phenotype since 32/33 cKp strains were able to do so; the hypervirulent phenotype also requires the possession of hvKp-specific virulence factors.

Conversely, the presence of some, but not all, of the established hvKp-specific virulence factor(s) does not necessarily confer the hypervirulent phenotype. Quantitative siderophore production was reasonably accurate for differentiating the hvKp-rich and cKp-rich cohorts in our initial study in which strains were identified by clinical syndrome irrespective of their antimicrobial susceptibility profile (4), but all 5 biomarkers were present in 94% of those strains. However, in this study, many strains designated as cKp possessed *iucA* and produced high levels of siderophores but possessed an incomplete repertoire of additional virulence factors and perhaps had decreased biofitness. So, the high level of siderophore production in the absence of a full armamentarium of hvKp-specific virulence factors was not sufficient to confer a hypervirulent phenotype. Hopefully future studies can begin to delineate the full repertoire of virulence factors and the genomic markers/changes that could predict an effect on biofitness, which would enable a more refined prediction of pathotypes based solely on genomics.

The Kleborate virulence score is commonly used to assess a strain's virulence. The score ranges from 0-5 and is based on the presence of yersiniabactin (1 point), colibactin (1 point), and aerobactin (3 points) (18). However, this metric performed less well than the presence of all five biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* for differentiating the hvKp and cKp strain cohorts in this study. Using a cut-off of >4.05 derived from the CART model, the Kleborate score's sensitivity was 44% (7/16), its specificity was 94% (31/33), and its accuracy was 78% (38/49). This may be due to the weight given to aerobactin in the Kleborate score and the fact that not all hvKp strains possess yersiniabactin or colibactin and some cKp strains possess these genes. Therefore, the Kleborate virulence score is not optimal for predicting the hypervirulent pathotype.

With the emergence of MDR and extensively drug resistant (XDR) *K. pneumoniae*, including carbapenem-resistant (CR)-Kp, an increasing number of studies have assessed for the presence of MDR-hvKp, XDR-hvKp and CR-hvKp from various strain collections (49). Such strains would represent the ultimate “superbug”, possessing both a hypervirulent pathotype and antimicrobial resistance. Many of these studies have only assessed for genotypic markers and not uncommonly only some, but not all, of the five biomarkers markers *iucA*, *iroB*, *peg-344*, *rmpA*, *rmpA2* that are evaluated in this report (5-7, 12). Although imperfect, as discussed due to potential biofitness issues, assessing for the presence of all five of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* will improve pathotype designation, as opposed to assessing for a more limited number of markers (43). In our study, all 24 strains that possessed only *iucA* and/or *rmpA/A2* were designated as cKp (Supplementary Table S1). Although, less pragmatic, ascites growth curves further enhance accuracy and murine infection models remain the most accurate means to identify hvKp strains (4, 50). Of note, although commonly used, the *Galleria mellonella* infection model is unable to differentiate hvKp from cKp (51).

This study has several limitations. First, these data assessed strains of *K. pneumoniae* that have acquired varying degrees of antimicrobial resistance. It is unclear whether these findings can be extrapolated to antimicrobial susceptible strains. Since antimicrobial resistance may affect biofitness, which in turn could affect a strain’s overall pathogenic potential, it is possible that antimicrobial susceptible strains may not require the possession of all 5 biomarkers to maintain a hypervirulent phenotype. Second, although we were only able to calculate Mash/Jaccard distances for 29 strains, these isolates did capture observed plasmid diversity. Although these genomic markers hold promise, calculation requires accurate plasmid construction, which cannot be achieved with Illumina sequencing alone. Lastly, the biomarkers chosen for this study were based on early data that demonstrated their utility for predicting pathotypes. However additional or alternative genotypic or phenotypic markers, both known and

yet to be identified, may prove to be superior in future studies. The search for the optimal pragmatic test for resolving *K. pneumoniae* pathotypes remains a fluid process.

Acknowledgments

This work was supported by NIH R21 AI123558-01 and 1R21AI141826-01A1 (Dr. Russo) and the Department of Veterans Affairs VA Merit Review (I01 BX004677-01) (Dr. Russo). The funders had no role in the decision to publish or the preparation of this manuscript. This study was also partially funded by the U.S. Army Medical command and the Defense Medical Research and Development Program. The authors are thankful to all the staff of the MRSN. The manuscript has been reviewed by the Walter Reed Army Institute of Research and there is no objection to its presentation. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or reflecting the views of the Department of the Army or the Department of Defense. CDC Disclaimer: The findings and conclusions in this presentation are those solely of the CDC author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry. They would like to thank and acknowledge the Clinical and Environmental Microbiology Branch staff in the CDC Division of Healthcare Quality Promotion who supported this effort.

Ethics approval and consent to participate

A subset of the isolates was collected as part of the public health surveillance activities of the MRSN, as determined by the WRAIR branch director and Human Subjects Protection Branch (HSPB) which granted ethical approval. Informed patient consent was waived as samples were taken under a hospital surveillance framework for routine sampling. The research conformed to the principles of the Helsinki Declaration. A subset of the isolates was collected under CDC surveillance and other public health activities.

Availability of data and materials

Both genomic assemblies and raw sequencing data of the isolates analyzed in this study are publicly available in NCBI database under the Bio Project numbers PRJNA981468 and PRJNA994967.

Competing interests

The authors declare that they have no competing interests.

References:

1. Russo TA, Marr CM. 2019. Hypervirulent Klebsiella pneumoniae. Clin Microbiol Rev 32:e00001-19.
2. Liu C, Du P, Xiao N, Ji F, Russo TA, Guo J. 2020. Hypervirulent Klebsiella pneumoniae is emerging as an increasingly prevalent K. pneumoniae pathotype responsible for nosocomial and healthcare-associated infections in Beijing, China. Virulence 11:1215-1224.
3. Sellick JA, Russo TA. 2018. Getting hypervirulent Klebsiella pneumoniae on the radar screen. Curr Opin Infect Dis doi:10.1097/qco.0000000000000464.
4. Russo TA, Olson R, Fang CT, Stoesser N, Miller M, MacDonald U, Hutson A, Barker JH, La Hoz RM, Johnson JR. 2018. Identification of Biomarkers for Differentiation of Hypervirulent Klebsiella pneumoniae from Classical K. pneumoniae. J Clin Microbiol 56.
5. Spadar A, Perdigão J, Campino S, Clark TG. 2023. Large-scale genomic analysis of global Klebsiella pneumoniae plasmids reveals multiple simultaneous clusters of carbapenem-resistant hypervirulent strains. Genome Med 15:3.
6. Banerjee T, Wangkheimayum J, Sharma S, Kumar A, Bhattacharjee A. 2021. Extensively Drug-Resistant Hypervirulent Klebsiella pneumoniae From a Series of Neonatal Sepsis in a Tertiary Care Hospital, India. Front Med (Lausanne) 8:645955.
7. Wang Y, Hua M, Wang J, Xing W, Chen J, Liu J, Du P. 2022. Clonal Dissemination of Multidrug-Resistant and Hypervirulent Klebsiella pneumoniae Clonal Complex in a Chinese Hospital. Pathogens 11.
8. Kot B, Piechota M, Szweda P, Mitrus J, Wicha J, Grużewska A, Witeska M. 2023. Virulence analysis and antibiotic resistance of Klebsiella pneumoniae isolates from hospitalised patients in Poland. Sci Rep 13:4448.

9. Wu H, Li D, Zhou H, Sun Y, Guo L, Shen D. 2017. Bacteremia and other body site infection caused by hypervirulent and classic Klebsiella pneumoniae. Microb Pathog 104:254-262.

10. Zhang R, Lin D, Chan EW, Gu D, Chen GX, Chen S. 2016. Emergence of Carbapenem-Resistant Serotype K1 Hypervirulent Klebsiella pneumoniae Strains in China. Antimicrob Agents Chemother 60:709-11.

11. Shankar C, Nabarro LE, Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Daniel JL, Doss CG, Veeraraghavan B. 2016. Draft Genome Sequences of Three Hypervirulent Carbapenem-Resistant Klebsiella pneumoniae Isolates from Bacteremia. Genome Announc 4.

12. Liu C, Du P, Yang P, Lu M, Shen N. 2023. Fusion plasmid enhanced the endemic extensively drug resistant Klebsiella pneumoniae clone ST147 harbored bla(OXA-48) to acquire the hypervirulence and cause fatal infection. Ann Clin Microbiol Antimicrob 22:11.

13. Chen YT, Chang HY, Lai YC, Pan CC, Tsai SF, Peng HL. 2004. Sequencing and analysis of the large virulence plasmid pLVPK of Klebsiella pneumoniae CG43. Gene 337:189-98.

14. Nassif X, Honore N, Vasselon T, Cole ST, Sansonetti PJ. 1989. Positive control of colanic acid synthesis in Escherichia coli by rmpA and rmpB, two virulence-plasmid genes of Klebsiella pneumoniae. Mol Microbiol 3:1349-59.

15. Nassif X, Sansonetti PJ. 1986. Correlation of the virulence of Klebsiella pneumoniae K1 and K2 with the presence of a plasmid encoding aerobactin. Infect Immun 54:603-8.

16. Tang HL, Chiang MK, Liou WJ, Chen YT, Peng HL, Chiou CS, Liu KS, Lu MC, Tung KC, Lai YC. 2010. Correlation between Klebsiella pneumoniae carrying pLVPK-derived loci and abscess formation. Eur J Clin Microbiol Infect Dis 29:689-98.

17. Russo TA, MacDonald U, Hassan S, Camanzo E, LeBreton F, Corey B, McGann P. 2021. An Assessment of Siderophore Production, Mucoviscosity, and Mouse Infection Models for Defining the Virulence Spectrum of Hypervirulent Klebsiella pneumoniae. mSphere 6.

18. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. 2021. A genomic surveillance framework and genotyping tool for Klebsiella pneumoniae and its related species complex. Nat Commun 12:4188.

19. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132.

20. Chung NC, Miasojedow B, Startek M, Gambin A. 2019. Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data. BMC Bioinformatics 20:644.

21. Bulger J, MacDonald U, Olson R, Beanan J, Russo TA. 2017. Metabolite Transporter PEG344 Is Required for Full Virulence of Hypervirulent Klebsiella pneumoniae Strain hvKP1 after Pulmonary but Not Subcutaneous Challenge. Infect Immun 85.

22. Russo TA, Olson R, Macdonald U, Metzger D, Maltese LM, Drake EJ, Gulick AM. 2014. Aerobactin mediates virulence and accounts for increased siderophore production under iron-limiting conditions by hypervirulent (hypermucoviscous) Klebsiella pneumoniae. Infect Immun 82:2356-67.

23. Nassif X, Fournier JM, Arondel J, Sansonetti PJ. 1989. Mucoid phenotype of Klebsiella pneumoniae is a plasmid-encoded virulence factor. Infect Immun 57:546-52.

24. Walker KA, Miller VL. 2020. The intersection of capsule gene expression, hypermucoviscosity and hypervirulence in Klebsiella pneumoniae. Curr Opin Microbiol 54:95-102.

25. Russo TA, Manohar A, Beanan JM, Olson R, MacDonald U, Graham J, Umland TC. 2016. The Response Regulator BfmR Is a Potential Drug Target for Acinetobacter baumannii. mSphere 1.

26. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. 2021. Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biology 22:266.

27. Martin MJ, Stribling W, Ong AC, Maybank R, Kwak YI, Rosado-Mendez JA, Preston LN, Lane KF, Julius M, Jones AR, Hinkle M, Waterman PE, Lesho EP, Lebreton F, Bennett JW, Mc Gann PT. 2023. A panel of diverse Klebsiella pneumoniae clinical isolates for research and development. Microb Genom 9.

28. CLSI. 2018. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That Grow Aerobically. 11th ed.CLSI Standard M07, Wayne, PA.

29. CLSI. 2022. Performance Standards for Antimicrobial Susceptibility Testing. 32nd ed. CLSI Supplement M100, Wayne, PA.

30. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691-3.

31. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: Visualization of Intersecting Sets. IEEE Transactions on Visualization and Computer Graphics 20:1983-1992.

32. da Silva GJ, Mendonça N. 2012. Association between antimicrobial resistance and virulence in Escherichia coli. Virulence 3:18-28.

33. Hennequin C, Robin F. 2016. Correlation between antimicrobial resistance and virulence in Klebsiella pneumoniae. Eur J Clin Microbiol Infect Dis 35:333-41.

34. Espinal P, Pantel A, Rolo D, Marti S, López-Rojas R, Smani Y, Pachón J, Vila J, Lavigne JP. 2019. Relationship Between Different Resistance Mechanisms and Virulence in Acinetobacter baumannii. Microb Drug Resist 25:752-760.

35. Chowdhury FR, Findlay BL. 2023. Fitness Costs of Antibiotic Resistance Impede the Evolution of Resistance to Other Antibiotics. ACS Infectious Diseases doi:10.1021/acsinfecdis.3c00156.

36. Gonzalez-Ferrer S, Peñaloza HF, Budnick JA, Bain WG, Nordstrom HR, Lee JS, Van Tyne D. 2021. Finding Order in the Chaos: Outstanding Questions in Klebsiella pneumoniae Pathogenesis. Infect Immun 89.

37. DeLeo FR, Porter AR, Kobayashi SD, Freedman B, Hao M, Jiang J, Lin YT, Kreiswirth BN, Chen L. 2023. Interaction of multidrug-resistant hypervirulent Klebsiella pneumoniae with components of human innate host defense. mBio doi:10.1128/mbio.01949-23:e0194923.

38. Du Q, Pan F, Wang C, Yu F, Shi Y, Liu W, Li Z, He P, Han D, Zhang H. 2022. Nosocomial dissemination of hypervirulent Klebsiella pneumoniae with high-risk clones among children in Shanghai. Front Cell Infect Microbiol 12:984180.

39. Liao W, Huang QS, Wei D, Xiong Z, Du FL, Xiang TX, Zhang S, Wan LG, Zhang W, Liu Y. 2022. Nosocomial transmission and rearrangement of large resistance-virulence hybrid plasmids between two bacteremic ST11 carbapenem-resistant hypervirulent Klebsiella pneumoniae strains with low fitness cost. Microb Pathog 168:105593.

40. Du P, Liu C, Fan S, Baker S, Guo J. 2022. The Role of Plasmid and Resistance Gene Acquisition in the Emergence of ST23 Multi-Drug Resistant, Hypervirulent Klebsiella pneumoniae. Microbiol Spectr 10:e0192921.

41. Catalan-Najera JC, Garza-Ramos U, Barrios-Camacho H. 2017. Hypervirulence and hypermucoviscosity: Two different but complementary Klebsiella spp. phenotypes? Virulence 8:1111-1123.

42. Yang Q, Jia X, Zhou M, Zhang H, Yang W, Kudinha T, Xu Y. 2020. Emergence of ST11-K47 and ST11-K64 hypervirulent carbapenem-resistant Klebsiella pneumoniae in bacterial liver abscesses from China: a molecular, biological, and epidemiological study. Emerg Microbes Infect 9:320-331.

43. Yan C, Zhou Y, Du S, Du B, Zhao H, Feng Y, Xue G, Cui J, Gan L, Feng J, Fan Z, Fu T, Xu Z, Zhang Q, Zhang R, Cui X, Tian Z, Chen Y, Zhang T, Huang L, Yuan J. 2023. Recombinase-Aided Amplification Assay for Rapid Detection of Hypervirulent Klebsiella pneumoniae (hvKp) and Characterization of the hvKp Pathotype. Microbiol Spectr doi:10.1128/spectrum.03984-22:e0398422.

44. Yu WL, Lee MF, Chang MC, Chuang YC. 2015. Intrapersonal mutation of rmpA and rmpA2: A reason for negative hypermucoviscosity phenotype and low virulence of rmpA-positive Klebsiella pneumoniae isolates. J Glob Antimicrob Resist 3:137-141.

45. Yu WL, Lee MF, Tang HJ, Chang MC, Chuang YC. 2015. Low prevalence of rmpA and high tendency of rmpA mutation correspond to low virulence of extended spectrum β-lactamase-producing Klebsiella pneumoniae isolates. Virulence 6:162-72.

46. Wang S, Ding Q, Zhang Y, Zhang A, Wang Q, Wang R, Wang X, Jin L, Ma S, Wang H. 2022. Evolution of Virulence, Fitness, and Carbapenem Resistance Transmission in ST23 Hypervirulent Klebsiella pneumoniae with the Capsular Polysaccharide Synthesis Gene wcaJ Inserted via Insertion Sequence Elements. Microbiol Spectr 10:e0240022.

47. Naylor J, Cianciotto NP. 2004. Cytochrome c maturation proteins are critical for in vivo growth of Legionella pneumophila. FEMS Microbiol Lett 241:249-56.

48. Murray GGR, Charlesworth J, Miller EL, Casey MJ, Lloyd CT, Gottschalk M, Tucker AWD, Welch JJ, Weinert LA. 2021. Genome Reduction Is Associated with Bacterial Pathogenicity across Different Scales of Temporal and Ecological Divergence. Mol Biol Evol 38:1570-1579.

49. Yang X, Sun Q, Li J, Jiang Y, Li Y, Lin J, Chen K, Chan EW, Zhang R, Chen S. 2022. Molecular epidemiology of carbapenem-resistant hypervirulent Klebsiella pneumoniae in China. Emerg Microbes Infect 11:841-849.

50. Kochan TJ, Nozick SH, Medernach RL, Cheung BH, Gatesy SWM, Lebrun-Corbin M, Mitra SD, Khalatyan N, Krapp F, Qi C, Ozer EA, Hauser AR. 2022. Genomic surveillance for multidrug-resistant or hypervirulent Klebsiella pneumoniae among United States bloodstream isolates. BMC Infect Dis 22:603.

51. Russo TA, MacDonald U. 2020. The Galleria mellonella Infection Model Does Not Accurately Differentiate between Hypervirulent and Classical Klebsiella pneumoniae. mSphere 5.

Table 1. Values for genotypic and phenotypic markers

Variable	Hypervirulent Klebsiella (hvKP)	Classical Klebsiella (cKP)	p-value (hvKP vs. cKP)
Log_{10} (LD_{50})	4.059 ± 1.912	8.911 ± 0.3078	<0.0001, ****
Siderophore	135.8± 55.51	139.8± 84.92	0.8628, ns
Mucoviscosity LB	0.1938 ± 0.1714	0.1002 ± 0.1097	0.0113, *
Mucoviscosity c-M9-CA-te	0.3655 ± 0.2544	0.1269 ± 0.1893	<0.0001, ****
Marker Count	4.938 ± 0.2500	2.364 ± 1.168	<0.0001, ****
Kleborate Score	4.313 ±	3.485 ±	0.0010, ***

	0.7042	0.8704	
Mash	0.00855 ± 0.00995	0.04705 ± 0.04237	0.0002, ***
Jaccard	0.3303 ± 0.1615	0.6718 ± 0.2320	0.0001, ***

Mann-Whitney nonparametric test used due to non-normal data, except for Siderophore analyzed using an unpaired t test.

Table 2. Logistic regression and CART prediction models

Variable	N	Cut-Point**	Sensitivity°	Specificity°	Accuracy°	AUC#	P value*
Siderophore	49	>100$	.82 (13/16)	0.33 (11/33)	0.49 (24/49)	0.502	0.860
Mucoviscosity LB	49	>0.11	.63 (10/16)	0.76 (25/33)	0.71 (35/49)	0.723	0.034
Mucoviscosity c-M9-CA-te	49	>0.088	1.0 (16/16)	0.70 (23/33)	0.80 (39/49)	0.869	0.001
Marker count	49	5	0.94 (15/16)	0.94 (31/33)	0.94 (46/49)	0.962	0.004
Marker count	49	≥4***	1 (16/16)	0.76 (25/33)	0.84 (41/49)		
Kleborate Score	49	>4.05	0.44 (7/16)	0.94 (31/33)	0.78 (38/49)	0.768	0.003

Mash	29	<0.008	0.77 (7/9)	0.9 (18/20)	0.86 (25/29)	0.900	<0.001
Jaccard	29	<0.318	0.77 (7/9)	0.95 (19/20)	0.90 (26/29)	0.919	<0.001

*Exact test about the odds-ratio from logistic regression model

#Area under the ROC Curve

**derived from classification and regression tree (CART) prediction model

***Assessment at a potentially clinically relevant cut-point

ºComputed using cut-point values

$determined from ROC curve

Figures and Legends

Figure 1. Phenotypic markers

Forty-nine strains comprising the hvKp (n=16) and cKp (n=33) cohorts underwent assessment for LD_{50} in the outbred CD1 murine subcutaneous challenge infection model (Panel A), *in vitro* quantitative siderophore production when grown in c-M9-CA, black symbols represent strains that do not possess *iucA* (Panel B), and *in vitro* quantitative mucoviscosity when grown in either LB (Panel C), or c-M9-CA-te medium (Panel D). For Panel A see Supplementary Table S1 for challenge inocula and animal number for each strain. For Panels B a minimum of three biologic with two technical repeats was performed for each strain. For Panels C and D a minimum of three biologic repeats was performed for each strain. Strain cohorts were compared via Mann-

Whitney except siderophore production was compared by an unpaired t test (Prism version 9.5.1). Panel A, **** P < 0.0001. Panel C, * P = 0.0113, Panel D, **** P < 0.0001

Figure 2. Genotypic markers

Forty-nine strains comprising the hvKp (n=16) and cKp (n=33) cohorts underwent assessment for marker count (Panel A), 29 strains comprising the hvKp (n=9) and cKp (n=20) cohorts underwent assessment for Mash distance (Panel B), and for Jaccard distance (Panel C), and all 49 strains underwent assessment for Kleborate score (Panel D). Strain cohorts were compared via Mann-Whitney (Prism version 9.5.1). Panel A, **** P = <0.0001, Panel B, *** P = 0.0002, Panel C, *** P = 0.0001, Panel D, *** P = 0.001

Figure 3. Biomarker distribution

Panel A. UpSet graph for the 49 strains comprising hvKp (n=16) and cKp (n=33) was generated to facilitate the visualization of the presence or absence of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2*. Black circles designate the presence of a given marker whereas grey circles designate its absence. Pathotypes are color coded. The number above each bar represents the number of strains that possess that marker configuration. Panel B: Results are shown as the distribution in proportions of each biomarker count (1-5) and Kleborate virulence score (0-5) for the hvKp and cKp cohorts. The biomarkers count is the presence of some combination of *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2.* The Kleborate virulence score is calculated by the presence of the genes that encode aerobactin (3 points), yersiniabactin (1 point), and colibactin (1 point).

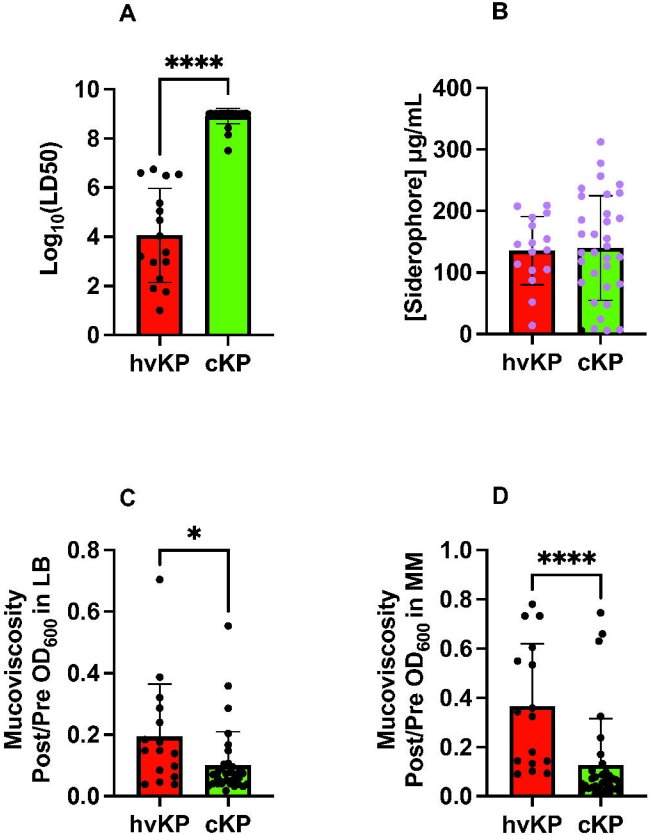
Figure 4. **Mash and Jaccard Distances**

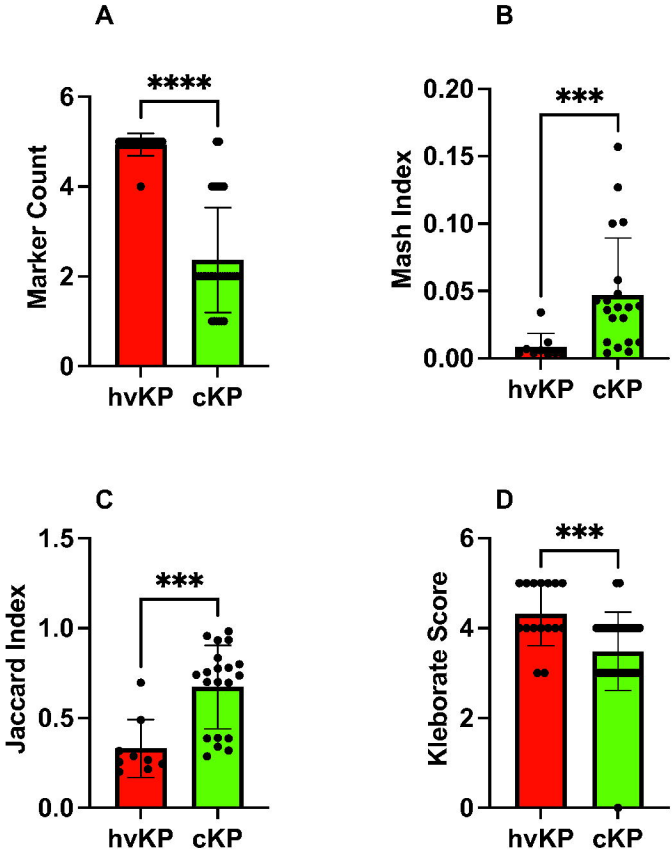
The pLVKP-like virulence plasmid of 29 strains comprising the hvKp (n=9) and cKp (n=20) cohorts, for which long-read sequencing was obtained thereby enabling a closed genome/plasmid, was compared to the canonical pLVPK sequence. Scatterplots of

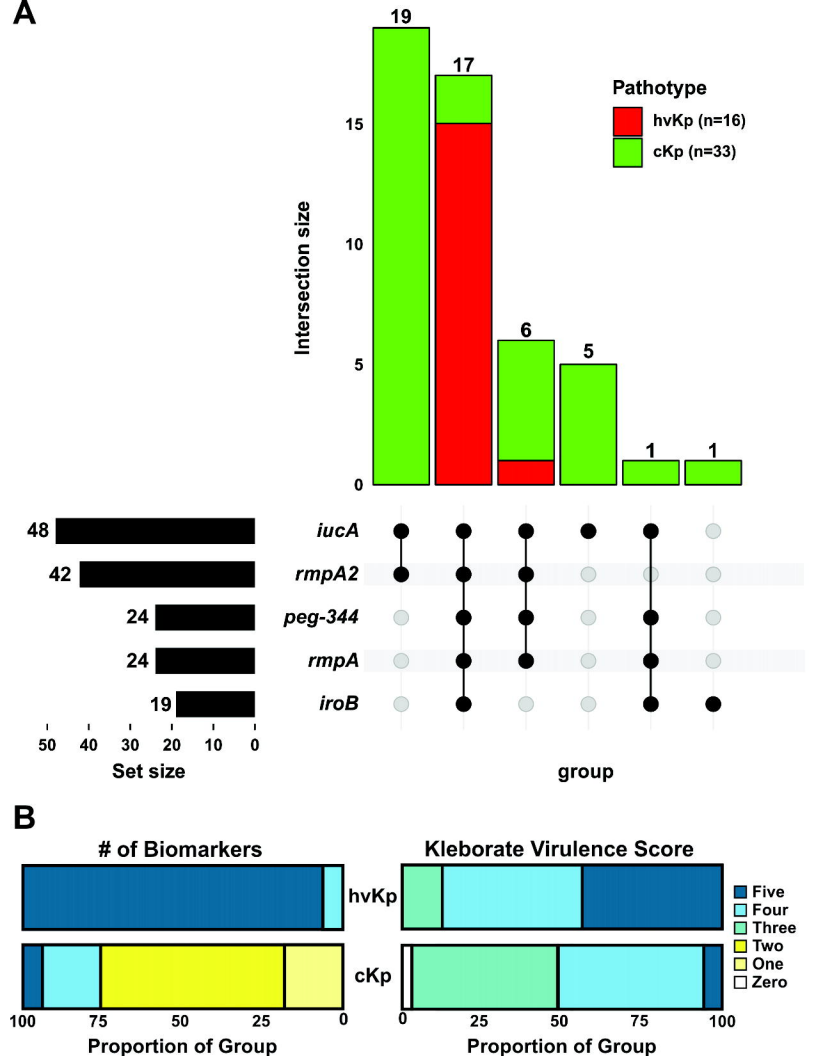
Mash/Jaccard distance to pLVPK were plotted for each strain and colored by pathotype, number of biomarkers (some combination of *iucA*, *iroB*, *peg-344*, *rmpA*, *rmpA2*), Kleborate virulence score (presence of genes that encode aerobactin (3 points), yersiniabactin (1 point), and colibactin (1 point)), and antimicrobial resistance (AMR, presence of ESBL/carbapenemase genes).

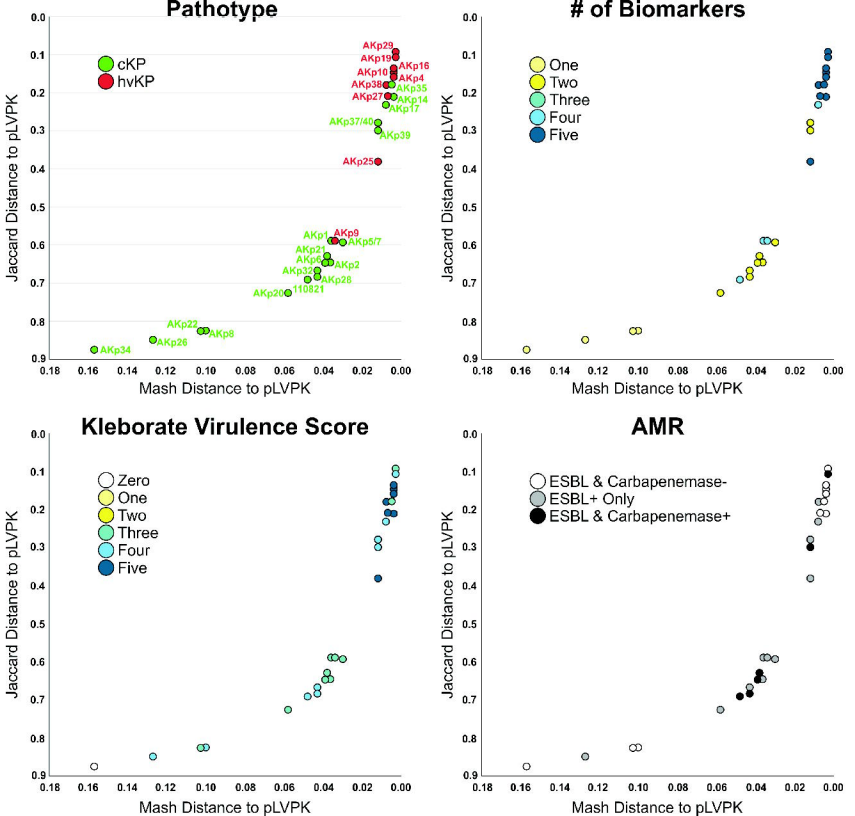
Figure 5. Proksee alignment to pLVPK

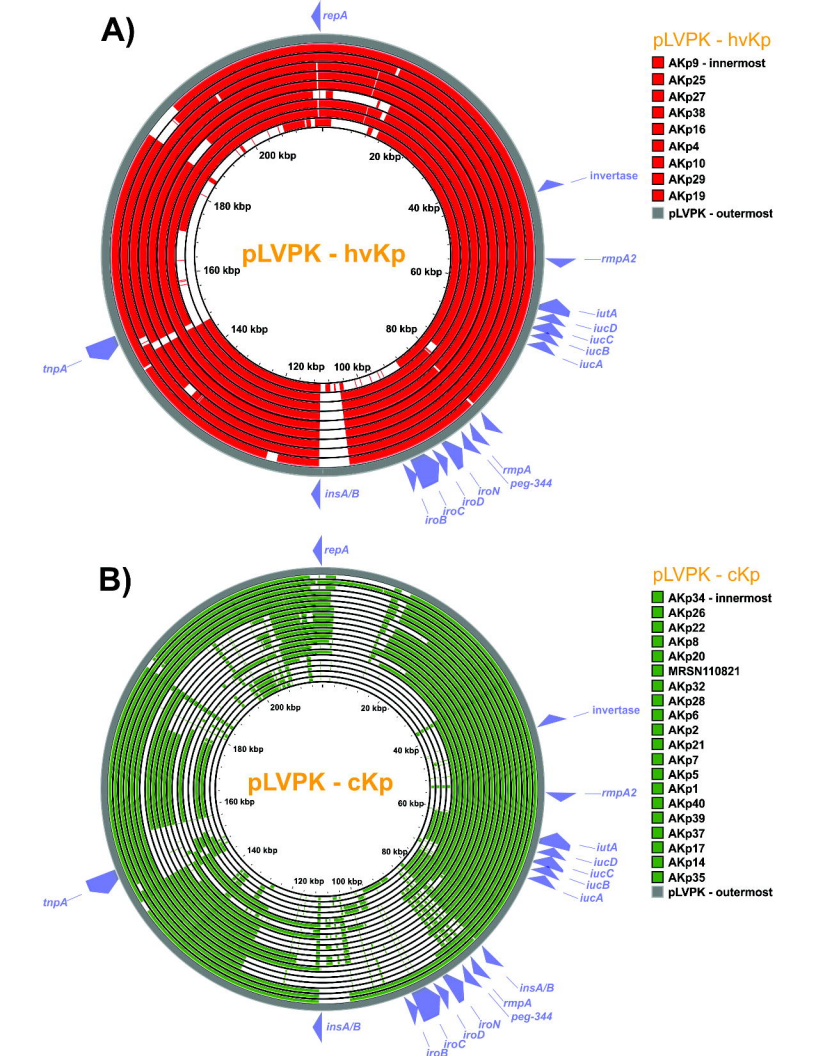
Twenty-nine strains with a closed plasmid harboring any virulence biomarker were compared to the canonical pLVPK. Alignment of the pLVPK sequence to all closed pLVPK-like plasmid sequences from **A)** hvKp (n=9) and **B)** cKp isolates (n=20) were produced with Proksee. Virulence and landmark genes from pLVPK are labeled in each panel. The five virulence biomarkers of interest are concentrated on a roughly 30kb region of pLVPK from *iucA* to *iroB*. Concentric circles are sorted by alignment fraction to pLVPK sequence with the outermost ring representing the isolate with the highest alignment fraction. The average alignment fraction of pLVPK-like plasmid to pLVPK is higher in hvKp isolates compared to cKp. Alignment covering the virulence biomarkers is more extensive in hvKp isolates. Most cKp isolates lack or have non-functioning *iroB, peg-344,* and *rmpA* while possessing a *rmpA2* and *iucA*.


A
****
Log10(LD50)
10
8
6
4
2
0
hvKP
cKP
B
[Siderophore] µg/mL
400
300
200
100
0
hvKP
cKP
C
*
Mucoviscosity
Post/Pre OD600 in LB
0.8
0.6
0.4
0.2
0.0
hvKP
cKP
D
****
Mucoviscosity
Post/Pre OD600 in MM
1.0
0.8
0.6
0.4
0.2
0.0
hvKP
cKP


A
****
Marker Count
6
4
2
0
hvKP
cKP
B
***
Mash Index
0.20
0.15
0.10
0.05
0.00
hvKP
cKP
C
***
Jaccard Index
1.5
1.0
0.5
0.0
hvKP
cKP
D
***
Kleborate Score
6
4
2
0
hvKP
cKP


A
Pathotype
hvKp (n=16)
cKp (n=33)
Intersection size
19
17
6
5
1
1
48
42
24
24
19
iucA
rmpA2
peg-344
rmpA
iroB
Set size
group
B
# of Biomarkers
Kleborate Virulence Score
hvKp
cKp
Five
Four
Three
Two
One
Zero
Proportion of Group
Proportion of Group


Pathotype
cKP
hvKP
AKp29
AKp19
AKp16
AKp10
AKp4
AKp38
AKp35
AKp27
AKp14
AKp17
AKp37/40
AKp39
AKp25
AKp1
AKp9
AKp5/7
AKp21
AKp6
AKp2
AKp32
AKp26
AKp20
110821
AKp22
AKp8
AKp26
AKp34
Jaccard Distance to pLVPK
Mash Distance to pLVPK
# of Biomarkers
One
Two
Three
Four
Five
Jaccard Distance to pLVPK
Mash Distance to pLVPK
Kleborate Virulence Score
Zero
One
Two
Three
Four
Five
Jaccard Distance to pLVPK
Mash Distance to pLVPK
AMR
ESBL & Carbapenemase-
ESBL+ Only
ESBL & Carbapenemase+
Jaccard Distance to pLVPK
Mash Distance to pLVPK


A)
repA
pLVPK - hvKp
AKp9 - innermost
AKp25
AKp27
AKp38
AKp16
AKp4
AKp10
AKp29
AKp19
pLVPK - outermost
200 kbp
20 kbp
180 kbp
40 kbp
160 kbp
60 kbp
140 kbp
80 kbp
120 kbp
100 kbp
pLVPK - hvKp
invertase
rmpA2
iutA
iucD
iucC
iucB
iucA
tnpA
rmpA
peg-344
iroN
iroD
iroC
iroB
insA/B
B)
repA
pLVPK - cKp
AKp34 - innermost
AKp26
AKp22
AKp8
AKp20
MRSN110821
AKp32
AKp28
AKp6
AKp2
AKp21
AKp7
AKp5
AKp1
AKp40
AKp39
AKp37
AKp17
AKp14
AKp35
pLVPK - outermost
200 kbp
20 kbp
180 kbp
40 kbp
160 kbp
60 kbp
140 kbp
80 kbp
120 kbp
100 kbp
pLVPK - cKp
invertase
rmpA2
iutA
iucD
iucC
iucB
iucA
tnpA
insA/B
rmpA
peg-344
iroN
iroD
iroC
iroB
insA/B