Measurement of *Klebsiella* Intestinal Colonization Density to Assess Infection Risk

Running Title: *Klebsiella* colonization density and infection risk

Yuang Sun[1], Alieysa Patel[1], John SantaLucia[2], Emily Roberts[3], Lili Zhao[3], Keith S. Kaye[4], Krishna Rao[4], and Michael A. Bachman[1,5]

[1] Department of Pathology, Michigan Medicine; [2] DNA Software, Inc.; [3] Department of Biostatistics, School of Public Health; [4] Division of Infectious Diseases, Department of Internal Medicine, [5] Department of Microbiology and Immunology, Michigan Medicine, University of Michigan, Ann Arbor, MI

Corresponding Author: Michael A. Bachman

Contact Information:

mikebach@med.umich.edu

7510E MSRB 1

1301 Catherine

Ann Arbor, MI 48109

Abstract

Background: *Klebsiella pneumoniae* and closely related species *K. variicola* and *K. quasipneumoniae* are common causes of healthcare-associated infections, and patients frequently become infected with their intestinal colonizing strain. To assess the association between *Klebsiella* colonization density and subsequent infections, a case-control study was performed.

Methods: A multiplex qPCR assay was developed and validated to quantify *Klebsiella (K. pneumoniae, K. variicola, and K. quasipneumoniae* combined) relative to total bacterial DNA copies in rectal swabs. Cases of *Klebsiella* infection were identified based on clinical definitions and having a clinical culture isolate and preceding or co-incident colonization isolate with the same *wzi* capsular sequence type. Controls were colonized patients without subsequent infection and were matched 2:1 to cases based on age, sex, and rectal swab collection date. Quantitative PCR (qPCR) from rectal swab samples was used to measure the association between relative abundance (RA) of *Klebsiella* and subsequent infections.

Results: *Klebsiella* RA by qPCR highly correlated with 16S sequencing (ρ=0.79; P <.001). The median *Klebsiella* RA in the study group was 2.6% (interquartile range (IQR) 0.1-22.5, n=238), and was higher in cases (15.7%, IQR 0.93-52.6%, n=83) than controls (1.01%, IQR 0.02-12.8%; n=155; P <0.0001). After adjusting for multiple clinical covariates using inverse probability of treatment weighting, subjects with a *Klebsiella* RA > 22% had a 2.87-fold (1.64-5.03, P =0.0003) increased odds of infection compared to those with lower colonization density levels.

Conclusions: Measurement of colonization density by qPCR could represent a novel approach to identify hospitalized patients at risk for *Klebsiella* infection.

Importance

Colonization by bacterial pathogens often precedes infection, and offers a window of opportunity to prevent these infections. *Klebsiella* colonization is significantly and reproducibly associated with subsequent infection, however factors that enhance or mitigate this risk in individual patients are unclear. This study developed an assay to measure the density of *Klebsiella* colonization, relative to total fecal bacteria, in rectal swabs from hospitalized patients. Applying this assay to 238 colonized patients, high *Klebsiella* density defined as >22% of total bacteria, was significantly associated with subsequent infection. Based on widely available polymerase chain reaction (PCR) technology, this type of assay could be deployed in clinical laboratories to identify patients at increased risk of *Klebsiella* infections. As novel therapeutics are developed to eliminate pathogens from the gut microbiome, a rapid *Klebsiella* colonization density assay could identify patients who would benefit from this type of infection prevention interventions.

Introduction

Klebsiella pneumoniae is a leading cause of healthcare-associated infections (HAIs)(1). Recent studies have shown that *Klebsiella variicola* and *Klebsiella quasipneumoniae* are closely related to, yet distinct species from, *K. pneumoniae* and cause indistinguishable infections (2, 3). These species are part of the *K. pneumoniae* complex that together pose a serious public health threat.

Klebsiella commonly colonizes hospitalized patients and can cause bacteremia, pneumonia, and urinary tract infections (UTIs). Prior studies show that *Klebsiella* colonization is significantly associated with subsequent infections and 80% of infections in colonized patients are caused by an intestinal colonizing strain (4, 5). Increased colonization density may increase the risk of subsequent infection. For example, intestinal domination (defined as at least 30% relative colonization density) of *Proteobacteria* was associated with subsequent Gram-negative bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation and relative and absolute abundance of *Enterobacterales* associate interactively with infection in intensive care patients (6, 7). In long term acute care patients, relative abundance of carbapenem-resistant *K. pneumoniae* above 22% was a risk factor for bacteremia (8). Similarly, increased relative abundance of *Escherichia* and *Enterococcus* in the gut are risk factors for corresponding bacteriuria or UTI in kidney transplant patients (9). These studies indicate that in many cases colonization is a necessary intermediate step before infection.

Understanding the association between *Klebsiella* colonization and subsequent infections could provide opportunities for identification of high-risk patients, intervention, and ultimately prevention of infection. Additionally, little is known about the association between *Klebsiella* gut

colonization density and specific infection types, such as bacteremia, pneumonia, and UTI. Measuring *Klebsiella* gut density and assessing gut density as a risk factor for various infections may also shed light on the mechanisms of dissemination from the colonized gut to various infection sites. However, the lack of a rapid and reliable assay to quantify *Klebsiella* relative abundance in the gut has been a hindrance to both research and potential clinical implementation. Here we report a qPCR-based assay that can quickly and accurately quantify *Klebsiella* from rectal swab specimens. We employed this assay in a case-control study to assess *Klebsiella* rectal relative abundance as a risk factor for bacteremia, pneumonia or UTI in colonized patients and found a significant association after adjusting for clinical variables.

Results

In silico Analysis

To design a quantitative PCR (qPCR) assay for the *K. pneumoniae* complex, 31 *K. pneumoniae*, *K. quasipeumoniae*, and *K. variicola* strains with complete genomes were selected as "inclusivity" for *in silico* analysis (**Supplemental Dataset 1**). Additionally, 8 *K. oxytoca* and *Raoultella* strains were selected to represent "near-neighbor exclusivity", and the human genome and common members of the gut microbiome were used as background sequences that should not be detected by the assay. PanelPlex™ *in silico* analysis was performed and the *ybiL* gene was identified as an optimal target for the assay (**Table 1**). An overall performance score, based on primer and probe thermodynamic stabilities with their targets, as well as any off-target bindings, were computed to constitute overall performance scores for each of 7 assay designs. The *ybiL* assay design 1 (overall score 99.9) has a predicted probe binding score of 99.5% with all 31 strains in the "inclusivity" set. In regard to the primers, all 31 predicted binding scores of the

forward primer (hereafter *ybiL*-F) are above 50.0%, with 29 above 86.0%. Twenty-four predicted binding scores of the reverse primer (hereafter *ybiL*-R) are above 95.0% while 6 are close to 50.0% and 1 below 50.0%. Additionally, *ybiL* assay design 1 was predicted to have no amplifications with any background genomes. Although its primer bindings have variations, its probe binding scores are uniformly excellent. Therefore, *ybiL* assay design 1 (hereafter *ybiL* assay) was chosen for further validation. To assess the relative abundance of the *K. pneumoniae* complex, the *ybiL* assay and a previously described panbacterial qPCR assay targeting the 23S rRNA gene (sepsis manuscript) were combined to construct a multiplex qPCR assay (hereafter Kp qPCR assay). Overall, the *ybiL* assay has good coverage of the *K. pneumoniae* complex and the Kp qPCR assay provides a possible solution to quantify the *K. pneumoniae* complex in clinical specimens.

K. pneumoniae **Complex Diversity Panel**

Eleven isolates with polymorphisms at sites of *ybiL* primer and probe binding were picked and grown overnight in LB broth. They were re-suspended in Amies media (BD ESwab™) and normalized based on colony forming units (CFU) for DNA extraction. The lab strain ATCC 43816 KPPR1, which contains a single mismatch identical to Kp8399, was set as reference and delta-delta-C_T (ddC_T) method was used to calculate the abundance of *Klebsiella* relative to KPPR1 (set as 100%) by qPCR. (**Figure 1)** Of the 11 isolates, 9 are within the range of 88 to >99% relative to KPPR1. Although they share the same polymorphism, the abundance calculation of Kp2950 was 72% relative to KPPR1 whereas Kp6966 was 88%. This suggests that technical imprecision may be greater than systematic errors caused by polymorphisms. Taken

127 together, the Kp qPCR assay should have accurate and consistent performance with most clinical
128 isolates despite the existence of binding variations.
129
130 **Specificity**
131 The Kp qPCR assay was designed to quantify *K. pneumoniae*, *K. quasipeumoniae*, and *K.*
132 *variicola* but not other *Klebsiella* species or other common bacteria in the gut microbiota. To
133 validate its specificity, *K. aerogenes, K. pneumoniae subsp. ozaenae, K. oxytoca, Raoultella*
134 *planticola*, *Raoultella ornithinolytica*, *Escherichia coli,* and *Pseudomonas aeruginosa* were
135 tested by the Kp qPCR assay. (**Supplemental Table 1**) *K. pneumoniae* KPPR1 was used as
136 positive control. Only KPPR1 and *K. pneumoniae subsp. ozaenae* were amplified by both *ybiL*
137 assay and 23S assay, whereas *K. aerogenes, K. oxytoca, Raoultella planticola, Raoultella*
138 *ornithinolytica, Escherichia coli, and Pseudomonas aeruginosa* strains were only amplified by
139 23S assay but not *ybiL* assay, demonstrating that *ybiL* assay specifically amplified the designated
140 targets but not its near-neighbors or background sequences.
141
142 **Linearity**
143 To assess Kp qPCR assay's linearity, KPPR1 was grown in LB broth overnight and re-suspended
144 in Amies media. A serial 10-fold dilution was made in triplicate and enumerated for CFU counts.
145 The CFU counts were close to a theoretical 10-fold dilution, as the slope was 0.9889 and R
146 square 0.9999. The slopes of *ybiL* and 23S assay were both 3.37 and R squares were both 0.9994,
147 demonstrating that both assays had good linearity and efficiency. (**Figure 3**)
148
149 **Precision and Accuracy**

To assess the precision and accuracy of Kp qPCR assay, a mixture of 89% *Bacteroides ovatus*, 10% KPPR1, and 1% *Serratia marcescens* by DNA quantifications was made and diluted 10-fold serially. The mixture was amplified by the Kp qPCR assay and the relative abundances of *Klebsiella* were calculated relative to KPPR1 using the ddC_T method. (**Figure 4**) The relative abundances of *Klebsiella* ranged from 15.0-18.1%, with a mean value of 17.0% (expect 10.0%). The standard deviations (hereafter SD) ranged from 0.246 to 7.386. The quantifications by Kp qPCR assay were consistent across concentrations of 4 $\log_{10}$ differences. When the total DNA concentrations of the mixture were over 1×10^{-2} ng/µl , the SDs of the relative abundances of *Klebsiella* were less than 1.5%. However, at 1×10^{-3} ng/µl total DNA concentration, the assay became less precise, as the SD increased to 7.39%. At 1×10^{-4} ng/µl total DNA concentration, *ybiL* assay did not detect *Klebsiella*. The copy number of 23S rRNA gene is organism-specific with 5, 7, and 8 copies in *Bacteroides ovatus*, *Serratia marcescens* and *K. pneumoniae* respectively. After adjustment for these differences, the calculated relative abundance of *Klebsiella* ranged from 10.0-12.0%, with a mean value of 11.3% (SD 0.163 to 4.91), demonstrating that Kp qPCR assay can accurately quantify *Klebsiella* from contrived samples. Fortunately, a bias in the calculation based on 23S copy number in the overall population relative to *Klebsiella* would not be expected to impact the ability to measure relative differences, as demonstrated below.

Accuracy: Relative differences

To assess Kp qPCR assay's ability to distinguish different relative abundances of *Klebsiella*, KPPR1 and *Escherichia coli* O6:K2:H1 CFT073 were mixed in Amies media according to CFU counts to make 10%, 20%, 30% and 40% *Klebsiella* mixtures. Ten-fold serial dilutions were

made from each mixture and genomic DNA of each serial dilution were isolated and then amplified by Kp qPCR assay. The relative abundance of *Klebsiella* was calculated relative to KPPR1 using the ddC_T method. (**Figure 5**) At total DNA concentrations from ~0.02-2 ng/μL the assay was able to detect relative differences between all dilutions, and at .002 ng/μL it can tell all differences except between 30 and 40% *Klebsiella.* At all but the lowest total bacterial concentration, the assay can reliably detect 10% differences in *Klebsiella* relative abundance.

Accuracy: In Comparison to 16S rRNA sequencing

To compare the relative abundance calculated by qPCR to the gold standard of 16S rRNA sequencing, 26 samples with 16S rRNA sequence data were analyzed (**Figure 6**). *Klebsiella* relative abundances by Kp qPCR were highly correlated with 16S rRNA sequencing (Spearman's $\rho=0.79$; $P < .001$). However, 16S sequencing does not have the resolution to separate all species and as a result, the OTU that contained *K. pneumoniae* complex also included other *Klebsiella* spp. as well as sequences from other genera such as *Enterobacter* spp. This might explain why in some cases *Klebsiella* relative abundances by Kp qPCR were significantly lower than that by 16S sequencing.

Case-control study

To assess the association between *Klebsiella* colonization density and subsequent infection, a nested case-control study was performed. We previously enrolled 1978 subjects from 2087 separate admissions with *Klebsiella* colonization in a rectal swab, collected as part of routine surveillance for vancomycin-resistant *Enterococcus* in intensive care and oncology wards (Companion manuscript). Of these colonized subjects, 83 cases were identified that met clinical

196 definitions of subsequent infection and had an infecting isolate that matched a colonizing isolate
197 on or prior to the day of infection by *wzi* sequence typing. This included 41 bloodstream
198 infections, 19 respiratory infections, and 23 urinary tract infections. Controls were defined as
199 colonized patients who had no documented *Klebsiella* infection but had a negative clinical
200 culture collected of the same type as the matching case. Each case was matched to two controls
201 based on age, sex, date of rectal swab collection and swab availability, for a total of 155 controls.
202 To find matches for every case, the criteria for age were modified for 2 cases (±20 years), and
203 the criteria for swab collection date were modified for 4 cases (±118 days). Cases had a
204 significantly higher comorbidity score, and were more likely to have exposure to diuretics,
205 vitamin D, a vasopressin blocker prior to the rectal swab. They were also more likely to have
206 been exposed to high-risk antibiotics associated with disruption of the intestinal microbiome
207 (10). Cases also had lower baseline albumin, and were more likely to have a urinary catheter or
208 feeding tube prior to the rectal swab.
209
210 To assess colonization density, the Kp qPCR assay was performed on all of the rectal swab
211 samples (Figure 7). The median relative abundance was 0.022 overall, with an interquartile range
212 (IQ) of 0.001-0.22 and an overall range of 0 to 1.38. In cases, the median was 0.15 (IQ 0.009-
213 0.53) and in controls the median was 0.01 (IQ 0.0002-0.13). To determine if dominance was
214 associated with infection, while accounting for the case-control matching, a cutoff for dominance
215 was applied. The 75th percentile of 0.22 in the overall dataset was chosen, consistent with cutoffs
216 of 0.22 and 0.3 used in previous studies (6, 8). Subjects with a *K. pneumoniae* gut colonization
217 density >22% had a 3.34-fold (1.95-5.72, P <0.0001) increased odds of infection compared to
218 those with lower colonization density levels.

To adjust for patient variables associated with *Klebsiella* infection, inverse probability of treatment weighting was used. In the overall cohort of 1978 subjects, an explanatory model for invasive infection was built using baseline clinical features at the time of colonization (Companion manuscript, Supplemental Table 2). The model built by purposeful selection selected the following variables for inclusion: Elixhauser score, depression, diuretic use, vitamin D use, use of pressors/inotropes, low serum albumin (<2.5 g/dL), and exposure to antibiotics with high risk of microbiome disruption. These variables were then used to model *K. pneumoniae* gut colonization density >22% to generate weights for the nested case-control study. Using weights derived from these clinical covariates (available in 228 of 240 subjects), patients with a *K. pneumoniae* gut colonization density >22% had a 2.87-fold (1.64-5.03, P =0.0003) increased odds of infection compared to those with lower colonization density levels. In a secondary analysis by site of infection, increased relative abundance was also significantly associated with bloodstream infection (OR 4.137, 95% CI 1.448-11.818, *P*=0.0084), whereas associations with urinary tract (OR 3.037, 95% CI 0.571-16.17, P=0.19) and respiratory infections (OR 1.32, 95% CI 0.38-4.565, P=0.66) did not reach significance.

Discussion

The goal of this study was to measure the association between *Klebsiella* colonization density and subsequent infection. To develop a robust method that accurately and precisely measured the relative abundance of *K. pneumoniae*, *K. quasipneumoniae* and *K. variicola* among the gut microbiota, we developed a novel qPCR assay for detecting these dominant members of the *K. pneumoniae* complex and combined it with measurement of 23S rRNA gene copies. Analytical

validation indicated that this assay is inclusive of multiple strains of each species and is able to distinguish as little as 10% differences in relative abundance between samples. Applying this assay to a case-control study of *Klebsiella* infections among colonized, intensive care patients indicated that increased *Klebsiella* density is associated with subsequent infection in both unadjusted and adjusted analysis.

The finding that *Klebsiella* colonization density is associated with subsequent infection raises several interesting possibilities. One is that infection risk is dictated by how much *Klebsiella* is present in the gut, independent of the varying gene content of *Klebsiella* strains. Indeed, we and others have shown that detectable colonization is associated with infection (4, 5), and the lower limit of detection for culture is an indirect measure of abundance in rectal swabs. However, we have also demonstrated that particular *Klebsiella* genes are associated with infection as opposed to asymptomatic colonization (11), indicating that which strain a patient is colonized with affects their risk. There is likely to be an important interaction between *Klebsiella* gene content and colonization density, where certain genes may increase gut fitness and therefore gut abundance. Alternatively, there may be strains where gut abundance is increased based on microbiome factors extrinsic to *Klebsiella* but the risk of infection is further increased by virulence genes that act at the site of infection. Finally, *Klebsiella* strains with fitness genes that increase abundance in the gut and virulence genes that act later in pathogenesis are likely to pose the greatest risk of infection in colonized patients.

The main limitation of this study was the relatively small number of cases (n=83). We compensated for this by using a case-control design and an inverse probability of treatment

265 weighting to account for clinical variables potentially associated with infection and intestinal
266 dominance without significant loss of statistical power (12). However, we were limited in our
267 ability to investigate associations by site of infection. Bloodstream infections were the largest
268 infection type and were independently associated with intestinal dominance. Intriguingly, the
269 point estimate for urinary tract infections (3.037) was similar to the overall odds ratio (2.87) for
270 infection. This may indicate that intestinal dominance is also associated with UTIs, perhaps
271 because a key step in pathogenesis is thought to be transit of intestinal bacteria across the
272 perineum to the urethra.
273
274 This study further supports the growing paradigm that intestinal dominance can be used to
275 predict infections in our hospitals. Previous studies have demonstrated that dominance of
276 carbapenemase-producing *K. pneumoniae* in long-term care patients (8), and *K. pneumoniae* in
277 allogeneic stem cell transplant patients (13) are associated with infection. This study evaluated a
278 more heterogeneous population of intensive care patients with a combined outcome of
279 bloodstream, respiratory or urinary tract infections and found the same association. The
280 successful use of qPCR demonstrates the feasibility of measuring relative abundance of targeted
281 pathogens in the gut using methods that are standard in clinical microbiology laboratories and
282 inexpensive relative to next-generation sequencing. This a key step in moving towards infection
283 prevention in hospitalized patients. A qPCR could be applied to detect colonization in rectal
284 swabs as well as quantify it in a single step, thereby incorporating two levels of *Klebsiella*
285 infection risk. Combined with the assessment of patient risk factors and perhaps targeted testing
286 for *Klebsiella* virulence genes, an integrated risk assessment could be performed. If this relative
287 risk is high enough, infection prevention interventions should be considered. Fortunately, safe

and effective therapeutic strategies to eliminate gut colonization by pathogens are emerging and results from fecal transplant studies are encouraging (14). In the near future, it may be possible to assess the risk of a carbapenem-resistant *Klebsiella* infection at the time of hospital admission and prevent it without the use of antibiotics.

Methods

Study design and subject enrollment. This study was approved by the Institutional Review Boards of the University of Michigan Medical School (IRBMED). To assess the role of *Klebsiella* colonization density on risk of subsequent invasive infection, we conducted a nested case-control study drawn from a larger cohort of 1978 patients consecutively enrolled from 2087 inpatient admissions. Subjects admitted to intensive care units and oncology wards at our hospital undergo routine surveillance by rectal swab culture for vancomycin-resistant *Enterococcus*. After such testing, we collected residual media from these swabs and enrolled subjects into our study if colonization with *K. pneumoniae* or *K. variicola* was detected by selective culture on MacConkey agar and confirmed by MALDI-TOF identification (Bruker MALDI Biotyper). Cases were identified from this larger cohort and matched to controls as described below.

Case definitions. Michigan Medicine patients from intensive care units (ICU) and select wards (hematology, oncology, and hematopoietic stem cell transplant) with *Klebsiella* colonization based on a rectal swab culture and a positive *Klebsiella* blood, respiratory, or urine cultures were identified as putative cases. Manual chart review was conducted by the study team to decide if

311 they met clinical definitions of pneumonia or urinary tract infections (15-19). All patients with a
312 *Klebsiella* blood culture were considered to have an infection. For those meeting clinical case
313 definitions of infection, the clinical isolate and preceding rectal swab isolates were evaluated for
314 concordance by *wzi* gene sequencing as previously described (4, 20). We have previously
315 demonstrated that *wzi* sequencing has similar discriminatory power to 7-gene multi-locus
316 sequence typing (4). Patients with concordant infection and colonizing isolates were considered a
317 case. Controls were defined as colonized patients who had no documented *Klebsiella* infection
318 but had a negative culture collected of the same type as the matching case. Cases and controls
319 were matched 1:2 based on age ±10 years, sex, and rectal swab collection date ±90 days. This
320 study was approved by the University of Michigan Institutional Review Board.
321
322 **Samples for PCR analysis.** Rectal swabs were collected using the ESwab™
323 Collection and Transport System (Becton Dickinson, Franklin Lakes, NJ, USA), which elutes the
324 sample into 1mL Amies medium. Unless specified otherwise, contrived samples that were used
325 in PCR analysis were eluted in the Amies medium as well. The 89% *Bacteroides ovatus*, 10%
326 KPPR1, and 1% *Serratia marcescens* mixtures were suspended in ddH_2O.
327
328 **Bacterial DNA extraction.** Genomic DNA was isolated using the MagAttract
329 PowerMicrobiome DNA/RNA kit (Qiagen, Germantown, MD, USA) on the epMotion ® 5075
330 liquid handler (Eppendorf, Hauppauge, NY, USA). A volume of 100 uL was added into the bead
331 plate for each rectal swab and contrived sample. Subsequent steps of DNA extraction were
332 conducted following the manufacturer's instructions. *Bacteroides ovatus*, KPPR1, and *Serratia*

marcescens cultures were extracted using DNeasy Blood & Tissue Kits (Qiagen, Germantown, MD, USA) following the manufacturer's instruction for Gram-negative bacteria.

***In silico* assay design.** PanelPlex™ (DNA Software, Ann Arbor, MI, USA) *in silico* analysis was performed. Thirty-One *K. pneumoniae*, *K. quasipeumoniae*, and *K. variicola* genomes were selected as "inclusivity", 8 *K. oxytoca* and *Raoultella* strains as "near-neighbor exclusivity", and the human genome and common members of the gut microbiome as background sequences that should not be detected by the assay (**Supplemental Dataset 1**). PanelPlex™ utilizes ThermoBLAST™ technology to scan for thermodynamically stable off-target hybridizations that cause false-positive tests.

Quantitative PCR assay. Real-time PCR was performed using primers (Integrated DNA Technologies, Coralville, IA, USA) and probes (Thermo Fisher Scientific, Waltham, MA, USA) with sequences and concentrations listed in Table 1 in combination with QuantiTect Multiplex PCR Kit (Qiagen, Germantown, MD, USA). PMAxx (Biotium, Fremont, CA, USA) with a final concentration of 6uM was added. A volume of 5 µL was used for each template. The final reaction volume was 25 µL. Prior to template addition, the reaction mixture was incubated for 10 minutes at room temperature and then treated in a Biotium PMA-Lite LED photolysis device for 10 minutes. PCR conditions were 50°C for 2 minutes, 95°C for 15 minutes, then 40 cycles of 94°C for 1 minute and 60°C for 1 minute on a QuantStudio 3 real-time thermocycler (Thermo Fisher Scientific, Waltham, MA, USA). KPPR1 genomic DNA was used as positive control and 100% reference for calculating *Klebsiella* relative abundances. Relative abundance was

355 calculated using the ddC$_T$ method, i.e. Relative abundance = $2^{(dC_{T\,Sample} - dC_{T\,KPPR1})} \times 100\%$,

356 where $dC_T = C_{T_{23S}} - C_{T_{ybiL}}$.

357

358 **Statistical Analysis.** Linearity was validated by linear regression. Spearman's rank correlation

359 coefficient was used for correlation between Kp qPCR and 16S rRNA sequencing. One-way

360 ANOVA and Tukey's post test was performed to compare each dilution of the KPPR1 and

361 *Escherichia coli* O6:K2:H1 CFT073 mixtures. Statistical analysis was performed by Prism 8

362 (GraphPad, San Diego, CA, USA).

363

364 **Clinical modeling.**

365 Conditional logistic regression was used to study the effect of relative abundance of colonization

366 on *Klebsiella* infection, while accounting for the case-control matching. Unadjusted analysis was

367 performed after dichotomizing the relative abundance at the third quartile of 22%. To adjust for

368 patient variables associated with *Klebsiella* infection, an inverse probability of treatment

369 weighting approach was used. However, given the smaller sample size in our nested case-control

370 study, we turned to the larger cohort from which our subjects were derived to identify most

371 robustly the clinical variables that best explain risk of infection (Companion manuscript). First,

372 using the increased power afforded by the overall cohort of 1978 subjects, an explanatory

373 unconditional logistic regression model for invasive infection was built using baseline clinical

374 features at the time of colonization. The model was built by purposeful selection, a common

375 technique(21). Briefly, purposeful selection begins with an unadjusted analysis of each variable

376 to select candidates with statistically significant associations with the outcome, and these are

377 included in the starting set of covariates for the multivariable model. Iteratively, covariates are

then removed from the model if they are non-significant (P >.05) and not a confounder (i.e. do not affect the estimate of other variables' coefficients by at least 20%). A change in a parameter estimate above the specified level indicates that the excluded variable was important in the sense of providing a needed adjustment for one or more of the variables remaining in the model (i.e. it should be retained even if not significant). The resulting model contains significant covariates and other confounders, and then variables not included are added back one at a time. Once again the model is iteratively reduced as before but only for the variables that were additionally added. At the end of this final step, we are left with a multivariable model for *Klebsiella* infection drawn from the larger cohort of subjects with rectal *Klebsiella* colonization. The variables selected for inclusion by this method were then used to generate propensity scores for *Klebsiella* colonization density >22%, but only for subjects in the nested case-control study, again via unconditional logistic regression. The propensity scores were then used to generate weights for the IPTW process and subsequent weighted conditional logistic regression for *Klebsiella* infection The inverse of these propensity scores were then used as weights in the subsequent weighted conditional logistic regression for *Klebsiella* infection with robust standard errors. Both unadjusted and adjusted analysis were conducted using *proc survveylogistic* procedure in SAS (version 9.4), and covariate balance was assessed using the *cobalt* package in R.

Data Availability:

16S sequencing samples from rectal swabs PR08714, PR11216, PR05497, PR09929, PR10907, PR05713, PR06107, PR08411, PR08133, PR05629, PR08147, PR07331, PR12066, PR07876, PR08427, PR07976, PR08661, PR05017, PR08962, PR09113, PR08102, PR09612, PR08748, PR08048, PR06316, PR10214 are available in the sequence read archive (PRJNA641414).

401 Deidentified data from human subjects may be made available upon request, pending approval

402 from the University of Michigan Institutional Review Board.

403

404

405 **Acknowledgements**

406 MAB would like to thank Caitlyn Holmes for assistance with figure formatting.

407

408 Research reported in this publication was supported by the National Institute Of Allergy And

409 Infectious Diseases of the National Institutes of Health under Award Number R01AI125307. The

410 content is solely the responsibility of the authors and does not necessarily represent the official

411 views of the National Institutes of Health.

412 **Table 1. Primers and probes used in the study**

Oligonucleotide name	Final Rx Conc.	Sequence (5' to 3')
*ybiL*_Probe	200 nM	FAM-CGTCCACAGCGTAAAGGCATGTT-MGB
23S_Probe	200 nM	VIC-CCTAAGGTAGCGAAATTCCTTGT-MGB
ybiL-F	400 nM	AACGTAGCGCAGGATGGATCTTCCG
ybiL-R	400 nM	GACAGATCGCTGGTGGCCTGATA
23S-F	400 nM	ATTACGCCATTCGTGCAGGTCGGA
23S-R	400 nM	TAAACGGCGGCCGTAACTATAACGGT

413

414

415

416

417

418

419

420 Table 2. Patient Demographics

	No. (%) or mean +/- SD		
Variable	Cases (n=83)	Controls(n=155)	P-value (logistic regression)
Age	60.08±12.90	59.43±12.29	0.759
gender (Male)	44(53.0)	83(53.5)	>.99
Race (White)	70(84.3)	122(78.7)	0.368
Elixhauser Comorbidity Score	7.53±3.25	6.62±3.12	0.05
Weighted Elixhauser Score	22.40±11.53	19.36±11.89	0.104
Depression	29(34.9)	40(25.8)	0.203
Diuretic	30(36.1)	36(23.2)	0.03
Vitamin D	18(21.7)	18(11.6)	0.032
Vasopressin blocker	19(22.9)	15(9.7)	0.008
Broad-spectrum antibiotic[1]	30(36.1)	30(19.4)	0.005
Baseline albumin (g/dL)	2.53±0.71	2.78±0.73	0.008
Albumin >=2.5 (g/dL)	46(55.4)	112(72.3)	0.003
urinary catheter	60(72.3)	86(55.5)	0.016
feeding tube	7(8.43)	3(1.94)	0.023
ventilator	35(42.2)	66(42.6)	>.99
central line	34(41.0)	74(47.7)	0.301

[1]Defined as exposure to any of the following in the 90 days prior to *Klebsiella* colonization: third- or fourth-generation cephalosporins, fluoroquinolones, lincosamides, β-lactam / β-lactamase inhibitor combinations, oral vancomycin, and carbapenems (10).

421

References

1. Magill SS, Edwards JR, Bamberg W, Beldavs ZG, Dumyati G, Kainer MA, Lynfield R, Maloney M, McAllister-Hollod L, Nadle J, Ray SM, Thompson DL, Wilson LE, Fridkin SK, Emerging Infections Program Healthcare-Associated I, Antimicrobial Use Prevalence Survey T. 2014. Multistate point-prevalence survey of health care-associated infections. N Engl J Med 370:1198-208.

2. 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:E3574-81.

3. Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, Olsen RJ. 2017. Whole-Genome Sequencing of Human Clinical *Klebsiella pneumoniae* Isolates Reveals Misidentification and Misunderstandings of *Klebsiella pneumoniae*, *Klebsiella variicola*, and *Klebsiella quasipneumoniae*. mSphere 2:e00290-17.

4. Martin RM, Cao J, Brisse S, Passet V, Wu W, Zhao L, Malani PN, Rao K, Bachman MA. 2016. Molecular Epidemiology of Colonizing and Infecting Isolates of *Klebsiella pneumoniae*. mSphere 1:e00261-16.

5. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, Pratt NF, Garlick JS, Watson KM, Pilcher DV, McGloughlin SA, Spelman DW, Jenney AWJ, Holt

KE. 2017. Gastrointestinal Carriage Is a Major Reservoir of *Klebsiella pneumoniae* Infection in Intensive Care Patients. Clin Infect Dis 65:208-215.

6. Taur Y, Xavier JB, Lipuma L, Ubeda C, Goldberg J, Gobourne A, Lee YJ, Dubin KA, Socci ND, Viale A, Perales MA, Jenq RR, van den Brink MR, Pamer EG. 2012. Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. Clin Infect Dis 55:905-14.

7. Rao K, Seekatz A, Bassis C, Sun Y, Mantlo E, Bachman MA. 2020. *Enterobacterales* Infection after Intestinal Dominance in Hospitalized Patients. mSphere 5:pii: 5/4/e00450-20.

8. Shimasaki T, Seekatz A, Bassis C, Rhee Y, Yelin RD, Fogg L, Dangana T, Cisneros EC, Weinstein RA, Okamoto K, Lolans K, Schoeny M, Lin MY, Moore NM, Young VB, Hayden MK, Centers for Disease C, Prevention Epicenters P. 2019. Increased Relative Abundance of *Klebsiella pneumoniae* Carbapenemase-producing *Klebsiella pneumoniae* Within the Gut Microbiota Is Associated With Risk of Bloodstream Infection in Long-term Acute Care Hospital Patients. Clin Infect Dis 68:2053-2059.

9. Magruder M, Sholi AN, Gong C, Zhang L, Edusei E, Huang J, Albakry S, Satlin MJ, Westblade LF, Crawford C, Dadhania DM, Lubetzky M, Taur Y, Littman E, Ling L, Burnham P, De Vlaminck I, Pamer E, Suthanthiran M, Lee JR. 2019. Gut uropathogen abundance is a risk factor for development of bacteriuria and urinary tract infection. Nat Commun 10:5521.

10. Baggs J, Jernigan JA, Halpin AL, Epstein L, Hatfield KM, McDonald LC. 2018. Risk of Subsequent Sepsis Within 90 Days After a Hospital Stay by Type of Antibiotic Exposure. Clin Infect Dis 66:1004-1012.

468 11. Martin RM, Cao J, Wu W, Zhao L, Manthei DM, Pirani A, Snitkin E, Malani PN, Rao K,
469 Bachman MA. 2018. Identification of Pathogenicity-Associated Loci in *Klebsiella*
470 *pneumoniae* from Hospitalized Patients. mSystems 3:e00015-18.
471 12. Haukoos JS, Lewis RJ. 2015. The Propensity Score. JAMA 314:1637-8.
472 13. Stoma I, Littmann ER, Peled JU, Giralt S, van den Brink MRM, Pamer EG, Taur Y.
473 2020. Compositional flux within the intestinal microbiota and risk for bloodstream
474 infection with gram-negative bacteria. Clin Infect Dis doi:10.1093/cid/ciaa068.
475 14. Gargiullo L, Del Chierico F, D'Argenio P, Putignani L. 2019. Gut Microbiota Modulation
476 for Multidrug-Resistant Organism Decolonization: Present and Future Perspectives. Front
477 Microbiol 10:1704.
478 15. American Thoracic S, Infectious Diseases Society of A. 2005. Guidelines for the
479 management of adults with hospital-acquired, ventilator-associated, and healthcare-
480 associated pneumonia. Am J Respir Crit Care Med 171:388-416.
481 16. CDC. 2016. Urinary tract infection (catheter-associated urinary tract infection [CAUTI]
482 and non-catheter-associated urinary tract infection [UTI]) and other urinary system
483 infection [USI]) events. Centers for Disease Control and Prevention, Atlanta, GA.
484 17. Hooton TM, Bradley SF, Cardenas DD, Colgan R, Geerlings SE, Rice JC, Saint S,
485 Schaeffer AJ, Tambayh PA, Tenke P, Nicolle LE, Infectious Diseases Society of A.
486 2010. Diagnosis, prevention, and treatment of catheter-associated urinary tract infection
487 in adults: 2009 International Clinical Practice Guidelines from the Infectious Diseases
488 Society of America. Clin Infect Dis 50:625-63.
489 18. Mandell LA, Wunderink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, Dowell
490 SF, File TM, Jr., Musher DM, Niederman MS, Torres A, Whitney CG, Infectious

Diseases Society of A, American Thoracic S. 2007. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis 44 Suppl 2:S27-72.

19. CDC. 2014. National Healthcare Safety Network device-associated module: ventilator-associated events.Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, Atlanta, GA.

20. Brisse S, Passet V, Haugaard AB, Babosan A, Kassis-Chikhani N, Struve C, Decre D. 2013. wzi Gene sequencing, a rapid method for determination of capsular type for *Klebsiella* strains. J Clin Microbiol 51:4073-8.

21. Bursac Z, Gauss CH, Williams DK, Hosmer DW. 2008. Purposeful selection of variables in logistic regression. Source Code Biol Med 3:17.

Figure Legends.

Figure 1. The Kp qPCR assay accurately quantifies *K. pneumoniae*, *K. variicola* and *K. quasipneumoniae*. Eleven *Klebsiella* clinical isolates were extracted and amplified by Kp qPCR assay, each with 3 technical replicates. A. Quantification of each isolate relative to KPPR1 set as 100%. B. The alignment of the amplicons of the 11 isolates with Kp qPCR assay primers and probe.

Figure 2: The Kp qPCR assay has the accuracy, precision and linearity to distinguish small differences in *Klebsiella* relative abundance. Linearity was assessed with serial dilutions of KPPR1 genomic DNA (n=3 technical replicates). Linear regression of Ct values from the *ybiL* assay (blue) and 23S (red) is shown (A). Precision and accuracy was assessed with serial dilutions of a mixture of 89% *Bacteroides ovatus*, 10% *K. pneumoniae* KPPR1, and 1% *Serratia marcescens* genomic DNA (3 technical replicates). Mean and SD of both direct and adjusted quantifications, after consideration of 23S gene copy number variations, are shown B). The ability to discern relative differences is shown using serial dilutions of mixtures of KPPR1 and *Escherichia coli* CFT073 (3 technical replicates). For each dilution, one-way ANOVA was performed ($P<0.0001$ for all) and Tukey's post test was performed (* for each comparison out of six with $P<0.05$) (C). Accuracy was compared to 16S rRNA sequencing using OTU 4 that contains *Klebsiella*. The correlation between *Klebsiella* relative abundance by Kp qPCR and OTU4 of 16S rRNA sequencing analysis that contains *Klebsiella* was measured by Spearman's rank correlation coefficient on 26 rectal swab samples (D).

526 Figure 3. Increased relative abundance of *Klebsiella* is associated with subsequent infection. The
527 relative abundance of *Klebsiella* in rectal swabs as measured by the Kp qPCR assay is shown for
528 238 specimens, with 83 cases matched 1:2 to 155 controls based on age, sex and date of swab
529 collection. Median and interquartile ranges are shown. **** P <0.0001 by unadjusted conditional
530 logistic regression.
531

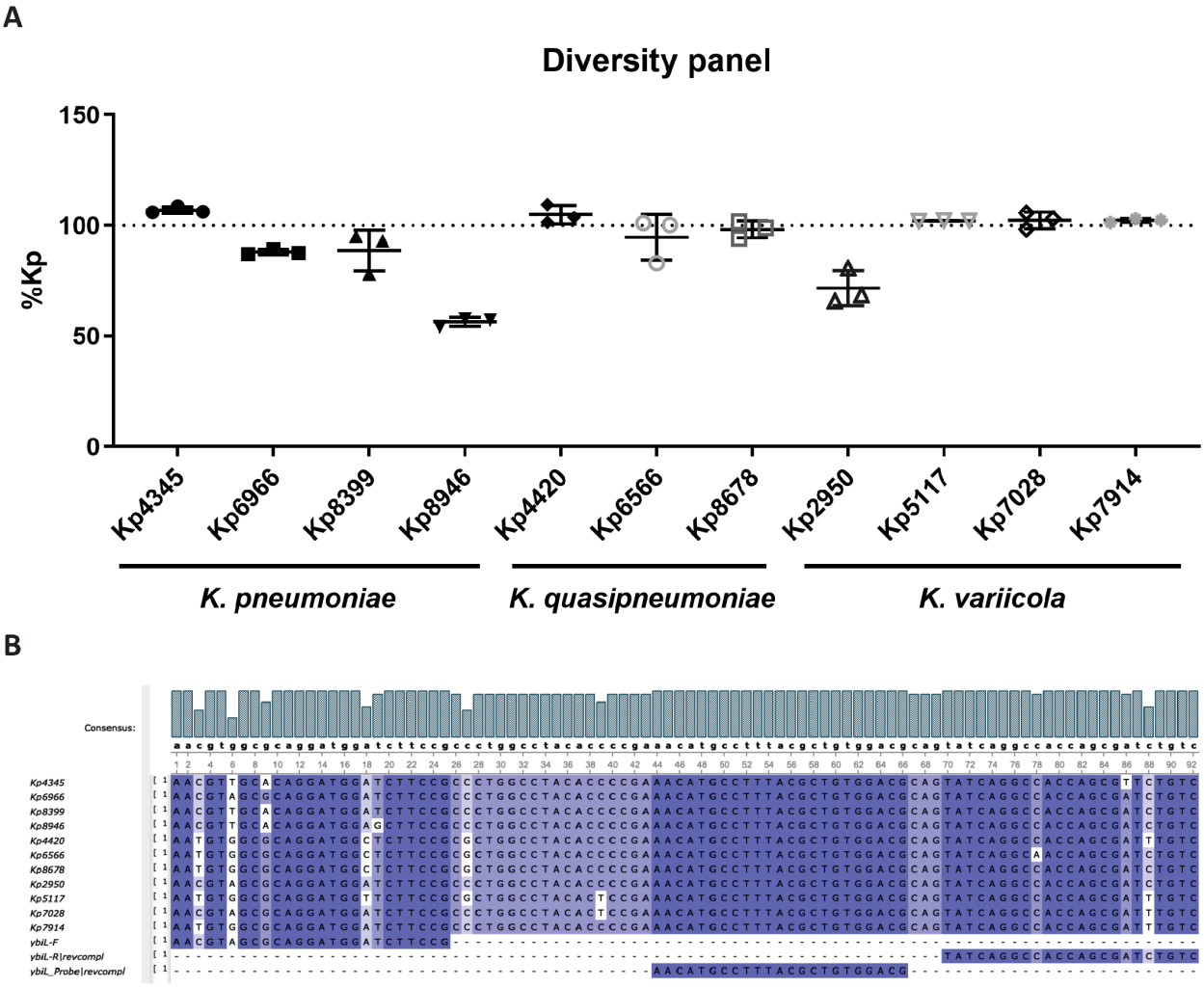
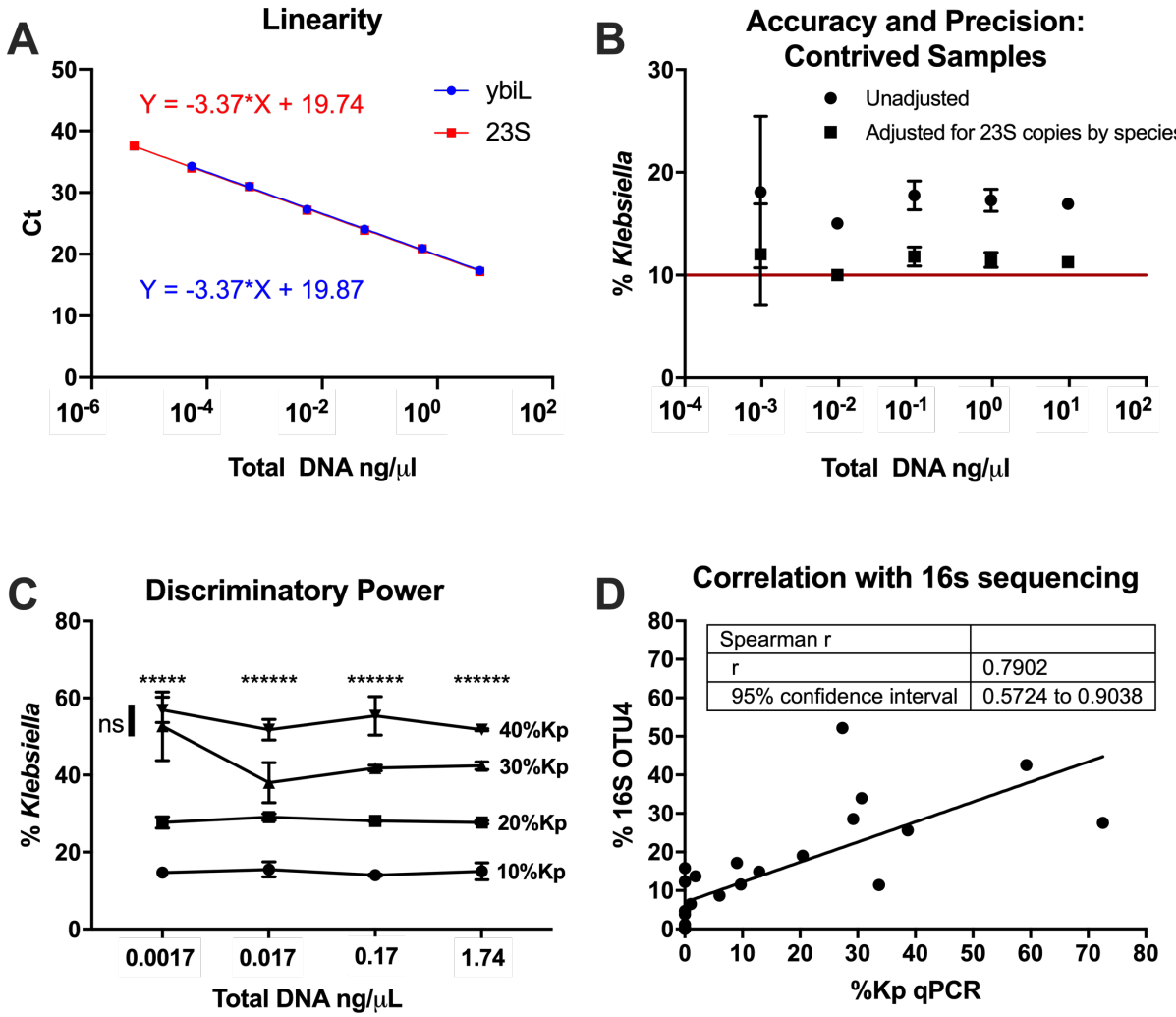


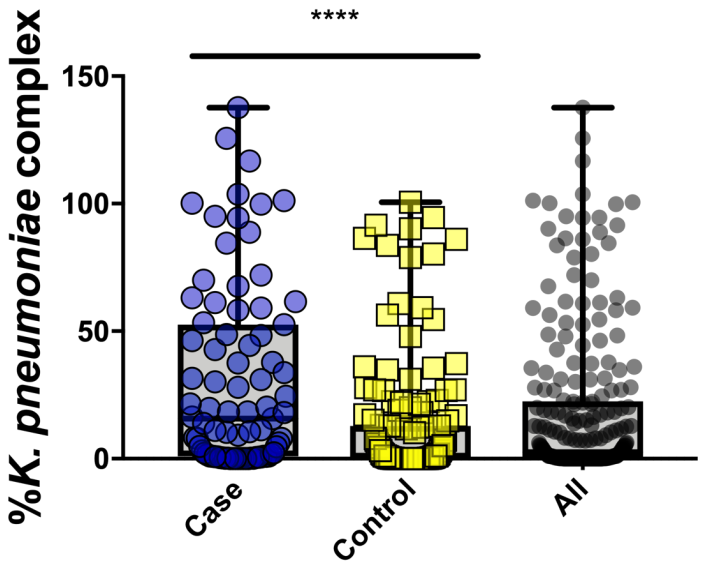
Figure 1. The Kp qPCR assay accurately quantifies *K. pneumoniae*, *K. variicola* and *K. quasipneumoniae*. Eleven *Klebsiella* clinical isolates were extracted and amplified by Kp qPCR assay, each with 3 technical replicates. A. Quantification of each isolate relative to KPPR1 set as 100%. B. The alignment of the amplicons of the 11 isolates with Kp qPCR assay primers and probe.



542

543 Figure 2: The Kp qPCR assay has the accuracy, precision and linearity to distinguish small
544 differences in *Klebsiella* relative abundance. Linearity was assessed with serial dilutions of
545 KPPR1 genomic DNA (n=3 technical replicates). Linear regression of Ct values from the *ybiL*
546 assay (blue) and 23S (red) is shown (A). Precision and accuracy were assessed with serial
547 dilutions of a mixture of 89% *Bacteroides ovatus*, 10% *K. pneumoniae* KPPR1, and 1% *Serratia*
548 *marcescens* genomic DNA (3 technical replicates). Mean and SD of both direct and adjusted
549 quantifications, after consideration of 23S gene copy number variations, are shown B). The
550 ability to discern relative differences is shown using serial dilutions of mixtures of KPPR1 and

551 *Escherichia coli* CFT073 (3 technical replicates). For each dilution, one-way ANOVA was
552 performed (P<0.0001 for all) and Tukey's post test was performed (* for each comparison out of
553 six with P<0.05) (C). Accuracy was compared to 16S rRNA sequencing using OTU 4 that
554 contains *Klebsiella*. The correlation between *Klebsiella* relative abundance by Kp qPCR and
555 OTU4 of 16S rRNA sequencing analysis that contains *Klebsiella* was measured by Spearman's
556 rank correlation coefficient on 26 rectal swab samples (D).
557



	Case	Control	All
Number of values	83	155	238
Minimum	0	0	0
25% Percentile	0.932	0.017	0.101
Median	15.74	1.009	2.605
75% Percentile	52.61	12.8	22.54
Maximum	137.7	100.6	137.7
Mean	30.45	11.9	18.37
Std. Deviation	36.09	22.79	29.44
Std. Error of Mean	3.961	1.831	1.908

Figure 3. Increased relative abundance of *Klebsiella* is associated with subsequent infection. The relative abundance of *Klebsiella* in rectal swabs as measured by the Kp qPCR assay is shown for 238 specimens, with 83 cases matched 1:2 to 155 controls based on age, sex and date of swab collection. Median and interquartile ranges are shown. **** P <0.0001 by unadjusted conditional logistic regression.