The ZKIR assay, a novel real-time PCR method for the detection of Klebsiella pneumoniae and closely related species in environmental Samples

Elodie Barbier, Carla Rodrigues, Géraldine G. Depret, Virginie Passet,
Laurent Gal, Pascal Piveteau, Sylvain Brisse

HAL Id: hal-02464159

https://institut-agro-dijon.hal.science/hal-02464159

Submitted on 7 Feb 2020

1 **The ZKIR Assay, a novel Real-Time PCR Method for the Detection of**

2 ***Klebsiella pneumoniae* and Closely Related Species**

3 **in Environmental Samples**

4

5

6 Elodie Barbier[a]#, Carla Rodrigues[b], Geraldine Depret[a], Virginie Passet[b], Laurent Gal[a], Pascal

7 Piveteau[a] and Sylvain Brisse[b]#

8

9 [a] Agroécologie, AgroSup Dijon, CNRS, INRA, Univ. Bourgogne Franche-Comté, Dijon, France

10 [b] Institut Pasteur, Biodiversity and Epidemiology of Bacterial Pathogens, Paris, France

11

12 Running title: Detection of *Klebsiella pneumoniae* by real-time PCR

13

14 # Corresponding authors

15 Sylvain Brisse:

16 Institut Pasteur, Biodiversity and Epidemiology of Bacterial Pathogens, 28 rue du Docteur Roux,

17 F-75724 Paris, France; E-mail: sylvain.brisse@pasteur.fr; Phone +33 1 45 68 83 34

18 Elodie Barbier:

19 Agroécologie, AgroSup Dijon, CNRS, INRA, Univ. Bourgogne Franche-Comté, 17 rue Sully, F-

20 21000 Dijon, France; E-mail: elodie.barbier@inrae.fr; Phone +33 3 80 69 33 24

21

23

24

25 **ABSTRACT**

26 *Klebsiella pneumoniae* (Kp) is of growing public health concern due to the emergence of strains

27 that are multidrug-resistant, virulent, or both. Taxonomically, Kp includes seven phylogroups,

28 with Kp1 (*K. pneumoniae sensu stricto*) being medically prominent. Kp can be present in

29 environmental sources such as soils and vegetation, which could act as reservoirs of animal and

30 human infections. However, the current lack of screening methods to detect Kp in complex

31 matrices limits research on Kp ecology. Here we analysed 1,001 genome sequences and found

32 that existing molecular detection targets lack specificity for Kp. A novel real-time PCR method,

33 the ZKIR assay, was developed and used to detect Kp in 96 environmental samples. Results were

34 compared to a culture-based method using Simmons citrate agar with 1% inositol (SCAI)

35 medium coupled to MALDI-TOF mass spectrometry identification. Whole-genome sequencing

36 of environmental Kp was performed.

37 The ZKIR assay was positive for the 48 tested Kp reference strains, whereas 88 non-Kp strains

38 were negative. The limit of detection of Kp in spiked soil microcosms was 1.5×10^{-1} CFU g^{-1}

39 after enrichment for 24 h in LB supplemented with ampicillin, and 1.5×10^3 to 1.5×10^4 CFU g^{-1}

40 directly after soil DNA extraction. The ZKIR assay was more sensitive than the culture method.

41 Kp was detected in 43% of environmental samples. Genomic analysis of the isolates revealed a

42 predominance of phylogroups Kp1 (65%) and Kp3 (32%), a high genetic diversity (23 MLST

43 sequence types), a quasi-absence of antibiotic resistance or virulence genes, and a high frequency

44 (50%) of O-antigen type 3. This study shows that the ZKIR assay is an accurate, specific and

45 sensitive novel method to detect the presence of Kp in complex matrices, and indicates that Kp

46 isolates from environmental samples differ from clinical isolates.

47

48

49 **IMPORTANCE**

50 The *Klebsiella pneumoniae* species complex (Kp) includes human and animal pathogens, some of

51 which are emerging as hypervirulent and/or antibiotic resistant strains. These pathogens are

52 diverse and classified into seven phylogroups, which may differ in their reservoirs and

53 epidemiology. Proper management of this public health hazard requires a better understanding of

54 Kp ecology and routes of transmission to humans. So far, detection of these microorganisms in

55 complex matrices such as food or the environment has been difficult due to a lack of accurate and

56 sensitive methods. Here, we describe a novel method based on real-time PCR, which enables

57 detection of all Kp phylogroups with high sensitivity and specificity. We used this method to

58 detect Kp isolates from environmental samples, and show based on genomic sequencing that they

59 differ in antimicrobial resistance and virulence gene content, from human clinical Kp isolates.

60 The ZKIR PCR assay will enable rapid screening of multiple samples for Kp presence and will

61 thereby facilitate tracking the dispersal patterns of these pathogenic strains across environmental,

62 food, animal and human sources.

3

INTRODUCTION

Klebsiella pneumoniae (Kp) is one of the leading causes of multidrug resistant (MDR) healthcare-acquired infections, with increasing rates of resistance to carbapenems and other last resort antibiotics being reported (1, 2). Furthermore, Kp is also an important agent of severe community-acquired infections (so-called 'hypervirulent' strains) in healthy persons (3), with recent worrisome reports of convergence between hypervirulent and MDR phenotypes (1, 4). Kp is recognized as a colonizer of the throat and the intestinal tract of humans and animals (5–7).

The main sources of human exposure to Kp are not well defined. Previous studies highlighted the large distribution of Kp in outdoor environments, including water, sewage, soil and plants (8–13). Animal and human food, particularly retail meat or salad may also be contaminated (6, 14, 15). Many studies suggest that food, water and/or environmental exposure may be associated with virulent and/or antibiotic resistant Kp to humans (13, 14, 16, 17). However, little is known on the relative contribution of these different sources of transmission. Although such information is a prerequisite to control efficiently transmission routes and reduce exposure, the ecology of Kp is currently poorly understood.

The systematics of Kp has been refined through recent taxonomic updates, which highlighted the existence of seven phylogroups (Kp1 to Kp7), corresponding to distinct taxa, within *Klebsiella pneumoniae sensu lato*. The *K. pneumoniae* species complex includes 5 different species: *K. pneumoniae sensu stricto* (phylogroup Kp1), *K. quasipneumoniae* [subsp. *quasipneumoniae* (Kp2) and subsp. *similipneumoniae* (Kp4)], *K. variicola* [subsp. *variicola* (Kp3) and subsp. *tropica* (Kp5)], '*K. quasivariicola*' (Kp6) and *K. africana* (Kp7) (8, 18–20). Most of these taxa are still often misidentified as "*K. pneumoniae*" or *"K. variicola"* due to the unsuitability of traditional clinical microbiology methods to distinguish among members of the Kp complex (21). Henceforth, we will use the 'Kp' abbreviation to refer collectively to the seven

4

87 taxa of the *K. pneumoniae* species complex, and will reserve '*K. pneumoniae*' for *K. pneumoniae*

88 *sensu stricto* (*i.e.,* phylogroup Kp1). As all Kp are potentially pathogenic for humans and animals

89 and can share acquired resistance and virulence genes, it is important that the seven taxa be

90 considered together when investigating the routes of transmission and ecology of Kp.

91 Detection of Kp is not well integrated in food or environmental microbiological surveillance

92 programs and there is a general lack of tools and procedures for its detection and quantification.

93 Culture-based laboratory methods used for the detection of microorganisms in complex matrices

94 are time consuming and have a low throughput. Moreover, Kp culture methods have not been

95 validated so far for food safety screening. Some molecular methods (without need of sequencing)

96 have been proposed over the years for the rapid detection of Kp (22–26). They target the 16S-23S

97 rRNA internal transcribed spacer sequence (ITS) (22), coding sequences of *tyrB* (26), *khe* (25,

98 27), chromosomal beta-lactamase (*bla*) genes (28, 29) or other molecular targets (23). Some of

99 these targets are described as able to detect the Kp complex (22), whereas others were designed

100 for specific members of the Kp complex, such as Kp1 and Kp3 (23, 24).

101 Real-time PCR is a powerful approach for the rapid detection and quantification of micro-

102 organisms in complex matrices (30). This approach presents multiple advantages including easy

103 standardization and high throughput. The aims of this work were (i) To define the phylogenetic

104 distribution of previously proposed molecular targets for Kp detection, in light of recent

105 taxonomic updates; and (ii) To develop a real-time PCR method for the rapid, specific and

106 sensitive detection of all Kp members. (iii) In addition, we used the novel qPCR method to detect

107 Kp in environmental samples and explored the genomic features (including antibiotic resistance

108 and virulence genes) of the recovered Kp isolates.

5

MATERIALS AND METHODS

Bacterial strains and culture conditions

A panel of 136 strains from the collections of the Institut Pasteur, Paris (France) and INRA Dijon

(France) was used (Table 1). It included the *K. pneumoniae* type strain ATCC13883[T] and 47

other strains of the Kp complex (Kp1 to Kp7 phylogroups, including type strains) isolated from

patients, animals and outdoor environments from multiple geographic locations. Eighty-eight

non-Kp bacterial strains from other species of the genus *Klebsiella*, closely related genera

(*Raoultella, Enterobacter*), and from other species that have similar environmental niches were

included for comparison. All strains were regenerated by streaking on tryptone soya agar (TSA,

Conda, Spain). After 24h at 37°C, single colonies were transferred to 10 ml tryptone soya broth

(TSB, Conda, Spain). Bacterial suspensions were collected after 24 h incubation at 37°C for

further experiments.

Bacterial DNA extraction

The bacterial DNA was extracted according to a homemade protocol. One ml of the 24h bacterial

culture was centrifuged (7000 g, 10 min, 4°C), the pellet washed with sterile water and suspended

in 800 μl Tris (50 mM, pH 8) in which 10 μl EDTA (500 mM, pH 8), 115 μl lysozyme (10

mg/ml, incubation 37°C, 30 min), 57 μl SDS 20% (incubation 65°C, 15 min), 5 μl RNAse (10

mg/ml, incubation 37°C, 30 min) and 23 μl proteinase K (6 mg/ml, incubation 37°C, 1h30) were

added. Finally, a 1:1 volume of chloroform/ isoamyl alcohol (25/24/1) was added. The

suspension was shaken 3 min at 15 rpm, then centrifuged 3 min at 20 000 g. The supernatant was

transferred in a new tube and a 1:1 volume of chloroform/ isoamyl alcohol (24/1) was added.

After shaking (3 min, 25 rpm) and centrifugation (3 min, 20 000 g), this step was repeated once.

After transfer of the supernatant in a new tube, a 1:10 volume of NaCl (5 M) was added (shaking

6

133   1 min, 15 rpm). Finally, a 1:1 volume of isopropanol (shaking 3 min, 15 rpm) was added.

134   Precipitated DNA filaments were collected with a glass Pasteur pipette and transferred in 400 µl

135   frozen ethanol (70%). After shaking 3 min at 15 rpm, DNA filaments were collected, dried at

136   room temperature and finally dissolved in 100 µl TE (pH 8). DNA concentration was estimated

137   using NanoDrop 2000 Spectrophotometer (ThermoFischer Scientific, USA) and adjusted at 20

138   ng/µl. Purified DNA was stored at -20°C.

139   A simpler protocol was implemented for the PCR specificity assay and validation study. Bacteria

140   grown in TSB or enrichment medium were centrifuged 5 min at 13 000 g. Pellets were suspended

141   in 200 µl of Tris HCl (5 mM pH 8.2) supplemented with 13 µl of Proteinase K (1 mg/ml). After 2

142   h incubation at 55 °C, Proteinase K was inactivated during 10 min boiling. Cell debris were

143   discarded by centrifugation (5 min at 13 000 g) and the supernatant was stored at – 20°C before

144   use for real-time PCR.

145

146   **Soil DNA extraction**

147   Soil metagenomic DNA was extracted from 2 g aliquots of soil using the modified ISO standard

148   11063 method (31). Crude soil DNA extracts were first purified onto PVPP

149   (polyvinylpolypyrrolidone) minicolumns (BIORAD, France), then with the GeneClean Turbo kit

150   (MP Biomedicals, France) according to the manufacturer's protocol (31). Purified DNA

151   concentrations were determined using a Nanodrop 2000 spectrophotometer.

152

153   *In silico* **identification of the target DNA region**

154   The phylogenetic distribution of molecular targets that were previously described in the literature

155   for Kp detection, was investigated *in silico*. All available Kp genomes were downloaded from the

156   NCBI database (n=3552, February 2018) and combined with the ones of Institut Pasteur's

7

157   internal reference collection (n=670) (20, 32, 33), representing a total of 4,222 genomes which

158   included *Klebsiella* species and closely related *Raoultella* species. The average nucleotide

159   identity (ANI) metric was used to classify the Kp genomes in each of the phylogroups. To avoid

160   redundancy in the dataset and due to the difficulty of analyzing phylogenetically 4,222 genomes,

161   in the case of the genomes belonging to the Kp complex (which represented the majority of the

162   dataset), we selected unique representatives of each MLST sequence type (ST), as defined using

163   the *mlst* software tool (https://github.com/tseemann/mlst). This resulted in a final collection of

164   1,001 genomes that were tested for the presence of targets from the literature. Nucleotide BLAST

165   (BLASTN) was used to detect the presence of the target sequences in the 1,001 genomes, with

166   80% nucleotide identity (also with 92% in the case of bla_{SHV}, bla_{OKP} and bla_{LEN} targets) and 80%

167   length coverage as cutoffs. Primer-blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was

168   used (with default parameters) to check the distribution of primer target sites in the genomic

169   sequences. Specificity was defined as the proportion of target organisms among those expected to

170   be amplified by the target assay. Sensitivity was defined as the proportion of target organisms

171   expected to be amplified by the target assay. A phylogenetic tree was constructed (34) and the

172   output was visualized using iTOL (https://itol.embl.de/; Figure 1; Figure S1).

173

174   **Design of primers**

175   The Kp *tyrB* and *khe* gene sequences (accession numbers: AF074934.1 and AF293352.1) were

176   aligned against the genome sequences of *K. pneumoniae* (CP012744.1, CP012743.1,

177   CP028787.1), *K. variicola* (CP008700.1, CP013985.1, CP017289.1), *K. quasipneumoniae*

178   (CP014071.1, CP023478.1, CP029432.1) and *K. quasivariicola* (CP022823.1) using BLASTN to

179   identify conserved genomic regions within the Kp complex. Sequences of closely related

180   bacterial species *K. oxytoca* (CP027426.1), *Raoultella ornithinolytica* (CP010557.1),

8

181 *R. planticola* (CP026047.1) and *K. aerogenes* (CP014029.2) were added in the alignments to

182 confirm the specificity of these regions for the Kp complex.

183 Primer sets were designed using Primer3Plus (http://www.bioinformatics.nl/cgi-

184 bin/primer3plus/ primer3plus.cgi). The specificity of the predicted primer sets and amplicons was

185 checked by applying BLASTN on the GenBank nucleotide collection (nr/nt) from the NCBI

186 database and Multiple primer analyzer (https://www.thermofisher.com) was run to check for

187 dimer formation. The best candidate primer sets defined by this *in silico* approach were ordered

188 for synthesis at Eurogentec; among these was primer pair ZKIR_F (5'-CTA-AAA-CCG-CCA-

189 TGT-CCG-ATT-TAA- 3') and ZKIR_R (5'-TTC-CGA-AAA-TGA-GAC-ACT-TCA-GA– 3').

190

191 **Real-time PCR**

192 All real-time PCR assays were performed on an ABI StepOne™ Real-time thermocycler (Fisher

193 Scientific, France) with the following temperature program: 95 °C for 3 min and 40 cycles at

194 95 °C 10 s and 60 °C for 1 min. Melt curves were generated with temperature increments of

195 0.3 °C per cycle from 60 to 95 °C. DNA was amplified in a 20 µl PCR mix containing 10 µl of

196 Takyon Low Rox SYBR MasterMix dTTP Blue (Eurogentec, Belgium), 2 µl of each primer

197 (final concentration 300 nM), 2.5 µl of template DNA and 3.5 µl PCR grade water. For detection

198 of Kp from environmental samples, 0.5 µl of T4 Gene 32 Protein (Sigma Aldrich) was added to

199 the PCR mix. The specificity and cross-reactivity of the ZKIR assay was evaluated with 2 ng of

200 purified DNA of 48 Kp complex strains and 88 non-Kp isolates (Table 1).

201 Two universal primers targeting a 174 base pairs region of the *16S rRNA* gene of Eubacteria,

202 341f (5'-CCT-ACG-GGA-GGC-AGC-AG-3') and 515r (5'-ATT-CCG-CGG-CTG-GCA-3'),

203 were used for positive control PCR reactions as described in a previous report (35).

204

9

205 **Standard curve development and sensitivity assessment**

206 Aliquots of 2.5 µl adjusted to 7.5 ng, 750 pg, 75 pg, 7.5 pg, 750 fg, 375 fg, 45 fg and 15 fg of

207 genomic DNA of *K. pneumoniae* ATCC13883[T] were prepared in triplicate and amplified with the

208 optimized PCR conditions described above. Results were analysed with the StepOne™ Data

209 Analysis software. PCR efficiency was determined. The genome number expressed as logarithm

210 was plotted against Ct values and the correlation coefficient (R^2) of the standard curve was

211 calculated.

212

213 **Comparison of the ZKIR PCR and culture methods for detection of Kp in environmental**

214 **samples**

215 Ninety-six environmental samples collected between July and September 2018 (23 in July; 39 in

216 August and 34 in September) in Auxonne (Burgundy, France) were analysed for Kp presence

217 using the ZKIR assay and culture methods in parallel. Samples corresponded to bulk soils (n=32),

218 roots (31), leaves (29) and irrigation water (4) and were processed in the lab within 24 hours after

219 sampling. Ten g of soil were weighed in 180 ml pots (Dutscher, France). Plant leaves and roots

220 were properly cut, cleaned with sterile water and transferred in 180 ml pots. Processed samples

221 were suspended in 90 ml of lysogeny broth (LB, 5 g of NaCl, 5 g of yeast extract and 10 g of

222 Tryptone for a 1 l final volume) supplemented with ampicillin (10 mg/l, ampicillin sodium salt,

223 Sigma Aldrich). Five hundred millilitres of water samples were filtered through a 0.25 µm

224 membrane (Millipore, France). The membrane was incubated in 20 ml of LB supplemented with

225 ampicillin as described above. After 24 h of incubation at 30 °C, enrichments were vortexed and

226 500 µl aliquots were centrifuged (5 min at 5800 × g) and washed with sterile water. The pellet

227 was suspended in 500 µl of sterile water and boiled for 10 min. Boiled enrichments were ten-fold

228 diluted (1:10 and 1:100) and the dilutions were used as template for real-time PCR.

10

229   In parallel, enrichments were serially diluted (1:10 to 1:10 000) in sterile water before plating on

230   Simmons Citrate Agar enriched with 1% inositol (SCAI medium) (36). Plates were incubated 48

231   h at 37 °C. Each plate was screened for presumptive colonies of *Klebsiella* (large, yellow, dome-

232   shaped colonies) and ten candidate colonies were purified on SCAI medium and identified using

233   Matrix Assisted Laser Desorption Ionization Time of Flight (MALDI-TOF) Mass Spectrometry

234   (MALDI Biotyper, Bruker) according to the MALDI Biotyper Compass database version 4.1.80

235   (Bruker Daltonics, Germany). In addition, whole-genome sequencing (WGS) was performed for

236   environmental Kp isolates (n=31) using the NextSeq-500 sequencing platform (Nextera XT

237   library; 2×150 nt). Genomic assemblies were obtained using SPAdes v3.9. Multilocus sequence

238   typing (MLST) and core-genome MLST (cgMLST) were performed using the BIGSdb-Kp web

239   tool (https://bigsdb.pasteur.fr/klebsiella/klebsiella.html). This resource coupled with Kleborate

240   (https://github.com/katholt/Kleborate) was used to search for antibiotic resistance, virulence and

241   heavy metal tolerance genes, and to predict capsular types. PlasmidFinder was used to look for

242   plasmid replicons (https://cge.cbs.dtu.dk/services/PlasmidFinder/). To construct the tree based on

243   *nif* cluster genes (Figure S2), the DNA region between *nifQ* and *nifJ* genes was extracted from

244   the *nif* carrying strains and from a set of reference strains carrying this cluster (37). Sequences

245   were aligned with MUSCLE, and IQ-TREE (http://iqtree.cibiv.univie.ac.at) was used to

246   reconstruct the maximum likelihood phylogeny using the HKY+I+F+G4 model.

247

248   **Determination of the limit of detection of the ZKIR assay in artificially spiked soils**

249   Two soils with contrasted edaphic characteristics (Sandy soil A; clay soil V) and free of

250   indigenous Kp, according to the above procedure, were used. A bacterial suspension of

251   *K. pneumoniae* ATCC13883[T] was serially diluted in sterile water and dilutions were enumerated

252   on TSA plates. Five grams aliquots of each soil were spiked with these decreasing Kp dilutions,

11

253 resulting in final concentrations from 1.5×10^{4} cfu/g to 1.5×10^{-1} cfu/g. Soil samples were

254 prepared in triplicates with three independently grown inoculums at 4 dilutions (36 spiked

255 microcosms for each soil). Negative controls were prepared by adding the same volume of sterile

256 water (3 microcosms for each soil). All 78 spiked and control soil microcosms were enriched for

257 24 h at 30 °C in 45 ml of LB supplemented with ampicillin (10 mg/l, Sigma Aldrich). Five

258 hundred µl aliquots of enrichment broth were sampled before and after the 24-h enrichment step,

259 centrifuged 5 min at 5800 g and washed with sterile water. The pellet was suspended in 500 µl of

260 sterile water and boiled for 10 min. Boiled enrichments were serially ten-fold diluted (1:10 to

261 1:100 000) and the diluted suspensions were used for real-time PCR. Moreover, metagenomic

262 DNA was extracted from spiked soils as described above in 'Soil DNA extraction' paragraph.

263

264 **Data availability**

265 The detailed ZKIR qPCR operating procedure was made publicly accessible to the scientific

266 community through the protocols.io platform (https://dx.doi.org/10.17504/protocols.io.7n6hmhe).

267 Genomic sequences generated in this study were submitted to the European Nucleotide Archive

268 and are accessible under the BioProject number PRJEB34643.

269   **RESULTS**

270   **Revisiting the phylogenetic distribution of proposed molecular targets for Kp detection**

271   Fourteen molecular targets were found in the literature (Figure 1; Table 2) (22–28, 38). Four of

272   them were proposed to detect the Kp complex, while others were designed for specific members

273   of this complex (*K. pneumoniae*, *K. variicola* or *K. quasipneumoniae*; Table 2). Mapping of the

274   presence/absence of the sequence region expected to be amplified by the primers was performed

275   across *Klebsiella* phylogenetic diversity (Figure 1). Regarding the targets proposed for the entire

276   Kp complex, this *in-silico* analysis revealed that only *khe* and *tyrB* (25, 26) presented both a high

277   sensitivity and a high specificity. We further analysed *in silico* the primers target sites for these

278   two genes using primer-blast. The results showed that *khe* primers would be expected to amplify

279   the same region in *Raoultella* spp. (77 bp, with a sequence identity of the homologous region <

280   80%, hence not visible on Figure 1), as well as an additional region (348 bp) in these organisms.

281   Furthermore, *tyrB* primers also appear to be able to amplify the target region in *K. aerogenes* and

282   *Raoultella* spp., consistent with the distribution of the target region (Figure 1). Regarding the

283   targets that were proposed to be specific for Kp1, target KpI50233a (24) proved to be the most

284   specific and sensitive one (although with ability to also amplify *K. aerogenes* isolates), whereas

285   the ones proposed for Kp3 (23) appeared unspecific or to lack sensitivity in our *in silico* analysis

286   (**Figure 1**). Regarding the targeted chromosomal class A beta-lactamases genes bla_{SHV}, bla_{OKP}

287   and bla_{LEN} (28), there appeared to be a lack of specificity (Figure 1), which is explained by the

288   high degree of sequence identity between these *bla* genes, even though they represent distinct

289   targets. To clarify their distribution, we mapped the expected amplicon sequences of bla_{SHV},

290   bla_{OKP} and bla_{LEN} using a higher threshold (92%, defined based on observed sequence similarity

291   within and between the three gene families) (Figure S1). Using this threshold, we observed a high

292   degree of specificity of the three families for their respective phylogroup/species, as expected

13

293   (39). However, it was evident that horizontal gene transfer events implicating the beta-lactamase

294   genes have occurred. In particular, the bla_{SHV} gene was often observed in non-Kp1 phylogroups

295   (particularly Kp2 and Kp4) and even in species outside the Kp complex, such as *E. coli* or

296   *Salmonella* spp. This can be attributed to the presence of bla_{SHV} on plasmids and was previously

297   described (21). The gene bla_{LEN} was also observed in a few Kp1 genomes. These observations

298   clearly render the use of the chromosomal beta-lactamase gene targets unreliable as

299   phylogroup/species identification markers (Figure S1).

300

301   **ZKIR primers design, PCR assay development and optimization**

302   The *tyrB* and *khe* genes were previously proposed as targets for the specific detection of Kp (25,

303   40). As shown above, *tyrB* and *khe* targets were not totally specific for the Kp complex. We

304   therefore attempted to define primers within the coding sequence of these genes but external to

305   the previously proposed target fragments. However,   high identity was observed with non-Kp

306   complex species when blasting the entire coding sequence (data not shown). Consequently, Kp

307   complex-specific primers could not be designed within the coding sequence of these two genes.

308   Interestingly, investigation of sequences upstream of *khe* did reveal a sequence that was highly

309   conserved within the Kp complex. This region located in the intergenic region (IR) upstream of

310   *khe* is partly deleted in other species such as *K. oxytoca* and *Raoultella* spp. As this 249 bp non-

311   coding IR is located between *zur* (zinc uptake regulator) and *khe* (annotated as a putative

312   hemolysin) genes (Figure 2), it was named ZKIR for *zur-khe* intergenic region.

313   A pair of primers (ZKIR_F and ZKIR_R, Figure 2) targeting the ZKIR region was designed and

314   tested *in silico.* When implemented in the SYBR green PCR assay using reference strain ATCC

315   700603 (known as *K. pneumoniae*, but belonging in fact to *K. quasipneumoniae* subsp.

316   *similipneumoniae*), these primers successfully amplified a 78 bp sequence, with a melting

14

317   temperature of 80.3 °C. Sequencing of PCR products confirmed amplification of the target

318   sequence.

319   The sensitivity and specificity of the primers were experimentally tested on 2 ng of purified DNA

320   of representatives of Kp phylogroups Kp1 to Kp7 (Table 1) and non-Kp strains. The ZKIR_F and

321   ZKIR_R primers amplified the target sequence in all tested isolates of the Kp complex, with a

322   Threshold Cycle (Ct) value ranging from 13 to 26 and melting temperatures of 80.1 to 80.7 °C. In

323   contrast, when the ZKIR PCR was performed on 88 non-Kp isolates, no amplification was

324   observed, showing that the PCR did not yield false positives. Late, non-specific amplification

325   was recorded with few isolates in the last cycles of the reaction (after cycle 35) and melting

326   temperatures were clearly lower or higher than 80°C, indicating a non-specific amplification. All

327   assays performed on non-Kp samples can therefore be considered negative.

328

329   **Analytical sensitivity of the ZKIR PCR**

330   PCR reactions were performed with quantities of genomic DNA of *K. pneumoniae* ATCC13883^{T}

331   ranging from 7.5 ng (2.5×10^{6} genomes) to 15 fg (5 genomes) (Figure 3 A and B). Linearity of

332   the assay was evaluated by plotting the Ct values against the $\log_{10}$ calculated genome number

333   (calculated genome size is about 3.0 fg of DNA per cell, based on a 5.54 Mb *K. pneumoniae*

334   genome). Ct values were closely proportional to the logarithm of the genome number (R^{2}=0.99)

335   (Figure 4). The sensitivity of the assay was determined as 15 genomes (45 fg) per reaction mix.

336   No amplification was observed at the lower DNA concentration of 15 fg corresponding to 5

337   genomes.

338

339   **Analytical sensitivity of the ZKIR assay on soil samples**

15

340   In order to assess the performance of the ZKIR assay for detection of Kp directly from soil

341   samples, two soils (A and V) were spiked with bacterial concentrations ranging from 1.5×10^{-1}

342   cfu/g to 1.5×10^{4} cfu/g. When soil samples were enriched in LB for 24 h and processed as

343   described above, Kp was detected in all spiked microcosms except in two out of the 3

344   microcosms of soil A inoculated with the lowest Kp concentration. When the soil/LB suspension

345   was tested prior to incubation, Kp was not detected in any of the spiked microcosms of soil V,

346   whereas in soil A, Kp could only be detected at the highest concentration (1.5×10^{4} cfu/g).

347   Finally, when the ZKIR assay was performed using purified metagenomic DNA from soil,

348   positive results were only observed in microcosms spiked with 1.5×10^{4} and 1.5×10^{3} cfu/g in

349   both soils, whereas no positive signal was observed at lower concentrations. The enrichment step

350   thus appeared critical to reach high sensitivity.

351

352   **Comparison of the ZKIR real-time PCR and culture-based methods for the detection of Kp**

353   **in environmental samples**

354   After enrichment and ZKIR real-time PCR, Kp was detected in 41 out of 96 (42.7 %) assayed

355   environmental samples, when the 1:10 dilution was used as template DNA (Table 3). When the

356   1:100 dilution was used, Kp was detected in a lower number of positive samples (38/96). The Ct

357   values were in the range 18.4 - 36.2. The 96 samples were processed in parallel with the culture-

358   based method. Kp was not detected in any of the ZKIR-negative samples: when presumptive

359   colonies were detected, they were always identified by MALDI-TOF MS as non-Kp, and

360   belonged to closely related species such as *K. oxytoca*, *Raoultella* spp. and *Serratia* spp. In

361   contrast, Kp was isolated in 37 out of the 41 ZKIR positive samples. These isolates were all

362   identified as either *K. pneumoniae* or *K. variicola* by MALDI-TOF MS.

16

363   Genomic characterization of 31 of these isolates (1 was contaminated and 5 were not yet

364   available) revealed a dominance of Kp1 (n= 20, 65%), followed by Kp3 (n=10, 32%) and Kp4

365   (n=1, 3%; Table S1). Population diversity analysis based on MLST (41) and cgMLST (24)

366   revealed a high genotypic diversity, with 23 STs and 25 cgMLST types. In five cases, the strain

367   detected in the soil was the same as the one present in the leaves and/or roots (*e.g.*, SB6439 and

368   SB6440; Table S1), showing colonization of several plant sites by the same strain. Only 2 isolates

369   belonging to STs commonly found in the clinical settings were detected (ST37 and ST76).

370   Interestingly, from the predicted O-types, the O3-type represented 50% of Kp population, which

371   contrasts with the nosocomial situation, where types O1 and O2 are dominant (2, 42), but not

372   with human carriage where the O3-type is also dominant (O3-31%, O1-19% and O2-17%; our

373   unpublished results). Also contrasting with the clinical epidemiology of Kp, a low level of

374   antibiotic resistance and virulence genes was observed, with 93.5% of the strains presenting an

375   ancestral 'wild-type' susceptibility genotype (Table S1). The number of strains harboring

376   plasmids detected in these environmental samples was also low (n=9, 29%), as well as the

377   number of plasmid-encoded heavy metal tolerance genes (mainly the silver and copper tolerance

378   clusters). These results contrast with clinical and animal isolates, where plasmids and metal

379   tolerance clusters are common (42, Rodrigues, unpublished results). As expected (1, 37), all Kp3

380   isolates harbored the *nif* cluster responsible for nitrogen fixation. Interestingly, the *nif* cluster was

381   also present in one Kp1 isolate from soil (SB6181). The phylogenetic analysis (Figure S2) of the

382   *nif* cluster from our environmental isolates compared to a panel of reference strains (37) revealed

383   that strain SB6181 (Kp1) branched within *K. variicola* strains (Kp3), showing that the *nif* cluster

384   in this Kp1 strain was acquired via horizontal gene transfer from a *K. variicola* donor. Kp3 was

385   also the inferred donor of the *nif* gene cluster for Kp5 and Kp6 *nif*-positive strains (Figure S2).

17

386   **Discussion**

387   Routes of transmission of ecologically generalist human pathogens are usually complex and

388   poorly understood. Proper risk management requires a holistic approach which has been theorised

389   as the One Health concept (44). Despite the fact that the number of human infections caused by

390   members of the Kp complex are on the rise and are increasingly resistant to antimicrobial

391   treatment (1, 45), the ecology of Kp remains poorly understood. Identification of the various

392   habitats in which Kp strives, and the routes of transmission to humans, for example through

393   specific types of food, is critical in order to limit exposure. In this prospect, large-scale sampling

394   is needed to define the sources of Kp contamination and such surveys will require sensitive,

395   reliable and cost-efficient detection methods.

396   Several molecular assays for the detection and/or identification of *Klebsiella* from clinical or food

397   and environmental samples have been previously proposed. These methods allow detecting

398   mainly *K. pneumoniae* (*sensu stricto*) and *K. variicola* (Figure 1). However, the Kp complex

399   currently encompasses 7 taxa. Our *in-silico* analyses showed that some previously described PCR

400   assays could be useful for the identification of phylogenetic subsets of the Kp complex

401   (Kp150233a for Kp1, bla_{OKP} for Kp2/Kp4 and bla_{LEN} for Kp3/Kp5).

402

403   Given that all members of the Kp complex can cause infections in humans or animals, developing

404   a novel method with the ability to detect the Kp complex by targeting exhaustively all currently

405   known Kp phylogroups would be an important advance. Although several previous targets were

406   designed for the entire Kp complex, they predate recent taxonomic advances and have therefore

407   not been validated on the entire phylogenetic breadth of this bacterial group. Here, we found by

408   *in-silico* approaches that *khe* and *tyrB* assays lack specificity, which may negatively affect Kp

18

409   detection efforts especially when testing microbiologically complex samples from the

410   environment, such as soil.

411   We therefore aimed to develop a novel real-time PCR assay, and used an intergenic region

412   located adjacent to the previously proposed target gene *khe*. Using well-defined reference strains

413   representative of current taxonomy (20, 32, 33), we show that the novel ZKIR PCR real-time

414   assay allows the accurate detection of all members of the Kp complex. We found complete

415   specificity and sensitivity (*i.e*., phylogenetic coverage) of this assay based on our reference

416   strains panel. To improve analytical sensitivity, a 24 h enrichment culture followed by an easy,

417   fast and cost-effective sample processing was implemented, upstream of molecular detection

418   using the ZKIR assay. This was necessary for complex matrices such as soil or sewage, where

419   microbial diversity and abundance are high, in which case direct detection of one particular group

420   of organisms, which might be present only at low abundance, is challenging. This procedure

421   turned out to be sensitive enough to detect 1 single Kp bacterium in 5 g of soil. The ZKIR assay

422   also appeared slightly more sensitive than the SCAI culture-based method, which is based on the

423   ability of *Klebsiella* strains to utilize citrate and inositol, itself previously shown to be highly

424   sensitive and to recover most Kp members (36, 46). Therefore, this fast and easy novel molecular

425   method represents a powerful approach for screening large numbers of samples. This will spare

426   the time-consuming handling of numerous presumptive colonies necessary for confirmation of

427   their identification, given that *K. oxytoca*, *Raoultella, Serratia* and *Enterobacter* are able to form

428   colonies on SCAI agar, with morphological characteristics similar to Kp (36, 47). As such, the

429   ZKIR protocol was significantly faster than culture, as results were available on average 26 h

430   after sampling (24 h enrichment, sample treatment, PCR) while up to 96 h were necessary when

431   using the culture-based method (24 h enrichment, plate incubation, colony purification, MALDI-

432   TOF identification).

19

433   Interestingly, the implementation of the ZKIR PCR real-time assay evidenced a high (43%)

434   detection rate of Kp from environmental samples, which represent niches that are still

435   underexplored for Kp presence and biological characteristics. The high detection rate was largely

436   confirmed by culture, suggesting a low rate of false positive qPCR results. This novel and

437   original sampling strategy allowed us to explore the genomic features of environmental Kp

438   populations. A strong contrast with Kp isolates typically recovered in the clinical setting was

439   observed. First, a high genetic diversity was observed, with no predominant sublineage. This

440   situation contrasts with MDR or hypervirulent Kp populations, from which frequent sublineages

441   (so-called high-risk clones) are recovered. Second, environmental Kp isolates were almost devoid

442   of antibiotic resistance and virulence genes, in conspicuous contrast with clinical samples (48).

443   Finally, the genomic characteristics of environmental Kp isolates revealed interesting biological

444   features, such as a frequent O3 O-antigen type and the horizontal transfer of nitrogen fixation

445   gene cluster into Kp1, the main phylogroup associated with human infections. These data call for

446   further studies into the biology of soil Kp isolates, which may reveal interesting novel adaptive

447   strategies of this important generalist pathogen. Finally, our findings suggest that environmental

448   Kp populations differ from clinical Kp populations, which implies indirect and possibly complex

449   epidemiological links between environmental and clinical Kp. The ZKIR PCR real-time assay

450   developed herein is expected to enable future large-scale studies into this important question.

451

452   In conclusion, the ZKIR assay is a new tool for Kp detection that is highly specific, sensitive,

453   reliable and cost-effective. To accelerate uptake of this method, the corresponding protocol was

454   released publicly on protocols.io (dx.doi.org/10.17504/protocols.io.7n6hmhe). This simple

455   method can easily be implemented in laboratories equipped with real-time PCR thermocyclers

456   (MedVetKlebs consortium, unpublished results). Using the ZKIR method after a short culture

20

457   enrichment step greatly enhances sensitivity. As this rapid screen of samples allows one to focus

458   only on ZKIR positive samples for more labour-intensive downstream microbiological isolation

459   and characterization, it is our hope that it will contribute to advance knowledge on the biology of

460   environmental Kp and on the reservoirs and transmission routes of this increasingly important

461   group of pathogens.

462 **SUPPLEMENTAL MATERIAL**

463 **SUPPLEMENTAL FILE:** 2 PDF files, 1 Excel file

464

465

466 **ACKNOWLEDGEMENTS**

473 **REFERENCES**

474   1.   Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu

475   LY, Severin J, Brisse S, Cao H, Wilksch J, Gorrie C, Schultz MB, Edwards DJ, Nguyen KV,

476   Nguyen TV, Dao TT, Mensink M, Minh VL, Nhu NTK, Schultsz C, Kuntaman K, Newton PN,

477   Moore CE, Strugnell RA, Thomson NR. 2015. Genomic analysis of diversity, population structure,

478   virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health.

479   Proc Natl Acad Sci 112:E3574–E3581.

480   2.   David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, Abudahab K, Goater R, Giani T,

481   Errico G, Aspbury M, Sjunnebo S, Koraqi A, Lacej D, Apfalter P, Hartl R, Glupczynski Y, Huang

482   T-D, Strateva T, Marteva-Proevska Y, Andrasevic AT, Butic I, Pieridou-Bagatzouni D, Maikanti-

483   Charalampous P, Hrabak J, Zemlickova H, Hammerum A, Jakobsen L, Ivanova M, Pavelkovich A,

484   Jalava J, Österblad M, Dortet L, Vaux S, Kaase M, Gatermann SG, Vatopoulos A, Tryfinopoulou K,

485   Tóth Á, Jánvári L, Boo TW, McGrath E, Carmeli Y, Adler A, Pantosti A, Monaco M, Raka L, Kurti

486   A, Balode A, Saule M, Miciuleviciene J, Mierauskaite A, Perrin-Weniger M, Reichert P, Nestorova

487   N, Debattista S, Mijovic G, Lopicic M, Samuelsen Ø, Haldorsen B, Zabicka D, Literacka E, Caniça

488   M, Manageiro V, Kaftandzieva A, Trajkovska-Dokic E, Damian M, Lixandru B, Jelesic Z, Trudic A,

489   Niks M, Schreterova E, Pirs M, Cerar T, Oteo J, Aracil B, Giske C, Sjöström K, Gür D, Cakar A,

490   Woodford N, Hopkins K, Wiuff C, Brown DJ, Feil EJ, Rossolini GM, Aanensen DM, Grundmann

491   H, the EuSCAPE Working Group, the ESGEM Study Group. 2019. Epidemic of carbapenem-

492   resistant Klebsiella pneumoniae in Europe is driven by nosocomial spread. Nat Microbiol 4:1919–

493   1929.

494   3.   Ko W-C. 2002. Community-acquired *Klebsiella pneumoniae* bacteremia: global differences in

495   clinical patterns. Emerg Infect Dis 8:160–166.

23

496   4.   Zhang Y, Zeng J, Liu W, Zhao F, Hu Z, Zhao C, Wang Q, Wang X, Chen H, Li H, Zhang F, Li S,

497   Cao B, Wang H. 2015. Emergence of a hypervirulent carbapenem-resistant *Klebsiella pneumoniae*

498   isolate from clinical infections in China. J Infect 71:553–560.

499   5.   Gorrie CL, Mirčeta M, Wick RR, Edwards DJ, Thomson NR, Strugnell RA, Pratt NF, Garlick JS,

500   Watson KM, Pilcher DV, McGloughlin SA, Spelman DW, Jenney AWJ, Holt KE. 2017.

501   Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care

502   patients. Clin Infect Dis 65:208–215.

503   6.   Brisse S, Grimont F, Grimont PAD. 2006. The Genus *Klebsiella*, p. 159–196. *In* Dworkin, M,

504   Falkow, S, Rosenberg, E, Schleifer, K-H, Stackebrandt, E (eds.), The Prokaryotes. Springer New

505   York.

506   7.   Munoz MA, Ahlström C, Rauch BJ, Zadoks RN. 2006. Fecal shedding of *Klebsiella pneumoniae* by

507   dairy cows. J Dairy Sci 89:3425–3430.

508   8.   Rosenblueth M, Martínez L, Silva J, Martínez-Romero E. 2004. *Klebsiella variicola*, a novel species

509   with clinical and plant-associated isolates. Syst Appl Microbiol 27:27–35.

510   9.   Lee D-Y, Shannon K, Beaudette LA. 2006. Detection of bacterial pathogens in municipal

511   wastewater using an oligonucleotide microarray and real-time quantitative PCR. J Microbiol

512   Methods 65:453–467.

513   10.   Benami M, Gross A, Herzberg M, Orlofsky E, Vonshak A, Gillor O. 2013. Assessment of

514   pathogenic bacteria in treated graywater and irrigated soils. Sci Total Environ 458–460:298–302.

515   11.   Barati A, Ghaderpour A, Chew L, Bong C, Thong K, Chong V, Chai L. 2016. Isolation and

516   characterization of aquatic-borne *Klebsiella pneumoniae* from tropical estuaries in Malaysia. Int J

517   Environ Res Public Health 13:426.

518 12. Falomir MP, Gozalbo D, Rico H. 2010. Coliform bacteria in fresh vegetables: from cultivated lands
519   to consumers, volume 2. *In* A. Mendez-Vilas, Current Research, Technology and Education Topics
520   in Applied Microbiology and Microbial Biotechnology.

521 13. Podschun R, Pietsch S, Holler C, Ullmann U. 2001. Incidence of *Klebsiella* species in surface waters
522   and their expression of virulence factors. Appl Environ Microbiol 67:3325–3327.

523 14. Davis GS, Waits K, Nordstrom L, Weaver B, Aziz M, Gauld L, Grande H, Bigler R, Horwinski J,
524   Porter S, Stegger M, Johnson JR, Liu CM, Price LB. 2015. Intermingled *Klebsiella pneumoniae*
525   populations between retail meats and human urinary tract infections. Clin Infect Dis 61:892–899.

526 15. Zadoks RN, Griffiths HM, Munoz MA, Ahlstrom C, Bennett GJ, Thomas E, Schukken YH. 2011.
527   Sources of *Klebsiella* and *Raoultella* species on dairy farms: Be careful where you walk. J Dairy Sci
528   94:1045–1051.

529 16. Runcharoen C, Moradigaravand D, Blane B, Paksanont S, Thammachote J, Anun S, Parkhill J,
530   Chantratita N, Peacock SJ. 2017. Whole genome sequencing reveals high-resolution epidemiological
531   links between clinical and environmental *Klebsiella pneumoniae*. Genome Med 9:6.

532 17. Gregg S. Davis, Lance B. Price. 2016. Recent research examining links among *Klebsiella*
533   *pneumoniae* from food, food animals, and human extraintestinal infections. Curr Envir Health Rpt
534   3:128–135.

535 18. Brisse S, Passet V, Grimont PAD. 2014. Description of *Klebsiella quasipneumoniae* sp. nov.,
536   isolated from human infections, with two subspecies, *Klebsiella quasipneumoniae* subsp.
537   *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae* subsp. nov.,
538   and demonstration that *Klebsiella singaporensis* is a junior heterotypic synonym of *Klebsiella*
539   *variicola*. Int J Syst Evol Microbiol 64:3146–3152.

25

540   19.   Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, Olsen RJ. 2017. Whole-
541   genome sequencing of a human clinical isolate of the novel species *Klebsiella quasivariicola* sp.
542   nov. Genome Announc 5:e01057-17.

543   20.   Rodrigues C, Passet V, Rakotondrasoa A, Diallo TA, Criscuolo A, Brisse S. 2019. Description of
544   *Klebsiella africanensis* sp. nov., *Klebsiella variicola* subsp. *tropicalensis* subsp. nov. and
545   *Klebsiella variicola* subsp. *variicola* subsp. nov. Res Microbiol 170:165-170.

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

550   22.   Liu Y, Liu C, Zheng W, Zhang X, Yu J, Gao Q, Hou Y, Huang X. 2008. PCR detection of
551   *Klebsiella pneumoniae* in infant formula based on 16S–23S internal transcribed spacer. Int J Food
552   Microbiol 125:230–235.

553   23.   Garza-Ramos U, Silva-Sánchez J, Martínez-Romero E, Tinoco P, Pina-Gonzales M, Barrios H,
554   Martínez-Barnetche J, Gómez-Barreto RE, Tellez-Sosa J. 2015. Development of a Multiplex-PCR
555   probe system for the proper identification of *Klebsiella variicola*. BMC Microbiol 15:64.

556   24.   Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Nicolas-Chanoine M-H, Decre D, Brisse S.
557   2014. Development of a multiplex PCR assay for identification of Klebsiella pneumoniae
558   hypervirulent clones of capsular serotype K2. J Med Microbiol 63:1608–1614.

559   25.   Yin-Ching C, Jer-Horng S, Ching-Nan L, Ming-Chung C. 2002. Cloning of a gene encoding a
560   unique haemolysin from *Klebsiella pneumoniae* and its potential use as a species-specific gene
561   probe. Microb Pathog 33:1–6.

562   26.   Jeong E-S, Lee K-S, Heo S-H, Seo J-H, Choi Y-K. 2013. Rapid Identification of *Klebsiella*

563   *pneumoniae*, *Corynebacterium kutscheri*, and *Streptococcus pneumoniae* using triplex Polymerase

564   Chain Reaction in rodents. Exp Anim 62:35–40.

565   27.   Hartman LJ, Selby EB, Whitehouse CA, Coyne SR, Jaissle JG, Twenhafel NA, Burke RL, Kulesh

566   DA. 2009. Rapid real-time PCR assays for detection of *Klebsiella pneumoniae* with the *rmpA* or

567   *magA* genes associated with the hypermucoviscosity phenotype. J Mol Diagn 11:464–471.

568   28.   Fonseca EL, Ramos N da V, Andrade BGN, Morais LLCS, Marin MFA, Vicente ACP. 2017. A one-

569   step multiplex PCR to identify *Klebsiella pneumoniae* , *Klebsiella variicola*, and *Klebsiella*

570   *quasipneumoniae* in the clinical routine. Diagn Microbiol Infect Dis 87:315–317.

571   29.   Singh K, Mangold KA, Wyant K, Schora DM, Voss B, Kaul KL, Hayden MK, Chundi V, Peterson

572   LR. 2012. Rectal screening for *Klebsiella pneumoniae* carbapenemases: comparison of Real-Time

573   PCR and culture using two selective screening agar plates. J Clin Microbiol 50:2596–2600.

574   30.   Bilodeau G. 2011. Quantitative polymerase chain reaction for the detection of organisms in soil.

575   CAB Rev Perspect Agric Vet Sci Nutr Nat Resour 6.

576   31.   Plassart P, Terrat S, Thomson B, Griffiths R, Dequiedt S, Lelievre M, Regnier T, Nowak V, Bailey

577   M, Lemanceau P, Bispo A, Chabbi A, Maron P-A, Mougel C, Ranjard L. 2012. Evaluation of the

578   ISO Standard 11063 DNA extraction procedure for assessing soil microbial abundance and

579   community structure. PLoS ONE 7:e44279.

580   32.   Rodrigues C, Passet V, Rakotondrasoa A, Brisse S. 2018. Identification of Klebsiella pneumoniae,

581   Klebsiella quasipneumoniae, Klebsiella variicola and Related Phylogroups by MALDI-TOF Mass

582   Spectrometry. Front Microbiol 9:3000.

583 33. Passet V, Brisse S. 2018. Description of *Klebsiella grimontii* sp. nov. Int J Syst Evol Microbiol
584   68:377–381.

585 34. Criscuolo A. 2019. A fast alignment-free bioinformatics procedure to infer accurate distance-based
586   phylogenetic trees from genome assemblies. Res Ideas Outcomes 5:e36178.

587 35.   López-Gutiérrez et al. - 2004 - Quantification of a novel group of nitrate-reducing bacteria in the
588   environment by real-time PCR. J. Microbiol. Methods 57:399-407.

589 36. Van Kregten E, Westerdaal NAC, Willers JMN. 1984. New, simple medium for selective recovery
590   of *Klebsiella pneumoniae* and *Klebsiella oxytoca* from human feces. J Clin Microbiol 20:6.

591 37. Blin C, Passet V, Touchon M, Rocha EPC, Brisse S. 2017. Metabolic diversity of the emerging
592   pathogenic lineages of *Klebsiella pneumoniae*: Metabolic diversity of *Klebsiella pneumoniae*.
593   Environ Microbiol 19:1881–1898.

594 38. Tian Y, Wang L, Zhang J, Han Q, Xia X, Song Y, Yang G. 2019. CelB is a suitable marker for rapid
595   and specific identification of Klebsiella pneumoniae by the loop-mediated isothermal amplification
596   (LAMP) assay. Braz J Microbiol 50:961–967.

597 39. Haeggman S, Lofdahl S, Paauw A, Verhoef J, Brisse S. 2004. Diversity and Evolution of the Class
598   A Chromosomal Beta-Lactamase Gene in *Klebsiella pneumoniae*. Antimicrob Agents Chemother
599   48:2400–2408.

600 40. Jeong E-S, Lee K-S, Heo S-H, Seo J-H, Choi Y-K. 2013. Rapid Identification of Klebsiella
601   pneumoniae, Corynebacterium kutscheri, and Streptococcus pneumoniae Using Triplex Polymerase
602   Chain Reaction in Rodents. Exp Anim 62:35–40.

603 41. Diancourt L, Passet V, Verhoef J, Grimont PAD, Brisse S. 2005. Multilocus Sequence Typing of
604   *Klebsiella pneumoniae* Nosocomial Isolates. J Clin Microbiol 43:4178.

28

605 42. Follador R, Heinz E, Wyres KL, Ellington MJ, Kowarik M, Holt KE, Thomson NR. 2016. The
606   diversity of Klebsiella pneumoniae surface polysaccharides. Microb Genomics 2.

607 43. Sütterlin S, Dahlö M, Tellgren-Roth C, Schaal W, Melhus Å. 2017. High frequency of silver
608   resistance genes in invasive isolates of Enterobacter and Klebsiella species. J Hosp Infect 96:256–
609   261.

610 44. King LJ, Anderson LR, Blackmore CG, Blackwell MJ, Lautner EA, Marcus LC, Meyer TE, Monath
611   TP, Nave JE, Ohle J, Pappaioanou M, Sobota J, Stokes WS, Davis RM, Glasser JH, Mahr RK. 2008.
612   Executive summary of the AVMA One Health Initiative Task Force report. J Am Vet Med Assoc
613   233:259–261.

614 45. Nordmann P, Naas T, Poirel L. 2011. Global spread of carbapenemase-producing
615   *Enterobacteriaceae*. Emerg Infect Dis 17:1791–1798.

616 46. Passet V, Brisse S. 2015. Association of Tellurite Resistance with Hypervirulent Clonal Groups of
617   *Klebsiella pneumoniae*. J Clin Microbiol 53:1380–1382.

618 47. Ohtomo R, Saito M. 2003. A New selective medium for detection of *Klebsiella* from dairy
619   environments. Microbes Environ 18:138–144.

620 48. Wyres KL, Holt KE. 2016. *Klebsiella pneumoniae* Population Genomics and Antimicrobial-
621   Resistant Clones. Trends Microbiol 24:944–956.

622

623

624

625 **TABLES**

626 **TABLE 1** Bacterial strains tested to develop the ZKIR PCR assay.

29

627   **TABLE 2** Molecular target sequences previously proposed for *Klebsiella pneumoniae*

628   identification.

629   **TABLE 3** Comparison between ZKIR PCR and culture results in samples collected in Auxonne,

630   France, between July and September 2018.

631

632   **FIGURES**

633   **FIG 1** Phylogenetic distribution of the molecular targets for *K. pneumoniae* detection described

634   in the literature. The ZKIR target sequence corresponds to target 15; other targets are given in

635   Table 2. The inner circle colored sectors correspond to *K. pneumoniae* phylogroups or other

636   *Klebsiella* species (see color key). Molecular targets were detected in the corresponding genomes

637   using BLASTN with 80% nucleotide identity and 80% length coverage.

638

639   **FIG 2** Genetic context of the ZKIR region on the genome of strain *K. pneumoniae* ATCC 13883[T]

640   (GenBank accession number GCA_000742135.1) and detailed location of ZKIR primers and

641   amplicon within the 249-base pair (bp) region that is specific for the *K. pneumoniae* species

642   complex (boxed area).

643

644   **FIG 3** Amplification curves (A) and melt curve peaks (B) established using real-time PCR

645   targeting the ZKIR region with serial dilutions of *K. pneumoniae* ATCC13883[T] DNA. Triplicate

646   of DNA concentrations of 7.5 ng, 750 pg, 75 pg, 7.5 pg, 750 fg are presented, but lower dilutions

647   are not.

648

649   **FIG 4** Standard curve established using real-time ZKIR PCR with serial dilutions of

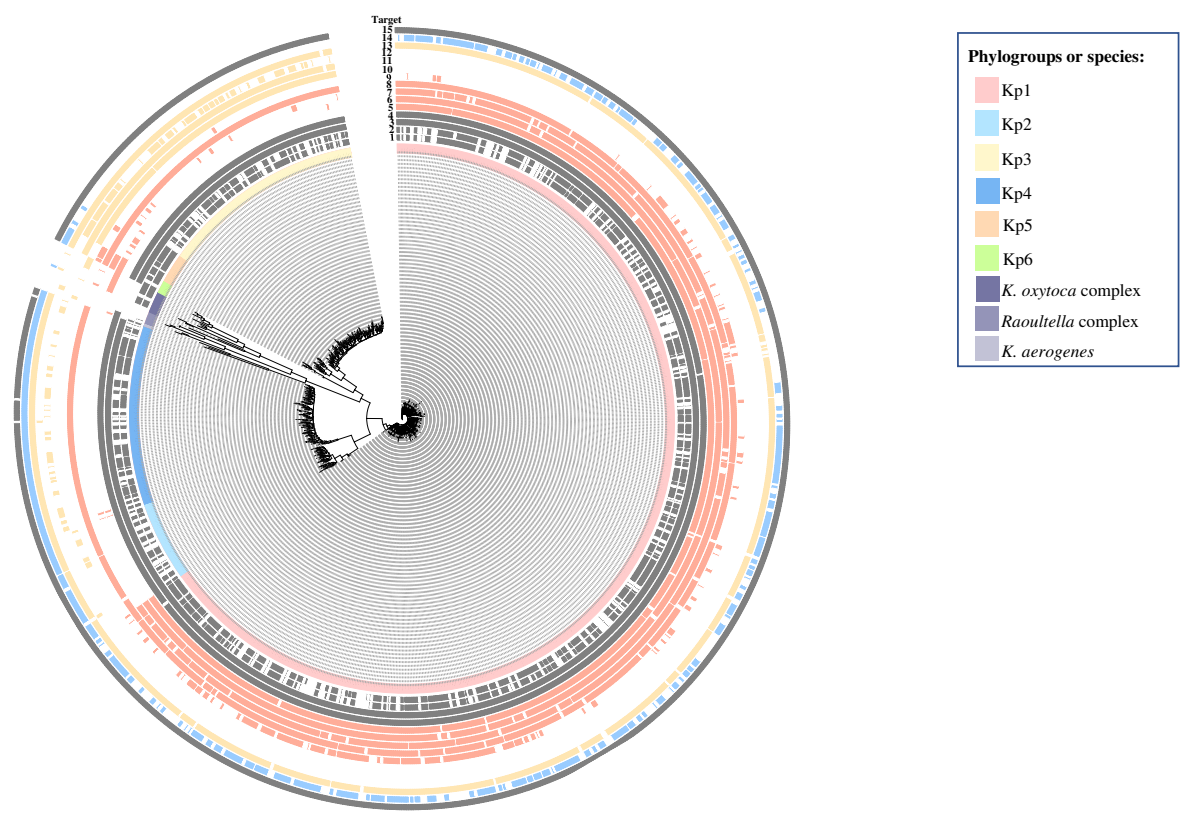650   *K. pneumoniae* ATCC13883[T] DNA from 7.5 ng to 45 fg.

30

Figure 1. Phylogenetic distribution of the molecular targets for *K. pneumoniae* detection described in the literature. The ZKIR target sequence corresponds to target 15; other targets are given in Table 2. The inner circle colored sectors correspond to *K. pneumoniae* phylogroups or other *Klebsiella* species (see color key). Molecular targets were detected in the corresponding genomes using BLASTN with 80% nucleotide identity and 80% length coverage.
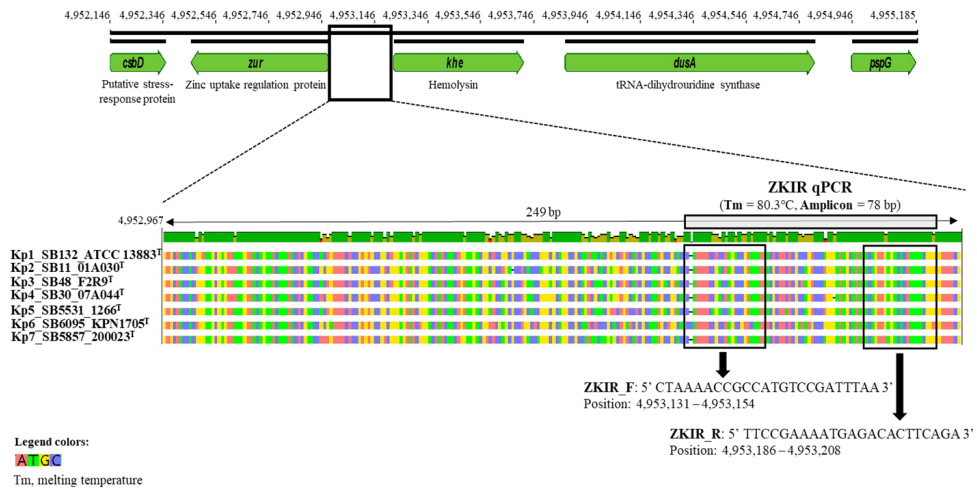
Figure 2 Genetic context of the ZKIR region on the genome of strain *K. pneumoniae* ATCC 13883[T] (GenBank accession number GCA_000742135.1) and detailed location of ZKIR primers and amplicon within the 249-base pair (bp) region that is specific for the *K. pneumoniae* species complex (boxed area).
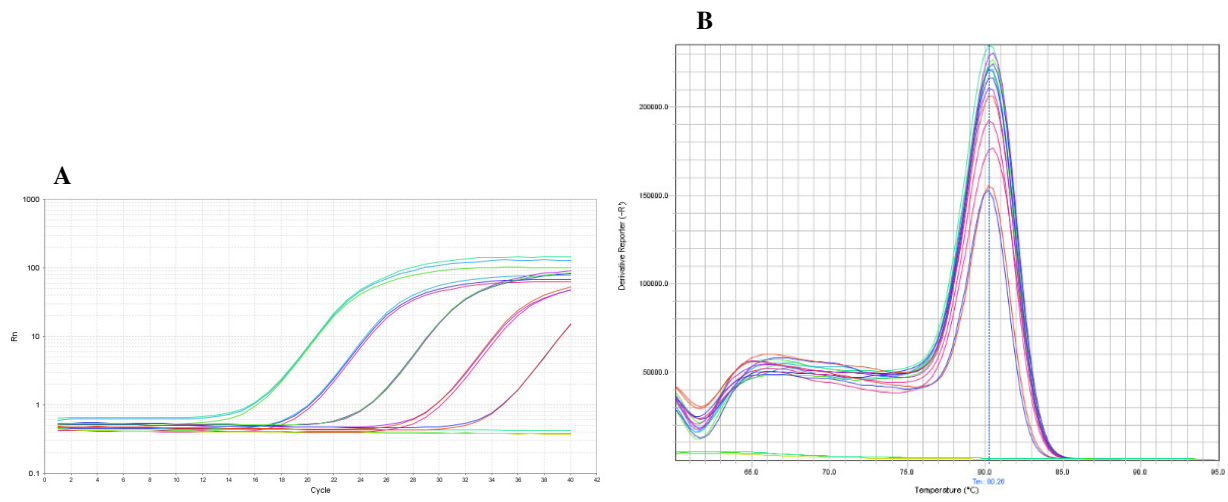
Figure 3 Amplification curves (A) and melt curve peaks (B) established using real-time PCR targeting the ZKIR region with serial dilutions of *K. pneumoniae* ATCC13883[T] DNA. Triplicates of DNA concentrations of 7.5 ng, 750 pg, 75 pg, 7.5 pg and 750 fg are presented, whereas lower dilutions are not.
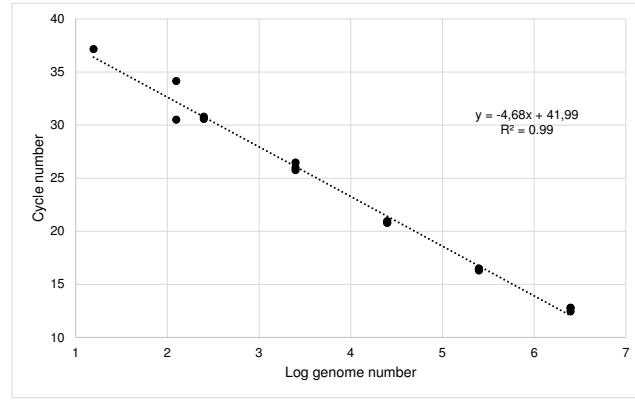
Figure 4 Standard curve established using real-time ZKIR PCR with serial dilutions of *K. pneumoniae* ATCC13883[T] DNA from 7.5 ng to 45 fg.

TABLE 1 Bacterial strains tested to develop the ZKIR PCR assay.

Organism name (Phylogroup)	Strain name	Strain bank ID*	Country	Isolation year
Klebsiella pneumoniae (Kp1)	SB4-2	SB1067	Netherlands	2002
K. pneumoniae (Kp1)	ATCC 13883[T]	SB132	n.a.	n.a.
K. pneumoniae (Kp1)	MGH 78578	SB107	n.a.	1994
K. pneumoniae (Kp1)	none	SB1139	Netherlands	2002
K. pneumoniae (Kp1)	5-2	SB617	Netherlands	2000
K. pneumoniae (Kp1)	04A025	SB20	France	1997
K. pneumoniae (Kp1)	2-3	SB612	Netherlands	2000
K. pneumoniae (Kp1)	BJ1-GA	SB4496	France	2011
K. pneumoniae (Kp1)	n.a.	MIAE07651	France	2015
K. *quasipneumoniae* subsp. quasipneumoniae (Kp2)	01A030[T]	SB11	Austria	1997
K. *quasipneumoniae* subsp. quasipneumoniae (Kp2)	none	SB1124	Netherlands	2002
K. *quasipneumoniae* subsp. quasipneumoniae (Kp2)	U41	SB2110	Germany	1990
K. *quasipneumoniae* subsp. quasipneumoniae (Kp2)	10A442	SB224	Italy	1998
K. *quasipneumoniae* subsp. quasipneumoniae (Kp2)	99-1002	SB2478	Netherlands	1999
K. *quasipneumoniae* subsp. quasipneumoniae (Kp2)	18A451	SB255	Spain	1998
K. *quasipneumoniae* subsp. quasipneumoniae (Kp2)	11128	SB3445	n.a.	n.a.
K. *quasipneumoniae* subsp. quasipneumoniae (Kp2)	18A069	SB59	Spain	1997
K. *quasipneumoniae* subsp. quasipneumoniae (Kp2)	Kleb Ali 0320584	SB98	n.a.	n.a.
K. *variicola* subsp. *variicola* (Kp3)	01A065	SB1	Austria	1997
K. *variicola* subsp. *variicola* (Kp3)	07A058	SB31	Germany	1997
K. *variicola* subsp. *variicola* (Kp3)	IPEUC-1516	SB3278	France	1988
K. *variicola* subsp. *variicola* (Kp3)	CIP 53.24	SB3295	n.a.	n.a.
K. *variicola* subsp. *variicola* (Kp3)	Ørskov 1756/51	SB3301	n.a.	n.a.
K. *variicola* subsp. *variicola* (Kp3)	F2R9[T]	SB48	Mexico	n.a.
K. *variicola* subsp. *variicola* (Kp3)	6115	SB489	n.a.	n.a.
K. *variicola* subsp. *variicola* (Kp3)	Ørskov 4425/51	SB497	n.a.	n.a.
K. *variicola* subsp. *variicola* (Kp3)	Kp342	SB579	USA	n.a.
K. *quasipneumoniae* subsp. *similipneumoniae* (Kp4)	09A323	SB164	Greece	1997
K. *quasipneumoniae* subsp. *similipneumoniae* (Kp4)	12A476	SB203	Netherlands	1998
K. *quasipneumoniae* subsp. *similipneumoniae* (Kp4)	07A044[T]	SB30	Germany	1997
K. *quasipneumoniae* subsp. *similipneumoniae* (Kp4)	325	SB3233	France	1975
K. *quasipneumoniae* subsp. *similipneumoniae* (Kp4)	Ørskov 1303/50	SB3297	Turkey	n.a.
K. *quasipneumoniae* subsp. *similipneumoniae* (Kp4)	Ørskov 4463/52	SB500	n.a.	n.a
K. *quasipneumoniae* subsp. *similipneumoniae* (Kp4)	CIP 110288	SB4697	China	2010
K. *quasipneumoniae* subsp. *similipneumoniae* (Kp4)	1-1	SB610	Netherlands	2000
K. variicola subsp. *tropica* (Kp5)	Gal12	SB824	Mexico	n.a.
K. variicola subsp. *tropica* (Kp5)	CDC 4241-71	SB94	n.a.	n.a.
K. variicola subsp. *tropica* (Kp5)	885	SB5439	Madagascar	2016
K. variicola subsp. *tropica* (Kp5)	1266[T]	SB5531	Madagascar	2016
K. variicola subsp. *tropica* (Kp5)	1283	SB5544	Madagascar	2016
K. variicola subsp. *tropica* (Kp5)	1375	SB5610	Madagascar	2016
K. variicola subsp. *tropica* (Kp5)	814	SB5387	Madagascar	2015
'*K. quasivariicola*' (Kp6)	08A119	SB33	Germany	1997
'*K. quasivariicola*' (Kp6)	10982	SB6071	USA	2005
'*K. quasivariicola*' (Kp6)	01-467-2ECBU	SB6094	Madagascar	2015
'*K. quasivariicola*' (Kp6)	01-310A	SB6095	Madagascar	2013

'K. quasivariicola' (Kp6)	KPN1705[T]	SB6096	USA	2014
K. africana (Kp7)	200023[T]	SB5857	Senegal	2016
K. michiganensis (Ko1)	CIP 110787[T]	SB4934	USA	2010
K. michiganensis (Ko1)	05A071	SB71	France	1997
K. michiganensis (Ko1)	09A029	SB78	Greece	1997
K. grimontii (Ko6)	07A479	SB324	Germany	1998
K. grimontii (Ko6)	06D090	SB352	France	1998
K. grimontii (Ko6)	06D021[T]	SB73	France	1997
K. oxytoca (Ko2)	ATCC 13182[T]	SB175	n.a.	n.a.
K. oxytoca (Ko2)	02A067	SB131	Belgium	1997
K. oxytoca (Ko2)	NCTC 49131	SB136	n.a	n.a
K. aerogenes	none	MIAE07652	France	n.a.
K. aerogenes	CIP 60.86[T]	SB3629	France	n.a.
K. aerogenes	01A089	SB538	Austria	1997
K. aerogenes	02A002	SB539	Belgium	1997
Raoultella terrigena	ATCC 33257[T]	SB170	n.a.	n.a.
R. terrigena	17C143	SB313	Spain	1998
R. terrigena	V9813596	SB2796	Netherlands	1998
R. planticola	01A041	SB7	Austria	1997
R. planticola	ATCC 33531[T]	SB174	n.a.	n.a.
R. planticola	12C169	SB303	Netherlands	1998
R. ornithinolytica	ATCC 31898	SB171	n.a.	n.a.
Acinetobacter calcoaceticus	ATCC 14987	none	USA	n.a.
A. lwoffii	none	MIAE07654	France	1998
A. johnsonii	ATCC 17909[T]	none	n.a.	n.a.
A. pittii	none	MIAE07655	France	1998
Aeromonas caviae	none	MIAE07656	France	n.a.
A. hydrophila	none	MIAE07657	France	n.a.
Agrobacterium tumefaciens	C.58	MIAE07675	France	1996
Alcaligenes faecalis	none	MIAE07658	France	n.a.
Bacillus cereus	ATCC 53522	none	USA	n.a.
B. cereus	none	MIAE07659	France	n.a.
B. circulans	none	MIAE07660	France	n.a.
B. megaterium	none	MIAE07661	France	n.a.
B.s subtilis	none	MIAE07662	France	n.a.
Edwardsiella hoshinae	DSM 13771[T]	none	France	n.a.
E. tarda	DSM 30052[T]	none	USA	n.a.
Enterobacter hafnia	none	MIAE07653	France	n.a.
Enterococcus casseliflavus	DSM 20680[T]	none	n.a.	1984
E. faecalis	DSM 12956	none	USA	n.a.
E. faecalis	DSM 20376	none	n.a.	n.a.
E. faecium	DSM 25389	none	Netherlands	n.a.
E. faecium	DSM 6177	none	n.a.	n.a.
E. faecium	DSM 25644	none	Georgia	n.a.
E. faecium	DSM 25697	none	South Korea	n.a.
E. faecium	DSM 13590	none	Germany	n.a.
E. faecium	DSM 20477[T]	none	n.a.	n.a.
E. galinarum	DSM 20628	none	n.a.	n.a.
E. hirae	DSM 20160[T]	none	n.a.	n.a.
E. raffinosus	DSM 5633[T]	none	USA	1979

Escherichia coli	DSM 499	none	n.a.	n.a.
E. coli	none	MIAE02388	France	2015
E. coli	none	MIAE02376	France	2015
E. coli	none	MIAE02510	France	2015
E. coli	none	MIAE02381	France	2015
E. coli	none	MIAE02198	France	2015
Hafnia alvei	DSM 30163[T]	none	n.a.	n.a.
Listeria monocytogenes	ATCC BAA-679, (EGDe)	none	n.a.	n.a.
L. monocytogenes	DSM 15675	none	n.a.	n.a.
L. monocytogenes	DSM 19094	none	United Kingdom	n.a.
L. innocua	none	MIAE07668	France	2018
Morganella morganii	none	MIAE07669	France	2018
Pantoea agglomerans	DSM 3493[T]	none	Zimbabwe	1956
Proteus mirabilis	none	MIAE07670	France	n.a.
P. vulgaris	none	MIAE07671	France	n.a.
Providencia rettgeri	none	MIAE07672	France	n.a.
P. stuartii	none	MIAE07673	France	n.a.
Pseudomonas brassicacearum	DSM 13227[T]	none	France	n.a.
P. fluorescens	DSM 50106	none	n.a.	n.a.
P. gessardii	CIP 105469[T]	none	France	n.a.
P. jessenii	CFBP4842	none	France	1995
P. kilonensis	DSM 13647[T]	none	Germany	n.a.
P. libanenis	DSM 17149[T]	none	Lebanon	1995
P. lini	DSM 16768[T]	none	France	n.a.
P. monteilii	DSM 14164[T]	none	France	1990
P. putida	ATCC 12633[T]	none	n.a.	n.a.
P. rhodesiae	DSM 14020[T]	none	France	n.a.
P. thivervalensis	CFBP5754	none	France	2000
Salmonella bongori	DSM 13772[T]	none	n.a.	n.a.
S. enterica	DSM 554	none	n.a.	n.a.
S. enteritidis	none	MIAE07674	France	2007
S.subterranea	DSM 16208[T]	none	USA	n.a.
Serratia fonticola	DSM 4576[T]	none	n.a.	n.a.
S. liquefaciens	DSM 4487[T]	none	Ireland	1997
Shigella boydii	DSM 7532[T]	none	India	n.a.
Staphylococcus aureus	DSM 2569	none	South Korea	n.a.
S. capitis	DSM 20326[T]	none	n.a.	n.a.
S. epidermidis	DSM 20044 [T]	none	n.a.	n.a.
Stenotrophomonas maltophilia	DSM 50170[T]	none	USA	1961
Vibrio chagasii	DSM 17138[T]	none	Norway	n.a.

n.a. : information not avalaible, [T] type strain

*SB: Internal strain collection number of the Biodiversity and Epidemiology of Bacterial Pathogens unit, Institut Pasteur ; MIAE : Internal strain collection number of the UMR Agroecologie, INRA

TABLE 2 Molecular target sequences previously proposed for *Klebsiella pneumoniae* identification.

Target	Annotation	Amplicon (bp)	Targeted group	Reference (PubMed ID)
1. 16S-23S rRNA ITS	tDNA-Ala	130	Kp complex	18579248
2. 16S-23S rRNA ITS	tDNA-Ala – 23S rDNA	260	Kp complex	18579248
3. *khe*	hemolysin	77	Kp complex	19644019
4. *tyrB*	tyrosine aminotransferase	931	Kp complex	23357944
5. KpI50233a	putative acyltransferase	484	Kp1	25261063
6. KP878	transferase	878	Kp1	25886267
7. KP888	phosphohydrolase	888	Kp1	25886267
8. *bla*$_{SHV}$	chromosomal SHV	995	Kp1	28139276
9. *celB*	cellobiose-specific PTS family enzyme IIC component	180	Kp1	31456171
10. KV770	phosphoglycerate mutase	449	Kp3/5	25886267
11. KV1000	thiopurine S-methyltransferase	499	Kp3/5	25886267
12. KV1615	N-acetyltransferase	438	Kp3/5	25886267
13. *bla*$_{LEN}$	chromosomal LEN	485	Kp3/5	28139276
14. *bla*$_{OKP}$	chromosomal OKP	348	Kp2/4	28139276
15. ZKIR	*zur-khe* intergenic region	78	Kp complex	This study

ITS, internal transcribed spacer; PTS, phosphotransferase system; *bla*, beta-lactamase; bp, base pairs.

TABLE 3 Comparison between ZKIR qPCR and culture results obtained using samples collected in Auxonne, France, between July and September 2018.

Source	No. of samples	qPCR positive samples (%)	Culture positive samples (%)
Bulk soil	32	13 (40.6%)	13 (40.6%)
Roots	31	19 (61.3%)	16 (51.6%)
Leaves	29	8 (27.6%)	7 (24.1%)
Water	4	1 (25.0%)	1 (25.0%)
Total	96	41/96 (42.7 %)	37/96 (38.5%)