

Forensic genomics of novel *Klebsiella* strain

Forensic genomics of a novel *Klebsiella quasipneumoniae* type from an NICU in China reveals patterns of genetic diversity, evolution and epidemiology

Laura Perlaza-Jiménez^{1,2}, Qing Wu¹, Von Vergel L. Torres², Xiaoxiao Zhang³, Jiahui Li¹, Andrea Rocker², PhD., Trevor Lithgow^{2,*}, Tieli Zhou^{1,*}, Dhanasekaran Vijaykrishna^{1,2,*}

¹The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China

² Monash University, Victoria, Australia

³ Women's Hospital School of Medicine Zhejiang University, Hangzhou, China

Correspondence:

vijay.dhanasekaran@monash.edu

trevor.lithgow@monash.edu

wyztli@163.com

Keywords:

Klebsiella quasipneumoniae subsp. *similipneumoniae*; genetic diversity; virulence determinants; antimicrobial resistance

Data summary:

Genome sequences generated in this study are available in NCBI under BioProject ID PRJNA610124. All bioinformatic protocols used to process the genomic data are available at https://github.com/vjlab/KpIIB_ST2727.

1 **ABSTRACT**

2 During March of 2017 a neonate patient suffered severe diarrhea and subsequently
3 developed septicemia and died, with *Klebsiella* isolated as the causative microorganism.
4 Coincident illness of an attending staff member and three other neonates with *Klebsiella*
5 triggered a response, leading to a detailed microbiological and genomics investigation of
6 isolates collected from the staff member and all 21 co-housed neonates. Multilocus
7 sequence typing and genomic sequencing identified that the *Klebsiella* from all 21
8 neonates was a new MLST ST2727, and belonged to a less frequently detected
9 subspecies *K. quasipneumoniae* subsp. *similipneumoniae* (KpIIB). Genomic
10 characterization showed that the isolated ST2727 strains had diverged from other KpIIB
11 strains at least >90 years ago, whereas the neonate samples were highly similar with a
12 genomic divergence of 3.6 months and not related to the staff member, indicating that
13 transmission did not occur from staff to patient or between patient to patient, but were
14 acquired from a common hospital source. The genomes revealed that the isolates
15 contained the ubiquitous *ampH* gene responsible for resistance to penicillin G, cefoxitin
16 and cephalosporin C, and all Kp-IIB strains were competent for host cell adhesion. Our
17 results highlight the clinical significance and genomic properties of relatively mild, but
18 persistent MLST types such as ST2727, and urges for genomic surveillance and
19 eradication within hospital environments.

20 INTRODUCTION

21 Species of *Klebsiella* are widespread in the environment, found in soil and
22 ground-water, as commensal organisms on plants and are carried widely in animal hosts
23 (1). *Klebsiella pneumoniae*, is recognized as a leading cause of healthcare-associated
24 urinary tract infections, surgical site infections and pneumonia, increasingly threatening
25 neonates and the immunocompromised (2). However, in recent times, an increasing
26 diversity of *Klebsiella* species have been identified to cause infection in humans (3, 4).
27 Genome sequencing has revealed that bacteria historically classified as *K. pneumoniae*
28 belonged to the *K. pneumoniae* species complex containing multiple phylogenetically
29 distinct but closely related bacterial species. These include *K. pneumoniae* (previously
30 classified as phylogroup KpI), *K. quasipneumoniae* (subsp. *quasipneumoniae* (KpIIA)
31 and subsp. *similipneumoniae* (KpIIB)), *K. variicola* (KpIII) (5), and two strains forming
32 divergent lineages indicating potentially novel species within this species complex (6).
33 Whole genome sequencing (WGS) has shown that the predominant cause of clinical
34 infection worldwide has been due to KpI (5), although there is an increasing trend, and
35 increasing concern, in the detection of severe cases due to the newly described species
36 (7). Different studies report *K. quasipneumoniae* isolates causing infections resistant to
37 treatments with ceftazidime and other oxyimino-betalactam antibiotics, and presenting
38 carbapenemase (KPC) genes (3, 4, 8-11).

39 Due to the emergence of a greater diversity of *Klebsiella* causing severe
40 infections, their ability of colonizing environmental niches such as sinks and ventilators
41 in hospitals (12), and increased detection of hybrid strains with drug resistant and
42 hypervirulent phenotypes, especially in regions such as China where the earliest reports
43 of hvKp were observed about three decades ago (13), an early warning system for the
44 detection of *Klebsiella* species is promoted in hospitals in China (14). Routine
45 surveillance is conducted using Multilocus Sequence Typing (MLST), followed by
46 antimicrobial and virulence testing providing an avenue to investigate potential outbreaks
47 and develop better control health measures.

48 Here we report a retrospective study of an incident that occurred in March 2017,
49 in the NICU of a major hospital with >3000 beds in China. A neonate patient suffered
50 severe diarrhea and subsequently developed septicemia and died, with *Klebsiella* isolated
51 as the causative microorganism from a blood sample. During the same period, severe
52 symptoms of diarrhea were found in three other neonate patients as well as an attending
53 NICU staff member, triggering a response to quell the apparent outbreak. In addition to
54 intensified hygiene procedures, the response included MLST and microbiological AMR
55 screening of stool samples from patients (21 neonates) and the staff member. Preliminary
56 microbiological characterization of the samples revealed that 20 of the isolates taken
57 from the infants belonged to a new MLST designation ST2727 and one infant was
58 infected with ST477. The sample from the staff member was of ST23 type, a well-
59 defined hypervirulent clonal group frequently detected in severe cases of infection,
60 whereas little was known of ST2727 and ST477. To forensically dissect the details of this
61 case, we generated whole genome data for all bacterial isolates and carried out a detailed
62 genomic characterization of virulence and antimicrobial resistance. Using a globally
63 sampled dataset comprising 3611 genomes we elucidate their genetic diversity, evolution
64 and epidemiology illuminating elements of nosocomial *Klebsiella* transmission and
65 ecological colonization that impact generally on understanding hospital-acquired
66 infections.

67 **METHODS**

68 **Collection of clinical isolates and clinical data**

69 *Klebsiella* was isolated from the blood sample taken prior to death of a neonate,
70 and stool samples were collected from all 21 infants in the NICU and a staff member that
71 suffered an episode of enteritis. These isolates were characterized using a Vitek II system
72 (bioMérieux, Marcy l'Etoile, France). Relevant clinical characteristics of the patients and
73 staff were extracted from their medical records, including demographic characteristics,
74 underlying medical conditions, clinical manifestations, treatments and outcomes. This
75 retrospective study was designed in accordance with the Declaration of Helsinki, 2013
76 (15) and been approved by the First Affiliated Hospital of Wenzhou Medical University
77 (China).

78 **Molecular genotyping and sequence typing of isolates**

79 Bacterial isolates were characterized by pulsed-field gel electrophoresis (PFGE)
80 after digestion of genomic DNA samples with XbaI under the following conditions:
81 temperature 14°C, voltage 6 V/cm, pulse angle 120°, and pulse duration of 5-35 seconds
82 for 18h, using a well-characterized XbaI digest (Takara Bio, Inc. Japan) of *Salmonella*
83 *enterica* serotype H9812 as molecular marker. DNA digest patterns were analysed and
84 interpreted according to Tenover *et al.* (16). Multilocus sequence typing (MLST) was
85 carried out by PCR amplifying seven housekeeping genes: *gapA*, *infB*, *mdh*, *pgi*, *phoE*,
86 *rpoB* and *tonB* (17), and the data used to assign sequence types (STs) using the Institute
87 Pasteur MLST database (<https://bigsd.b.pasteur.fr/>). Through this process, the new
88 ST2727 was established. Additionally, a standard multiplex-PCR was used to confirm
89 that the isolates did not contain any of the five virulent extracellular polysaccharide
90 capsules (CPS) (K antigen status; K-type) K1, K2, K20, K54 and K57.

91 **Bacterial genome sequencing and assembly**

92 DNA was extracted using a Biospin Bacterial Genomic DNA extraction kit
93 (Hangzhou Bioer Technology Co. Ltd.). Libraries were prepared using the TruePrep™
94 DNA Library Prep Kit V2 for Illumina (Vazyme). Briefly, the DNA sample was

95 fragmented and tagged with adapters using a single transposase enzymatic reaction,
96 followed by amplification using an optimized, limited-cycle PCR protocol and indexing.
97 Individual libraries were assessed on the QIAxcel Advanced Automatic nucleic acid
98 analyzer, and quantitated through qPCR using KAPA SYBR FAST qPCR Kit. Illumina
99 Next Generation Sequencing technology (HiSeq PE150) was used to generate a paired-
100 end library of short read sequences (150nt) for each isolate. Sequence quality controls
101 were performed using FastQC version 10-01-18 (Andrews, n.d.) and trimmed and refined
102 using Fastx toolkit version 0.0.13 (Lab, n.d.). *De novo* assembly was performed using
103 SPAdes version 3.13.1 (18). Scaffold generation and ordering were performed using
104 MeDuSa version that was available on July 2018 and Mauve version 2.4.0, respectively
105 (19, 20). Closely related genomes (NTUH-K2044 (Kp9), CP027602 (Kp10), HKUOPLC
106 (Kp1-8, Kp11-22) from NCBI were identified using genome identity. Upon assembly and
107 alignment of all ST2727 genomes, single nucleotide variations (SNV) were inferred using
108 Parsnp tool from Harvest Suite version 1.1.2 using a reference Kp1 isolate genome (21).

109 **Genomic analysis**

110 Genes contributing to drug-resistance and virulence were detected using Kaptive
111 version available on July 2018 and Kleborate version 0.3.0 (22, 23). VRprofile version
112 2.0 was used to detect genes clusters characteristic of protein secretion systems of type 3
113 (T3SS), type 4 (T4SS) or type 6 (T6SS) (24). The presence of genes encoding a type 2
114 secretion system (T2SS) was detected by BLAST version 2.7.1, using as queries the
115 known T2SS genes (25). Phylogenetic analysis was conducted using 3611 *Klebsiella*
116 genomes downloaded from the NCBI database (downloaded on 20/06/2018) to find the
117 closest related genomes to the novel strains (Data S1). Core genome alignment of
118 *Klebsiella* genomes (including ST2727 isolates) was generated using Roary version
119 3.11.2. Excluding poorly assembled genomes the final dataset constituted 3611 genomes
120 comprising 3251 genes that were in at least 99% of the sequences (26). Phylogenetic
121 relationships were estimated using the maximum likelihood (ML) method in RAxML
122 version 8.2.12 (27) using the General Time Reversible (GTR) nucleotide substitution
123 model with a gamma (Γ) distribution of among-site rate. Branch support was estimated

124 using a ML bootstrap analysis with replicates ranging from 10 to >1000 replicates for the
125 different datasets analysed.

126 Finally, to detect signatures of natural selection in the ST2727 outbreak samples,
127 we used McDonald-Kreitman Test (MKT) implemented in the Rpackage PopGenome
128 (28). MKT was calculated for the whole genome, and gene-wise. The MKT is the ratio
129 (neutrality index, $NI=(Pn/Ps)/(Dn/Ds)$) of the relation of nonsynonymous/synonymous
130 mutations between strains (Pn/Ps) and the relation of nonsynonymous/synonymous
131 mutations within strains (Dn/Ds). When this ratio is 1 it means that the same number of
132 mutations are happening between and within strains, which suggests that there is not
133 selective pressure or neutral selection occurring within strains. When the ratio is more
134 than 1 ($NI>1$) it means that $Pn/Ps>Dn/Ds$, which is characteristic of negative selective
135 pressure occurring within strains. The opposite, a ration less than 1 ($NI<1$), when
136 $Pn/Ps<Dn/Ds$, is characteristic of positive selection occurring within strains (29).

137 To confirm that the genome datasets contained sufficient genetic change between
138 sampling times, necessary for the reconstruction of the time-scale of evolution, we used a
139 root-to-tip regression of sampling years against genetic diversity in TempEst v1.5 (30),
140 optimizing the best fit for the root to maximize the determination coefficient, R^2 . The
141 slope of the regression was positive, showing that the genomic data reflects temporal
142 signal, so a molecular clock was estimated using a least-squares dating (LSD version 0.3)
143 method (31) with 1000 samples for the confidence interval. The reliability of the analysis
144 was confirmed using a permutation test in which the sampling years were randomized for
145 100 replication, with a Zscore test (32). The nucleotide substitution rate per year and the
146 branch lengths were used to determine the time of divergence of each node.

147 **RESULTS**

148 **Clinical scenario and diagnosis**

149 This study was initiated when outbreak status was called in a NICU due to a fatal
150 case of a neonate patient presenting with symptoms of respiratory and gastrointestinal
151 infections. Despite treatment with meropenem, vancomycin, cefoperazone and sulbactam,
152 the neonate patient died from neonatal septicemia, septic shock, gastrointestinal
153 hemorrhage and multiple organ dysfunction syndrome. Days after the fatal case, 20 more
154 neonates were sampled along with an NICU staff member who presented symptoms of
155 infection (Fig 1C). According to a Vitek II system (see Methods), blood and stool
156 samples collected from the 21 neonates co-housed in the single-room NICU and the staff
157 member were positive for *K. pneumoniae*. The neonates presented with a wide range of
158 symptoms: three with respiratory and digestive symptoms, thirteen with respiratory
159 disease such as dyspnea, severe asphyxia, tachypnea and/or cyanosis severe, one with
160 digestive symptoms alone, while three presented with normal health (Table 1). The staff
161 member presented with digestive symptoms including vomiting and severe diarrhea.
162 MLST type classification using the *gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB* showed that 20/21
163 neonate isolates were novel with a new MLST designation ST2727, whereas one neonate
164 was of ST477 and the staff member was infected with a well-characterised hypervirulent
165 type ST23, suggesting that the fatality was caused by the ST2727, and demonstrating that
166 the staff member was not the cause of the event. Here, we sought to understand the
167 carriage and genomic characteristics of ST2727 in the neonates and test the hypothesis of
168 transmission in this closed scenario.

169 **Genomics-based evaluation of the NICU strains**

170 To determine the genomic characteristics of the bacteria recovered in this study,
171 the complete genomes of all 20 ST2727 isolates were sequenced using a paired-end short
172 read sequencing methodology. The lengths of the assembled genomes ranged from
173 5,192,865 bp to 5,202,717 bp that encoded between 4706 and 5191 annotated genes
174 (Table S2). The GC content of all genomes was 57%. Genomic analysis showed the

175 isolates exhibited high genomic similarity (0.018% sequence divergence), with 1,066
176 single nucleotide variations (SNV) along their genomes corresponding to an average of
177 one SNV every 2,392 bases. The variation observed among the ST2727 samples were at
178 the lower range to the median divergence of 0–0.08% observed among the same MLST
179 types (5).

180 Phylogenies generated from a core genome of 3611 globally sampled *Klebsiella*
181 strains (Fig S1, [Data S1](#)), including isolates from the present study made evident that the
182 ST2727 strains belong to *K. quasipneumoniae* subsp. *similipneumoniae* (KpIIB) (33)
183 (Fig. 1A, Fig S2). The high similarity of the ST2727 strains suggests a clonal population.
184 A root-to-tip regression of sampling dates and genetic distance of 54 KpIIB genomes
185 sampled during 1994–2017, including the ST2727 genomes, showed a positive
186 correlation (correlation coefficient: 0.64) between time and mutation, indicating that the
187 dataset contained sufficient signal for the estimation of a molecular clock. We used the
188 least-square dating method, to estimate an evolutionary rate for the KpIIB clade of 4.596
189 $\times 10^{-05}$ nucleotide substitutions per site per year (nt subs/s/y) (95% CI of 4.180 to $4.980 \times$
190 10^{-05}). The ST2727 group exhibited a genomic divergence of >90 years (mean time of
191 most recent common ancestor (TMRCA) 1926; 95% confidence interval (95% CI), 1917
192 to 1933) to the most closely related isolate collected in 2016 from a patient from Thailand
193 (KPPSTH03; MLST type ST929-1LV) (Fig 1A).

194 The TMRCA of all KpIIB lineages was 1874 (95%CI, 1861 – 1886), whereas, the
195 mean TMRCA of ST2727 collected within the hospital was approximately 3 months prior
196 to the first case Kp1 (mean TMRCA 2016.94, 95%CI 2016.88 – 2016.97) (Fig 1B). This
197 showed a divergence time much greater than the duration of sampling across 22 days in
198 February–March 2017 (Fig 1C). These results suggest that what was designated as an
199 outbreak was instead the colonization of infants by a community of bacteria that were
200 transmitted to the neonates during this period. Due to the variation present between
201 ST2727 isolates, it is unlikely that the colonization occurred through direct transmission
202 between one neonate to another. We suggest instead that the isolates were introduced to
203 the hospital or the ward in the months prior to the outbreak and sustained within the

204 hospital during this period. The precise mechanism of colonization of the neonates is not
 205 clear without comprehensive sampling of the built environment and the equipment used
 206 in treatment and care of the neonates.

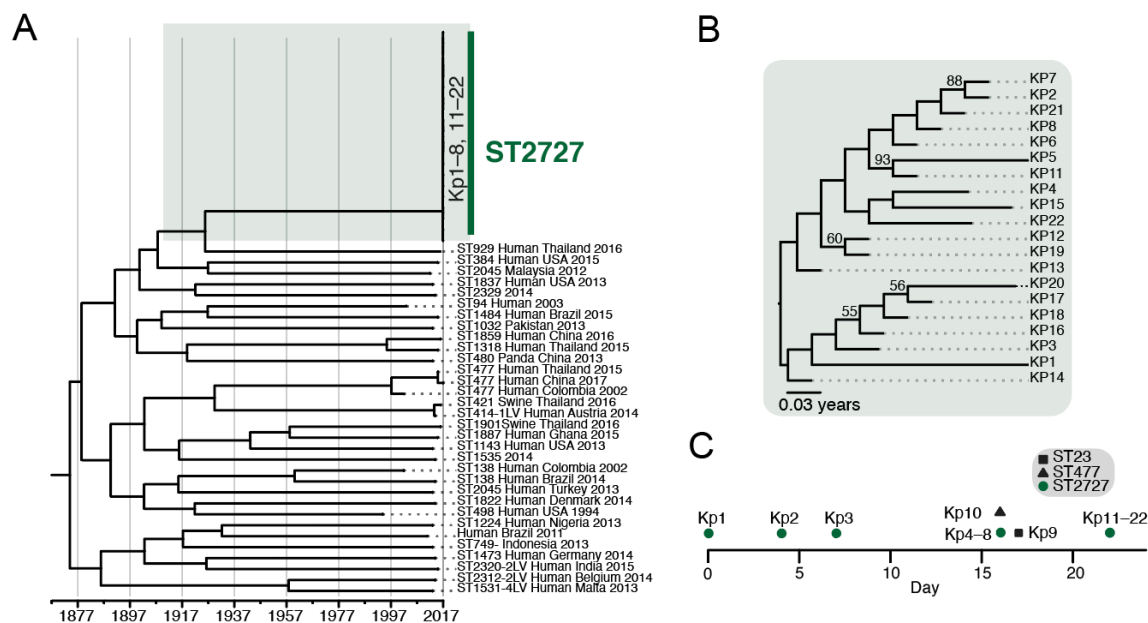


Figure 1. Genomic epidemiology and timeline of *K. quasipneumoniae* subsp.

***similipneumoniae* (KpIIB) sample collection in NICU.** (A) Evolutionary relationships and time-scale of evolution of 54 *K. quasipneumoniae* subsp. *similipneumoniae* (KpIIB) strains, and (B) ST2727 isolates detected in NICU during February and March 2017, generated using the least-square dating method. (C) Timeline of sample collection in NICU, where the calendar dates have been replaced by a timeline where the first sample was collected during day 0.

207 The core genome of ST2727 clade exhibited positive selection in comparison with
 208 strains in the KpIIB clade (MKT, neutrality index: 0.35, $NI < 1$), indicating that ST2727
 209 isolates are under selection pressure. Gene-wise estimates of selection pressure in the
 210 ST2727 clade in comparison with KpIIB strains showed that 21 genes were under
 211 negative selection ($NI > 1$) (Table S4), 10 were under positive selection (Table S5),
 212 whereas, mutations in the remaining genes were either strongly deleterious or were under
 213 neutral selection pressure (29). Many of these genes are fundamental to bacterial cell

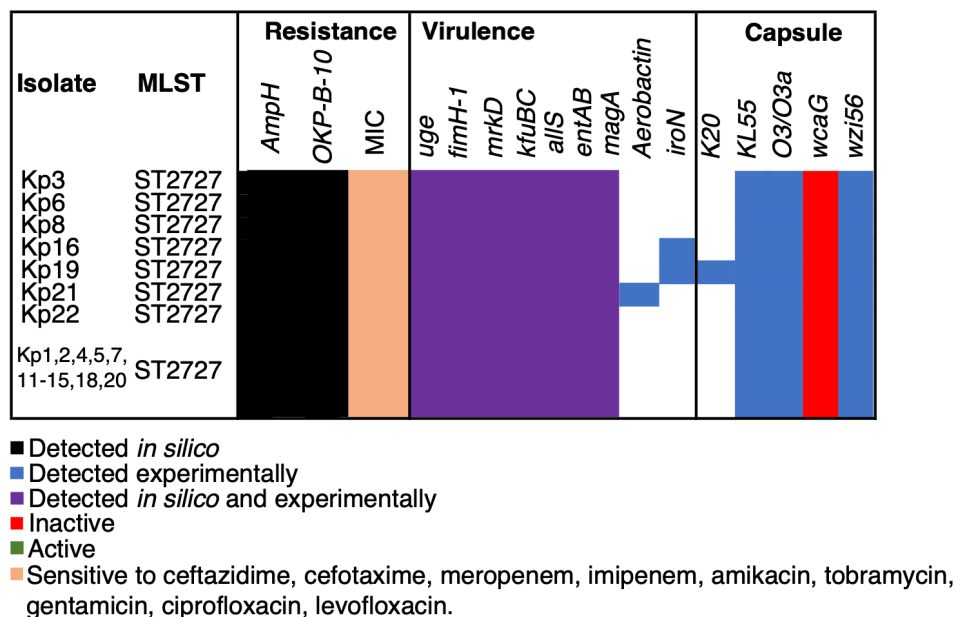
214 biology and have been functionally characterized in various *K. pneumoniae* strains, and
215 these functions can be expected to be conserved in *K. quasipneumoniae* subsp.
216 *similipneumoniae* (KpIIB). Genes under purifying or negative selection included genes
217 coding for transcriptional factors, metabolic enzymes, a structural protein, and proteins
218 coding for active transport of solutes (Table S4). Transporter proteins found to be under
219 purifying selection conform to a set of protein families: the ATP-Binding Cassette (ABC)
220 transporters, Major Facilitator Superfamily (MFS) transporters, and antiporters and
221 symporters. ABC transporter under negative selection were those designated as
222 responsible for the transport of thiamine, nickel and carbohydrates, all of which are
223 required for bacterial survival in nutrient rich environments (34). MFS transporters
224 uptake nutrients from rich environments using chemiosmotic ion gradients (34), while
225 antiporters and symporters maintain ion homeostasis for these and other functions (34).
226 Taken together, this suggests that the ST2727 population has adapted for growth in a
227 nutrient poor environment. Conversely, while the genes under positive selection also
228 include some for active transportation of solutes (Table S5), these transporters are
229 responsible for export of toxins and antibiotics as efflux systems. For example, genes
230 coding for efflux RND transporter permease subunits (drug efflux) and HlyD toxin
231 secretion (34). This reliance on drug-efflux is consistent too with the observation that the
232 structural protein peptidoglycan glycosyltransferase/peptidoglycan DD-transpeptidase
233 (PBP1A) encoded by the gene *mrcA*, was under negative selection: PBP1A catalyses the
234 transglycosylation and transpeptidation of murein that is a major component of the cell
235 wall, but it is also the target of beta-lactam antibiotics (34).

236 **O- and K- antigen serogroups of NICU isolates**

237 WGS analysis using Kaptive from Kleborate (22, 35) identified that ST2727
238 strains have a KL55 capsular type (*wzi56*). O-typing of the surface lipopolysaccharide
239 revealed all ST2727 were O3/O3a. This serotype has a strong adjuvant effect compared
240 with other lipopolysaccharide types (36), but is a rarely reported serotype with no known

241 clinical impact (37). Genome sequence data revealed that all isolates were *wcaG*- (Fig
 242 2A).

A



B

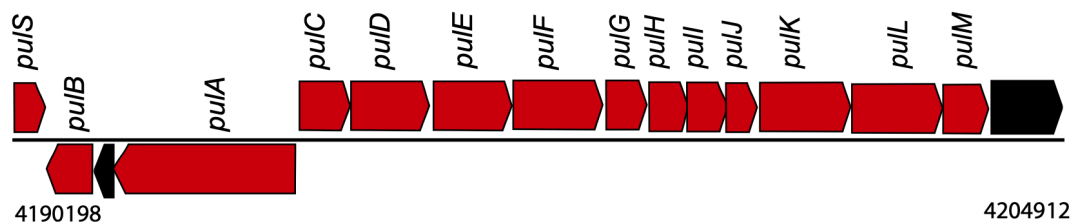


Figure 2. Major resistance, virulence and capsular genes detected in ST2727. (A)

Summary of genes contributing to drug-resistance, iron-acquisition and other virulence-enhancing features. (B) Genomic architecture of the T2SS cassette in ST2727.

243 Genes associated with virulence and antimicrobial resistance

244 All ST2727 isolates shared three genes considered diagnostic as virulence factors
 245 (Fig 2A, Table S3): *uge* (uridine diphosphate galacturonate 4-epimerase), *fimH* (type 1
 246 fimbrial adhesion) and *mrkD* (type 3 fimbrial adhesion). *Uge* is responsible for the
 247 conversion of UDP-GlcA to UDP-GalA and is key for the production of polysaccharides
 248 containing GalA residues, a feature which has been shown to promote colonization of

249 human tissues (38). Fimbriae are an important feature of *Klebsiella*, and the gene *fimH*
250 encodes a well-conserved adhesin subunit that is found in around 90% of *Klebsiella*
251 strains (39). This gene has high mobility within clones through horizontal transfer (39).
252 The gene encoding *mrkD* is associated with “type III” fimbriae, important for adhesion to
253 promote establishment of bacterial biofilms in harsh environments (40).

254 In terms of antimicrobial resistance, two genes encoding beta-lactamases were detected
255 (Fig 2A, Table S3): the ubiquitous *ampH*, responsible for resistance to penicillin G,
256 cefoxitin and cephalosporin C, and the Kp-IIB restricted *bla*OKP-B-10 (41).

257 **Secretion Systems**

258 Protein secretion systems are key components of virulence in many bacterial
259 pathogens (42), and in *Klebsiella* the type 2 secretion system (T2SS) is the best
260 characterized of these virulence determinants (43). *In silico* inspection of the genomes
261 showed that all ST2727 isolates collected have a similar T2SS architecture, typically
262 composed of *pulS* and *pulA-pulM* (44) (Fig 2B). HlyD, is a diagnostic component of the
263 type 1 secretion system (T1SS), and was detected in the ST2727 strains (Table S5). Other
264 common secretion systems (T3SS, T4SS, T5SS and T6SS) were not found in any of
265 ST2727 isolates.

266 **Evolutionary relationship of ST2727 to globally isolated *Klebsiella* strains**

267 *Klebsiella quasipneumoniae* was previously considered of low clinical risk (3-5).
268 Mapping of virulence and resistance phenotypes observed among *K. quasipneumoniae*
269 subsp. *similipneumoniae* (KpIIB) and globally collected *Klebsiella* genomes showed that
270 two KpIIB MLST types, ST1901 (GCF_002248055.1) and ST2320 (GCF_001729665.1)
271 and a KPIII (*K. variicola*) type (ST981) (GCF_001463685.1) (45) contained virulence
272 markers, that were previously only observed among *K. pneumoniae* strains (22, 23). The
273 ST1910 genome collected in Thailand in 2016 contained the virulence marker aerobactin
274 (*iuc3*), whereas the ST2320 strain collected in India during 2015 can express aerobactin
275 (*iuc1*), yerseniabactin (*ybt 9*; *ICEKp3*) and salmochelin (*iro1*), as well as producing a
276 hypermucoidy phenotype (*rmpA_4*, *rmpA2_3*). The emergence of virulence markers
277 among KPII genomes raises concerns as there has been a rise in documentation of

278 hospital-acquired infections caused by *K. quasipneumoniae* (5, 7). Also, Figure 3 shows
 279 that ST2727 is closely related to ST498, a group that has been found to have resistance
 280 genes including *aadB*, *sull*, *blaOKP-B-7*, and the genes *blaOXA-2* and *blaSHV-18* that
 281 encode ESBL (8) (Fig. 3).

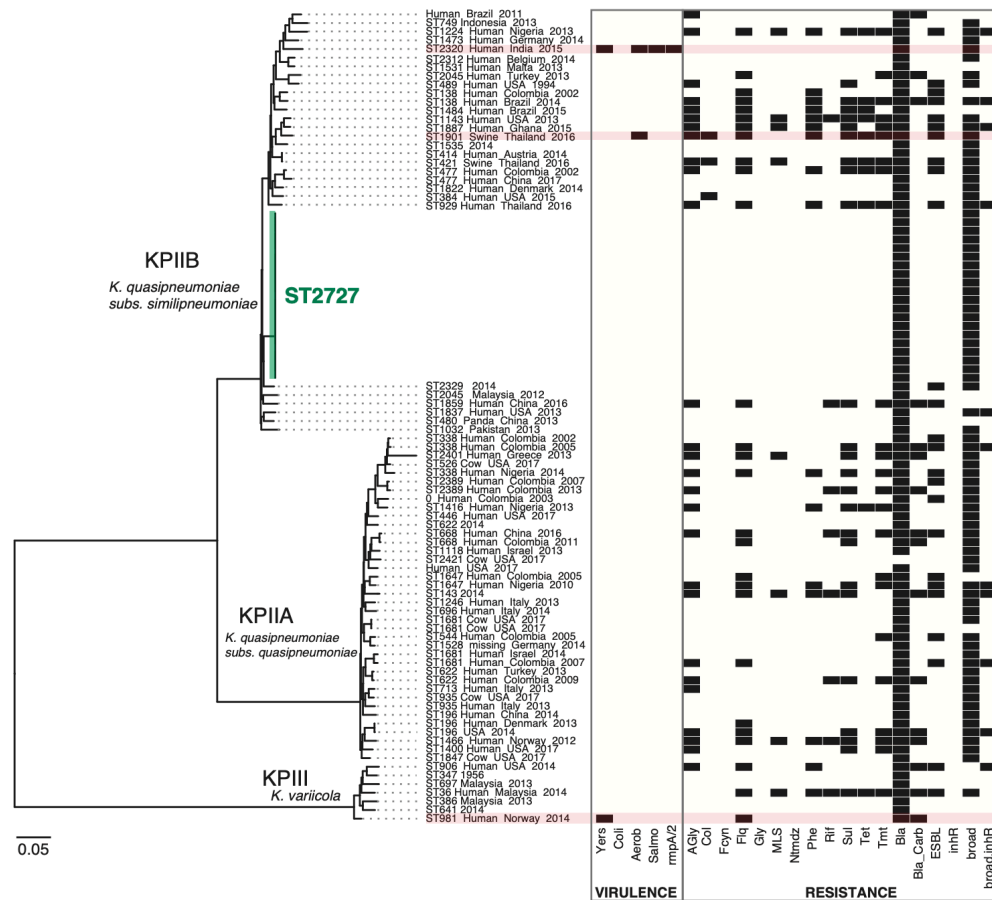


Figure 3. Phylogenetic distribution of virulence and resistance genes among KPII and KPIII genomes. The presence of virulent and resistance genes is shown in black. Virulent genes are grouped in categories: Yers (yersiniabactin), Coli (colibactin), Aerob (aerobactin), Salmo (Salmochelins) and *rmpA*, *rmpA_2* (hypermucoic). Resistant genes are grouped in categories: Gly (glycopeptides), MLS (macrolides), Phe (phenicols), Rif (rifampin), Sul (sulphonamides), Tet (tetracyclines), Tmt (trimethoprim), Bla (beta-lactamases), Bla_Carb (carbapenemase), ESBL (extended spectrum beta-lactamases), inhR (extended spectrum beta-lactamases with resistance to beta-lactamase inhibitors), broad (broad spectrum beta-lactamases), broad.inhR (broad spectrum beta-lactamases with resistance to beta-lactamase inhibitors). Virulent MLST types are highlighted in red.

282 DISCUSSION

283 The colonization of the neonate gut is a complex process with contributions from the
284 mother's microbiome and environmental sources, that can not easily be disentangled (46-
285 48). In this study, 21 neonates housed in a single room were sampled across 22 days for
286 the carriage of *Klebsiella* in response to an outbreak-status event. It is clear from
287 genomics data the hyper-virulent ST23 isolated from the staff member was an irrelevant
288 coincidence to the sepsis event in the neonate patients. The only *Klebsiella* isolated from
289 the sick neonates carried a previously unreported sequence type (ST2727), with core-
290 genome phylogenies showing they belonged to *K. quasipneumoniae* subsp.
291 *similipneumoniae* (33). Despite the availability of few KPIIB genomes, our analysis of 54
292 available genomes including the 20 new ST2727 samples, showed positive correlation
293 between genetic distances and date of sampling (30). This indicates that our dataset was
294 adequate to show that the NICU samples had diverged months prior to the detection of
295 the first case, and showing that ST2727 was not transmitted between patients during the
296 outbreak. These results were further supported by a permutation test where the sampling
297 dates were randomized (32, 49). Furthermore, the lower sampling of KPIIB genomes, as
298 compared to the greatly sampled *K pneumoniae* genomes, would not have an effect on the
299 estimation of the molecular clock for the outbreak samples because bacterial phylogenies
300 in general show a large genetic distance between MLST types – signified by the long
301 phylogenetic branch lengths between MLST types (star-shaped species level phylogeny),
302 indicating that adequate sampling of closely related bacteria would have an improvement
303 in estimates rather than greater sampling of highly divergent genomes within a bacterial
304 species. Overall, this study suggests that a population of ST2727 with some sequence
305 variability is likely endemic within the hospital environment, and that distinct members
306 within this population have infected the neonates in their first days of life.

307 It is probable that ST2727 is under antibiotic selective pressure due to their
308 prevalence in the hospital environment for the months prior to the events analysed here.
309 This is further confirmed with a MKT indicating positive selective pressure ongoing for
310 genes related with efflux systems that are a key part for the transport of toxic substances

311 as antibiotics. However, further investigation on evolutionary mechanism selecting for
312 resistant genes and genes related with resistance mechanisms would need to be done in
313 order to assert these findings. Previously, estimates of selection pressure on the genes
314 encoding beta-lactamases in the KpI, KpII and KpIII clades have showed neutral
315 selection of these genes (45). In agreement with this, we did not find either of the genes
316 encoding beta-lactamases in ST2727 to be under natural selective pressure.

317 Although made highly unlikely, the data in this study cannot rule out a maternal
318 contribution to the *Klebsiella* carriage, but we note that (i) 12/20 neonates were delivered
319 via Caesarian section (Table 1) where maternal contribution to microflora should be
320 inconsistent and minimal, and that (ii) ST2727 is probably not a dominant sequence type
321 in the community as it has not been sampled previously, despite the hospital having more
322 than 1500 *Klebsiella* isolates collected from adult patients over 15 years. We suggest that
323 a substantial contribution to carriage comes from the endemic microflora in the NICU,
324 with this specific built environment colonized by a population corresponding to ST2727.
325 A suggestion for future analyses would be inclusion of more detailed samplings of the
326 NICU environment, including the built environment itself and systematic sampling of
327 staff members and visitors.

328 Previous analysis estimated that at least half of *Klebsiella* infections result from
329 patients' own microbiota (50). That the neonate who died likely had a bowel perforation
330 to initiate sepsis from a strain carried in his gut supports the contention that *Klebsiella*
331 *quasipneumoniae* carriage contributes to the risk of infection in hospital environments.
332 Thus, carriage of drug-resistant or hypervirulent clones of *Klebsiella*, including *K.*
333 *quasipneumoniae* in the gut would be of great concern in people with lowered immune
334 systems – such as neonates - or those with conditions requiring surgical treatment where
335 bowel perforation is a risk factor (46). Our study highlights the importance of WGS for
336 elucidating the transmission and carriage of hospital-resident bacterial populations, urges
337 for the early identification and eradication of persistent strains that increase the rate of
338 hospital-acquired infections.

FUNDING

This work was supported by research grants from the National Natural Science Foundation of China (no. 81171614), the Health Department of Zhejiang Province of the People's Republic of China (no. 2011KYA106), the Zhejiang Provincial Program for the Cultivation of High-level Innovative Health Talents (no. 2012, 241) and Program Grant 1092262 from the National Health and Medical Research Council, Australia.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest.

ETHICAL APPROVAL

This study was approved by the Ethics Committee of the First Affiliated Hospital of Wenzhou Medical University (China).

ACKNOWLEDGEMENTS

We thank Dr. Jiangning Song for his input in the early stages of this analysis.

REFERENCES

1. Wyres KL, Holt KE. *Klebsiella pneumoniae* as a key trafficker of drug resistance genes from environmental to clinically important bacteria. *Curr Opin Microbiol.* 2018;45:131-9.
2. Zhan L, Wang S, Guo Y, Jin Y, Duan J, Hao Z, et al. Outbreak by Hypermucoviscous *Klebsiella pneumoniae* ST11 Isolates with Carbapenem Resistance in a Tertiary Hospital in China. *Frontiers in Cellular and Infection Microbiology.* 2017;7.
3. Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, et al. Whole-Genome Sequencing of Human Clinical *Klebsiella pneumoniae* Isolates Reveals Misidentification and Misunderstandings of *Klebsiella pneumoniae*, *Klebsiella variicola*, and *Klebsiella quasipneumoniae*. *mSphere.* 2017;2(4).
4. Long SW, Olsen RJ, Eagar TN, Beres SB, Zhao P, Davis JJ, et al. Population Genomic Analysis of 1,777 Extended-Spectrum Beta-Lactamase-Producing *Klebsiella pneumoniae* Isolates, Houston, Texas: Unexpected Abundance of Clonal Group 307. *MBio.* 2017;8(3).
5. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, et al. 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.* 2015;112(27):E3574-81.
6. Blin C, Passet V, Touchon M, Rocha EPC, Brisse S. Metabolic diversity of the emerging pathogenic lineages of *Klebsiella pneumoniae*. *Environ Microbiol.* 2017;19(5):1881-98.
7. Ejaz H, Wang N, Wilksch JJ, Page AJ, Cao H, Gujran S, et al. Phylogenetic analysis of *Klebsiella pneumoniae* from hospitalized children, Pakistan. *Emerging Infectious Diseases.* 2017;23:1872-5.
8. Elliott AG, Ganesamoorthy D, Coin L, Cooper MA, Cao MD. Complete Genome Sequence of *Klebsiella quasipneumoniae* subsp. *similipneumoniae* Strain ATCC 700603. *Genome Announc.* 2016;4(3).
9. Maatallah M, Vading M, Kabir MH, Bakhrouf A, Kalin M, Naucier P, et al. *Klebsiella variicola* is a frequent cause of bloodstream infection in the stockholm area, and associated with higher mortality compared to *K. pneumoniae*. *PLoS One.* 2014;9(11):e113539.
10. Marisa NF, Ramos PIP, Marques de Carvalho F, Camargo DRA, de Fatima Morais Alves C, Loss de Morais G, et al. Comparative Genomic Analysis of a Clinical Isolate of *Klebsiella quasipneumoniae* subsp. *similipneumoniae*, a KPC-2 and OKP-B-6 Beta-Lactamases Producer Harboring Two Drug-Resistance Plasmids from Southeast Brazil. *Front Microbiol.* 2018;9:220.
11. Rodrigues C, Passet V, Rakotondrasoa A, Brisse S. Identification of *Klebsiella pneumoniae*, *Klebsiella quasipneumoniae*, *Klebsiella variicola* and Related Phylogroups by MALDI-TOF Mass Spectrometry. *Front Microbiol.* 2018;9:3000.
12. Lowe C, Willey B, O'Shaughnessy A, Lee W, Lum M, Pike K, et al. Outbreak of extended-spectrum beta-lactamase-producing *Klebsiella oxytoca* infections associated with contaminated handwashing sinks(1). *Emerg Infect Dis.* 2012;18(8):1242-7.
13. Chang FY, Chou MY. Comparison of pyogenic liver abscesses caused by *Klebsiella pneumoniae* and non-*K. pneumoniae* pathogens. *J Formos Med Assoc.* 1995;94(5):232-7.
14. Gu D, Dong N, Zheng Z, Lin D, Huang M, Wang L, et al. A fatal outbreak of ST11 carbapenem-resistant hypervirulent *Klebsiella pneumoniae* in a Chinese hospital: a molecular epidemiological study. *Lancet Infect Dis.* 2018;18(1):37-46.
15. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-4.
16. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol.* 1995;33(9):2233-9.
17. Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S. Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. *J Clin Microbiol.* 2005;43(8):4178-82.
18. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-77.

19. Bosi E, Donati B, Galardini M, Brunetti S, Sagot MF, Lio P, et al. MeDuSa: a multi-draft based scaffold. *Bioinformatics*. 2015;31(15):2443-51.
20. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. 2010;5(6):e11147.
21. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biology*. 2014;15(11):524.
22. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, et al. Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom*. 2016;2(12):e000102.
23. Lam MM, Wyres KL, Duchene S, Wick RR, Judd LM, Gan Y-H, et al. Population genomics of hypervirulent *Klebsiella pneumoniae* clonal-group 23 reveals early emergence and rapid global dissemination. *Nature Communications*. 2018;9:2703.
24. Li J, Tai C, Deng Z, Zhong W, He Y, Ou HY. VRprofile: gene-cluster-detection-based profiling of virulence and antibiotic resistance traits encoded within genome sequences of pathogenic bacteria. *Brief Bioinform*. 2018;19(4):566-74.
25. Dunstan RA, Heinz E, Wijeyewickrema LC, Pike RN, Purcell AW, Evans TJ, et al. Assembly of the Type II Secretion System such as Found in *Vibrio cholerae* Depends on the Novel Pilotin AspS. *PLOS Pathogens*. 2013;9(1):e1003117.
26. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015;31(22):3691-3.
27. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312-3.
28. Pfeifer B. An introduction to the PopGenome package. *Cran-R*. 2015:1-18.
29. Stoletzki N, Eyre-Walker A. Estimation of the neutrality index. *Mol Biol Evol*. 2011;28(1):63-70.
30. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol*. 2016;2(1):vew007.
31. To TH, Jung M, Lycett S, Gascuel O. Fast Dating Using Least-Squares Criteria and Algorithms. *Syst Biol*. 2016;65(1):82-97.
32. Duchene S, Duchene DA, Geoghegan JL, Dyson ZA, Hawkey J, Holt KE. Inferring demographic parameters in bacterial genomic data using Bayesian and hybrid phylogenetic methods. *BMC Evolutionary Biology*. 2018;18:1-11.
33. Brisse S, Passet V, Grimont PAD. Description of *Klebsiella quasipneumoniae* sp. nov., Isolated from human infections, With two subspecies, *Klebsiella quasipneumoniae* subsp. *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae* subsp. nov., And demonstration th. *International Journal of Systematic and Evolutionary Microbiology*. 2014;64:3146-52.
34. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 2019;47(D1):D351-D60.
35. Wick RR, Heinz E, Holt KE, Wyres KL. Kaptive Web: User-Friendly Capsule and Lipopolysaccharide Serotype Prediction for *Klebsiella* Genomes. *J Clin Microbiol*. 2018;56(6).
36. Kato N, Kido N, Ohta M, Naito S, Nakashima I. Adjuvant Activity of *Klebsiella* O3 Lipopolysaccharide: Comparative Study Using Defined Uniform Salt Forms. *Microbiology and Immunology*. 1984;28(6):659-66.
37. Rollenske T, Szijarto V, Lukasiewicz J, Guachalla LM, Stojkovic K, Hartl K, et al. Cross-specificity of protective human antibodies against *Klebsiella pneumoniae* LPS O-antigen. *Nat Immunol*. 2018;19(6):617-24.
38. Regue M, Hita B, Pique N, Izquierdo L, Merino S, Fresno S, et al. A gene, *uge*, is essential for *Klebsiella pneumoniae* virulence. *Infect Immun*. 2004;72(1):54-61.
39. Stahlhut SG, Chattopadhyay S, Struve C, Weissman SJ, Aprikian P, Libby SJ, et al. Population variability of the FimH type 1 fimbrial adhesin in *klebsiella pneumoniae*. *Journal of Bacteriology*. 2009;191:1941-50.

40. Wilksch JJ, Yang J, Clements A, Gabbe JL, Short KR, Cao H, et al. MrkH, a novel c-di-GMP-dependent transcriptional activator, controls *Klebsiella pneumoniae* biofilm formation by regulating type 3 fimbriae expression. *PLoS Pathog.* 2011;7(8):e1002204.
41. Henderson TA, Young KD, Denome SA, Elf PK. AmpC and AmpH, proteins related to the class C beta-lactamases, bind penicillin and contribute to the normal morphology of *Escherichia coli*. *J Bacteriol.* 1997;179(19):6112-21.
42. Gunasinghe SD, Webb CT, Elgass KD, Hay ID, Lithgow T. Super-Resolution Imaging of Protein Secretion Systems and the Cell Surface of Gram-Negative Bacteria. *Front Cell Infect Microbiol.* 2017;7:220.
43. Cianciotto NP, White RC. The expanding role of type II secretion in bacterial pathogenesis and beyond. *Infection and Immunity.* 2017;85:1-17.
44. Nivaskumar M, Francetic O. Type II secretion system: A magic beanstalk or a protein escalator. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research.* 2014;1843(8):1568-77.
45. Samuelsen O, Overballe-Petersen S, Bjornholt JV, Brisse S, Doumith M, Woodford N, et al. Molecular and epidemiological characterization of carbapenemase-producing Enterobacteriaceae in Norway, 2007 to 2014. *PLoS One.* 2017;12(11):e0187832.
46. Raveh-Sadka T, Thomas BC, Singh A, Firek B, Brooks B, Castelle CJ, et al. Gut bacteria are rarely shared by co-hospitalized premature infants, regardless of necrotizing enterocolitis development. *Elife.* 2015;4.
47. Brooks B, Olm MR, Firek BA, Baker R, Thomas BC, Morowitz MJ, et al. Strain-resolved analysis of hospital rooms and infants reveals overlap between the human and room microbiome. *Nat Commun.* 2017;8(1):1814.
48. Yassour M, Jason E, Hogstrom LJ, Arthur TD, Tripathi S, Siljander H, et al. Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host & Microbe.* 2018;24(1):146-54.e4.
49. Mir D, Jung M, Delatorre E, Vidal N, Peeters M, Bello G. Phylodynamics of the major HIV-1 CRF02_AG African lineages and its global dissemination. *Infect Genet Evol.* 2016;46:190-9.
50. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, et al. Gastrointestinal Carriage Is a Major Reservoir of *Klebsiella pneumoniae* Infection in Intensive Care Patients. *Clin Infect Dis.* 2017;65(2):208-15.

Table 1. Patient and *Klebsiella pneumoniae* strains.

Sample id	MLST ¹	Patient	Birth	Symptoms ²	Antibiotics	Invasive procedure
KP1	ST2727	Neonate	vaginal	Digestive, Respiratory	yes	Yes
KP2	ST2727	Neonate	cesarean	Digestive, Respiratory	yes	Yes
KP3	ST2727	Neonate	cesarean	Digestive, Respiratory	yes	Yes
KP4	ST2727	Neonate	vaginal	Respiratory	no	Yes
KP5	ST2727	Neonate	vaginal	Respiratory	no	Yes
KP6	ST2727	Neonate	cesarean	Respiratory	yes	Yes
KP7	ST2727	Neonate	cesarean	Respiratory	yes	Yes
KP8	ST2727	Neonate	cesarean	Respiratory	no	No
KP9	ST23	Staff	N/A	Digestive	-	-
KP10	ST477	Neonate	vaginal	Digestive	yes	No
KP11	ST2727	Neonate	cesarean	Respiratory	yes	Yes
KP12	ST2727	Neonate	vaginal	Respiratory	yes	No
KP13	ST2727	Neonate	cesarean	Asymptomatic	no	no
KP14	ST2727	Neonate	cesarean	Respiratory	yes	Yes
KP15	ST2727	Neonate	cesarean	Respiratory	no	No
KP16	ST2727	Neonate	vaginal	Respiratory	no	Yes
KP17	ST2727	Neonate	cesarean	Respiratory	yes	Yes
KP18	ST2727	Neonate	vaginal	Noninfection	no	Yes
KP19	ST2727	Neonate	vaginal	Asymptomatic	no	No
KP20	ST2727	Neonate	vaginal	Respiratory	yes	Yes
KP21	ST2727	Neonate	cesarean	Respiratory	yes	Yes
KP22	ST2727	Neonate	cesarean	Asymptomatic	no	No

¹Multilocus Sequence Typing. ² Symptoms are classified depending of what type of infection they are related to, “noninfections” suggests the patient has symptoms related to other diseases that are non-infectious disease for example symptoms related with cardio-respiratory chronic diseases, “Asymptomatic” refers to no symptoms at all. “Respiratory” and “Digestive” refers to symptoms related to respiratory infections and digestive infections, respectively.

FIGURE LEGENDS

Figure 1. Genomic epidemiology and timeline of *K. quasipneumoniae* subsp.

***similipneumoniae* (KpIIB) sample collection in NICU.** (A) Evolutionary relationships and time-scale of evolution of 54 *K. quasipneumoniae* subsp. *similipneumoniae* (KpIIB) strains, and (B) ST2727 isolates detected in NICU during February and March 2017, generated using the least-square dating method. (C) Timeline of sample collection in NICU, where the calendar dates have been replaced by a timeline where the first sample was collected during day 0.

Figure 2. Major resistance, virulence and capsular genes detected in ST2727. (A)

Summary of genes contributing to drug-resistance, iron-acquisition and other virulence-enhancing features. (B) Genomic architecture of the T2SS cassette in ST2727.

Figure 3. Phylogenetic distribution of virulence and resistance genes among KPII and KPIII genomes.

The presence of virulent and resistance genes is shown in black. Virulent genes are grouped in categories: Yers (yersiniabactin), Coli (colibactin), Aerob (aerobactin), Salmo (Salmochelin) and *rmpA*, *rmpA_2* (hypermucoid). Resistant genes are grouped in categories: Gly (glycopeptides), MLS (macrolides), Phe (phenicols), Rif (rifampin), Sul (sulphonamides), Tet (tetracyclines), Tmt (trimethoprim), Bla (beta-lactamases), Bla_Carb (carbapenemase), ESBL (extended spectrum beta-lactamases), inhR (extended spectrum beta-lactamases with resistance to beta-lactamase inhibitors), broad (broad spectrum beta-lactamases), broad.inhR (broad spectrum beta-lactamases with resistance to beta-lactamase inhibitors). Virulent MLST types are highlighted in red.