JCM Accepts, published online ahead of print on 2 October 2013
J. Clin. Microbiol. doi:10.1128/JCM.01924-13

wzi gene sequencing, a rapid method to determine the capsular type of *Klebsiella* strains

Sylvain Brisse[1,2*], Virginie Passet[1,2], Anita Björk Haugaard[3,4], Anamaria Babosan[5], Najiby Kassis-Chikhani[6], Carsten Struve[3,4] and Dominique Decré[5]

1 Institut Pasteur, Microbial Evolutionary Genomics, Paris, France

2 CNRS, UMR3525, Paris,75015, France

3 Department of Microbiology and Infection Control, Statens Serum Institut, Artillerivej 5, Copenhagen, Denmark

4 WHO Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella* Statens Serum Institut, Copenhagen, Denmark

5 Département de Bactériologie, Université Pierre et Marie Curie, Faculté de Médecine; and Assistance Publique-Hôpitaux de Paris, Hôpitaux Universitaires Est Parisiens, Paris, France

6 Laboratoire de Bactériologie, Hôpital Paul Brousse, Assistance Publique-Hôpitaux de Paris, Villejuif, France

* Corresponding author: Sylvain Brisse. Microbial Evolutionary Genomics, Institut Pasteur, 28 rue du Dr Roux, F-75724 Paris, France. E-mail: sylvain.brisse@pasteur.fr; Phone +33 1 40 61 36 58; Fax: +33 1 45 68 87 27

Keywords: *Klebsiella pneumoniae*, capsular serotyping, molecular method, clonal diversity, emerging pathogen

Abstract

Pathogens of the genus *Klebsiella* have been classified into distinct capsular (K) types for nearly one century. K-typing of *Klebsiella* still has important applications in epidemiology and clinical microbiology, but the serological method has strong practical limitations. Our objective was to evaluate sequencing of *wzi*, a gene conserved in all capsular types coding for an outer membrane protein involved in capsule attachment to the cell surface, as a simple and rapid method to predict the K-type. Sequencing of a 447-nucleotides region of *wzi* distinguished K-type reference strains with only nine exceptions. A reference *wzi* sequence database was created by inclusion of multiple strains representing K-types associated with high virulence and multidrug resistance. A collection of 119 prospective clinical isolates of *K. pneumoniae* was then analyzed in parallel by *wzi* sequencing and classical K-typing. Whereas K-typing achieved 81% typeability and 94.4% discrimination, these figures were 98.1%, and 98.3% for *wzi* sequencing. Prediction of K-type once knowing the *wzi* allele was 94%. *wzi* sequencing is as a rapid and simple method to determine the K-type of most *K. pneumoniae* clinical isolates.

Introduction

Bacteria of the genus *Klebsiella* are important pathogens. *K. pneumoniae*, the most frequent species in human infections, is emerging as an agent of severe community infections, including pyogenic liver abscess (9, 35, 42). Besides, multidrug resistant infections by *K. pneumoniae* are increasingly reported and are responsible for countless health-care settings outbreaks (43). Strain typing is necessary to recognize clonal groups, to assist epidemiological investigations and to link genetic diversity with pathophysiological specificities. Pulsed-field gel electrophoresis (PFGE), multilocus sequence typing (MLST) and capsular polysaccharide characterisation (K-typing) are widely used to characterize *K. pneumoniae* clinical isolates (11, 18, 41, 43). MLST has been widely adopted as the common language to characterize *K. pneumoniae* and recognize important emerging sequence types (ST), such as ST23 associated with liver-abscess (40) and ST15 associated with multidrug resistance worldwide (8, 19). K-typing was pioneered in 1926 (21), was further developed until the 1970's (24) and still is a widely-used characterization method. At least 78 K-types can be distinguished, some of which are strongly associated with virulence in experimental models (22) and with particular human infections (13, 16). However, K-typing is technically difficult and the production of the necessary reagents is complex. Molecular methods to deduce the K-type from genomic sequence have been developed based on PCR-RFLP (3) or allele-specific PCR amplification (13, 27, 37, 47). Although the former method can be used to characterize most *K. pneumoniae* clinical isolates, it is technically complex. On the contrary, allele-specific PCRs are easy to perform but were mainly developed for serotypes that are frequent among community-acquired bacteremia and liver-abscess isolates, namely K1, K2, K5, K20, K54 and K57 (13, 27, 45, 46). However, these serotypes represent a small proportion (less than 20%) of *K. pneumoniae* isolates from other sources than liver abscess or community-acquired pneumonia (7, 18, 39), thus leaving most of the classical *K. pneumoniae* isolates untypeable. Although it should be feasible to develop specific PCR assays for most K-types, it is not practical given the limitations of PCR multiplexing. Furthermore, the occurrence of horizontal transfer between unrelated STs of the *cps* cluster responsible for capsular antigen synthesis and export, implies that the K-type cannot be predicted reliably from the ST despite strong concordance between both methods (2).

K-type variation has been linked to the presence or absence of particular genes at the *cps* locus (1, 14, 27). The *cps* locus has a mosaic structure, with a group of six conserved genes (*galF, orf2, wzi, wza, wzb* and *wzc*) at its 5' end. Among these, gene *wzi* codes for outer

membrane protein Wzi involved in the attachment to the capsular polysaccharide to the outer membrane (6, 29). The objectives of this work were to evaluate the sequence polymorphism of gene *wzi* for strain typing and for inferring the serotype of *K. pneumoniae* clinical isolates.

Materials and Methods

Strains

Reference strains of the 77 K-types of the International serotyping scheme (24) were included. Strains of K-types K1 to K54 were obtained from Collection de l'Institut Pasteur. Strains from K-types above 54 were obtained from the WHO reference centre. Strain A1517 (27) was a gift of J.-T. Wang. 75 strains representing important K types or STs were gathered from our previous studies (2, 3, 5, 9, 11). One hundred and nineteen *K. pneumoniae* isolates from infected patients hospitalized in five University hospitals between 2008 and 2012 were included. Infections were community acquired (54%) or nosocomial (46%), as defined by isolation before or after 48 h of hospitalization, respectively. Clinical isolates were recovered from urine (n = 42, 36%), blood (n = 28, 24%), low respiratory tract (n = 19, 16%), abscesses (n = 9, 8%), peritonitis, cholangitis or other digestive system infections (n = 6), cerebrospinal fluid (n = 1) or other sources (n = 4).

DNA preparation

DNA templates were prepared by suspending a freshly grown colony in 200 microliters of purified water, heating at 94°C for 10 min, and submitting the extracts to minicentrifugation at 7,500 g for 5 min. Supernatants were stored at -20°C until use.

K-typing

K-typing was performed at the WHO Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella* by counter current immunoelectrophoresis (CCIE) using a modified version of the method described by Palfreyman (26). An extract was used as antigen instead of a whole cell suspension (23). The extract was only heated once for 1 h at 100°C before centrifugation. K-typing was performed blindly with respect to the results of the *wzi* sequencing.

rpoB sequence analysis

The sequence of a 501 base pairs (bp) internal portion of the *rpoB* gene was obtained as described (33).

PCR amplification and sequencing *of gene wzi*

PCR amplification of a 580 bp DNA fragment was achieved using primers wzi_for2 (GTG CCG CGA GCG CTT TCT ATC TTG GTA TTC C) and wzi_rev (GAG AGC CAC TGG TTC CAG AA[C or T] TT[C or G] ACC GC). Two µL of the DNA extract were used as template for PCR. The 50-µL PCR mixture contained 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl2, 0.2 µM each primer, 100 µM each deoxynucleoside triphosphate and 0.17 µL Taq polymerase (invitrogen, France). After denaturation at 94°C for 2 min, 30 cycles were performed (94°C, 30 sec; 55°C, 40 sec; 72°C, 30 sec), followed by a final elongation step (72°C, 5 min). Sanger sequencing was performed on both strands using the PCR primers. An internal portion of 447 bp, which was well covered by sequence chromatograms on both strands, was retained for sequence comparisons. This selected region corresponded to the 5' region of gene *wzi* starting at the ATG initiation codon. The template used for *wzi* characterization thus corresponds to the 149 N-terminal amino-acids of the outer membrane protein.

Data analysis

Forward and reverse chromatograms were assembled using BIONUMERICS v6.6 (Applied-Maths, Belgium). Phylogenetic trees were constructed using MEGA v5.1 (38) based on the neighbour-joining method and Jukes-Cantor distance. Simpson index of discrimination and the adjusted Wallace coefficient of concordance were computed at http://comparingpartitions.info.

Results and Discussion

Identification of the K-type reference strains

The International *Klebsiella* K typing scheme was constituted between 1916 and 1977 (21, 24). Subsequently, the taxonomy of *Klebsiella* has evolved with the proposal to reclassify three species into genus *Raoultella* and with description of the novel species *K. variicola* and *K. michiganensis* (12, 17, 31, 32). To our knowledge, the identification at the species level of the reference strains of the 77 K-types has not been re-evaluated at the light of these taxonomic changes. To better characterize the phylogenetic range of the *wzi* primers used herein (see below), we sequenced the *rpoB* gene of all K-type reference strains and compared them to *rpoB* sequences previously described for taxonomic type strains (10, 12, 25, 31-33). A portion of 501 nucleotides was sequenced from the 77 K-type reference strains. Based on the phylogenetic analysis of the *rpoB* gene sequence (**Supplementary Figure 1**), the 77 reference strains could be attributed to *K. pneumoniae* (n = 54), including three strains (K50, K60 and K80) corresponding to phylogenetic group KpII-B (15) and three strains (K4, K5 and K6) belonging to *K. pneumoniae* subsp. *ozaenae*. Other reference strains were attributed to *K. variicola* (K48, K49, K53, K54, K56, K57, K58 and K71), *K. oxytoca* (K26, K29, K74), *K. michiganensis* (K41, K59, K66, K70), *K.* (*Raoultella) planticola* (K35, K79), *K. (Raoultella) ornithinolytica* (K32, K44, K72) and *K. (Raoultella) terrigena* (K65, K67, K68, K69).

wzi sequencing of K-type reference strains

Gene *wzi* is one of the few genes present in the capsular polysaccharide cluster of all capsular types of *K. pneumoniae*. Alignment of previously published capsular polysaccharide cluster sequences (1, 14, 27, 44)(36) allowed us to identify within gene *wzi*, a region of high sequence variability flanked by conserved motifs suitable for primer design. The 78 K-type reference strains including strain A1517 (27), were analyzed by PCR using primers *wzi_for2* and *wzi_rev*. For 64 (80%) reference strains, a PCR product of the expected size (580 bp) was obtained. The 14 reference strains that were PCR negative included the *K. oxytoca*, *K. michiganensis* except K59, and *K. terrigena* strains. In contrast, PCR amplification was negative for only three (K33, K40 and K82) *K. pneumoniae,* and none of *K. variicola*, *K. planticola* and *K. ornithinolytica* K-type reference strains. Negative PCR results could be due to priming site sequence variation or other alterations of the *wzi* gene. Note that PCR

amplification of *Escherichia coli* isolates that harbor group 1 capsule gene clusters can be positive, given the sequence similarity observed with *Klebsiella* clusters (30).

The *wzi* sequences obtained in this study are provided as **Supplementary Material 1.** Sequencing of the 64 *wzi* PCR products of the reference strains distinguished 58 *wzi* alleles (**Supplementary Table 1**). Reference strains of all K-types had distinct *wzi* sequences, with the following exceptions: K46 and K61, K14 and K23, and the group of K15, K17, K50, K51 and K52. To our knowledge, these nine serotypes were not associated with severe infections or multidrug resistance.

Phylogenetic analysis showed three main branches (**Figure 1**). The first branch (branch A) corresponded to all but one *K. pneumoniae* reference strains, and also included *K. variicola* reference strains. Strain A1517 had a unique *wzi* sequence, which clustered close to the K54 reference strain. Branch B included the *wzi* sequences from *K. planticola* and *K. ornithinolytica* (K32, K35, K44, K72 and K79). Finally, branch C comprised solely the K42 reference, which differed from those of group B by 14% on average. The *wzi* sequences of groups B and C diverged from group A by 22% on average.

It was noticeable that *wzi* sequences of serotypes K1 and K4 were very closely related, differing by a single nucleotide position (*wzi*-128 and *wzi*-127, **Figure 1 and Supplementary Material 1**). This result is consistent with the similarity of the capsular polysaccharide (*cps*) clusters of these two serotypes (3). Serotype K4 is a hallmark of *K. pneumoniae* subsp. *ozaenae*, which is associated with the chronic condition ozena. Other pairs of closely related *wzi* sequences were (1) K24 and K45; (2) K7 and K39; and (3) K8 and K19 (**Figure 1 and Supplementary Material 1**).

***wzi* sequencing of multiple strains of medically important K-types**

In order to determine the diversity of *wzi* allele sequences among strains of the same K-type, we sequenced *wzi* for a total of 75 strains of previously determined K-type. These included 43 multiple isolates of serotypes known to be associated with rhinoscleroma and ozaenae (K3, K4), with liver abscess and other community-acquired invasive infections (K1, K2, K5, K20, K54 and K57) and with multidrug resistance (ST15). **Table S1** provides details on the correspondence between K-types, STs and *wzi* alleles.

Four K1 isolates including strain NTUH-K2044, were included. These four isolates belong to ST23, the liver-abscess associated clone (2, 40), whereas the reference strain of serotype K1 (CIP 52.208 = A5054) belongs to the non-virulent clone ST82 (2). The four ST23 K1 isolates had identical *wzi* sequences (*wzi*-1), which differed by two nucleotides from

CIP 52.208 (*wzi*-128; **Supplementary Material 1)**. These results suggest that *wzi* sequencing could distinguish ST23 strains.

Seven K2 isolates were sequenced at *wzi*. These isolates were selected to represent ST86 and ST380 associated with severe infections (9) and the non-virulent ST14 clone (2). These K2 isolates also differed by their *cps* PCR-RFLP pattern, C-patterns C2b to C2e (3). Nevertheless, the seven *wzi* sequences turned out to be totally identical (allele *wzi*-2). Surprisingly, this *wzi* sequence was very distinct (27 nt differences out of 447, 6.0%) from the *wzi* sequence (allele *wzi*-4) of the K2 reference strain B5055 (= CIP 52.145). The seven *wzi*-2 isolates differed by a single nucleotide difference from eight isolates of serotype K24 and the reference strain of this serotype, which were identical (allele *wzi*-24). The K24 isolates belonged to ST15, a widely distributed multiresistant clone of *K. pneumoniae* (2, 8, 19). Therefore, *wzi* sequencing could help in the identification of the international clone ST15 of serotype K24.

Six K3 isolates were included. These included two *K. pneumoniae* subsp. *rhinoscleromatis* strains of ST67, and four *K. pneumoniae* subsp. *pneumoniae* strains of ST8, 13, 71 and 153. These isolates represented all known K3 variant C-patterns, C3b to C3d (3). Remarkably, all six isolates clustered in a single branch together with the K3 reference strain CIP52.146 (ST3). This branch was specific to K3 strains and differed by more than 4.5% from any other *wzi* sequence.

Three K4 strains of *K. pneumoniae* subsp. *ozaenae* were included. Their *wzi* sequence (*wzi*-129) clustered in a tight branch (99.7%) with the K4 reference strain (CIP 52.211 = D5050) and with the K1 reference strain. This cluster was closely related (99.34%) from *wzi*-1, the K1 ST23 *wzi* sequence.

Four K5 isolates were included and had an identical *wzi* sequence (allele *wzi*-5) to that of the K5 reference strain. The *wzi*-5 sequence differed by any other *wzi* sequence by more than 5%, underlining its distinctness.

We included three K20 isolates, one of which showed a *wzi* sequence identical to that of the K20 reference strain, whereas the two others had very distinct *wzi* alleles (31 nt differences, 6.9%). Both *wzi* alleles were unique to K20 isolates.

Of six K54 isolates, two isolates had *wzi*-115, which differed by 7% from the *wzi* sequence (*wzi*-54) of the K54 reference. The *wzi* sequence (*wzi*-66) of the four other isolates was distinct, with 4% divergence to any other *wzi* allele.

Two isolates of K57 had identical *wzi* sequences (*wzi*-77), which were distinct from the *wzi* sequence of K57 reference strain identified as *K. variicola*. Allele *wzi*-77 was closely

related (1 nucleotide difference, 99.78%) to the *wzi* sequence of the reference strain of K34 (**Figure 1 and Supplementary Material 1**).

***wzi* sequencing of additional strains of known K-types**

In order to supplement the database of reference *wzi* sequences, we included 32 strains that represented a variety of C-patterns (3) and were of known K-type. These strains showed 25 distinct *wzi* alleles, 17 (68%) of which were novel compared to the 109 strains described above, thus supplementing the database of reference *wzi* sequences. The *wzi* alleles from these strains and those of the medically important K-types described in the previous paragraph are given in **Supplementary Table 1** in column "Previously characterized strains".

***wzi* sequencing and K-typing of prospective clinical isolates**

We aimed to compare the ability of *wzi* sequencing and K-typing to characterize clinical isolates. To this purpose, 119 isolates from community-acquired and nosocomial infections prospectively collected in five university hospitals were included. Of the 119 clinical isolates, 23 (19%) could not be typed by the serological method, resulting in a typeability of 81% for K-typing. There were 35 distinct K-types among the 96 typeable isolates, the most frequent ones being K2 (n = 15), K14 (n = 11), K1 (n = 9), and K51 (n = 7). Simpson's discrimination index was 94.4% (95% confidence interval, 92.3-96.6%). The wzi alleles obtained are described in the column "Clinical isolates" of **Supplementary Table 1**.

Of the 119 isolates, only two isolates (T2, non typeable; and T17, a K2 isolate) could not be PCR amplified for *wzi*; the 117 remaining isolates were successfully sequenced. Therefore, the typeability of *wzi* sequencing was 98.3%. A total of 72 distinct *wzi* alleles were found among the 117 clinical isolates, with a Simpson's discrimination index of 98.1% (97.1-99.0). For the 95 isolates for which both the *wzi* sequence and the K-type could be determined, there were 35 distinct K-types (Simpson, 94.6% ; CI 0.926 - 0.967) and 55 *wzi* alleles (Simpson: 97.4%; CI 96.1 – 98.7%). Therefore, *wzi* sequencing was significantly more discriminatory than K-typing, in addition to being able to type a much higher proportion of clinical isolates.

Of the 117 isolates, 61 (52%) had a *wzi* sequence that was identical to an allele in the reference database, corresponding to 25 distinct *wzi* alleles. Seven (11%) of them were non-typeable by K-typing and the remaining isolates were split into 20 distinct K-types, among which K1 (n = 9), K14 (n = 9), K2 (n = 8) and K51 (n = 7) were the most common. There were six (10%) isolates for which *wzi* sequencing and K-typing disagreed. First, one isolate

identified as K29 had a *wzi* sequence identical to that of the K12 reference strain; this discrepancy was confirmed upon retesting. Second, five isolates had allele *wzi*-64, which was the allele of K64 reference strain, whereas initial K-typing (performed blindly) suggested they were K14 (four strains) or K22 (one isolate). However, as all five *wzi*-64 isolates consistently reacted with K14 antiserum or with K64 antiserum (two isolates) upon retesting, we consider *wzi*-64 as indicative of K14 or K64 (thus denoted K14.K64).

The remaining 56 isolates for which the *wzi* sequence was not matched to the reference database represented 48 novel *wzi* alleles. Phylogenetic analysis showed that they all fell in branch A. Only six alleles were represented by more than one isolate (2: 3 isolates; 4: 2 isolates), showing that they were all uncommon. K-typing showed that these isolates included 15 (26%) non-typeable isolates and 22 distinct K-types. Importantly, among those, there were six K2 isolates. Their *wzi* sequence was very similar to the K2 *wzi* reference sequences *wzi*-4 (four isolates, 2 or 3 nucleotide differences) or *wzi*-2 (2 isolates, single nucleotide difference), showing that K2 isolates can be recognized despite their sequence heterogeneity.

Considering the 48 novel *wzi* alleles as novel reference sequences of their respective K-type for future studies, the overall concordance among the two methods was estimated using the Wallace coefficient adjusted for agreement due to chance alone (34). The resulting directional agreement of K-type classification once knowing the *wzi* allele was 94.4% (CI: 91.1 – 97.7%). This result indicates that K-type can be predicted from *wzi* allele in a large majority of cases. Conversely, knowing the K-type predicted the *wzi* allele only weakly (0.441; CI: 0.335-0.547), as expected by the higher diversity of *wzi* alleles as compared to K-types.

Conclusions

This study reports on the development of a rapid method, *wzi* sequencing, to predict the K-type of *K. pneumoniae* isolates. PCR amplification was negative for K-type reference strains of several *Klebsiella* species but was positive for most strains of *K. pneumoniae* and *K. variicola*, which are the most frequent *Klebsiella* species encountered among clinical isolates (4, 5, 15, 28). The *wzi* sequence of most K-type reference strains could be distinguished, with few exceptions. Besides, *wzi* sequences from multiple strains of important K-types fell into recognizable clusters. The typeability of *wzi* sequencing was 98.3%, which is much higher than allele-specific PCR assays, which specifically target serotypes associated with liver abscess or other community-acquired infections (13, 27, 45, 46). Based on the

present database, the ability to predict the K-type from the *wzi* sequence was estimated at nearly 95%. Furthermore, it should be noted that K-type reference strains are old and likely to represent poorly the current population of *K. pneumoniae* clinical isolates. Therefore, it can be expected that typeability will improve by including additional clinical isolates into the reference database. Given that isolates with distinct K-types can have closely related *wzi* alleles, we recommend that isolates with novel alleles should be K-typed and incorporated into the reference database. We are currently developing a BIGSdb genome database (20) into which the wzi/K-type correspondence will be incorporated and continuously updated. This will allow automated extraction of *wzi* allele information and deduced K-type from genomic sequences obtained using high-throughput sequencing technologies. Unfortunately, the K-type cannot be always deduced based on *wzi* given sequence identity of some strains with distinct K types. It is expected that sequence information from other loci of the capsular polysaccharide synthesis cluster could be used to resolve these cases. Nevertheless, *wzi* sequencing represents a clear improvement compared to molecular methods that are currently available for approximating the capsular serotype of *K. pneumoniae* isolates. Irrespective of K-typing correspondence, *wzi* sequencing also represents a powerful strain typing method *per se*, given its high discrimination and typeability achievements. It is our hope that this method will contribute to better understand the epidemiology of groups of strains of particular medical importance, including hypervirulent and multidrug resistant clonal groups, and will at the same time allow linking future work to the large historical corpus of knowledge accumulated based on classical K-typing.

Acknowledgments

We thank Jin-Town Wang for providing strain A1517 and Karen A. Krogfelt and Flemming Scheutz for helpful advice.

Funding

This work received financial support from a grant from Assistance Publique-Hôpitaux de Paris, contract CRC no. 08007 (KPath) and from the French Government's Investissement d'Avenir program, Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (grant n°ANR-10-LABX-62-IBEID).

References

1. **Arakawa, Y., R. Wacharotayankun, T. Nagatsuka, H. Ito, N. Kato, and M. Ohta.** 1995. Genomic organization of the *Klebsiella pneumoniae cps* region responsible for serotype K2 capsular polysaccharide synthesis in the virulent strain Chedid. J Bacteriol **177:**1788-96.
2. **Brisse, S., C. Fevre, V. Passet, S. Issenhuth-Jeanjean, R. Tournebize, L. Diancourt, and P. Grimont.** 2009. Virulent clones of *Klebsiella pneumoniae*: identification and evolutionary scenario based on genomic and phenotypic characterization. PLoS One **4:**e4982.
3. **Brisse, S., S. Issenhuth-Jeanjean, and P. A. Grimont.** 2004. Molecular serotyping of *Klebsiella* species isolates by restriction of the amplified capsular antigen gene cluster. J Clin Microbiol **42:**3388-98.
4. **Brisse, S., T. van Himbergen, K. Kusters, and J. Verhoef.** 2004. Development of a rapid identification method for *Klebsiella pneumoniae* phylogenetic groups and analysis of 420 clinical isolates. Clin. Microbiol. Infect. **10:**942-945.
5. **Brisse, S., and J. Verhoef.** 2001. Phylogenetic diversity of *Klebsiella pneumoniae* and *Klebsiella oxytoca* clinical isolates revealed by randomly amplified polymorphic DNA, *gyrA* and *parC* genes sequencing and automated ribotyping. Int J Syst Evol Microbiol **51:**915-24.
6. **Bushell, S. R., I. L. Mainprize, M. A. Wear, H. Lou, C. Whitfield, and J. H. Naismith.** 2013. Wzi is an outer membrane lectin that underpins group 1 capsule assembly in Escherichia coli. Structure **21:**844-53.
7. **Cryz, S. J., Jr., Mortimer, P. M., Mansfield, V., and Germanier, R.** 1986. Seroepidemiology of *Klebsiella pneumoniae* bacteremic isolates and implications for vaccine development. Journal of Clinical Microbiology **23:**687-690.
8. **Damjanova, I., A. Toth, J. Paszti, G. Hajbel-Vekony, M. Jakab, J. Berta, H. Milch, and M. Fuzi.** 2008. Expansion and countrywide dissemination of ST11, ST15 and ST147 ciprofloxacin-resistant CTX-M-15-type beta-lactamase-producing *Klebsiella pneumoniae* epidemic clones in Hungary in 2005--the new 'MRSAs'? J Antimicrob Chemother **62:**978-85.
9. **Decré, D., C. Verdet, A. Emirian, T. Le Gourrierec, J. C. Petit, G. Offenstadt, E. Maury, S. Brisse, and G. Arlet.** 2011. Emerging severe and fatal infections due to *Klebsiella pneumoniae* in two university hospitals in France. J Clin Microbiol **49:**3012-4.
10. **Deletoile, A., D. Decre, S. Courant, V. Passet, J. Audo, P. Grimont, G. Arlet, and S. Brisse.** 2009. Phylogeny and identification of *Pantoea* species and typing of *Pantoea agglomerans* strains by multilocus gene sequencing. J Clin Microbiol **47:**300-10.
11. **Diancourt, L., V. Passet, J. Verhoef, P. A. Grimont, and S. Brisse.** 2005. Multilocus sequence typing of *Klebsiella pneumoniae* nosocomial isolates. J Clin Microbiol **43:**4178-82.
12. **Drancourt, M., C. Bollet, A. Carta, and P. Rousselier.** 2001. Phylogenetic analyses of *Klebsiella* species delineate *Klebsiella* and *Raoultella* gen. nov., with description of *Raoultella ornithinolytica* comb. nov., *Raoultella terrigena* comb. nov. and *Raoultella planticola* comb. nov. Int J Syst Evol Microbiol **51:**925-32.
13. **Fang, C. T., S. Y. Lai, W. C. Yi, P. R. Hsueh, K. L. Liu, and S. C. Chang.** 2007. *Klebsiella pneumoniae* genotype K1: an emerging pathogen that causes septic ocular or central nervous system complications from pyogenic liver abscess. Clin Infect Dis **45:**284-93.
14. **Fevre, C., V. Passet, A. Deletoile, V. Barbe, L. Frangeul, A. S. Almeida, P. Sansonetti, R. Tournebize, and S. Brisse.** 2011. PCR-based identification of

Klebsiella pneumoniae subsp. *rhinoscleromatis*, the agent of rhinoscleroma. PLoS Negl Trop Dis **5:**e1052.

15. **Fevre, C., V. Passet, F. X. Weill, P. A. Grimont, and S. Brisse.** 2005. Variants of the *Klebsiella pneumoniae* OKP chromosomal beta-lactamase are divided into two main groups, OKP-A and OKP-B. Antimicrob Agents Chemother **49:**5149-52.
16. **Fung, C. P., F. Y. Chang, S. C. Lee, B. S. Hu, B. I. Kuo, C. Y. Liu, M. Ho, and L. K. Siu.** 2002. A global emerging disease of *Klebsiella pneumoniae* liver abscess: is serotype K1 an important factor for complicated endophthalmitis? Gut **50:**420-4.
17. **Grimont, P. A. D., and F. Grimont.** 2005. Genus *Klebsiella*, p. 685-693. *In* D. J. Brenner, N. R. Krieg, and J. T. Staley (ed.), Bergey's manual of Systematic Bacteriology. Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria. Springer-Verlag, New York.
18. **Hansen, D. S., R. Skov, J. V. Benedi, V. Sperling, and H. J. Kolmos.** 2002. *Klebsiella* typing: pulsed-field gel electrophoresis (PFGE) in comparison with O:K-serotyping. Clin Microbiol Infect **8:**397-404.
19. **Hu, L., Q. Zhong, J. Tu, Y. Xu, Z. Qin, C. Parsons, B. Zhang, X. Hu, L. Wang, F. Yu, and J. Pan.** 2013. Emergence of blaNDM-1 among *Klebsiella pneumoniae* ST15 and novel ST1031 clinical isolates in China. Diagn Microbiol Infect Dis **75:**373-6.
20. **Jolley, K. A., and M. C. Maiden.** 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics **11:**595.
21. **Julianelle, L. A.** 1926. Immunological relationships of encapsulated and capsule-free strains of *Encapsulatus pneumoniae* (Friedlander's bacillus). J Exp Med **44:**683-696.
22. **Mizuta, L. G., Ohta, M., Mori, M. Hasegawa, T., Nakashima, I., and Kato, N.** 1983. Virulence for mice of *Klebsiella* strains belonging to the O1 group: relationship to their capsular (K) types. Infect Immun **40:**56-61.
23. **Orskov, F., and I. Orskov.** 1972. Immunoelectrophoretic patterns of extracts from *Escherichia coli* O antigen test strains O1 to O157 examinations in homologous OK sera. Further comments on the classification of *Escherichia* K antigens. Acta Pathol Microbiol Scand B Microbiol Immunol **80:**905-10.
24. **Orskov, I., and F. Orskov.** 1984. Serotyping of *Klebsiella*. Methods Microbiol **14:**117-133.
25. **Paauw, A., M. P. Caspers, F. H. Schuren, M. A. Leverstein-van Hall, A. Deletoile, R. C. Montijn, J. Verhoef, and A. C. Fluit.** 2008. Genomic diversity within the *Enterobacter cloacae* complex. PLoS ONE **3:**e3018.
26. **Palfreyman, J. M.** 1978. *Klebsiella* serotyping by counter-current immunoelectrophoresis. Journal of Hygiene (Cambridge) **81:**219-225.
27. **Pan, Y. J., H. C. Fang, H. C. Yang, T. L. Lin, P. F. Hsieh, F. C. Tsai, Y. Keynan, and J. T. Wang.** 2008. Capsular polysaccharide synthesis regions in *Klebsiella pneumoniae* serotype K57 and a new capsular serotype. J Clin Microbiol **46:**2231-40.
28. **Podschun, R., and U. Ullmann.** 1998. *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clin Microbiol Rev **11:**589-603.
29. **Rahn, A., K. Beis, J. H. Naismith, and C. Whitfield.** 2003. A novel outer membrane protein, Wzi, is involved in surface assembly of the *Escherichia coli* K30 group 1 capsule. J Bacteriol **185:**5882-90.
30. **Rahn, A., J. Drummelsmith, and C. Whitfield.** 1999. Conserved organization in the cps gene clusters for expression of *Escherichia coli* group 1 K antigens: relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. J Bacteriol **181:**2307-13.

31. **Rosenblueth, M., L. Martínez, J. Silva, and E. Martínez-Romero.** 2004. *Klebsiella variicola*, A Novel Species with Clinical and Plant-Associated Isolates. System Appl Microbiol **27:**27-35.
32. **Saha, R., C. E. Farrance, B. Verghese, S. Hong, and R. S. Donofrio.** 2013. *Klebsiella michiganensis* sp. nov., a new bacterium isolated from a tooth brush holder. Curr Microbiol **66:**72-8.
33. **Salerno, A., A. Deletoile, M. Lefevre, I. Ciznar, K. Krovacek, P. Grimont, and S. Brisse.** 2007. Recombining population structure of *Plesiomonas shigelloides (Enterobacteriaceae)* revealed by multilocus sequence typing. J Bacteriol **189:**7808-7818.
34. **Severiano, A., F. R. Pinto, M. Ramirez, and J. A. Carrico.** 2011. Adjusted Wallace coefficient as a measure of congruence between typing methods. J Clin Microbiol **49:**3997-4000.
35. **Shon, A. S., R. P. Bajwa, and T. A. Russo.** 2013. Hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*: a new and dangerous breed. Virulence **4:**107-18.
36. **Shu, H. Y., C. P. Fung, Y. M. Liu, K. M. Wu, Y. T. Chen, L. H. Li, T. T. Liu, R. Kirby, and S. F. Tsai.** 2009. Genetic diversity of capsular polysaccharide biosynthesis in *Klebsiella pneumoniae* clinical isolates. Microbiology.
37. **Struve, C., M. Bojer, E. M. Nielsen, D. S. Hansen, and K. A. Krogfelt.** 2005. Investigation of the putative virulence gene magA in a worldwide collection of 495 *Klebsiella* isolates: *magA* is restricted to the gene cluster of *Klebsiella pneumoniae* capsule serotype K1. J Med Microbiol **54:**1111-3.
38. **Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol **28:**2731-9.
39. **Toivanen, P., D. S. Hansen, F. Mestre, L. Lehtonen, J. Vaahtovuo, M. Vehma, T. Mottonen, R. Saario, R. Luukkainen, and M. Nissila.** 1999. Somatic serogroups, capsular types, and species of fecal *Klebsiella* in patients with ankylosing spondylitis. J Clin Microbiol **37:**2808-12.
40. **Turton, J. F., H. Englender, S. N. Gabriel, S. E. Turton, M. E. Kaufmann, and T. L. Pitt.** 2007. Genetically similar isolates of *Klebsiella pneumoniae* serotype K1 causing liver abscesses in three continents. J Med Microbiol **56:**593-7.
41. **Vimont, S., B. Mnif, C. Fevre, and S. Brisse.** 2008. Comparison of PFGE and multilocus sequence typing for analysis of *Klebsiella pneumoniae* isolates. J Med Microbiol **57:**1308-10.
42. **Wang, J. H., Y. C. Liu, S. S. Lee, M. Y. Yen, Y. S. Chen, J. H. Wang, S. R. Wann, and H. H. Lin.** 1998. Primary liver abscess due to *Klebsiella pneumoniae* in Taiwan. Clin Infect Dis **26:**1434-8.
43. **Woodford, N., J. F. Turton, and D. M. Livermore.** 2011. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. FEMS Microbiol Rev **35:**736-55.
44. **Wu, K. M., L. H. Li, J. J. Yan, N. Tsao, T. L. Liao, H. C. Tsai, C. P. Fung, H. J. Chen, Y. M. Liu, J. T. Wang, C. T. Fang, S. C. Chang, H. Y. Shu, T. T. Liu, Y. T. Chen, Y. R. Shiau, T. L. Lauderdale, I. J. Su, R. Kirby, and S. F. Tsai.** 2009. Genome sequencing and comparative analysis of *Klebsiella pneumoniae* NTUH-K2044, a strain causing liver abscess and meningitis. J Bacteriol **191:**4492-501.
45. **Yeh, K. M., A. Kurup, L. K. Siu, Y. L. Koh, C. P. Fung, J. C. Lin, T. L. Chen, F. Y. Chang, and T. H. Koh.** 2007. Capsular serotype K1 or K2, rather than *magA* and *rmpA*, is a major virulence determinant for *Klebsiella pneumoniae* liver abscess in Singapore and Taiwan. J Clin Microbiol **45:**466-71.

46. **Yu, V. L., D. S. Hansen, W. C. Ko, A. Sagnimeni, K. P. Klugman, A. von Gottberg, H. Goossens, M. M. Wagener, and V. J. Benedi.** 2007. Virulence characteristics of *Klebsiella* and clinical manifestations of *K. pneumoniae* bloodstream infections. Emerg Infect Dis **13:**986-93.

47. **Yu, W. L., C. P. Fung, W. C. Ko, K. C. Cheng, C. C. Lee, and Y. C. Chuang.** 2007. Polymerase chain reaction analysis for detecting capsule serotypes K1 and K2 of *Klebsiella pneumoniae* causing abscesses of the liver and other sites. J Infect Dis **195:**1235-6; author reply 1236.

Figure 1. Dendrogram of 135 *wzi* alleles.

The tree was built using BioNumerics v6.6 based on the uncorrected *p*-distance. Each sequence corresponds to a distinct *wzi* allele, as indicated. The allele number is followed by the corresponding capsular type(s). Dots separating several K-types indicate cross-reactions, whereas successions of distinct K-types with no dot (e.g., K15K17K50K51K52) indicate that the *wzi* allele was found in strains reacting individually with one of these distinct K-types. The three main branches A, B and C are labelled. K P-2008 indicates the novel serotype described by Pan et al., 2008. Bootstrap values of remarkable nodes are given.

Figure 1