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: 4835

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 0.955 (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 88% (14/16), a specificity of 94% (31/33), and an overall accuracy of 92% (45/49). 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 (92%); 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.

77 **Introduction:**

78 Hypervirulent *Klebsiella pneumoniae* (hvKp) is capable of causing a variety of severe infectious
79 syndromes in healthy hosts from the community as well as in hospitalized hosts (1, 2). By
80 contrast, classical *K. pneumoniae* (cKp) primarily causes healthcare-associated infections (1).
81 The ability to accurately identify hvKp has important clinical implications, such as identifying
82 occult abscesses to enable source control and monitoring for endophthalmitis (3). Presently,
83 clinical microbiology laboratories are unable to differentiate between these two pathotypes.
84

85 As a first step to be able to identify hvKp, our group previously designed a study in which two
86 strain cohorts were assembled. Clinical criteria were used to assemble an hvKp-rich strain
87 cohort; random blood isolates from geographical regions predicted to have a low prevalence of
88 hvKp were used to create the cKp-rich strain cohort. Subsequent assessment of these strains
89 identified 5 genotypic biomarkers that achieved a diagnostic accuracy of ≥0.95 for differentiating
90 the hvKp-rich and cKp-rich strain cohorts, namely *iucA, iroB, peg-344, rmpA,* and *rmpA2.* This
91 epidemiological analysis was experimentally validated in a murine sepsis model (4).
92

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

103 this initial study were identified by clinical syndrome irrespective of their antimicrobial
104 susceptibility profile. As a result, these data did not adequately address the impact of
105 possessing fewer than five of these markers or the effect of acquired antimicrobial resistance on
106 the hypervirulent phenotype.
107
108 To address this knowledge gap, two collections of *K. pneumoniae* strains that were assembled
109 for antimicrobial resistance surveillance projects and public health efforts, and possessed
110 acquired antimicrobial resistance, other than the ubiquitous ampicillin resistance mediated by
111 chromosomal SHV ß-lactamase, were screened for the presence of the biomarkers *iucA, iroB,*
112 *peg-344, rmpA,* and *rmpA2* using previously generated short-read sequence data (Illumina, San
113 Diego, CA). Forty-nine strains were identified that possessed 1-5 of these genes. The virulence
114 of these 49 strains was assessed in a CD1 murine subcutaneous challenge model and these
115 strains were categorized as hvKp (LD_{50} of ≤ 1 x 10^7 CFU) or cKp (LD_{50} of > 1 x 10^7 CFU) (17).
116 Next, we hypothesized that some combination of biomarker number, total siderophore
117 production, mucoviscosity, Kleborate virulence score (18), and Mash/Jaccard distances (19, 20)
118 to the canonical pLVPK could accurately determine the pathotype of *K. pneumoniae* strains that
119 have acquired varying degrees of antimicrobial resistance. Analysis by both stepwise logistic
120 regression and a machine-learning based prediction model using a classification and regression
121 tree (CART) demonstrated that the biomarker count alone proved to be the strongest predictor
122 and was highly accurate. These data are of value for informing the development of a diagnostic
123 test for the clinical microbiology laboratory as well as guiding surveillance, translational and
124 clinical research, and patient care.
125
126 **Methods and Materials:**
127 Bacterial strains. Forty-nine strains used in this study and selected genotypic and phenotypic
128 characteristics are listed in Supplementary Tables 1 and 2. Fifteen strains were obtained from

129 the U.S. Centers for Disease Control and Prevention and 34 were obtained from the Walter
130 Reed Army Institute of Research (WRAIR). All strains were stored at -70°C or -80°C prior to
131 use.
132
133 CD1 mouse subcutaneous (SQ) challenge infection models.
134 Animal studies were reviewed and approved by the Veterans Administration Institutional Animal
135 Care Committee and the University at Buffalo-SUNY and were carried out in strict accordance
136 with the recommendations in the guidelines delineated in the "NIH Guide for the Care and Use
137 of Laboratory Animals"(revised 1985) and the "Ethics of Animal Experimentation Statement"
138 (Canadian Council on Animal Care, July 1980) as monitored by the Institutional Animal Care
139 and Use Committee (IACUC). All efforts were made to minimize suffering. Veterinary care for
140 the animals was supplied by the staff of the Veterans Administration Animal Facility under the
141 direction of a fully licensed veterinarian. Studies were performed as described (4). The various
142 challenge inocula and number of mice used for each strain are delineated in Supplemental
143 Table S1. The animals were closely monitored for 14 days after challenge for the development
144 of the study endpoints, survival or severe illness (*in extremis* state)/death, which was recorded
145 as a dichotomous variable. Signs that were monitored and which resulted in immediate
146 euthanasia using methods consistent with the recommendations of the American Veterinary
147 Medical Association Guidelines included hunched posture, ruffled fur, labored breathing,
148 reluctance to move, photophobia, and dehydration.
149
150 Quantitative siderophore assay.
151 Strains were grown individually overnight at 37°C in iron-chelated M9 minimal media containing
152 casamino acids (c-M9-CA) (21) and culture supernatants were assessed using the
153 chromeazurol S dye assay as described (22). A minimum of 3 biological assays with 3 technical
154 repeats were performed and the results were reported as the mean ± the SD.

155

156 Mucoviscosity assay.

157 Strains grown overnight in lysogeny broth (LB) were used to inoculate 10 mL of either LB or c-
158 M9-CA plus added trace elements (5 µg/mL $CaCl_2$, 1 µg/mL $CoCl_2$, 20 µg/mL $MgCl_2$, 10 µg/mL
159 $MnCl_2$) (c-M9-CA-te) to a starting OD_{600} of approximately 0.2. Strains were grown for 24 hours,
160 and the assay was performed as described (17). A minimum of 3 biological assays were
161 performed and the results were reported as the mean ± the SD.

162

163 Growth/survival in human ascites fluid

164 Growth/survival experiments in human ascites and serum were performed as previously
165 described (23). The procedures for obtaining human ascites fluid and serum were reviewed and
166 approved by the Western New York Veterans Administration Institutional Review Board;
167 informed consent for ascites fluid was waived because it was collected from de-identified
168 patients who were undergoing therapeutic paracentesis for symptoms due to abdominal
169 distension. These individuals were not being treated with antimicrobials and were not infected
170 with human immunodeficiency virus, hepatitis B virus, or hepatitis C virus. The ascites fluid was
171 cultured to confirm sterility, divided into aliquots, and stored at -80°C.

172

173 Sequencing and Assembly.

174 All strains were sequenced from Illumina MiSeq or NextSeq platforms. CDC strains were
175 assembled with Spades (Phoenix v1.) WRAIR strains were assembled with de novo assembly
176 using Newbler (v2.7) with minimum thresholds for contig size and coverage set at 200 bp and
177 49.5×, respectively. Long-read sequencing was completed on a subset of WRAIR isolates using
178 with PacBio (Pacific Biosciences: Menlo Park, CA. USA) or MinION (Oxford Nanopore
179 Technologies: Oxford, United Kingdom) platforms. Assembly was performed using the Trycycler
180 tool (https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02483-z). After

181 filtering the bottom 5% of reads by quality score, reads were subsampled, with replacement, 12
182 times. Four assemblies were generated using Flye, Miniasm, and Raven using the 12
183 independent subset of reads. The consensus assembly was constructed with the assistance of
184 Trycycler and was finally polished with short reads from Illumina MiSeq or NextSeq platforms.
185
186 Identification of acquired antimicrobial resistance and virulence genes.
187 Study strains possessed acquired antimicrobial resistance and some combination of the
188 biomarkers *iucA, iroB, peg-344, rmpA,* and *rmpA2.* Kleborate (v2.3.2) was used to screen each
189 isolate for any relevant virulence and antimicrobial resistance determinants (18). An
190 agglomerative virulence score was determined by the presence or absence of the five genes:
191 *iroB, iucA, peg-344*, and plasmid-borne *rmpA* and *rmpA2*. All biomarkers except for *peg-344*
192 are within the interrogative scope of Kleborate. The metabolite transporter *peg-344* was
193 included in the agglomerative scoring due to its high association with fully virulent *K.*
194 *pneumoniae* isolates (4). Strains also underwent antimicrobial susceptibility testing using
195 reference broth microdilution with panels that were prepared in-house according to the Clinical
196 and Laboratory Standards Institute (CLSI) guidelines (24) and stored at -70°C until use. Results
197 of testing were interpreted according to guidance provided in CLSI document M100 (25), with
198 the exception of tigecycline, for which criteria established by the U.S. Food and Drug
199 Administration were applied. *E. coli* ATCC 25922, *P. aeruginosa* ATCC 27853, and *K.*
200 *pneumoniae* ATCC 700603 were used for quality control.
201
202 Mash and Jaccard distance indices.
203 To determine genetic relatedness of each isolates' virulence plasmid to the canonical pLVPK
204 virulence plasmid (13), Mash distance (19) and Jaccard similarity indices (20) were calculated
205 between closed plasmid sequences containing virulence biomarkers and pLVPK. Mash distance
206 calculation was performed with the dependency-free binary provided freely by the developers

207 (https://github.com/marbl/mash). The Jaccard similarity is defined by the proportion of genes
208 shared by two comparator plasmids over the set of all observed genes in the two plasmids.
209 Hence, the Jaccard distance is the complement of the similarity value and represents the
210 proportion of all observed genes that are not shared in the two comparator plasmids. The
211 Jaccard metric was calculated from Roary output (26), which provided a gene presence or
212 absence table from a set of annotated input genome assemblies. The Vegan package in R was
213 used to calculate Jaccard distance between each pairwise comparison of virulence plasmids to
214 pLVPK.
215
216 Statistical Analyses.
217 The LD_{50} was estimated using a logistic regression model with the factors for strain and
218 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,
219 $\hat{\beta}_0$ is the intercept and $\hat{\beta}_1$ is the regression parameter associated with inoculum. The k-means
220 cluster analysis was employed to examine whether distinct groupings could be formed based on
221 the LD_{50} values, using two, three, or more clusters. The analysis revealed two well-defined
222 clusters that were categorized as cKP (LD_{50} of > 1 x 10^7 CFU) and hvKP (LD_{50} of ≤ 1 x 10^7
223 CFU).
224
225 Mucoviscosity when grown in LB or c-M9-CA-te, biomarker numbers (*iucA*, *iroB*, *peg-344*, *rmpA*,
226 *rmpA2*), Mash and Jaccard distances, and Kleborate scores for the hvKp and cKp strain cohorts
227 were initially compared using a Mann-Whitney test. Siderophore production for the two cohorts
228 was compared utilizing an unpaired t test (Prism version 9.5.1).
229
230 Exact logistic regression was used to model the association of class with each marker of
231 virulence and to estimate the area under the ROC curve (AUC) in both a univariate and

232 multivariate fashion. The AUC ranges from 0.5 to 1 and is a measure of the percent agreement
233 between the observed and predicted values. A value of AUC=0.5 implies no predictive utility. A
234 value of 1.0 implies 100% accuracy in terms of prediction (SAS).
235
236 A machine-learning based prediction model using a classification and regression tree (CART)
237 was used to examine the prediction of the strain cohorts cKP (LD_{50} of > 1 x 10^7 CFU) and hvKP
238 (LD_{50} of ≤ 1 x 10^7 CFU) as a function of count (sum of positive *iucA*, *iroB*, *peg-344*, *rmpA* and
239 *rmpA2* markers), quantitative siderophore production, mucoviscosity when grown in LB or c-M9-
240 CA-te, Mash Distance, Jaccard Distance, and Kleborate Score. The method of cost-complexity
241 pruning was utilized to determine the final tree fit (SAS).
242
243 **Results:**
244 Strain cohorts
245 Forty-nine *K. pneumoniae* strains representing 28 sequence types that possessed some
246 combination of the biomarkers *iucA, iroB, peg-344, rmpA,* and *rmpA2* and antimicrobial
247 resistance other than the ubiquitous ampicillin resistance were chosen for study. Of these 49
248 strains 33 (67%) possessed an extended-spectrum ß-lactamase (ESBL) and 20 (41%)
249 possessed a carbapenemase; the mean number of resistance genes was 7.51 (range 1-14)
250 (Supplementary Table S2). Of these 49 strains, 16 (33%) possessed all 5 biomarkers, 8 (16%)
251 possessed 4 biomarkers, 19 (39%) possessed 2 biomarkers, and 6 (12%) possessed 1
252 biomarker (Supplementary Table S1).
253
254 Next, these 49 strains underwent pathotype characterization as defined by virulence in the
255 outbred CD1 SQ challenge model. hvKp were defined by an LD_{50} of ≤1 x 10^7 CFU and cKp was
256 defined by an LD_{50} of > 1 x 10^7 CFU (17). Sixteen (33%) strains were categorized as hvKp

257 (mean log LD_{50} 4.06 ± 1.91) and 33 (67%) as cKp (mean log LD_{50} 8.91 ± 0.31) (Figure 1A, Table
258 1).
259
260 Phenotypic assessment of the hvKp and cKp strain cohorts
261 Previous studies have demonstrated that quantitative siderophore production and mucoviscosity
262 assessment held promise for differentiating hvKp from cKp strains (4, 17). Therefore, we
263 evaluated these phenotypes for all 49 strains in the hvKp and cKp strain cohorts.
264
265 Quantitative siderophore production was similar for the hvKp (135.8 ± 55.5 µg/mL) and cKp
266 (139.8 ± 84.9 µg/mL) strain cohorts (Figure 1B, Table 1). When grown in LB medium, the hvKp
267 strain cohort (0.194 ± 0.171) was significantly more mucoviscous than the cKp strain cohort
268 (0.1002 ± 0.1097) (P = 0.0113) (Figure 1C, Table 1). When grown in c-M9-CA-te medium, the
269 hvKp (0.366 ± 0.254) was significantly more mucoviscous than the cKp strain cohort (0.1269 ±
270 0.1893) (P < 0.0001) (Figure 1D, Table 1). These data demonstrate that quantitative
271 siderophore production was unable to distinguish the hvKp and cKp strain cohorts, but
272 quantitative mucoviscosity could.
273
274 Genotypic Assessment of the hvKp and cKp strain cohorts
275 Previous studies demonstrated that the presence of *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2*,
276 which are linked on the hvKp-specific virulence plasmid possessed a >95% accuracy for
277 differentiating the hvKp-rich and cKp-rich strain cohorts (4). Therefore, all 49 strains in the hvKp
278 and cKp strain cohorts were assessed for these genes and their Kleborate virulence score was
279 calculated.
280
281 The hvKp strain cohort has a significantly higher marker count (some combination of *iucA*, *iroB*,
282 *peg-344*, *rmpA*, and *rmpA2*) (4.88 ± 0.34) compared to the cKp strain cohort (2.36 ± 1.17) (P <

283 0.0001) (Figure 2A, Table 1). Of note, all 25 strains that possessed only one or two markers
284 were categorized as cKp based on the murine infection model (Figure 3A, Supplementary Table
285 S1) (27). Of these, 19/25 (76%) possessed *iucA* and *rmpA2* and 5/25 (20%) possessed *iucA*
286 alone. In addition, the Kleborate virulence score was assessed, which is defined as the genes
287 for aerobactin (3 points), yersiniabactin (1 point), and colibactin (1 point). The average Kleborate
288 score was significantly different for the hvKp (4.3 ± 0.70) compared to the cKp (3.5 ± 0.87) (P=
289 0.001) strain cohort (Figure 2D, Table 1). Individually, within the 16 hvKp strains, 14 (87%)
290 received a score of five or four, and the remaining two scored as three on the virulence scale
291 (Figure 3B, Supplementary Table S1). However, a virulence score of five or four also
292 represented 50% of cKp isolates (17/33)(Figure 3B, Supplementary Table S1).
293
294 Beyond individual biomarkers, it was hypothesized that carriage of an hvKp virulence plasmid
295 would also be predictive of the hvKp pathotype as measured by Mash/Jaccard distances
296 compared to the canonical virulence plasmid pLVKP. Mash/Jaccard distances for pLVLP-like
297 virulence plasmids from the 29 closed genomes (hvKp N=9, cKp N=20) were measured. The
298 Mash distance was significantly different for the hvKp (0.0086 ± 0.001) compared to cKp
299 (0.04705 ± 0.042) strain cohort (P = 0.0002) (Figure 2B, Table 1). Likewise, the Jaccard
300 distance was significantly different for the hvKp (0.33 ± 0.16) compared to the cKp (0.6718 ±
301 0.2320) (P = 0.0001) (Figure 2C, Table 1). Each pLVPK-like plasmid can be compared to the
302 canonical pLVPK sequence by its Mash and Jaccard distance. The two distances are highly
303 correlated, but do not scale linearly (Figure 4). Figure 4 show scatterplots constructed using
304 these two distances whereby each point (n=29) represents a strain's pLVPK-like plasmid. As
305 Mash and Jaccard distances approach 0, (i.e., the strain's pLVPK-like sequence and shared
306 gene content is more similar to pLVPK), the strain is more likely to belong to a hvKp pathotype
307 (Figure 4, top left panel, Pathotype). Viewing from different perspectives, we observe that
308 pLVPK-like plasmids that are more similar to pLVPK, are also more likely to carry all five

309 biomarkers within that plasmid (Figure 4 top right panel, # of Biomarkers). Within the same
310 scope, we can extend generalizations at the genome level from the sequences of their
311 respective pLVPK-like plasmids. In general, strains that carry a pLVPK-like that is highly similar
312 to pLVPK are more likely to score ≥3 in Kleborate virulence (Figure 4 bottom left panel,
313 Kleborate Virulence Score) and are also less likely to carry ESBLs/carbapenemases (Figure 4
314 bottom right panel, AMR) elsewhere in the genome.
315
316 Logistic regression and CART prediction models
317 Biomarker number, total siderophore production, mucoviscosity, Kleborate virulence score, and
318 Mash/Jaccard distances to the canonical pLVPK were analyzed by both stepwise logistic
319 regression and a CART model to determine which factor or combination of factors were most
320 accurate for predicting the hvKp and cKp strain cohorts. The strongest predictor in both models
321 was the presence of all five of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* (Table 2).
322 In the logistic regression model the AUC for biomarker count was 0.955 (P = 0.004). In that
323 model the Jaccard distance was the next most predictive with an AUC of 0.919 (P < 0.001).
324 Likewise, the CART model predicted biomarker count alone as most accurate. Using a cut-point
325 of possession of all 5 biomarkers resulted in a sensitivity for predicting hvKP of 88% (14/16), a
326 specificity of 94% (31/33), and an overall accuracy of 92% (45/49). A count of ≥ 4 resulted in a
327 100% (16/16) sensitivity for predicting hvKp, but specificity was decreased to 76% (25/33) and
328 accuracy decreased to 84% (41/49). The Jaccard distance, Mash distance, mucoviscosity (c-
329 M9-CA-te), Kleborate score, mucoviscosity (LB), and siderophore concentration demonstrated
330 accuracies of 90%, 86%, 80%, 78%, 71%, 49% respectively (Table 2).
331
332 Assessment of cKp strains that possessed all five of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*,
333 and *rmpA2*.

334 Although imperfect, the presence of all five of the biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and
335 *rmpA2* was 92% accurate for predicting a strain that exhibited the hvKp phenotype. However,
336 two strains, AKp14 and AKp35, both of which possessed all 5 biomarkers, exhibited a cKp
337 phenotype as defined by the murine infection model (Supplementry Table S1). Further, both
338 strains possessed a nearly complete hvKp-specific virulence plasmid based on Mash/Jaccard
339 distances (Figure 2B and 2C) with the exception of an approximately 15 kb region of pLVPK
340 missing in AKp35 as well as 18 other cKp isolates (Figure 5). To gain further insight into this
341 phenotypic and genotypic virulence disparity a growth/survival analysis was performed for all 49
342 strains from this study in human ascites ex vivo. This medium was chosen because it mimics
343 the in vivo growth environment and possesses complement and likely antimicrobial peptide
344 mediated antimicrobial activity. Interestingly, by 24 hours 48/49 strains were able to
345 grow/survive in ascites. But AKp14 was unable to do so (Supplementary Figure 1). These data
346 support that AKp14 has a biofitness deficiency that is independent of the possession of
347 biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* and a nearly complete hvKp-specific
348 virulence plasmid.
349

350 **Discussion:**

351 The goal of this report was to test the hypothesis that some combination of pre-selected
352 genotypic and phenotypic factors could differentiate hvKp and cKp strains that have acquired
353 varying degrees of antimicrobial resistance. Data from this study support that only the presence
354 of all five biomarkers (*iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2*) was most accurate in both the
355 logistic regression model (AUC 0.958) and the CART model (92% accuracy) (Table 2). In
356 support of this concept, Jaccard and Mash distances in comparison with the canonical pLVPK
357 were the second-most accurate factors for predicting the hvKp strain cohort with an AUC of
358 0.919 and 0.90 respectively in the logistic regression model and accuracies of 90% and 86% in
359 the CART model using the defined cut-points (Table 2). However, it is important to note that all

360 the strains studied in this report had acquired antimicrobial resistance beyond the ubiquitous
361 ampicillin resistance in *K. pneumoniae*. This comparison may not be valid for strains that have
362 not acquired additional antimicrobial resistance. These data are consistent with our earlier study
363 in which 94% of the strains in the hvKp-rich cohort possessed all 5 biomarkers, which likely
364 reflects the presence of a complete hvKp virulence plasmid. Notably, these strains were
365 identified by clinical syndrome irrespective of their antimicrobial susceptibility profile (4).
366
367 The reasonable accuracy of differentiating hvKp from cKp strains using only the five biomarkers
368 *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* was particularly encouraging. Assays for these markers
369 could be readily developed and validated for use by clinical microbiology laboratories. It is
370 important to note that *iutA* and *iroE* can be present on the chromosome and should not be used
371 to differentiate hvKp from cKp (1). Phenotypic assays, such as those assessed in this study, are
372 more challenging to implement outside of the research setting.
373
374 The findings from this report hold promise for clinical utility. The ability to differentiate hvKp from
375 cKp is needed for optimal clinical management. For example, hvKp infection requires vigilance
376 for endophthalmitis and assessment for occult abscesses that could require source control.
377 Further, the sites of infection due to hvKp could dictate modifications of the antimicrobial
378 regimen to optimize tissue concentrations (e.g. prostate, central nervous system (CNS)) and
379 may affect the duration of therapy. This is even more important for antimicrobial resistant hvKp
380 (e.g. CR-hvKp) due to more limited treatment options. Although the classical clinical syndrome
381 of an otherwise healthy individual from the community presenting with hepatic abscess, multiple
382 sites of infection, or unusual sites of infection for cKp (e.g. endophthalmitis, CNS infection, or
383 necrotizing fasciitis) is highly suggestive of hvKp, an increasing number of non-classical
384 presentations, and healthcare-associated infections are being reported (28-30). This dictates
385 the need for an accurate clinical microbiology laboratory-based test to identify hvKp. Although

386 some laboratories have used the string test as a phenotype to identify hvKp, it lacks optimal
387 sensitivity or specificity (4, 31, 32). The genotypic markers described in this study could form the
388 basis of such a test, but these data demonstrate the need for all 5 biomarkers to be included
389 (33). Although the presence of all 5 biomarkers was most accurate, using a cut-point of 4
390 markers versus 5 increases the sensitivity to 100% (versus 88%), albeit this comes at the cost
391 of decreased specificity (76% versus 94%) and accuracy (84% versus 92%). Nonetheless for
392 the clinical perspective, over-diagnosing hvKp infection could be considered more pragmatic
393 since the consequences of missing such an infection may be significant (e.g., loss of vision).
394
395 Although the accuracy of predicting a strain to be either hvKp or cKp strain was 88% based on
396 the possession of all 5 biomarkers, it was imperfect. This is due to the complexity of
397 pathogenesis, which is a multifactorial process. The possession of defined virulence factors
398 requisite for the hypervirulent phenotype does not necessarily reflect its overall pathogenic
399 potential. It may be missing undefined virulence traits that were not measured, produce less
400 surface polysaccharides which may facilitate the acquisition of resistance genes (34), or have
401 decreased biofitness due to a variety of other factors such as antimicrobial resistance,
402 increased sensitivity to antimicrobial factors, or other occult metabolic changes. This proved to
403 be the case for one of the strains in our study (AKp14). Despite possessing all 5 biomarkers,
404 this strain demonstrated significantly decreased growth/survival in human ascites compared to
405 the other 48 strains assessed in this report (Supplementary Figure S1). Although the
406 mechanism for decreased growth/survival in ascites for AKp14 has not yet been defined, this
407 strain possessed a nearly complete hvKp-specific virulence plasmid, suggesting a chromosomal
408 alteration(s) is responsible. For AKp14 it appears that this significant decrease in biofitness
409 trumps the presence of hvKp specific virulence factors as reflected by a lack of lethality in the
410 murine infection model (Supplementary Table S1). A similar decrease in biofitness is
411 hypothesized for AKp35, but the assessment in ascites was unrevealing. It is important to note

412 that the ability to grow/survive in ascites does not equate to a hypervirulent phenotype since
413 32/33 cKp strains were able to do so; the hypervirulent phenotype also requires the possession
414 of hvKp-specific virulence factors.
415
416 Conversely, the presence of some, but not all, of the established hvKp-specific virulence
417 factor(s) does not necessarily confer the hypervirulent phenotype. Quantitative siderophore
418 production was reasonably accurate for differentiating the hvKp-rich and cKp-rich cohorts in our
419 initial study in which strains were identified by clinical syndrome irrespective of their
420 antimicrobial susceptibility profile (4), but all 5 biomarkers were present in 94% of those strains.
421 However, in this study, many strains designated as cKp possessed *iucA* and produced high
422 levels of siderophores but possessed an incomplete repertoire of additional virulence factors
423 and perhaps had decreased biofitness. So, the high level of siderophore production in the
424 absence of a full armamentarium of hvKp-specific virulence factors was not sufficient to confer a
425 hypervirulent phenotype. These examples demonstrate an unfortunate limitation of a
426 bioinformatic approach when attempting to define virulence. Hopefully future studies can begin
427 to delineate the full repertoire of virulence factors and the genomic markers/changes that could
428 predict an effect on biofitness, which would enable a more refined prediction of pathotypes
429 based solely on genomics.
430
431 The Kleborate virulence score is commonly used to assess a strain's virulence. The score
432 ranges from 0-5 and is based on the presence of yersiniabactin (1 point), colibactin (1 point),
433 and aerobactin (3 points) (18). However, this metric performed less well than the presence of all
434 five biomarkers *iucA*, *iroB*, *peg-344*, *rmpA*, and *rmpA2* for differentiating the hvKp and cKp
435 strain cohorts in this study. Using a cut-off of >4.05 derived from the CART model, the Kleborate
436 score's sensitivity was 44% (7/16), its specificity was 94% (31/33), and its accuracy was 78%
437 (38/49). This may be due to the weight given to aerobactin in the Kleborate score and the fact

438 that not all hvKp strains possess yersiniabactin or colibactin and some cKp strains possess
439 these genes. Therefore, the Kleborate virulence score is not optimal for predicting the
440 hypervirulent pathotype.
441
442 With the emergence of multi- and extensively drug resistant (MDR and XDR, respectively) *K.*
443 *pneumoniae*, including CR-Kp, an increasing number of studies have assessed for the presence
444 of MDR-hvKp, XDR-hvKp and CR-hvKp from various strain collections (35). Such strains would
445 represent the ultimate "superbug", possessing both a hypervirulent pathotype and antimicrobial
446 resistance. Many of these studies have only assessed for genotypic markers and not
447 uncommonly only some, but not all, of the five biomarkers markers *iucA*, *iroB*, *peg-344*, *rmpA*,
448 *rmpA2* that are evaluated in this report (5-7, 12). Although imperfect, as discussed due to
449 potential biofitness issues, assessing for the presence of all five of the biomarkers *iucA*, *iroB*,
450 *peg-344*, *rmpA*, and *rmpA2* will improve pathotype designation, as opposed to assessing for a
451 more limited number of markers (33). In our study, all 24 strains that possessed only *iucA*
452 and/or *rmpA/A2* were designated as cKp (Supplementary Table S1). Although, less pragmatic,
453 ascites growth curves further enhance accuracy and murine infection models remain the most
454 accurate means to identify hvKp strains (4, 36). Of note, although commonly used, the *Galleria*
455 *mellonella* infection model is unable to differentiate hvKp from cKp (37).
456
457 This study has several limitations. First, these data assessed strains of *K. pneumoniae* that
458 have acquired varying degrees of antimicrobial resistance. It is unclear whether these findings
459 can be extrapolated to antimicrobial sensitive strains. Since antimicrobial resistance may affect
460 biofitness, which in turn could affect a strain's overall pathogenic potential, it is possible that
461 antimicrobial sensitive strains may not require the possession of all 5 biomarkers to maintain a
462 hypervirulent phenotype. Second, although we were only able to calculate Mash/Jaccard
463 distances for 29 strains, these isolates did capture observed plasmid diversity. Although these

464 genomic markers hold promise, calculation requires accurate plasmid construction, which
465 cannot be achieved with Illumina sequencing alone. Lastly, the biomarkers chosen for this
466 study were based on early data that demonstrated their utility for predicting pathotypes.
467 However additional or alternative genotypic or phenotypic markers, both known and yet to be
468 identified, may prove to be superior in future studies. The search for the optimal pragmatic test
469 for resolving *K. pneumoniae* pathotypes remains a fluid process.
470

471 **Acknowledgments**

472 This work was supported by NIH R21 AI123558-01 and 1R21AI141826-01A1 (Dr. Russo) and
473 the Department of Veterans Affairs VA Merit Review (I01 BX004677-01) (Dr. Russo). The
474 funders had no role in the decision to publish or the preparation of this manuscript. This study
475 was also partially funded by the U.S. Army Medical command and the Defense Medical
476 Research and Development Program. The authors are thankful to all the staff of the MRSN. The
477 manuscript has been reviewed by the Walter Reed Army Institute of Research and there is no
478 objection to its presentation. The opinions or assertions contained herein are the private views
479 of the authors and are not to be construed as official or reflecting the views of the Department of
480 the Army or the Department of Defense. CDC Disclaimer: The findings and conclusions in this
481 presentation are those solely of the CDC author(s) and do not necessarily represent the views
482 of the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease
483 Registry. They would like to thank and acknowledge the Clinical and Environmental
484 Microbiology Branch staff in the CDC Division of Healthcare Quality Promotion who supported
485 this effort.
486

487 **Ethics approval and consent to participate**

488 A subset of the isolates was collected as part of the public health surveillance activities of the
489 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. These were shared following HRPO Exemption Determination for Protocol #7218 approved by the CDC Human Research Protection Office.

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests.

516 **References:**

517 1. Russo TA, Marr CM. 2019. Hypervirulent Klebsiella pneumoniae. Clin Microbiol Rev
518 32:e00001-19.
519 2. Liu C, Du P, Xiao N, Ji F, Russo TA, Guo J. 2020. Hypervirulent Klebsiella pneumoniae
520 is emerging as an increasingly prevalent K. pneumoniae pathotype responsible for
521 nosocomial and healthcare-associated infections in Beijing, China. Virulence 11:1215-
522 1224.
523 3. Sellick JA, Russo TA. 2018. Getting hypervirulent Klebsiella pneumoniae on the radar
524 screen. Curr Opin Infect Dis doi:10.1097/qco.0000000000000464.
525 4. Russo TA, Olson R, Fang CT, Stoesser N, Miller M, MacDonald U, Hutson A, Barker JH,
526 La Hoz RM, Johnson JR. 2018. Identification of Biomarkers for Differentiation of
527 Hypervirulent Klebsiella pneumoniae from Classical K. pneumoniae. J Clin Microbiol 56.
528 5. Spadar A, Perdigão J, Campino S, Clark TG. 2023. Large-scale genomic analysis of
529 global Klebsiella pneumoniae plasmids reveals multiple simultaneous clusters of
530 carbapenem-resistant hypervirulent strains. Genome Med 15:3.
531 6. Banerjee T, Wangkheimayum J, Sharma S, Kumar A, Bhattacharjee A. 2021.
532 Extensively Drug-Resistant Hypervirulent Klebsiella pneumoniae From a Series of
533 Neonatal Sepsis in a Tertiary Care Hospital, India. Front Med (Lausanne) 8:645955.
534 7. Wang Y, Hua M, Wang J, Xing W, Chen J, Liu J, Du P. 2022. Clonal Dissemination of
535 Multidrug-Resistant and Hypervirulent Klebsiella pneumoniae Clonal Complex in a
536 Chinese Hospital. Pathogens 11.
537 8. Kot B, Piechota M, Szweda P, Mitrus J, Wicha J, Grużewska A, Witeska M. 2023.
538 Virulence analysis and antibiotic resistance of Klebsiella pneumoniae isolates from
539 hospitalised patients in Poland. Sci Rep 13:4448.

540 9. Wu H, Li D, Zhou H, Sun Y, Guo L, Shen D. 2017. Bacteremia and other body site
541 infection caused by hypervirulent and classic Klebsiella pneumoniae. Microb Pathog
542 104:254-262.
543 10. Zhang R, Lin D, Chan EW, Gu D, Chen GX, Chen S. 2016. Emergence of Carbapenem-
544 Resistant Serotype K1 Hypervirulent Klebsiella pneumoniae Strains in China. Antimicrob
545 Agents Chemother 60:709-11.
546 11. Shankar C, Nabarro LE, Devanga Ragupathi NK, Muthuirulandi Sethuvel DP, Daniel JL,
547 Doss CG, Veeraraghavan B. 2016. Draft Genome Sequences of Three Hypervirulent
548 Carbapenem-Resistant Klebsiella pneumoniae Isolates from Bacteremia. Genome
549 Announc 4.
550 12. Liu C, Du P, Yang P, Lu M, Shen N. 2023. Fusion plasmid enhanced the endemic
551 extensively drug resistant Klebsiella pneumoniae clone ST147 harbored bla(OXA-48) to
552 acquire the hypervirulence and cause fatal infection. Ann Clin Microbiol Antimicrob
553 22:11.
554 13. Chen YT, Chang HY, Lai YC, Pan CC, Tsai SF, Peng HL. 2004. Sequencing and
555 analysis of the large virulence plasmid pLVPK of Klebsiella pneumoniae CG43. Gene
556 337:189-98.
557 14. Nassif X, Honore N, Vasselon T, Cole ST, Sansonetti PJ. 1989. Positive control of
558 colanic acid synthesis in Escherichia coli by rmpA and rmpB, two virulence-plasmid
559 genes of Klebsiella pneumoniae. Mol Microbiol 3:1349-59.
560 15. Nassif X, Sansonetti PJ. 1986. Correlation of the virulence of Klebsiella pneumoniae K1
561 and K2 with the presence of a plasmid encoding aerobactin. Infect Immun 54:603-8.
562 16. Tang HL, Chiang MK, Liou WJ, Chen YT, Peng HL, Chiou CS, Liu KS, Lu MC, Tung KC,
563 Lai YC. 2010. Correlation between Klebsiella pneumoniae carrying pLVPK-derived loci
564 and abscess formation. Eur J Clin Microbiol Infect Dis 29:689-98.

565 17. Russo TA, MacDonald U, Hassan S, Camanzo E, LeBreton F, Corey B, McGann P.
566 2021. An Assessment of Siderophore Production, Mucoviscosity, and Mouse Infection
567 Models for Defining the Virulence Spectrum of Hypervirulent Klebsiella pneumoniae.
568 mSphere 6.
569 18. Lam MMC, Wick RR, Watts SC, Cerdeira LT, Wyres KL, Holt KE. 2021. A genomic
570 surveillance framework and genotyping tool for Klebsiella pneumoniae and its related
571 species complex. Nat Commun 12:4188.
572 19. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM.
573 2016. Mash: fast genome and metagenome distance estimation using MinHash.
574 Genome Biol 17:132.
575 20. Chung NC, Miasojedow B, Startek M, Gambin A. 2019. Jaccard/Tanimoto similarity test
576 and estimation methods for biological presence-absence data. BMC Bioinformatics
577 20:644.
578 21. Bulger J, MacDonald U, Olson R, Beanan J, Russo TA. 2017. Metabolite Transporter
579 PEG344 Is Required for Full Virulence of Hypervirulent Klebsiella pneumoniae Strain
580 hvKP1 after Pulmonary but Not Subcutaneous Challenge. Infect Immun 85.
581 22. Russo TA, Olson R, Macdonald U, Metzger D, Maltese LM, Drake EJ, Gulick AM. 2014.
582 Aerobactin mediates virulence and accounts for increased siderophore production under
583 iron-limiting conditions by hypervirulent (hypermucoviscous) Klebsiella pneumoniae.
584 Infect Immun 82:2356-67.
585 23. Russo TA, Manohar A, Beanan JM, Olson R, MacDonald U, Graham J, Umland TC.
586 2016. The Response Regulator BfmR Is a Potential Drug Target for Acinetobacter
587 baumannii. mSphere 1.
588 24. CLSI. 2018. Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria That
589 Grow Aerobically. 11th ed.CLSI Standard M07, Wayne, PA.

590 25. CLSI. 2022. Performance Standards for Antimicrobial Susceptibility Testing. 32nd ed.
591 CLSI Supplement M100, Wayne, PA.
592 26. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D,
593 Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis.
594 Bioinformatics 31:3691-3.
595 27. Lex A, Gehlenborg N, Strobelt H, Vuillemot R, Pfister H. 2014. UpSet: Visualization of
596 Intersecting Sets. IEEE Transactions on Visualization and Computer Graphics 20:1983-
597 1992.
598 28. Du Q, Pan F, Wang C, Yu F, Shi Y, Liu W, Li Z, He P, Han D, Zhang H. 2022.
599 Nosocomial dissemination of hypervirulent Klebsiella pneumoniae with high-risk clones
600 among children in Shanghai. Front Cell Infect Microbiol 12:984180.
601 29. Liao W, Huang QS, Wei D, Xiong Z, Du FL, Xiang TX, Zhang S, Wan LG, Zhang W, Liu
602 Y. 2022. Nosocomial transmission and rearrangement of large resistance-virulence
603 hybrid plasmids between two bacteremic ST11 carbapenem-resistant hypervirulent
604 Klebsiella pneumoniae strains with low fitness cost. Microb Pathog 168:105593.
605 30. Du P, Liu C, Fan S, Baker S, Guo J. 2022. The Role of Plasmid and Resistance Gene
606 Acquisition in the Emergence of ST23 Multi-Drug Resistant, Hypervirulent Klebsiella
607 pneumoniae. Microbiol Spectr 10:e0192921.
608 31. Catalan-Najera JC, Garza-Ramos U, Barrios-Camacho H. 2017. Hypervirulence and
609 hypermucoviscosity: Two different but complementary Klebsiella spp. phenotypes?
610 Virulence 8:1111-1123.
611 32. Yang Q, Jia X, Zhou M, Zhang H, Yang W, Kudinha T, Xu Y. 2020. Emergence of ST11-
612 K47 and ST11-K64 hypervirulent carbapenem-resistant Klebsiella pneumoniae in
613 bacterial liver abscesses from China: a molecular, biological, and epidemiological study.
614 Emerg Microbes Infect 9:320-331.

615 33. Yan C, Zhou Y, Du S, Du B, Zhao H, Feng Y, Xue G, Cui J, Gan L, Feng J, Fan Z, Fu T,
616 Xu Z, Zhang Q, Zhang R, Cui X, Tian Z, Chen Y, Zhang T, Huang L, Yuan J. 2023.
617 Recombinase-Aided Amplification Assay for Rapid Detection of Hypervirulent Klebsiella
618 pneumoniae (hvKp) and Characterization of the hvKp Pathotype. Microbiol Spectr
619 doi:10.1128/spectrum.03984-22:e0398422.
620 34. Wang S, Ding Q, Zhang Y, Zhang A, Wang Q, Wang R, Wang X, Jin L, Ma S, Wang H.
621 2022. Evolution of Virulence, Fitness, and Carbapenem Resistance Transmission in
622 ST23 Hypervirulent Klebsiella pneumoniae with the Capsular Polysaccharide Synthesis
623 Gene wcaJ Inserted via Insertion Sequence Elements. Microbiol Spectr 10:e0240022.
624 35. Yang X, Sun Q, Li J, Jiang Y, Li Y, Lin J, Chen K, Chan EW, Zhang R, Chen S. 2022.
625 Molecular epidemiology of carbapenem-resistant hypervirulent Klebsiella pneumoniae in
626 China. Emerg Microbes Infect 11:841-849.
627 36. Kochan TJ, Nozick SH, Medernach RL, Cheung BH, Gatesy SWM, Lebrun-Corbin M,
628 Mitra SD, Khalatyan N, Krapp F, Qi C, Ozer EA, Hauser AR. 2022. Genomic surveillance
629 for multidrug-resistant or hypervirulent Klebsiella pneumoniae among United States
630 bloodstream isolates. BMC Infect Dis 22:603.
631 37. Russo TA, MacDonald U. 2020. The Galleria mellonella Infection Model Does Not
632 Accurately Differentiate between Hypervirulent and Classical Klebsiella pneumoniae.
633 mSphere 5.
634
635
636
637
638
639
640

641 **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	4.875 ± 0.3416	2.364 ± 1.168	<0.0001, ****
Kleborate Score	4.313 ± 0.7042	3.485 ± 0.8704	0.0010, ***
Mash	0.00855 ± 0.00995	0.04705 ± 0.04237	0.0002, ***
Jaccard	0.3303 ± 0.1615	0.6718 ± 0.2320	0.0001, ***

642

643 Mann-Whitney nonparametric test used due to non-normal data, except for Siderophore

644 analyzed using an unpaired t test.

645

646

647

648

649

650

651

652

653 **Table 2. Logistic regression and CART prediction models**
654

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	0.88 (14/16)	0.94 (31/33)	0.92 (45/49)	0.955	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

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

656 #Area under the ROC Curve

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

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

659 °Computed using cut-point values

660 $determined from ROC curve

661

662

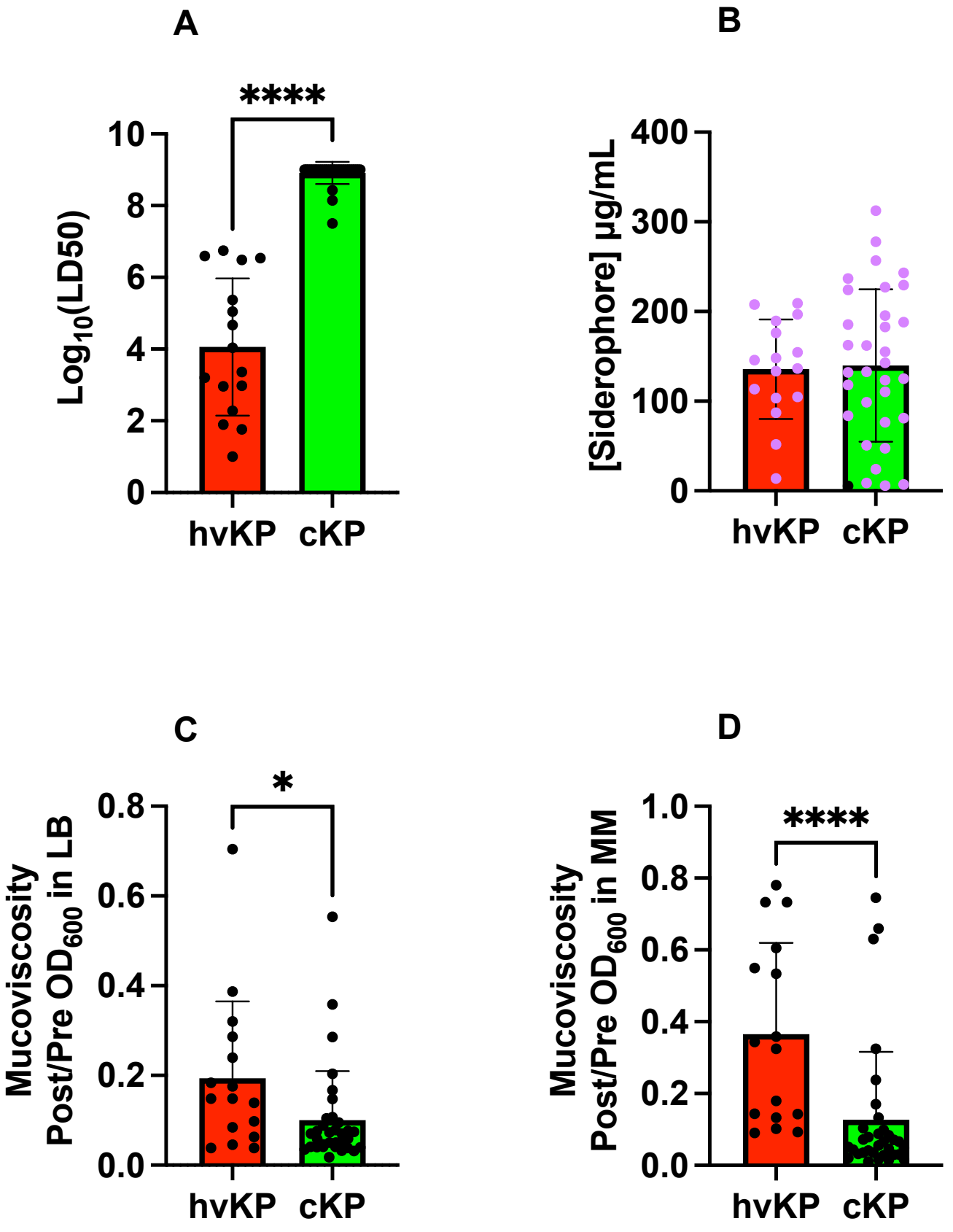
663

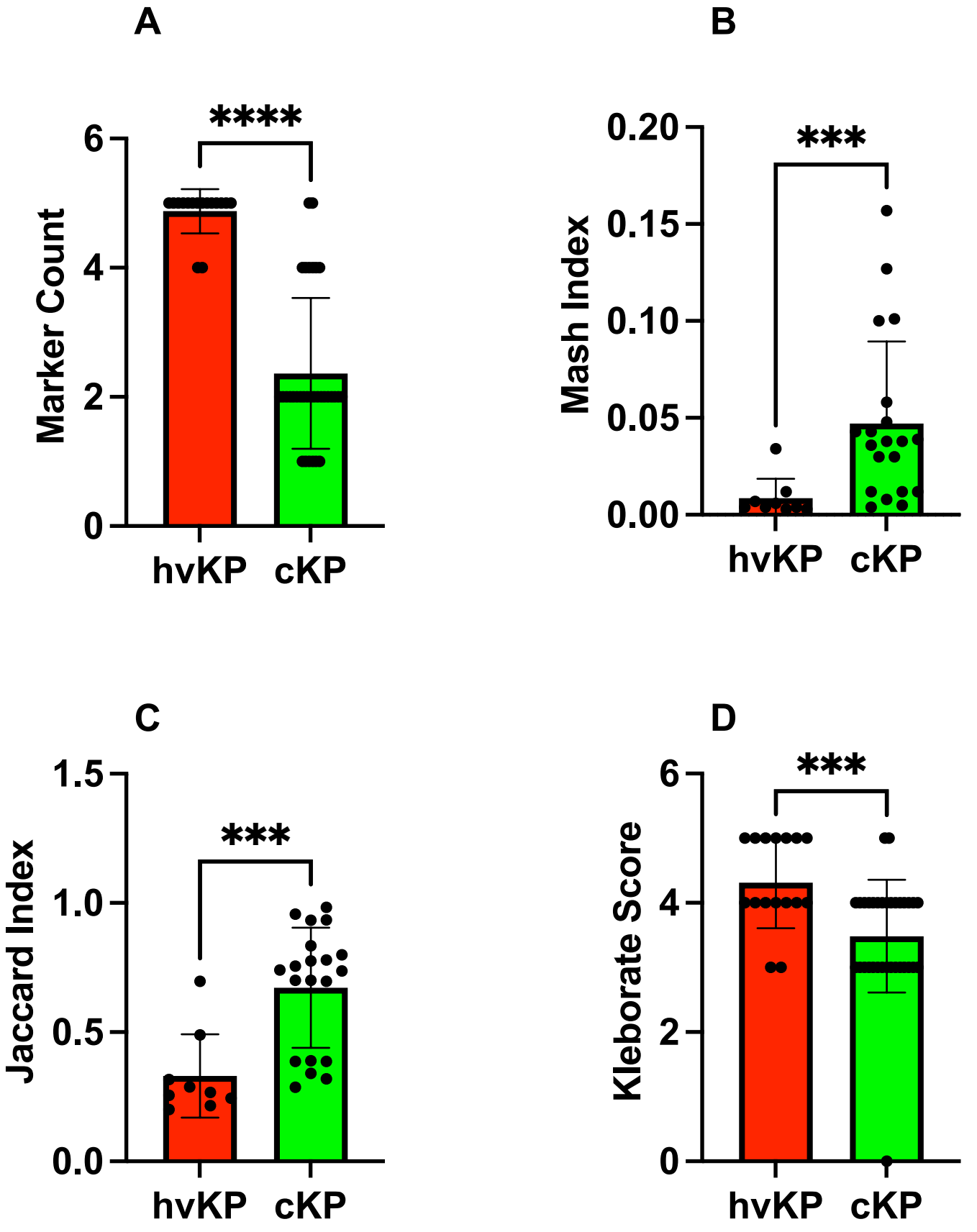
664
665
666
667

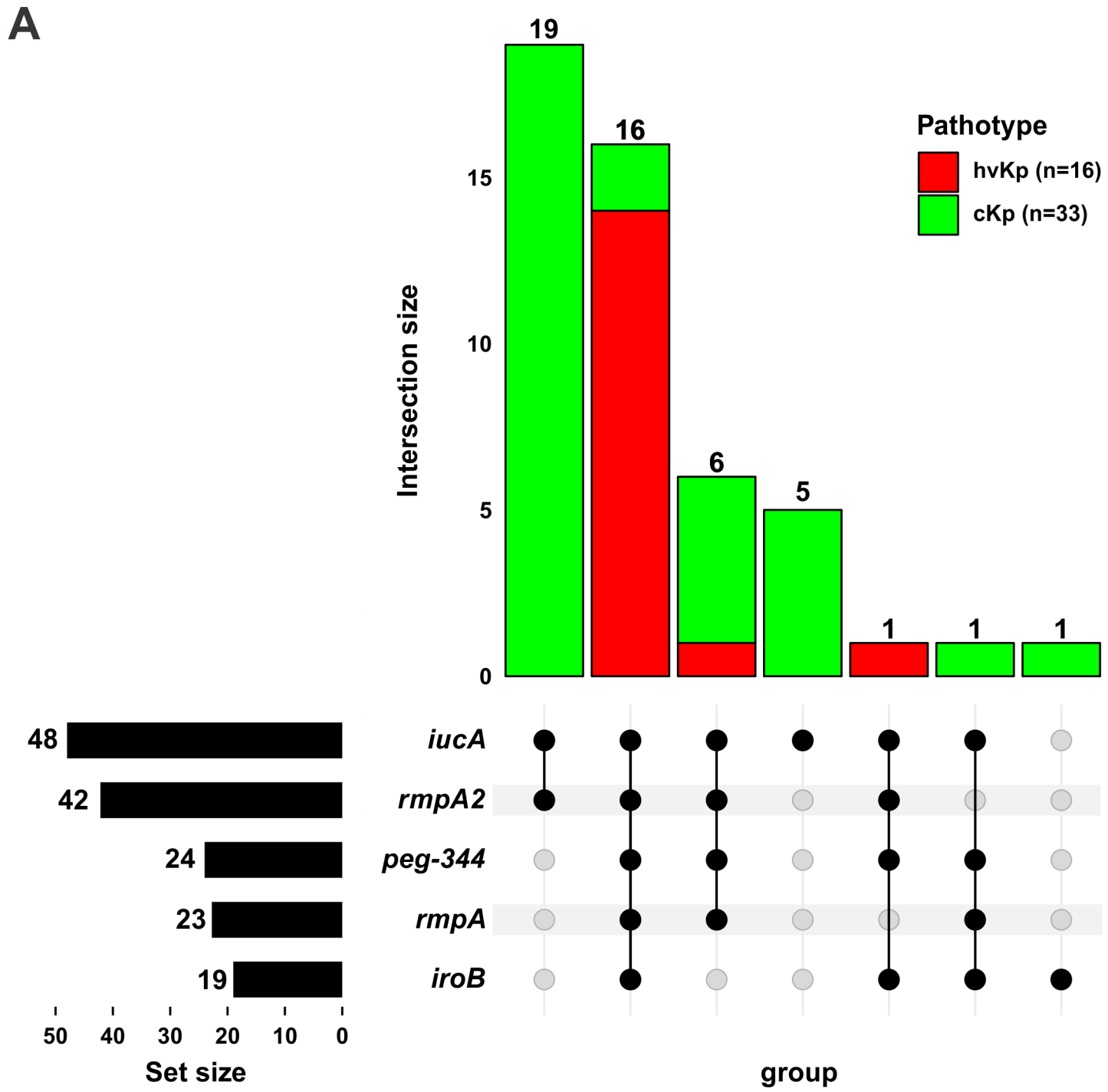
668 **Figures and Legends**
669
670
671 **Figure 1. Phenotypic markers**
672
673 Forty-nine strains comprising the hvKp (n=16) and cKp (n=33) cohorts underwent assessment
674 for LD_{50} in the outbred CD1 murine subcutaneous challenge infection model (Panel A), *in vitro*
675 quantitative siderophore production when grown in c-M9-CA, black symbols represent strains
676 that do not possess *iucA* (Panel B), and *in vitro* quantitative mucoviscosity when grown in either
677 LB (Panel C), or c-M9-CA-te medium (Panel D). For Panel A see Supplementary Table S1 for
678 challenge inocula and animal number for each strain. For Panels B a minimum of three biologic
679 with two technical repeats was performed for each strain. For Panels C and D a minimum of
680 three biologic repeats was performed for each strain. Strain cohorts were compared via Mann-
681 Whitney except siderophore production was compared by an unpaired t test (Prism version
682 9.5.1). Panel A, **** P < 0.0001. Panel C, * P = 0.0113, Panel D, **** P < 0.0001
683
684
685 **Figure 2. Genotypic markers**
686
687 Forty-nine strains comprising the hvKp (n=16) and cKp (n=33) cohorts underwent assessment
688 for marker count (Panel A), 29 strains comprising the hvKp (n=9) and cKp (n=20) cohorts
689 underwent assessment for Mash distance (Panel B), and for Jaccard distance (Panel C), and all
690 49 strains underwent assessment for Kleborate score (Panel D). Strain cohorts were compared
691 via Mann-Whitney (Prism version 9.5.1). Panel A, **** P = <0.0001, Panel B, *** P = 0.0002,
692 Panel C, *** P = 0.0001, Panel D, *** P = 0.001
693
694 **Figure 3. Biomarker distribution**
695 Panel A. UpSet graph for the 49 strains comprising hvKp (n=16) and cKp (n=33) was generated
696 to facilitate the visualization of the presence or absence of the biomarkers *iucA*, *iroB*, *peg-344*,
697 *rmpA*, and *rmpA2*. Black circles designate the presence of a given marker whereas grey circles

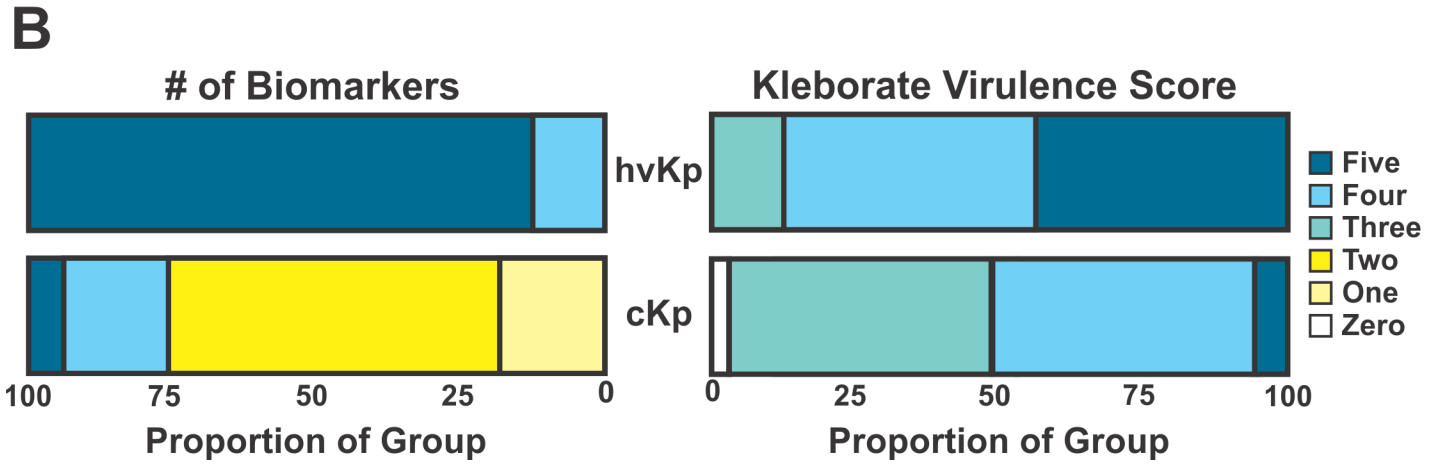
698 designate its absence. Pathotypes are color coded. The number above each bar represents the
699 number of strains that possess that marker configuration. Panel B: Results are shown as the
700 distribution in proportions of each biomarker count (1-5) and Kleborate virulence score (0-5) for
701 the hvKp and cKp cohorts. The biomarkers count is the presence of some combination of *iucA*,
702 *iroB*, *peg-344*, *rmpA*, and *rmpA2.* The Kleborate virulence score is calculated by the presence of
703 the genes that encode aerobactin (3 points), yersiniabactin (1 point), and colibactin (1 point).
704
705 **Figure 4**. **Mash and Jaccard Distances**
706 The pLVKP-like virulence plasmid of 29 strains comprising the hvKp (n=9) and cKp (n=20)
707 cohorts, for which long-read sequencing was obtained thereby enabling a closed
708 genome/plasmid, was compared to the canonical pLVPK sequence. Scatterplots of
709 Mash/Jaccard distance to pLVPK were plotted for each strain and colored by pathotype,
710 number of biomarkers (some combination of *iucA*, *iroB*, *peg-344*, *rmpA*, *rmpA2*), Kleborate
711 virulence score (presence of genes that encode aerobactin (3 points), yersiniabactin (1 point),
712 and colibactin (1 point)), and antimicrobial resistance (AMR, presence of
713 ESBL/carbapenemase genes).
714
715 **Figure 5. Proksee alignment to pLVPK**
716 Twenty-nine strains with a closed plasmid harboring any virulence biomarker were compared to
717 the canonical pLVPK. Alignment of the pLVPK sequence to all closed pLVPK-like plasmid
718 sequences from **A)** hvKp (n=9) and **B)** cKp isolates (n=20) were produced with Proksee.
719 Virulence and landmark genes from pLVPK are labeled in each panel. The five virulence
720 biomarkers of interest are concentrated on a roughly 30kb region of pLVPK from *iucA* to *iroB*.
721 Concentric circles are sorted by alignment fraction to pLVPK sequence with the outermost ring
722 representing the isolate with the highest alignment fraction. The average alignment fraction of

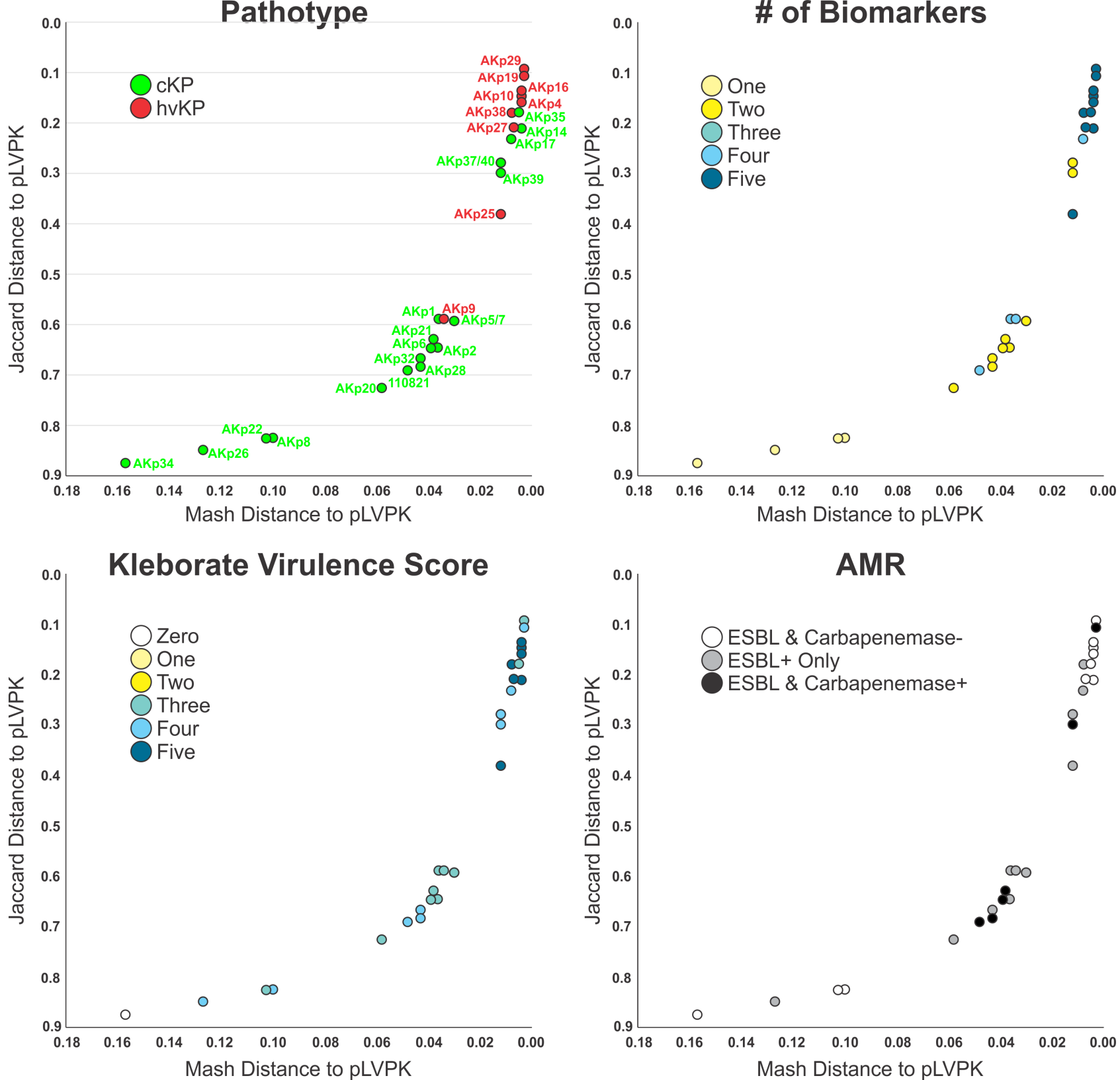
723 pLVPK-like plasmid to pLVPK is higher in hvKp isolates compared to cKp. Alignment covering
724 the virulence biomarkers is more extensive in hvKp isolates. Most cKp isolates lack or have
725 non-functioning *iroB, peg-344,* and *rmpA* while possessing a *rmpA2* and *iucA*.
726
727
728 **Supplementary Figure 1**
729 Forty-nine strains comprising the hvKp (n=16) and cKp (n=33) cohorts underwent assessment
730 for growth/survival in human ascites via enumeration of colony forming units. A minimum of two
731 biologic with one-three technical repeats was performed for each strain.
732
733

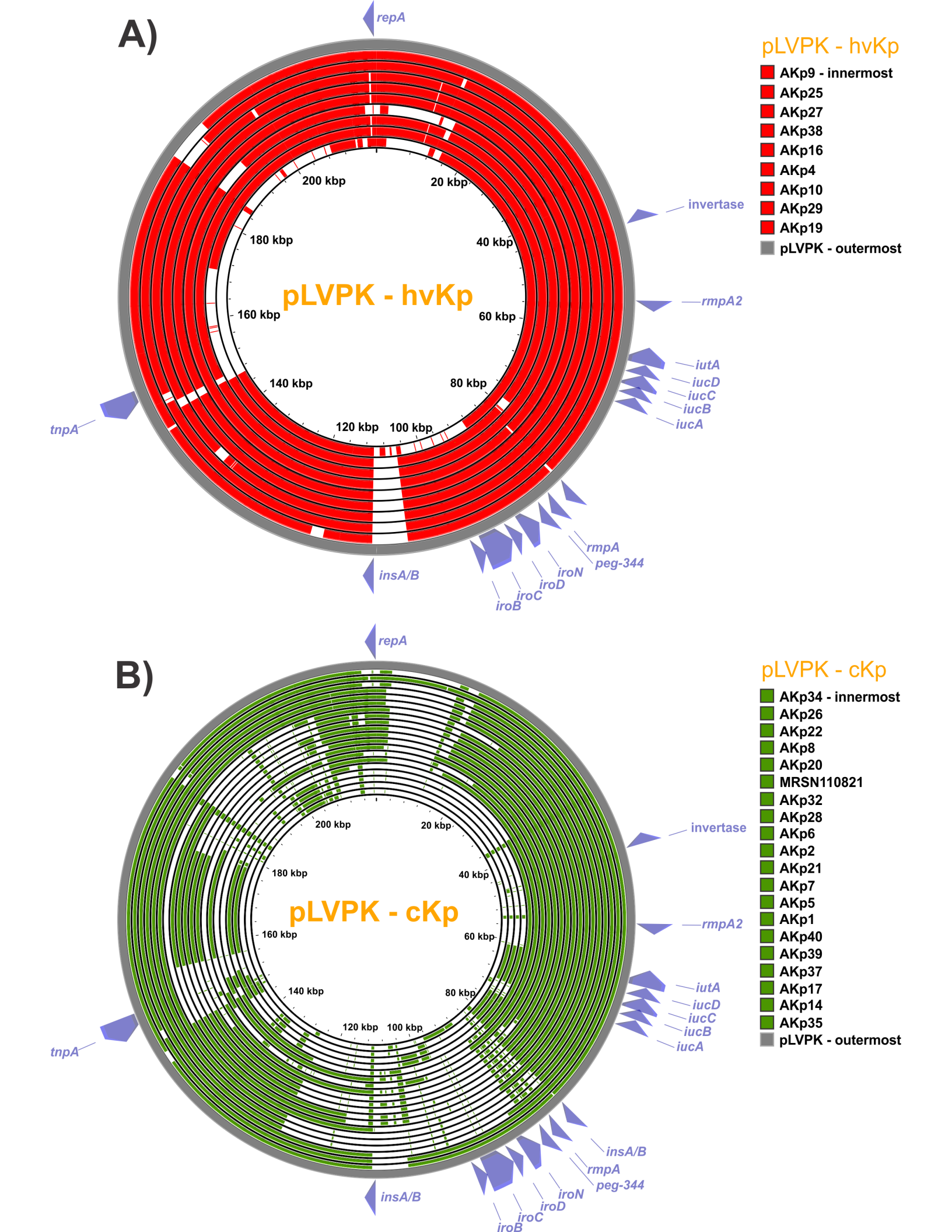

A
****
Log10(LD50)
10
8
6
4
2
0
hvKP
cKP
B
[Siderophore] µg/mL
400
300
200
100
0
hvKP
cKP
C
*
Mucoviscosity
Post/Pre OD600 in LB
0.8
0.6
0.4
0.2
0.0
hvKP
cKP
D
****
Mucoviscosity
Post/Pre OD600 in MM
1.0
0.8
0.6
0.4
0.2
0.0
hvKP
cKP


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
16
6
5
1
1
1
0
5
10
15
48
42
24
23
19
iucA
rmpA2
peg-344
rmpA
iroB
50
40
30
20
10
0
Set size
group


B
# of Biomarkers
Kleborate Virulence Score
hvKp
cKp
Five
Four
Three
Two
One
Zero
100
75
50
25
0
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
AKp28
AKp20
110821
AKp22
AKp8
AKp26
AKp34
Jaccard Distance to pLVPK
Mash Distance to pLVPK
# of Biomarkers
One
Two
Three
Four
Five
Kleborate Virulence Score
Zero
One
Two
Three
Four
Five
AMR
ESBL & Carbapenemase-
ESBL+ Only
ESBL & Carbapenemase+


A)
pLVPK - hvKp
AKp9 - innermost
AKp25
AKp27
AKp38
AKp16
AKp4
AKp10
AKp29
AKp19
pLVPK - outermost
repA
invertase
rmpA2
iutA
iucD
iucC
iucB
iucA
tnpA
insA/B
rmpA
peg-344
iroN
iroD
iroC
iroB
20 kbp
40 kbp
60 kbp
80 kbp
100 kbp
120 kbp
140 kbp
160 kbp
180 kbp
200 kbp
B)
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
repA
invertase
rmpA2
iutA
iucD
iucC
iucB
iucA
tnpA
insA/B
rmpA
peg-344
iroN
iroD
iroC
iroB
20 kbp
40 kbp
60 kbp
80 kbp
100 kbp
120 kbp
140 kbp
160 kbp
180 kbp
200 kbp