KlebSeq: A Diagnostic Tool for Healthcare Surveillance and Antimicrobial Resistance Monitoring of *Klebsiella pneumoniae.*

Jolene R. Bowers,[a]# Darrin Lemmer,[a] Jason W. Sahl,[b] Talima Pearson,[b] Elizabeth M. Driebe,[a] David M. Engelthaler,[a] Paul Keim[a,b].

Translational Genomics Research Institute, Flagstaff, AZ, USA[a]; Northern Arizona University, Flagstaff, AZ, USA[b]

Running Head: Characterizing *K. pneumoniae* for Surveillance

#Address correspondence to Jolene R. Bowers, jbowers@tgen.org.

ABSTRACT

Healthcare-acquired infections (HAIs) kill tens of thousands of people each year and add significantly to healthcare costs. Multidrug resistant and epidemic strains are a large proportion of HAI agents, and multidrug resistant strains of *Klebsiella pneumoniae*, a leading HAI agent, have become an urgent public health crisis. In the healthcare environment, patient colonization of *K. pneumoniae* precedes infection, and transmission via colonization leads to outbreaks. Periodic patient screening for *K. pneumoniae* colonization has cost-effective and life-saving potential. In this study, we describe the design and validation of KlebSeq, a highly informative screening tool that detects *Klebsiella* species and identifies clinically important strains and characteristics using highly multiplexed amplicon sequencing without a live culturing step. We demonstrate the utility of this tool on several complex specimen types including urine, wound swabs and tissue, several types of respiratory, and fecal, showing *K. pneumoniae* species and clonal group identification and antimicrobial resistance and virulence profiling, including capsule typing. Use of this amplicon sequencing tool can be used to screen patients for *K. pneumoniae* carriage to assess risk of infection and outbreak potential, and the expansion of this tool can be used for several other HAI agents or applications.

INTRODUCTION

Klebsiella pneumoniae has been a leading healthcare-acquired infection (HAI) agent for decades (1, 2). Emergence of multidrug-resistant *K. pneumoniae*, especially expanded-spectrum β-lactamase (ESBL) producers and carbapenemase producers, has elevated the morbidity and mortality rates and healthcare costs associated with *K. pneumoniae* to

highly significant levels (3-6). Healthcare- and outbreak-associated strain types of *K. pneumoniae* that appear highly transmissible and have a propensity for antimicrobial resistance (AMR) or virulence gene acquisition are a growing proportion of the *K. pneumoniae* species (7-18). ST258, the crux of the worldwide carbapenemase-producing Enterobacteriaceae (CPE) threat, disseminated rapidly around the world's healthcare systems despite its recent emergence (17). Its progenitor strains in clonal group (CG) 258 also cause outbreaks and carry many important ESBL and carbapenemase genes (9, 19-21). Several other strain types such as those in CG14, CG20, and CG37, also frequently appear as multidrug resistant and in outbreak situations (7, 10, 12, 15).

Host colonization is likely an important reservoir driving transmission of these strains. In the healthcare environment, intestinal colonization of *K. pneumoniae* is risk factor for infection (22-24), and carriers of CPE are at high risk for invasive disease (25). Rates of CPE and ESBL-producing *K. pneumoniae* colonization are rising in patient and healthcare worker populations, increasing the size of the reservoir, and increasing chances of transmission (26, 27). Asymptomatic transmission of multidrug resistant strains is rapid (16, 28), and transmission events that lead to outbreaks often go undetected (29, 30). Early detection of *K. pneumoniae* colonization of healthcare patients, especially multidrug resistant *K. pneumoniae* or epidemic strain type colonization, is now considered critical to infection control (24, 30-33).

Infection control programs that include detection and isolation of carriers have repeatedly been successful in markedly decreasing multidrug resistant or epidemic strain infections (31, 34-37), but this practice is uncommon. Many of these programs use culture-based methods such as antibiotic-containing broth enrichment or selective media

for detecting CPE or ESBL producers, which have several limitations including turn-around time, narrow application, lack of sensitivity and specificity, subjectivity, and extensive labor for high-throughput screening (31, 38). Automated systems also require up-front organism culture and isolation, with many of the same limitations (31). PCR-based assays are rapid, but often use DNA from culture, or if used on DNA extracted from specimens, may have low sensitivity. Additionally, a limited number of tests can be run simultaneously, and may miss important AMR genes not previously known to circulate in a given locale (31, 39).

In this study, we describe a new tool, KlebSeq, for screening and surveillance that detects and characterizes *Klebsiella* from complex samples such as wound and nasal swabs or fecal samples without culturing using relatively easy-to-use multiplex amplicon sequencing. KlebSeq includes a sizeable panel of assays for species identification, strain identification, and important virulence and AMR gene targets designed to generate information for hospital epidemiology and infection prevention. Results from screening a patient population with this system would rule in or rule out the possibilities of particular transmission events, and identify patients carrying high-risk strains like ST258 or other multidrug resistant *Klebsiella*. The highly multiplexed nature of the amplicon sequencing tool greatly expands the capacity of a single sequencing run, minimizing costs, and allows for high-throughput patient sample testing. Additionally, this innovation can serve as a model system for many other applications including targeting other HAI agents and their multiple AMR mechanisms.

METHODS

Samples. Isolates for target identification and assay validation, along with DNA extracted from clinical specimens were acquired through collaborations with a large hospital reference laboratory that receives specimens from ten system-wide medical centers in Arizona, and from a high volume private reference laboratory that receives specimens from regional inpatient, long-term care, and outpatient facilities. Clinical specimen types included various respiratory specimens (nasal, ear, and throat swabs, sputa, tracheal aspirates, and bronchial alveolar lavages), urine, and wound swabs or tissue. DNA was extracted from isolates by Qiagen DNeasy Blood and Tissue Kit with additional lytic enzymes when appropriate. DNA was extracted from clinical specimens by NucliSENS easyMAG (bioMerieux, Durham, NC). DNA from healthy donor fecal samples was acquired from a family microbiome study. Samples had been collected from members of seven families over multiple timepoints. DNA was extracted following the Earth Microbiome Project protocol (40).

Assay target identification and assay design. Figure 1 illustrates the methodologies and resources utilized, also described below, to amass a target library and develop several types of amplicon sequencing assays.

Whole genome sequencing, SNP detection, and phylogenetic analysis. In-house genome libraries were prepared from 130 *Klebsiella* isolates with a 500 base pair insert size using KAPA Library Preparation Kit with Standard PCR Library Amplification (Kapa Biosystems, Wilmington, MA) and sequenced on Illumina's GAIIx or MiSeq. Public genome sequence data from 277 *K. pneumoniae*, 18 K. *quasipneumoniae*, and 13 *K. variicola* isolates were downloaded from the SRA database

(http://www.ncbi.nlm.nih.gov/Traces/sra/) and 178 *K. pneumoniae*, four *K. quasipneumoniae*, and 11 *K. variicola* from the Assembly database (http://www.ncbi.nlm.nih.gov/assembly), and all passed filters for high quality. The publicly available bioinformatic tool NASP (41), developed for microbial genome analysis, was used to detect single nucleotide polymorphisms (SNPs) among the genomes. In brief, reads were aligned to a reference genome, either one from concatenated scgMLST alleles (10) or MGH 78578 (Genbank accession no. CP000647) using Novoalign (Novocraft.com) and SNPs were called with GATK (42). Data filtered out included SNP loci with less than 10X coverage or with less than 90% consensus in any one sample, regions duplicated in the reference genome as identified by Nucmer, and SNP loci that were not present in all genomes in the dataset. In NASP, results were output in a SNP matrix from a core genome common to all isolates in the analysis. Phylogenetic trees were generated from the NASP SNP matrices using MEGA 6.0 (43) and subsequently plotted by means of ITOL v2 or v3 (44).

Genomic target identification. To find whole gene targets for assay design, select genomes were assembled with UGAP (https://github.com/jasonsahl/UGAP), which uses the SPAdes genome assembler (45). Assemblies were then run through LS-BSR (46), which generates a list of ORFs that have high identity among target species genomes and that have low identity or are not present in non-target genomes. Alleles of the candidate target ORFs were collected by BLAST, including alleles from non-target genomes if present. Lastly, alleles of candidate ORFs were aligned for assay design. Two ORFs (M1 and M2) were collected for two *K. pneumoniae*-specific assays. Canonical

128 SNPs (canSNPs) were identified from the SNP matrix generated from NASP. Sequence
129 flanking each SNP was collected from the NASP reference genome.

130 **AMR and virulence gene target collection.** AMR and virulence gene sequences
131 were identified and collected in several ways, including from
132 http://www.lahey.org/studies/other.asp#table1, http://www.lahey.org/qnrstudies/, the
133 *Klebsiella* BIGSdb http://bigsdb.web.pasteur.fr/klebsiella/, public literature, and NCBI
134 http://www.ncbi.nlm.nih.gov. Public literature included Holt *et al.* (47), in which a
135 species-wide analysis of *K. pneumoniae* genomes revealed several siderophore systems
136 and other virulence factors more associated with infectious than colonizing strains. AMR
137 genes include the major ESBL and carbapenemase genes, plasmid-mediated quinolone
138 resistance determinants as well as the *gyrA* and *parC* chromosomal genes, several
139 aminoglycoside resistance genes, TMP-SMX, tetracycline, streptomycin,
140 chloramphenicol, and fosfomycin resistance genes, and the recently discovered plasmid-
141 mediated colistin resistance gene *mcr-1*. Virulence targets include several siderophore
142 systems, for which multiple genes from each were used as assay targets, the regulator of
143 mucoid phenotype (an indicator of a hypervirulence), and *wzi* gene for capsule typing, for
144 which we used the published assay (48), and two genes highly associated with invasive
145 infection pK2044_00025 and pK2044_00325 (47). For genes that consist of highly
146 diverse alleles, for example $bla_{\text{CTX-M}}$, *qnrB*, or *dfrA*, phylogenetic trees based on
147 nucleotide sequences were generated in order to group similar alleles for assay design.

148 **Assay design and validation.** Gene-based target alleles were aligned in SeqMan
149 (DNAStar, Madison, WI) to identify conserved regions for primer design, and assays
150 were designed with guidance from RealTimeDesign™ (Biosearch Technologies,

Petaluma, CA), or gene-based assays were generated with AlleleID® (Premier Biosoft, Palo Alto, CA), which designs assays to capture alleles in an alignment rather than to individual sequences. SNP assay primers were designed using RealTimeDesign™, and primer sequences were checked for conservation in the NASP SNP matrix. Lastly, assays were run through BLAST http://blast.ncbi.nlm.nih.gov/Blast.cgi, to check for cross-reactivity to other relevant targets or species, including human. Universal tails were added to each primer sequence for library preparation, as described in Amplicon library preparation below. The assays and their primer sequence are listed in Table S1.

Individual assays were screened across positive controls when accessible, and screened across several isolate gDNAs to increase confidence in robustness, especially when known positive controls were not available. Additionally, multiplex PCR was validated by initial gene-specific PCR (described below) followed by PCR product dilution, then screening of individual assays by Sybr green-based qPCR. For this, 10 µL reactions of 1X Platinum SYBR Green qPCR SuperMix (ThermoFisher Scientific, Waltham, MA), 200 nM forward and reverse primers of one assay, and 1 µL diluted multiplex PCR product were run at 95°C initial denaturation for 4 min, then 40 cycles of 95°C for 15 s and 60°C for 1 min. Lastly, several panels of known isolate DNAs were screened with the amplicon sequencing method to test for sensitivity and specificity of the species and strain identification assays. AMR and virulence gene assays were validated by comparing amplicon sequencing results with whole genome sequence data.

Amplicon library preparation and sequencing. Amplicon library preparation using universal tails was described in detail previously (49). Here, assays were pooled for multiplex PCR. Initial gene-specific PCR comprised 12.5 µL Kapa Multiplex PCR

Mastermix (Kapa Biosystems, Wilmington, MA), 10 µL primer mix (for final concentration of 200 nM each), and 2.5 µL DNA template from each sample, and was denatured at 95°C for 3 min, cycled 25 times at 95°C 15 s, 60°C 30 s, 72°C 1 min 30 s, with final extension 72°C 1 min. Multiple multiplex PCR products from the same sample were pooled, and PCR product pools were cleaned with 1X Agencourt AMPure XP beads (Beckman Coulter, Indianapolis, IN). A second PCR using the universal tail-specific primers added Illumina's sample-specific index and sequencing adapters. This PCR comprised 12.5 µL 2X Kapa HiFi HotStart Ready Mix (Kapa Biosystems), 400 nM each primer, and 1 to 10 µL cleaned gene-specific PCR product, and was denatured 98°C 2 min, cycled 6 to 12 times at 98°C 30 s, 65°C 20s, 72°C 30 s, with final extension 72°C 30 s. Final PCR product was cleaned with 0.8X Agencourt AMPure XP beads (Beckman Coulter).

Amplicon libraries from individual samples were quantified by qPCR using Kapa Library Quantification Kit (Kapa Biosystems). Samples were then pooled in equimolar concentration for sequencing on the Illumina MiSeq platform with 2x250bp version 2 kit.

Analysis. Amplicon sequencing results were automatically analyzed using a newly developed amplicon sequencing analysis pipeline (ASAP) (50), which uses a JavaScript Object Notation (JSON) file that describes all assays in the multiplex. Information in the JSON file included a category for each assay (presence/absence, SNP, gene variant, or region of interest) that dictates how ASAP will report results, and reference sequences for mapping. In ASAP, amplicon sequence reads were first trimmed of adapter or read-through sequences with Trimmomatic (51), then mapped to the reference sequences with an aligner. BAM alignment files were analyzed alongside the

197 JSON file assay descriptions to determine presence, breadth and depth of coverage, and
198 proportions of nucleotide polymorphisms for each amplicon. User defined parameters
199 included use of the bowtie2 aligner (52), and thresholds for determining results of
200 screening included 80% breadth and 100X depth of coverage for isolate DNA, and 80%
201 breadth and 20X depth of coverage and ≥10% proportion of polymorphism for
202 informative SNP loci for complex specimen DNA (meaning that at least 10% of the reads
203 had to share a SNP state at a given locus for it to be reported). ASAP output included an
204 XML file containing details of the analysis of each assay target for each sample, which
205 can be converted into a webpage interface using XSLT transformations. SeqMan NGEN
206 (DNAStar, Madison, WI) and Tablet (53) were used to verify results.
207
208 **RESULTS**
209 **Phylogenetic analysis and canSNP identification.** Using the *Klebsiella* strict core
210 genome MLST (scgMLST) (10) assembly as a reference, SNPs among a diverse set of
211 genomes from *K. pneumoniae* and genomes from the newly defined *K. quasipneumoniae*
212 (22 from the public databases and one from in-house isolates) and *K. variicola* (24 from
213 the public databases and five from in-house isolates) were identified with NASP. A
214 canonical SNP that differentiates *K. quasipneumoniae* and one that differentiates *K.*
215 *variicola* from *K. pneumoniae* were selected for assay development.
216 Using the reference genome MGH 78578 and 548 diverse *K. pneumoniae*
217 genomes, NASP generated a SNP matrix from which canonical SNPs for each of the
218 major clonal groups were selected for assay development. Clonal groups and locations of
219 canSNPs identifying 35 clonal groups and sequence types in the context of the *K.*

pneumoniae species are illustrated in Figure 2. Redundancy was intentionally included in the assays expected to be positive for the most epidemic strains of *K. pneumoniae* such as ST14, ST20, and strains in CG258 in order to increase confidence in positive results.

KlebSeq validation. For validation of the species and clonal group identification assays, genomic DNA from 73 *K. pneumoniae* isolates that had also been whole genome sequenced (4 of which were later identified as *K. quasipneumoniae* and *K. variicola*, see below), 22 *K. oxytoca* isolates, and 157 other enteric, opportunistic-pathogen isolates, which included *E. coli*, *Enterobacter aerogenes*, *E. amnigenus*, *E. cloacae*, *E. hormaechei*, *Enterococcus faecalis*, *E. faecium*, *E. sp.*, *Proteus mirabilis*, *Providencia stuartii*, and *Serratia marcescens* were screened with KlebSeq. Sensitivity and specificity results of the species identification assays compared with clinical microbiological identification are in Table 1. With the redundancy built into the multiplex by including two assays Kp-M1 and Kp-M2, which target two different *K. pneumoniae* species-specific genes (M1 and M2), 100% sensitivity is achieved. One isolate previously identified as *K. pneumoniae* typed as *K. quasipneumoniae* and two as *K. variicola*. These isolates' whole genomes were added to the phylogenetic analysis of these three species that was previously run to find the species-specific canSNPs (see Methods). The *K. quasipneumoniae* and *K. variicola* genomes identified by our assay clustered with their respective species in the phylogeny (Fig. 3). Clinical methods do not currently distinguish among all three of these species, so assay sensitivity and specificity were not calculated for *K. quasipneumoniae* and *K. variicola*.

Table 2 shows results of the *K. pneumoniae* clonal group identification and capsule typing assays. Each isolate's strain type was correctly captured by the appropriate

243 assays, or not captured in cases where no assay was designed for that clonal group.
244 Partial sequencing of the *wzi* gene for capsule typing gave surprisingly clear results,
245 given that approximately 75 bp of the informative region is missing in our sequence
246 output, as it was based on a previously published assay (48). The amplicon size of the *wzi*
247 PCR is approximately 580 bp, which is too long to cover with the Illumina version 2
248 sequencing chemistry. However, our data show the promise of full capsule typing by *wzi*
249 sequence with longer read chemistry (*i.e.* Illumina version 3 chemistry, which provides
250 2x300 bp reads). Results from screening non-target organisms showed that several of the
251 *K. pneumoniae* clonal group assays amplified DNA from other organisms, as was
252 expected, however all SNP states that define a particular clonal group are specific to that
253 clonal group. As such, sequence analysis by ASAP reported when a clonal group was
254 present only if the defining canSNP state was present, and reported nothing if it was not.
255 AMR gene detection by amplicon sequencing was validated by comparing ASAP
256 results with AMR gene screening of whole genome sequence with SRST2 (54). Results
257 showed good correlation with few discrepancies, which were almost all negative by
258 amplicon sequencing while positive by SRST2. This may have been due to PCR primer
259 discrepancy or inefficiency or failure in amplicon library preparation. Only six samples
260 account for a majority of the negatives, which points to the latter alternative. Virulence
261 gene detection was validated by comparing ASAP results from whole genome sequence
262 data with those from amplicon sequence data. Results showed almost perfect
263 concordance. In addition, by targeting genes that are part of the same virulence factors
264 (i.e. siderophore systems), sensitivity and confidence in results was increased.

These results also confirm that KlebSeq is applicable to pure isolates as well as complex specimens. Screening isolate DNA has the added benefit of traceability of the AMR and virulence genes, which are often carried on mobile genetic elements, to their host. Isolate screening could be used for surveillance and other purposes for identifying or characterizing *Klebsiella*.

Specimen sample results. KlebSeq was run on DNA from 87 respiratory specimens, 46 urine specimens, 40 wound specimens, and 89 fecal samples from healthy individuals. Sensitivity and specificity results of the species identification assays compared with clinical microbiological methods are shown in Table 1. In most cases, sensitivity was very high, except in the wound specimens where the one sample clinically identified as *K. pneumoniae* typed as *K. variicola*. Amplicon sequencing identified several samples with *Klebsiella* that went undetected with clinical microbiological methods, including several in which clonal groups were also detected.

Important clonal groups of *Klebsiella* were detected in multiple specimens. In the 18 urine samples positive for *K. pneumoniae*, clonal identifications included CG34, ST20, CG45, CG392 (which includes the NDM-producer ST147 (55), though these samples were negative for bla_{NDM}, n=2), ST133, and CG111. In wounds, one strain identification, CG29, was made from the three *K. pneumoniae*-positive samples. From respiratory specimens, groups CG37 (n=2), ST134 (n=1), ST258 (n=2), CG36 (n=3), and innerST14 (n=1) were identified. Interestingly, several clonal groups were identified in the healthy donor fecal specimens as well. Out of the eight *K. pneumoniae*-positive samples, groups included ST20 (an alignment of which is shown in Figure 4), CG37, and CG76, which are all members of multidrug-resistant outbreak strain types (11, 12, 15),

288 along with ST133 and ST380. ST380 is associated with a K2 capsule type and
289 hypervirulence, and causes pyogenic liver abscesses in healthy people, especially of
290 Asian ethnicity (56). Many Asians are colonized with hypervirulent, K1 or K2 capsule
291 strain types; however the level of risk of subsequent liver infection is unknown (56). For
292 this sample, no *wzi* gene sequence was obtained, thus the capsule type is unknown. A
293 majority of the *K. pneumoniae* in our samples did not fall into the major clonal groups
294 targeted by KlebSeq. Likely these strains all belong to lesser-known clonal groups, as
295 more studies are showing that many *K. pneumoniae* infections are caused by non-
296 epidemic, sporadic strains (57, 58).

297 Numerous and variable AMR genes were detected in the specimens, including
298 different variants of the same gene that confer different phenotypes. Using sequence-
299 based information, we demonstrate that many of the *K. pneumoniae* had key mutations in
300 the *gyr*A gene known to confer resistance to fluoroquinolones. Additionally, several
301 samples contained the *aac(6')-Ib* gene for aminoglycoside resistance, and some of those
302 contained the minor sequence variant *aac(6')-Ib-cr* for fluoroquinolone resistance;
303 mixtures of these two genes were also detected. A majority of the infection specimens
304 (non-healthy donor specimens), both positive and negative for *K. pneumoniae*, were
305 positive for other aminoglycoside resistance genes, as well as tetracycline, TMP-SMX,
306 streptomycin, fosfomycin, and chloramphenicol resistance genes. A few contained
307 plasmid-mediated quinolone resistance genes. Several samples, especially the respiratory
308 specimens, were also positive for KPC and CTX-M groups 1 and 9 genes. Most of the
309 healthy donor specimens were positive for TMP-SMX resistance genes, and many for
310 streptomycin, aminoglycoside, tetracycline, and fosfomycin resistance genes. Some also

311 contained plasmid-mediated quinolone resistance genes. Fortunately, none were found to
312 contain ESBL or carbapenemase genes. No complex specimens in the study were positive
313 for genes encoding the important carbapanemases OXA-48, VIM, or NDM, and none
314 were positive for the plasmid-mediated colistin resistance gene *mcr-1*.

315 These sets of samples did not appear to contain especially virulent strains of *K.*
316 *pneumoniae*. The yersiniabactin siderophore genes were by far the most prevalent of the
317 virulence genes tested, although positive samples made up less than half of the *K.*
318 *pneumoniae*-positive samples (46%). No samples were positive for *rmpA*, regulator of
319 mucoid phenotype gene, including the ST380-containing sample, and few were positive
320 for the salmochelin siderophore genes, which are associated with invasive *K. pneumoniae*
321 infection (47). One respiratory specimen that contained a ST14 strain was positive for a
322 K2 capsule type by partial *wzi* sequencing. K2 strains of *K. pneumoniae* are associated
323 with hypermucoviscosity and hypervirulence, as previously mentioned. However this
324 respiratory sample was not positive for *rmpA*, and a recent study proposed that the
325 presence of multiple siderophore system genes (linked to K1 or K2 capsule genes)
326 explains hypervirulence rather than capsule type (47). In our data, *K. pneumoniae*-
327 containing samples were positive for multiple siderophores or other virulence-associated
328 genes only 15% of the time. Sequencing of *wzi* revealed a variety of capsule types, and
329 incidences where the same clonal groups had different *wzi* genotypes and where they had
330 the same genotype. This character would help identify or rule out a transmission event
331 when patients carrying the same strain are found.

332 On an interesting note, in the healthy donor fecal samples collected from
333 members of the same families over time, out of the eight *K. pneumoniae*-positive

334 samples, only two came from the same person over time. The characterization assays
335 suggest that the same strain of *K. pneumoniae* was present at both time points. The fact
336 that there were not more cases of positive results from the same person in multiple rather
337 than single time points is interesting. This could be due to intermittent shedding of *K.*
338 *pneumoniae* in feces, intermittent colonization of *K. pneumoniae*, or heterogeneity in the
339 sample itself, underrepresenting the full microbial community when a small sample is
340 taken. *K. pneumoniae*-positive samples were found in multiple members of two of the
341 seven families. In one of these families, the positive members carried different strains
342 from one another, and in the other it appears two members shared a CG37 with the same
343 capsule type. The sample set is too small to draw conclusions from the data; however, the
344 data raise interesting questions about community *K. pneumoniae* carriage.
345

346 **DISCUSSION**

347 The frequency of HAI in the United States is estimated at one in 25 hospital patients,
348 totaling hundreds of thousands, with significant mortality (2). HAIs have a significant
349 impact on healthcare costs; a 2009 CDC report estimated upwards of $45 billion in
350 annual additional cost (59). Infections of AMR organisms cause significantly higher
351 mortality rates and ICU admissions, and significant excess healthcare costs including
352 hospitalization, medical care, and antimicrobial therapy over infections of susceptible
353 strains (5, 60). HAI prevention measures, although costly in and of themselves (61), have
354 the potential to save many lives and billions of dollars (59). Periodic patient screening
355 and isolation of AMR organism carriers have proven successful in controlling
356 transmission and outbreaks in several hospitals (31, 34-37). Use of a highly informative

357 screening and surveillance tool such as KlebSeq has cost-effective and life-saving
358 potential.

359 Early detection of *K. pneumoniae* colonization of healthcare patients, especially
360 multidrug resistant *K. pneumoniae*, would allow healthcare staff to make more informed
361 patient management decisions. In outbreak situations, rapid identification of
362 transmissions before subsequent infections would allow for proactive measures to curb an
363 outbreak. In non-outbreak situations, identification of particular strains and AMR genes
364 would help to assess the risk of *K. pneumoniae* carriage to the host patient as well as to
365 other patients, as some strains are more associated with outbreaks, HAI, AMR, and
366 treatment failure than others (7-13, 62). Likewise, identification of virulence genes also
367 informs risk, as particular virulence factors are more associated with pathogenic than
368 colonizing *K. pneumoniae* (47). Additionally, many *K. pneumoniae* infections, including
369 HAIs and non-multidrug resistant infections, are caused by non-epidemic, lesser-known
370 strain types (57, 58). Classifying the *K. pneumoniae* in each patient sample would help
371 an institution decide when and which intervention procedures should be enacted, and also
372 understand more about transmission dynamics and local strain type circulation.

373 The amplicon sequencing assay and analysis pipeline described here has several
374 characteristics that make it ideal as a healthcare screening tool. With a single assay,
375 enough information is garnered about a patient's *Klebsiella* carriage status to contribute
376 greatly to patient management or to infection control decisions. Indexing samples by
377 means of the universal tail during sample preparation allows characterization of a large
378 number of specimens in one run, minimizing sequencing costs per specimen and allowing
379 for high-throughput screening of hundreds of patient samples simultaneously. KlebSeq

uses DNA extracted directly from a specimen so targets from entire populations of a species are analyzed to capture different strains in the same sample, which can be numerous (63, 64). If culture-based methods are used for screening, different strains are missed when one genotype (*i.e.* colony) is chosen for characterization, and resulting information is limited. Additionally, culture-based methods can miss "silent" multidrug resistant *K. pneumoniae* that test negative for carbapanemases *in vitro* (16), and if used for high-throughput screening, they can be laborious, time-consuming, costly, and subjective (31, 38). If screening of large numbers of patients by amplicon sequencing is cost-prohibitive, it can be limited to the highest-risk groups of patients, *i.e.* long-term care facility patients (31, 65), travelers returning from endemic regions (66, 67), ICU patients (28), patients that previously carried (67-69), patients that shared a room with a known carrier (70) or case contacts of carriers (71), those who've recently taken antibiotics (72, 73), or patients on mechanical ventilation, enteral feeds, or that have had prior *Clostridium difficile* infections (74). Additionally, using ASAP makes the analysis in KlebSeq streamlined, and results are easily interpretable. Lastly, the amplicon sequencing and ASAP surveillance approach is customizable and updateable. Individual assays can be added or removed, adding only the cost of new primers.

The results we present here show that KlebSeq works on DNA from numerous sample types, including pure organism culture, complex, multi-organism samples, and swab samples with low-level microbial DNA in a presumably high human DNA background. Our data show that in addition to identifying different species of *Klebsiella*, clinically important clonal lineages of *K. pneumoniae* can be identified from culture or complex specimens without culture methods. Our method distinguishes *K. pneumoniae*

403 from *K. quasipneumoniae* and *K. variicola*, the latter of which appear to be lower risk
404 species with regard to infection and virulence (47), and we identify cases where these
405 species were previously misidentified as *K. pneumoniae*. We highlight several instances
406 where culture methods failed to produce a positive *K. pneumoniae* result, including one
407 sample that contained the critical ST258 strain. We identified dozens of AMR and
408 virulence genes within individual samples, demonstrating the additional function of
409 profiling for clinically important characteristics, and were able to distinguish minor
410 genotype differences that confer different phenotypes, i.e. the *gyrA* gene, *aac(6')-Ib*
411 versus *aac(6')-Ib-cr*, and the *wzi* gene.

412 Notably, our data show that several healthy individuals carry clinically important
413 strains of *K. pneumoniae* as well as many AMR genes and siderophore virulence systems.
414 For our purposes, these healthy donor fecal DNA samples were used to validate the usage
415 of our amplicon sequencing approach on highly complex gut metagenome samples. Much
416 more study is needed to elucidate the implications of healthy host carriage of known
417 pathogenic strains of *K. pneumoniae*. Furthermore, the fact that we observed carriage of
418 the hypervirulence-associated ST380 strain from a healthy person, and the
419 hypervirulence-associated K2 capsule type in a ST14 strain from a respiratory infection,
420 lends credence to the idea that we need much more information about avirulent *K.*
421 *pneumoniae* to be able to draw conclusions about these associations.

422 Overall, the KlebSeq method was able to accurately and consistently identify and
423 characterize *Klebsiella* from complex specimens. A limitation to our study is that clonal
424 group identification in the complex specimens was not confirmed by either whole
425 metagenomic sequencing or from isolation and whole genome sequencing of the

426 *Klebsiella* from the specimen. Additionally, profiling complex specimens directly for
427 AMR and virulence genes, most of which are on mobile elements, can be confounding, as
428 it can't be known which organism carries the genes. However, KlebSeq is designed for
429 screening and surveillance for high-risk situations using a rule-in/rule-out determination
430 of the possibility of transmission events and through identification of high-risk multidrug
431 resistant or epidemic strains of *Klebsiella*. For these purposes, KlebSeq is ideal. The
432 specimen types used to validate could be considered a limitation, as we did not test rectal
433 swabs, a common specimen type for CPE surveillance, due to unavailability. However,
434 we show KlebSeq works on different swab types and fecal specimens, which addresses
435 the challenges of detection in rectal swabs. Turnaround time from sample collection to
436 result, is dependent only on current technology (not organism culture). We recently ran a
437 proof of concept of a 24-hour sample-to-answer test using different targets (data not
438 shown). This test was done on an Illumina MiSeq using only 60 cycles. Other platforms
439 may allow for this turn around time to be decreased even further
440 Rapid amplicon sequencing with automated analysis and reporting is a promising
441 response to the proposal for constant surveillance for highly transmissible or highly drug
442 resistant pathogens. Our model system, directed at *Klebsiella*, can easily be adapted for
443 multiple other pathogens and for different purposes such as environmental sampling,
444 community host screening, and, as smaller, more on-demand next-generation systems
445 become available, for diagnostics and individual patient monitoring. For several reasons,
446 amplicon sequencing is an applicable tool for healthcare facility surveillance. As these
447 technologies are adopted, considerable coordination within the healthcare facility is
448 paramount to the success of infection and outbreak prevention, with the integration of

449 isolation and barrier precautions, excellent communication, and good stewardship.
450 Nevertheless, several institutions have shown that the combination of surveillance and
451 systematic response reduces outbreaks and multidrug resistant infections (31, 33-37).
452

453 **ACKNOWLEDGEMENTS**

454 We would like to thank Dr. Michael Saubolle of Laboratory Sciences of Arizona for
455 providing isolates and specimen DNA, Krystal Sheridan for technical assistance, and
456 Tricia O'Reilly for administrative support.
457

458 **REFERENCES**

459 1. **Montgomerie JZ.** 1979. Epidemiology of Klebsiella and hospital-associated
460 infections. Rev Infect Dis **1:**736-753.
461 2. **Magill SS, Edwards JR, Bamberg W, Beldavs ZG, Dumyati G, Kainer MA,**
462 **Lynfield R, Maloney M, McAllister-Hollod L, Nadle J, Ray SM, Thompson**
463 **DL, Wilson LE, Fridkin SK, Emerging Infections Program Healthcare-**
464 **Associated I, Antimicrobial Use Prevalence Survey T.** 2014. Multistate point-
465 prevalence survey of health care-associated infections. N Engl J Med **370:**1198-
466 1208.
467 3. **Centers for Disease Control and Prevention.** 2013. Antibiotic Resistance
468 Threats in the United States, 2013. Centers for Disease Control and Prevention,
469 Atlanta, GA, U S A. http://www.cdc.gov/drugresistance/threat-report-2013/.

4. **Xu Y, Gu B, Huang M, Liu H, Xu T, Xia W, Wang T.** 2015. Epidemiology of carbapenem resistant Enterobacteriaceae (CRE) during 2000-2012 in Asia. J Thorac Dis **7:**376-385.

5. **World Health Organization.** 2014. Antimicrobial resistance global report on surveillance. World Health Organization, Geneva, Switzerland. http://apps.who.int/iris/bitstream/10665/112642/1/9789241564748_eng.pdf?ua=1.

6. **Correa L, Martino MD, Siqueira I, Pasternak J, Gales AC, Silva CV, Camargo TZ, Scherer PF, Marra AR.** 2013. A hospital-based matched case-control study to identify clinical outcome and risk factors associated with carbapenem-resistant Klebsiella pneumoniae infection. BMC Infect Dis **13:**80.

7. **Kitchel B, Sundin DR, Patel JB.** 2009. Regional dissemination of KPC-producing Klebsiella pneumoniae. Antimicrob Agents Chemother **53:**4511-4513.

8. **Frasson I, Lavezzo E, Franchin E, Toppo S, Barzon L, Cavallaro A, Richter SN, Palu G.** 2012. Antimicrobial treatment and containment measures for an extremely drug-resistant Klebsiella pneumoniae ST101 isolate carrying pKPN101-IT, a novel fully sequenced bla(KPC-2) plasmid. J Clin Microbiol **50:**3768-3772.

9. **Lascols C, Peirano G, Hackel M, Laupland KB, Pitout JD.** 2013. Surveillance and molecular epidemiology of Klebsiella pneumoniae isolates that produce carbapenemases: first report of OXA-48-like enzymes in North America. Antimicrob Agents Chemother **57:**130-136.

10. **Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Delannoy-Vieillard AS, Garin B, Le Hello S, Arlet G, Nicolas-Chanoine MH, Decre D,**

493 **Brisse S.** 2014. Genomic definition of hypervirulent and multidrug-resistant
494 Klebsiella pneumoniae clonal groups. Emerg Infect Dis **20:**1812-1820.
495 11. **Nagasaka Y, Kimura K, Yamada K, Wachino J, Jin W, Notake S,**
496 **Yanagisawa H, Arakawa Y.** 2015. Genetic profiles of fluoroquinolone-
497 nonsusceptible Klebsiella pneumoniae among cephalosporin-resistant K.
498 pneumoniae. Microb Drug Resist **21:**224-233.
499 12. **Liu Y, Wan LG, Deng Q, Cao XW, Yu Y, Xu QF.** 2015. First description of
500 NDM-1-, KPC-2-, VIM-2- and IMP-4-producing Klebsiella pneumoniae strains in
501 a single Chinese teaching hospital. Epidemiol Infect **143:**376-384.
502 13. **Cubero M, Grau I, Tubau F, Pallares R, Dominguez MA, Linares J,**
503 **Ardanuy C.** 2016. Hypervirulent Klebsiella pneumoniae clones causing
504 bacteraemia in adults in a teaching hospital in Barcelona, Spain (2007-2013). Clin
505 Microbiol Infect **22:**154-160.
506 14. **Jacobson RK, Manesen MR, Moodley C, Smith M, Williams SG, Nicol MP,**
507 **Bamford CM.** 2015. Molecular characterisation and epidemiological
508 investigation of an outbreak of blaOXA-181 carbapenemase-producing isolates of
509 Klebsiella pneumoniae in South Africa. S Afr Med J **105:**1030-1035.
510 15. **Zhu J, Sun L, Ding B, Yang Y, Xu X, Liu W, Zhu D, Yang F, Zhang H, Hu F.**
511 2016. Outbreak of NDM-1-producing Klebsiella pneumoniae ST76 and ST37
512 isolates in neonates. Eur J Clin Microbiol Infect Dis doi:10.1007/s10096-016-
513 2578-z.
514 16. **Viau RA, Hujer AM, Marshall SH, Perez F, Hujer KM, Briceno DF, Dul M,**
515 **Jacobs MR, Grossberg R, Toltzis P, Bonomo RA.** 2012. "Silent" dissemination

of Klebsiella pneumoniae isolates bearing K. pneumoniae carbapenemase in a long-term care facility for children and young adults in Northeast Ohio. Clin Infect Dis **54:**1314-1321.

17. **Bowers JR, Kitchel B, Driebe EM, MacCannell DR, Roe C, Lemmer D, de Man T, Rasheed JK, Engelthaler DM, Keim P, Limbago BM.** 2015. Genomic Analysis of the Emergence and Rapid Global Dissemination of the Clonal Group 258 Klebsiella pneumoniae Pandemic. PLoS One **10:**e0133727.

18. **Zheng R, Zhang Q, Guo Y, Feng Y, Liu L, Zhang A, Zhao Y, Yang X, Xia X.** 2016. Outbreak of plasmid-mediated NDM-1-producing Klebsiella pneumoniae ST105 among neonatal patients in Yunnan, China. Ann Clin Microbiol Antimicrob **15:**10.

19. **Pena I, Picazo JJ, Rodriguez-Avial C, Rodriguez-Avial I.** 2014. Carbapenemase-producing Enterobacteriaceae in a tertiary hospital in Madrid, Spain: high percentage of colistin resistance among VIM-1-producing Klebsiella pneumoniae ST11 isolates. Int J Antimicrob Agents **43:**460-464.

20. **Voulgari E, Gartzonika C, Vrioni G, Politi L, Priavali E, Levidiotou-Stefanou S, Tsakris A.** 2014. The Balkan region: NDM-1-producing Klebsiella pneumoniae ST11 clonal strain causing outbreaks in Greece. J Antimicrob Chemother **69:**2091-2097.

21. **Woodford N, Turton JF, Livermore DM.** 2011. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. FEMS Microbiol Rev **35:**736-755.

22. **Swaminathan M, Sharma S, Poliansky Blash S, Patel G, Banach DB, Phillips M, LaBombardi V, Anderson KF, Kitchel B, Srinivasan A, Calfee DP.** 2013. Prevalence and risk factors for acquisition of carbapenem-resistant Enterobacteriaceae in the setting of endemicity. Infect Control Hosp Epidemiol **34:**809-817.

23. **Pena C, Pujol M, Ardanuy C, Ricart A, Pallares R, Linares J, Ariza J, Gudiol F.** 1998. Epidemiology and successful control of a large outbreak due to Klebsiella pneumoniae producing extended-spectrum beta-lactamases. Antimicrob Agents Chemother **42:**53-58.

24. **Tischendorf J, de Avila RA, Safdar N.** 2016. Risk of infection following colonization with carbapenem-resistant Enterobactericeae: A systematic review. Am J Infect Control doi:10.1016/j.ajic.2015.12.005.

25. **Amit S, Mishali H, Kotlovsky T, Schwaber MJ, Carmeli Y.** 2015. Bloodstream infections among carriers of carbapenem-resistant Klebsiella pneumoniae: etiology, incidence and predictors. Clin Microbiol Infect **21:**30-34.

26. **Thaden JT, Lewis SS, Hazen KC, Huslage K, Fowler VG, Jr., Moehring RW, Chen LF, Jones CD, Moore ZS, Sexton DJ, Anderson DJ.** 2014. Rising rates of carbapenem-resistant enterobacteriaceae in community hospitals: a mixed-methods review of epidemiology and microbiology practices in a network of community hospitals in the southeastern United States. Infect Control Hosp Epidemiol **35:**978-983.

559 27. **Bassyouni RH, Gaber SN, Wegdan AA.** 2015. Fecal carriage of extended-
560 spectrum beta-lactamase- and AmpC- producing Escherichia coli among
561 healthcare workers. J Infect Dev Ctries **9:**304-308.
562 28. **Papadimitriou-Olivgeris M, Marangos M, Fligou F, Christofidou M, Sklavou**
563 **C, Vamvakopoulou S, Anastassiou ED, Filos KS.** 2013. KPC-producing
564 Klebsiella pneumoniae enteric colonization acquired during intensive care unit
565 stay: the significance of risk factors for its development and its impact on
566 mortality. Diagn Microbiol Infect Dis **77:**169-173.
567 29. **Monari C, Merlini L, Nardelli E, Cacioni M, Repetto A, Mencacci A,**
568 **Vecchiarelli A.** 2016. Carbapenem-Resistant Klebsiella pneumoniae: Results of a
569 Laboratory Surveillance Program in an Italian General Hospital (August 2014-
570 January 2015) : Surveillance of Carbapenem-resistant Klebsiella pneumoniae.
571 Adv Exp Med Biol doi:10.1007/5584_2015_5018.
572 30. **Voor in 't Holt AF, Severin JA, Goessens WH, Te Witt R, Vos MC.** 2015.
573 Instant Typing Is Essential to Detect Transmission of Extended-Spectrum Beta-
574 Lactamase-Producing Klebsiella Species. PLoS One **10:**e0136135.
575 31. **Viau R, Frank KM, Jacobs MR, Wilson B, Kaye K, Donskey CJ, Perez F,**
576 **Endimiani A, Bonomo RA.** 2016. Intestinal Carriage of Carbapenemase-
577 Producing Organisms: Current Status of Surveillance Methods. Clin Microbiol
578 Rev **29:**1-27.
579 32. **Lee BY, Bartsch SM, Wong KF, McKinnell JA, Slayton RB, Miller LG, Cao**
580 **C, Kim DS, Kallen AJ, Jernigan JA, Huang SS.** 2016. The Potential Trajectory
581 of Carbapenem-Resistant Enterobacteriaceae, an Emerging Threat to Health-Care

582 Facilities, and the Impact of the Centers for Disease Control and Prevention
583 Toolkit. Am J Epidemiol **183:**471-479.
584 33. **Parker VA, Logan CK, Currie B.** 2014. Carbapenem-Resistant
585 Enterobacteriaceae (CRE) Control and Prevention Toolkit. Agency for
586 Healthcare Research and Quality, Rockville, MD, U S A. AHRQ Publication No.
587 14-0028.
588 34. **Lucet JC, Decre D, Fichelle A, Joly-Guillou ML, Pernet M, Deblangy C,**
589 **Kosmann MJ, Regnier B.** 1999. Control of a prolonged outbreak of extended-
590 spectrum beta-lactamase-producing enterobacteriaceae in a university hospital.
591 Clin Infect Dis **29:**1411-1418.
592 35. **Schwaber MJ, Carmeli Y.** 2014. An ongoing national intervention to contain the
593 spread of carbapenem-resistant enterobacteriaceae. Clin Infect Dis **58:**697-703.
594 36. **Hayden MK, Lin MY, Lolans K, Weiner S, Blom D, Moore NM, Fogg L,**
595 **Henry D, Lyles R, Thurlow C, Sikka M, Hines D, Weinstein RA, Centers for**
596 **Disease C, Prevention Epicenters P.** 2015. Prevention of colonization and
597 infection by Klebsiella pneumoniae carbapenemase-producing enterobacteriaceae
598 in long-term acute-care hospitals. Clin Infect Dis **60:**1153-1161.
599 37. **Epson EE, Pisney LM, Wendt JM, MacCannell DR, Janelle SJ, Kitchel B,**
600 **Rasheed JK, Limbago BM, Gould CV, Kallen AJ, Barron MA, Bamberg**
601 **WM.** 2014. Carbapenem-resistant Klebsiella pneumoniae producing New Delhi
602 metallo-beta-lactamase at an acute care hospital, Colorado, 2012. Infect Control
603 Hosp Epidemiol **35:**390-397.

38. **Pisney LM, Barron MA, Kassner E, Havens D, Madinger NE.** 2014. Carbapenem-resistant Enterobacteriaceae rectal screening during an outbreak of New Delhi metallo-beta-lactamase-producing Klebsiella pneumoniae at an acute care hospital. Infect Control Hosp Epidemiol **35:**434-436.

39. **Dhar S, Martin ET, Lephart PR, McRoberts JP, Chopra T, Burger TT, Tal-Jasper R, Hayakawa K, Ofer-Friedman H, Lazarovitch T, Zaidenstein R, Perez F, Bonomo RA, Kaye KS, Marchaim D.** 2016. Risk Factors and Outcomes for Carbapenem-Resistant Klebsiella pneumoniae Isolation, Stratified by Its Multilocus Sequence Typing: ST258 Versus Non-ST258. Open Forum Infect Dis **3:**ofv213.

40. **Gilbert JA, Meyer F, Jansson J, Gordon J, Pace N, Tiedje J, Ley R, Fierer N, Field D, Kyrpides N, Glockner FO, Klenk HP, Wommack KE, Glass E, Docherty K, Gallery R, Stevens R, Knight R.** 2010. The Earth Microbiome Project: Meeting report of the "1 EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6 2010. Stand Genomic Sci **3:**249-253.

41. **Sahl JW, Lemmer D, Travis J, Schupp J, Gillece J, Aziz M, Driebe E, Drees K, Hicks N, Williamson C, Hepp C, Smith D, Roe C, Engelthaler D, Wagner D, Keim P.** 2016. The Northern Arizona SNP Pipeline (NASP): accurate, flexible, and rapid identification of SNPs in WGS datasets. bioRxiv doi:10.1101/037267.

42. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA.** 2010. The

Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20:**1297-1303.

43. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725-2729.

44. **Letunic I, Bork P.** 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res **39:**W475-478.

45. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455-477.

46. **Sahl JW, Caporaso JG, Rasko DA, Keim P.** 2014. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ **2:**e332.

47. **Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J, Brisse S, Cao H, Wilksch J, Gorrie C, Schultz MB, Edwards DJ, Nguyen KV, Nguyen TV, Dao TT, Mensink M, Minh VL, Nhu NT, Schultsz C, Kuntaman K, Newton PN, Moore CE, Strugnell RA, Thomson NR.** 2015. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proc Natl Acad Sci U S A **112:**E3574-3581.

650 48. **Brisse S, Passet V, Haugaard AB, Babosan A, Kassis-Chikhani N, Struve C,**
651 **Decre D.** 2013. wzi Gene sequencing, a rapid method for determination of
652 capsular type for Klebsiella strains. J Clin Microbiol **51:**4073-4078.
653 49. **Colman RE, Schupp JM, Hicks ND, Smith DE, Buchhagen JL, Valafar F,**
654 **Crudu V, Romancenco E, Noroc E, Jackson L, Catanzaro DG, Rodwell TC,**
655 **Catanzaro A, Keim P, Engelthaler DM.** 2015. Detection of Low-Level Mixed-
656 Population Drug Resistance in Mycobacterium tuberculosis Using High Fidelity
657 Amplicon Sequencing. PLoS One **10:**e0126626.
658 50. **Colman RE, Anderson J, Lehmkuhl E, Lemmer D, Heaton H, Georghiou SB,**
659 **Wiggins K, Crudu V, Cohen T, Rodwell TC, Engelthaler DM.** 2016. Straight
660 from Sputum: Rapid Drug Susceptibility Testing of Mycobacterium tuberculosis
661 Directly from Clinical Samples using Amplicon Sequencing. Submitted to J Clin
662 Microb.
663 51. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: a flexible trimmer for
664 Illumina sequence data. Bioinformatics **30:**2114-2120.
665 52. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2.
666 Nat Methods **9:**357-359.
667 53. **Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD,**
668 **Marshall D.** 2013. Using Tablet for visual exploration of second-generation
669 sequencing data. Brief Bioinform **14:**193-202.
670 54. **Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J,**
671 **Holt KE.** 2014. SRST2: Rapid genomic surveillance for public health and
672 hospital microbiology labs. Genome Med **6:**90.

673 55. **Li JJ, Munoz-Price LS, Spychala CN, DePascale D, Doi Y.** 2016. New Delhi
674 Metallo-β-Lactamase-1–Producing Klebsiella pneumoniae, Florida, USA. Emerg
675 Infect Dis doi: 10.3201/eid2204.151176.
676 56. **Shon AS, Bajwa RP, Russo TA.** 2013. Hypervirulent (hypermucoviscous)
677 Klebsiella pneumoniae: a new and dangerous breed. Virulence **4:**107-118.
678 57. **Davis GS, Waits K, Nordstrom L, Weaver B, Aziz M, Gauld L, Grande H,**
679 **Bigler R, Horwinski J, Porter S, Stegger M, Johnson JR, Liu CM, Price LB.**
680 2015. Intermingled Klebsiella pneumoniae Populations Between Retail Meats and
681 Human Urinary Tract Infections. Clin Infect Dis **61:**892-899.
682 58. **Ocampo AM, Chen L, Cienfuegos AV, Roncancio G, Chavda KD, Kreiswirth**
683 **BN, Jimenez JN.** 2015. A Two-Year Surveillance in Five Colombian Tertiary
684 Care Hospitals Reveals High Frequency of Non-CG258 Clones of Carbapenem-
685 Resistant Klebsiella pneumoniae with Distinct Clinical Characteristics.
686 Antimicrob Agents Chemother **60:**332-342.
687 59. **Scott, RD II.** 2009. The Direct Medical Costs of Healthcare-Associated
688 Infections in U.S. Hospitals and the Benefits of Prevention. Centers for Disease
689 Control and Prevention, Atlanta, GA, U S A.
690 http://www.cdc.gov/HAI/pdfs/hai/Scott_CostPaper.pdf.
691 60. **Patel TS, Nagel JL.** 2015. Clinical outcomes of Enterobacteriaceae infections
692 stratified by carbapenem MICs. J Clin Microbiol **53:**201-205.
693 61. **Daroukh A, Delaunay C, Bigot S, Ceci JM, Siddhoun N, Bukreyeva I, Raisin**
694 **J, Porcheret H, Maisonneuve L, Bouldouyre MA.** 2014. Characteristics and

costs of carbapenemase-producing enterobacteria carriers (2012/2013). Med Mal Infect **44:**321-326.

62. **van Duin D, Cober E, Richter SS, Perez F, Kalayjian RC, Salata RA, Evans S, Fowler VG, Jr., Kaye KS, Bonomo RA.** 2015. Impact of therapy and strain type on outcomes in urinary tract infections caused by carbapenem-resistant Klebsiella pneumoniae. J Antimicrob Chemother **70:**1203-1211.

63. **Sahl JW, Sistrunk JR, Fraser CM, Hine E, Baby N, Begum Y, Luo Q, Sheikh A, Qadri F, Fleckenstein JM, Rasko DA.** 2015. Examination of the Enterotoxigenic Escherichia coli Population Structure during Human Infection. MBio **6:**e00501.

64. **Stoesser N, Sheppard AE, Moore CE, Golubchik T, Parry CM, Nget P, Saroeun M, Day NP, Giess A, Johnson JR, Peto TE, Crook DW, Walker AS, Modernizing Medical Microbiology Informatics G.** 2015. Extensive Within-Host Diversity in Fecally Carried Extended-Spectrum-Beta-Lactamase-Producing Escherichia coli Isolates: Implications for Transmission Analyses. J Clin Microbiol **53:**2122-2131.

65. **Lin MY, Lyles-Banks RD, Lolans K, Hines DW, Spear JB, Petrak R, Trick WE, Weinstein RA, Hayden MK, Centers for Disease C, Prevention Epicenters P.** 2013. The importance of long-term acute care hospitals in the regional epidemiology of Klebsiella pneumoniae carbapenemase-producing Enterobacteriaceae. Clin Infect Dis **57:**1246-1252.

66. **Janvier F, Delacour H, Tesse S, Larreche S, Sanmartin N, Ollat D, Rapp C, Merens A.** 2014. Faecal carriage of extended-spectrum beta-lactamase-producing

enterobacteria among soldiers at admission in a French military hospital after aeromedical evacuation from overseas. Eur J Clin Microbiol Infect Dis **33:**1719-1723.

67. **Allyn J, Angue M, Belmonte O, Lugagne N, Traversier N, Vandroux D, Lefort Y, Allou N.** 2015. Delayed diagnosis of high drug-resistant microorganisms carriage in repatriated patients: three cases in a French intensive care unit. J Travel Med **22:**215-217.

68. **Bart Y, Paul M, Eluk O, Geffen Y, Rabino G, Hussein K.** 2015. Risk Factors for Recurrence of Carbapenem-Resistant Enterobacteriaceae Carriage: Case-Control Study. Infect Control Hosp Epidemiol **36:**936-941.

69. **Kim KR, Lee JY, Park HY, Kwak SH, Lim YJ, Hong MJ, Kim SK, Park SY, Kim HJ, Choi HS, Kim MN, Choi HR, Jeong JS, Choi SH.** 2015. Clearance Rate of Carbapenemase-Producing Enterobacteriaceae Carriage Among Hospitalized Patients. Infect Control Hosp Epidemiol **36:**1361-1362.

70. **Torres-Gonzalez P, Cervera-Hernandez ME, Niembro-Ortega MD, Leal-Vega F, Cruz-Hervert LP, Garcia-Garcia L, Galindo-Fraga A, Martinez-Gamboa A, Bobadilla-Del Valle M, Sifuentes-Osornio J, Ponce-de-Leon A.** 2015. Factors Associated to Prevalence and Incidence of Carbapenem-Resistant Enterobacteriaceae Fecal Carriage: A Cohort Study in a Mexican Tertiary Care Hospital. PLoS One **10:**e0139883.

71. **Gagliotti C, Ciccarese V, Sarti M, Giordani S, Barozzi A, Braglia C, Gallerani C, Gargiulo R, Lenzotti G, Manzi O, Martella D, Moro ML.** 2013.

Active surveillance for asymptomatic carriers of carbapenemase-producing Klebsiella pneumoniae in a hospital setting. J Hosp Infect **83:**330-332.

72. **Ludden C, Cormican M, Vellinga A, Johnson JR, Austin B, Morris D.** 2015. Colonisation with ESBL-producing and carbapenemase-producing Enterobacteriaceae, vancomycin-resistant enterococci, and meticillin-resistant Staphylococcus aureus in a long-term care facility over one year. BMC Infect Dis **15:**168.

73. **Armand-Lefevre L, Angebault C, Barbier F, Hamelet E, Defrance G, Ruppe E, Bronchard R, Lepeule R, Lucet JC, El Mniai A, Wolff M, Montravers P, Plesiat P, Andremont A.** 2013. Emergence of imipenem-resistant gram-negative bacilli in intestinal flora of intensive care patients. Antimicrob Agents Chemother **57:**1488-1495.

74. **Prasad N, Labaze G, Kopacz J, Chwa S, Platis D, Pan CX, Russo D, LaBombardi VJ, Osorio G, Pollack S, Kreiswirth BN, Chen L, Urban C, Segal-Maurer S.** 2016. Asymptomatic rectal colonization with carbapenem-resistant Enterobacteriaceae and Clostridium difficile among residents of a long-term care facility in New York City. Am J Infect Control doi:10.1016/j.ajic.2015.11.021.

TABLES

Table 1. Sensitivity and specificity of the KlebSeq species identification assays on genomic DNA from known isolates, DNA from urine, wound, and respiratory specimens for which clinical culture results are known, and results of screening across specimens of unknown content.

		Amplicon sequencing assay[a]					
Isolate DNA n=252	Total screened	Kp-M1	Kp-M2	Kp-M1 + Kp-M2	Kquasi	Kvari	Koxy
K. pneumoniae	69	68	67	69	0	0	0
K. quasipneumoniae	2	0	0	0	2	0	0
K. variicola	2	2	0	0	0	2	0
K. oxytoca	22	0	0	0	0	0	21
Non-target species	157	2/88	0/88	2/88	0/155	0/155	0/135
Sensitivity		99%	97%	100%	100%	100%	95%
Specificity		98%	100%	98%	100%	100%	100%
Urine DNA n=46	Total screened	Kp-M1	Kp-M2	Kp-M1 + Kp-M2	Kquasi	Kvari	Koxy
K. pneumoniae	16	14	16	16	2 (1 mix[b])	1 (mix[b])	0
K. oxytoca	6	1	1	1 (CG34)	0	0	6
Other species	24	2	2	2 (1 CG392)	0	0	0
Unknown	0	0	0	0	0	0	0
Sensitivity		88%	100%	100%	-	-	100%
Specificity		90%	90%	90%	-	-	100%
Wound DNA n=40	Total screened	Kp-M1	Kp-M2	Kp-M1 + Kp-M2	Kquasi	Kvari	Koxy
K. pneumoniae	1	0	0	0	0	1	0
K. oxytoca	1	0	0	0	0	0	1
Other species	31	3	3	3 (1 CG29)	0	0	0
Unknown	7	1	1	1	0	0	1
Sensitivity		100%	0%	100%	-	-	100%
Specificity		94%	94%	94%	-	-	100%
Respiratory specimen DNA n=87	Total screened	Kp-M1	Kp-M2	Kp-M1 + Kp-M2	Kquasi	Kvari	Koxy
K. pneumoniae	6	6	6	6	0	0	0
K. oxytoca	1	0	0	0	0	0	1
Other species	77	7 (1 ST258)	5	7 (2 CG37, 1 CG36)	0	0	1
Unknown	3	0	0	0	0	0	0
Sensitivity		100%	100%	100%	-	-	100%
Specificity		92%	95%	92%	-	-	99%
Fecal specimen DNA n=89	Total screened	Kp-M1	Kp-M2	Kp-M1 + Kp-M2	Kquasi	Kvari	Koxy
Unknown	89	8	5	8	1 (mix[b])	3 (2 mix[b])	13

[a] Assay descriptions are as follows: Kp-M1 and Kp-M2 are K. pneumoniae species identification assays that detect targets M1 and M2 in the *K. pneumoniae* genome. Kquasi, Kvari, and Koxy are the *K. quasipneumoniae*-, *K. variicola*-, and *K. oxytoca*-specific assays, respectively.

[b] These species were found as mixtures with *K. pneumoniae*, based on a proportion (≥10%) of sequencing reads containing the species-defining SNP.

763 Table 2. Isolates used for assay validation and results of strain typing with amplicon

764 sequencing.

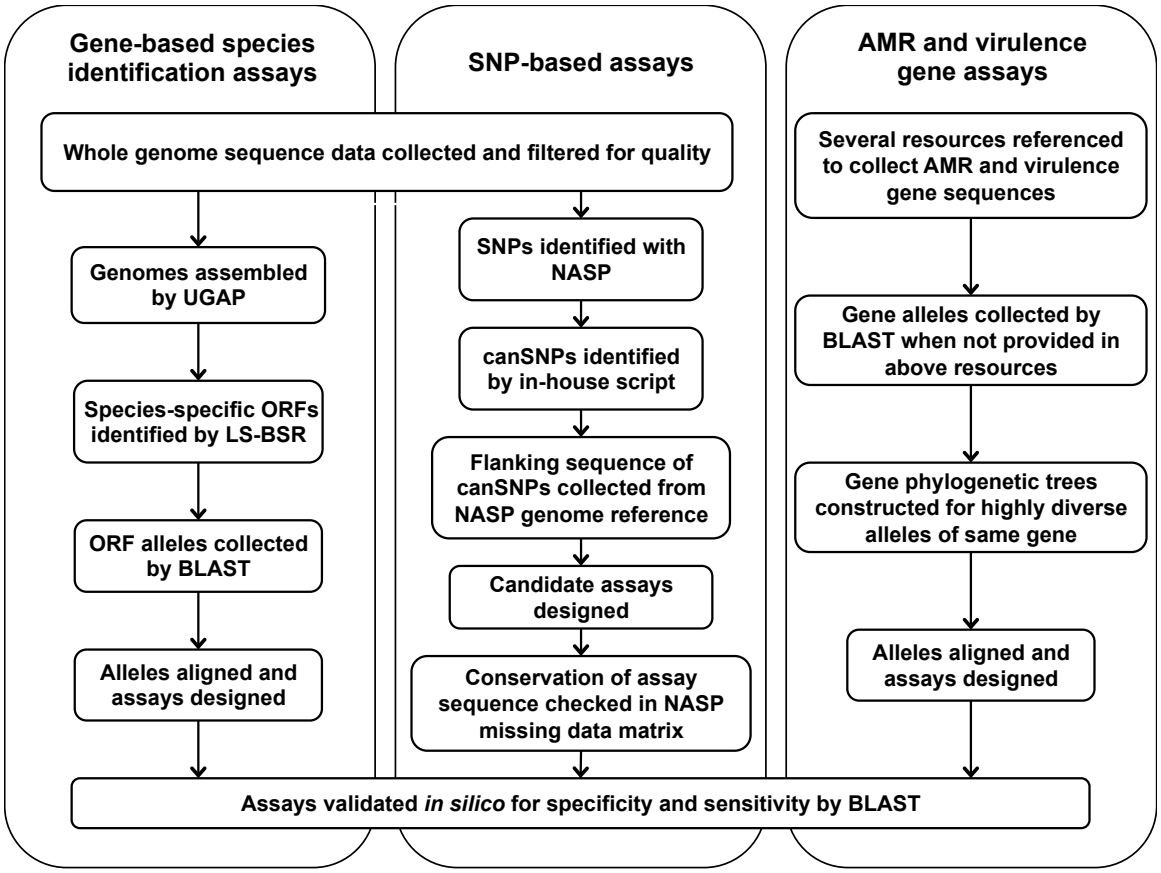
Isolate sequence type	No. isolates	ASAP Strain Typing Assay Results	Capsule Typing Results by Partial *wzi* Sequence[a]
ST11	3	CG258, CG258wo395	wzi-39 or -75, wzi-74, not typable
ST14	5	CG14, ST14, innerST14	wzi-2
ST14 SLV	1	CG14	wzi-16
ST15	2	CG14, ST15	wzi-24 or -45
ST20	2	CG20, ST20	wzi-84, wzi-118
ST23	8	ST23	wzi-1
ST34, ST34 SLV	2	CG34	wzi-114, wzi-12
ST36	2	CG36	wzi-27 or -79
ST37	2	CG37	wzi-50, wzi-39 or -75
ST39	1	No group	wzi-2
ST42	2	bigCG42, CG42	wzi-29
ST43	1	CG43	wzi-30
ST45	1	CG45	wzi-133
ST65	1	CG25	wzi-72
ST101	2	CG43	wzi-29, wzi-137
ST107	1	No group	wzi-74
ST111	1	CG111	wzi-63
ST147	1	CG392	wzi-64
ST152	1	CG105	wzi-150
ST228	1	CG34	wzi-116[b]
ST234	1	No group	wzi-114
ST249	2	No group	wzi-128
ST258 group, no clade	6	CG258, CG258wo395, 3xST258	wzi-154
ST258 group, clade 2	2	CG258, CG258wo395, 3xST258, clade 2	wzi-154
ST277	1	No group	wzi-97 or -185
ST334	1	*K. quasipneumoniae*	wzi-68
ST340	2	CG258, CG258wo395, ST340	wzi-50, wzi-173
ST376	1	bigCG42, CG42	wzi-2
ST380	1	ST380	wzi-203
ST636	1	No group	wzi-155
ST719	1	No group	wzi-192
ST776	1	No group	wzi-39[b] or -75[b] or -193[b]
ST833	1	CG258, CG258wo395	wzi-50
ST978	1	*K. quasipneumoniae*	wzi-212[b]
ST1401	1	No group	wzi-96
ST82	2	No group	wzi-128
ST260	1	No group	wzi-1
ST360 SLV	1	*K. variicola*	wzi-53
ST427 SLV	1	No group	wzi-64
ST513 SLV	1	No group	wzi-87
ST815 SLV	1	No group	wzi-114[b]
ST244 SLV	1	No group	wzi-162[b]
ST2006	1	*K. variicola*	wzi-227
ST2055	1	No group	wzi-14

[a] The Illumina version 2 chemistry used provides approximately 500 bp of sequence data. The amplicon size for the *wzi* assay (48) is approximately 580 bp.

[b] *wzi* allele represents best match when one or multiple SNPs were present.

765

766 **FIGURES**



767 Figure 1. Workflow of the amplicon sequencing target library and assay development pipeline.

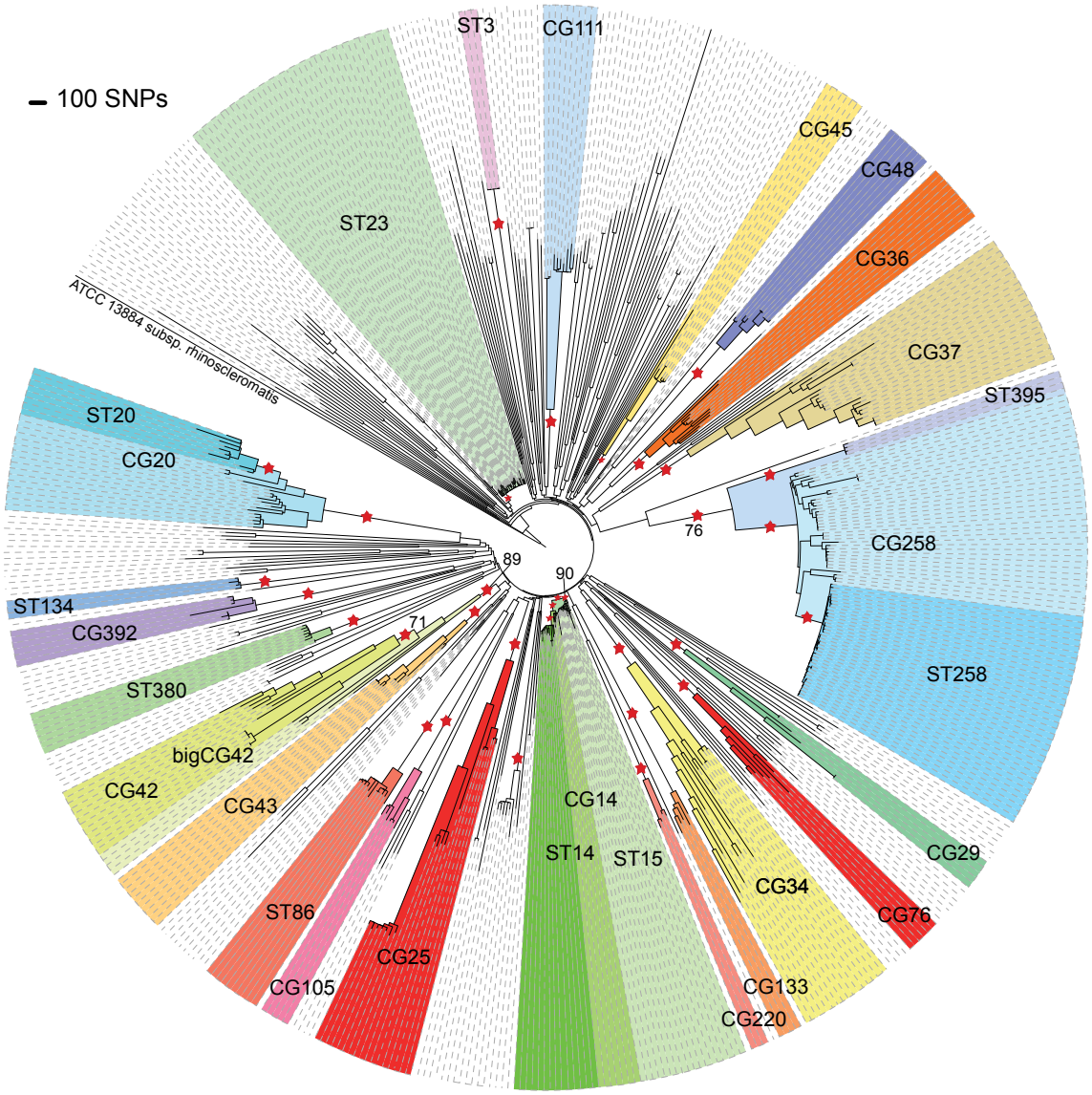


Figure 2. Maximum parsimony tree with 100 bootstraps of the SNPs among 548 *K. pneumoniae* genomes. Major clonal groups are colored, and locations of canonical SNPs for strain identification assays are marked with stars. All branches labeled with canonical SNPs had >99% bootstrap support, except on the three branches indicated.

768

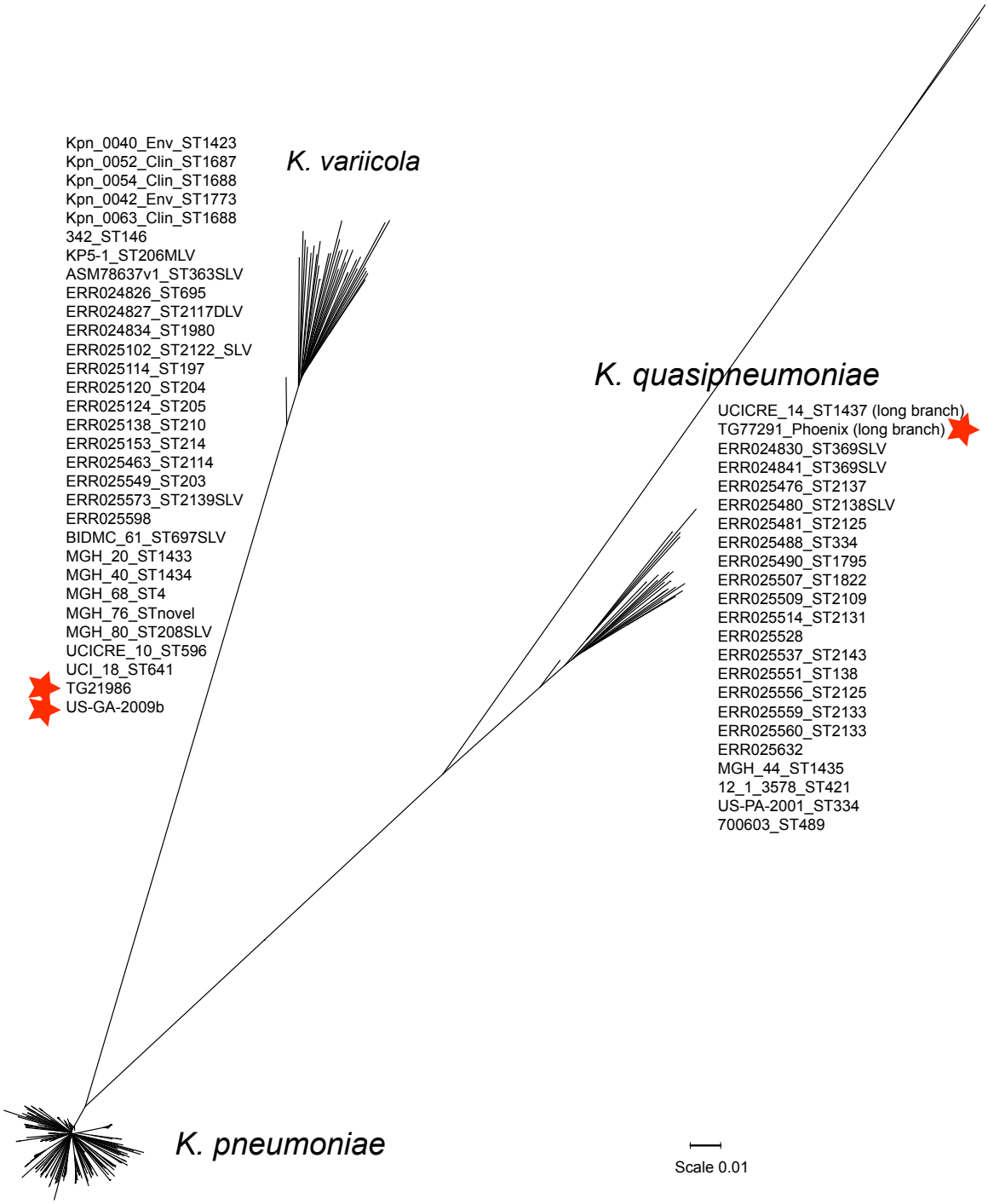


Figure 3. Neighbor-joining tree with 100 bootstraps of the SNPs among the diverse set of *K. pneumoniae*, the *K. variicola*, and *K. quasipneumoniae* genomes, with our unknown specimens that typed as *K. variicola* and *K. quasipneumoniae* labeled with stars.

769