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

Strains. Reference strains of the 77 K types of the international serotyping scheme (12) were included. Strains of the K. pneumoniae K types K1 to K54 were obtained from the Collection de l’Institut Pasteur. Strains from K types higher than 54 were obtained from the WHO reference center. K. pneumoniae strain A1517 (17) was a gift from J.-T. Wang. Seventy-five strains representing important K types or STs were gathered from our previous studies (1, 5, 16, 24, 29). One hundred 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 their isolation before or after 48 h of hospitalization, respectively. The clinical isolate specimens were recovered from urine (n = 42 [36%]), blood (n = 28 [24%]), lower 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 μl of purified water, heating at 94°C for 10 min, and submitting the extracts to microcentrifugation at 7,500 × g for 5 min. The 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 (30). An extract was used as the antigen instead of a whole-cell suspension (31). The extract was heated only 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-bp internal portion of the rpoB gene was obtained as described previously (32).

PCR amplification and sequencing of wzi gene. PCR amplification of a 580-bp DNA fragment was achieved using the primers wzi_for2 (GTG CGG GGA CCT TTC 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 microliters of the DNA extract was used as the template for PCR. The 30-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 (each 94°C for 30 s, 55°C for 40 s, and 72°C for 40 s), followed by a final elongation step (72°C for 5 min). The supernatants were stored at −20°C until use.

Data analysis. Forward and reverse chromatograms were assembled using BioNumerics version 6.6 (Applied Maths, Belgium). Phylogenetic trees were constructed using MEGA version 5.1 (33) based on the neighbor-joining method and Jukes-Cantor distance. The Simpson index of discrimination and the adjusted Wallace coefficient of concordance were computed using the Comparing Partitions tool (see http://comparingpartitions.info).

RESULTS AND DISCUSSION

Identification of the K-type reference strains. The international Klebsiella K-typing scheme was developed between 1916 and 1977 (11, 12). Subsequently, the taxonomy of Klebsiella has evolved with the proposal to reclassify three species into the genus Raoultella and with descriptions of the novel species Klebsiella varicola and Klebsiella michiganensis (34–37). To our knowledge, identification of the reference strains of the 77 K types at the species level has not been reevaluated in light of these taxonomic changes. To better characterize the phylogenetic range of the wzi primers used in this study (see below), we sequenced the rpoB gene of each K-type reference strain and compared them to the rpoB sequences previously described for taxonomic type strains (32, 34, 36–39). A portion of 501 nucleotides was sequenced from the 77 K-type reference strains. Based on the phylogenetic analysis of the rpoB gene sequence (see Fig. S1 in the supplemental material), the 77 reference strains were attributed to K. pneumoniae (n = 54), including three strains (K50, K60, and K80) corresponding to phylogenetic group KpII-B (40) 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), Klebsiella oxytoca (K26, K29, and K74), K. michiganensis (K41, K59, K66, and K70), K. (Raoultella) planticola (K35 and K79), K. (Raoultella) ornithinolytica (K32, K44, and K72), and K. (Raoultella) terrigena (K65, K67, K68, and K69).

wzi sequencing of K-type reference strains. wzi is one of the few genes present in the capsular polysaccharide cluster of all capsular types of K. pneumoniae. The alignment of previously published capsular polysaccharide cluster sequences (17, 25, 26, 41, 42) allowed us to identify within the wzi gene a region of high sequence variability flanked by conserved motifs that are suitable for primer design. The 78 K-type reference strains, including strain A1517 (17), were analyzed by PCR using the 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 strains and none of the K. variicola, K. planticola, and K. ornithinolytica K-type reference strains. Negative PCR results might be due to priming site sequence variation or other alterations of the wzi gene. Note that PCR amplification of the Escherichia coli isolates that harbor group 1 capsule gene clusters can be positive, given the sequence similarity observed with Klebsiella clusters (43).

The wzi sequences obtained in this study are provided in File S3 in the supplemental material. Sequencing of the 64 wzi PCR products of the reference strains distinguished 58 wzi alleles (see Table S1 in the supplemental material). The 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.

A phylogenetic analysis showed three main branches (Fig. 1). The first branch (branch A) corresponded to all but one K. pneumoniae reference strain 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 (strains K32, K35, K44, K72, and K79). Finally, branch C comprised the K42 reference strain only, which differed from the strains of group B by 14%, on average. The wzi sequences of groups B and C diverged from group A by 22%, on average.

It is significant that the wzi sequences of serotypes K1 and K4 are very closely related, differing by a single nucleotide position (wzi-128 and wzi-127) (Fig. 1; see also File S3 in the supplemental material). This result is consistent with the similarity of the capsular polysaccharide (cps) clusters of these two serotypes (16). 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 (i) K24 and K45, (ii) K7 and K39, and (iii) K8 and K19 (Fig. 1; see also File S3 in the supplemental material).

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

Four *K. pneumoniae* K1 isolates, including strain NTUH-K2044, were included. These four isolates belong to ST23, the liver abscess-associated clone (8, 24), whereas the reference strain of serotype K1 (CIP 52.208/H11005A5054) belongs to the nonvirulent clone ST82 (24). The four ST23 K1 isolates had identical \( wzi \) sequences (\( wzi-1 \)), which differed by two nucleotides from that of CIP 52.208 (\( wzi-128 \)) (see File S3 in the supplemental material). These results suggest that \( wzi \) sequencing can distinguish ST23 strains.

Seven *K. pneumoniae* K2 isolates were sequenced at the \( wzi \) gene. These isolates were selected to represent ST86 and ST380, which are associated with severe infections (1), and the nonvirulent ST14 clone (24). These K2 isolates also differed by their cps PCR-RFLP patterns, which were C patterns C2b to C2e (16). Nevertheless, the seven \( wzi \) sequences turned out to be totally identical (allele \( wzi-2 \)). Surprisingly, this \( wzi \) sequence was very distinct (27 nucleotide [nt] differences out of 447 [6.0%]) from the \( wzi \) sequence (allele \( wzi-4 \)) of the *K. pneumoniae* K2 reference strain B5055 (= CIP 52.145). The seven \( wzi-2 \) isolates differed by a single nucleotide from eight isolates of serotype K24 and the reference strain of this serotype, which were identical (allele \( wzi-24 \)). The K24 isolates belong to ST15, a widely distributed multiresistant clone of *K. pneumoniae* (9, 10, 24). Therefore, \( wzi \) sequencing might 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, ST13, ST71, and ST153. These isolates represented all known K3 variant C patterns, C3b to C3d (16). Remarkably, all six isolates clustered in a single branch together with the K3 reference strain CIP 52.146 (ST3). This branch was specific to K3 strains and differed by >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%) to \( wzi-1 \), the K1 ST23 \( wzi \) sequence.

**FIG 1** Dendrogram of 135 \( wzi \) alleles. The tree was built using BioNumerics version 6.6 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 labeled. K P-2008 indicates the novel serotype described by Pan et al. (17). Bootstrap values of the remarkable nodes are given.
Four K5 isolates were included and had an identical \textit{wzi} sequence (allele \textit{wzi-5}) to that of the K5 reference strain. The \textit{wzi-5} sequence differed from any other \textit{wzi} sequence by $>5\%$, underlining its distinctness.

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

Of six K34 isolates, two isolates had \textit{wzi-115}, which differed by 7\% from the \textit{wzi} sequence (\textit{wzi-54}) of the K54 reference. The \textit{wzi} sequence (\textit{wzi-66}) of the four other isolates was distinct, with 4\% divergence to any other \textit{wzi} allele.

Two isolates of K57 had identical \textit{wzi} sequences (\textit{wzi-77}), which were distinct from the \textit{wzi} sequence of the K57 reference strain identified as \textit{K. variicola}. Allele \textit{wzi-77} was closely related (1 nucleotide difference [99.78\%]) to the \textit{wzi} sequence of the reference strain of K34 (Fig. 1; also see File S3 in the supplemental material).

\textit{wzi} sequencing of additional strains of known K types. In order to supplement the database of the reference \textit{wzi} sequences, we included 32 strains that represented a variety of C patterns (16) and whose \textit{K} types were known. These strains showed 25 distinct \textit{wzi} alleles, 17 (68\%) of which were novel compared to the 109 strains described above, thus supplementing the database of reference \textit{wzi} sequences. The \textit{wzi} alleles from these strains and those of the medically important \textit{K} types described in the previous paragraph are given in Table S1 in the supplemental material.

\textit{wzi} sequencing and K typing of prospective clinical isolates. We aimed to compare the abilities of \textit{wzi} sequencing and \textit{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 rate of 81\% for \textit{K} typing. There were 35 distinct \textit{K} types among the 96 typeable isolates, the most frequent ones being \textit{K2} ($n = 15$), \textit{K14} ($n = 11$), \textit{K1} ($n = 9$), and \textit{K51} ($n = 7$). Simpson’s discrimination index was 94.4\% (95\% confidence interval, 92.3 to 96.6\%). The \textit{wzi} alleles obtained are described in Table S1 in the supplemental material.

Of the 119 isolates, only two isolates (T2 [nontypeable] and T17 [K2 isolate]) could not be PCR amplified for \textit{wzi}; the 117 remaining isolates were successfully sequenced. Therefore, the typeability rate of \textit{wzi} sequencing was 98.3\%. A total of 72 distinct \textit{wzi} alleles were found among the 117 clinical isolates, with a Simpson’s discrimination index of 98.1\% (confidence interval [CI], 97.1 to 99.0\%). For the 95 isolates for which both the \textit{wzi} sequence and the \textit{K} type could be determined, there were 35 distinct \textit{K} types (Simpson index, 94.6\%; CI, 92.6 to 96.7\%) and 55 \textit{wzi} alleles (Simpson index, 97.4\%; CI, 96.1 to 98.7\%). Therefore, \textit{wzi} sequencing was significantly more discriminatory and was able to type a much higher proportion of clinical isolates.

Of the 117 isolates, 61 (52\%) had a \textit{wzi} sequence that was identical to an allele in the reference database, corresponding to 25 distinct \textit{wzi} alleles. Seven (11\%) of them were nontypeable by \textit{K} typing, and the remaining isolates were split into 20 distinct \textit{K} types, among which \textit{K1} ($n = 9$), \textit{K14} ($n = 9$), \textit{K2} ($n = 8$), and \textit{K51} ($n = 7$) were the most common. There were six (10\%) isolates for which \textit{wzi} sequencing and \textit{K} typing had discrepant results. First, one isolate identified as K29 had a \textit{wzi} sequence identical to that of the K12 reference strain; this discrepancy was confirmed upon retesting. Second, five isolates had allele \textit{wzi-64}, which was the allele of K64 reference strain, whereas initial \textit{K} typing (performed blindly) suggested they were K14 (four strains) or K22 (one isolate). However, as all five \textit{wzi-64} isolates consistently reacted with K14 antiserum or with K64 antiserum (two isolates) upon retesting, we consider \textit{wzi-64} to be indicative of K14 or K64 (thus denoted K14,K64).

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

Considering the 48 novel \textit{wzi} alleles to be novel reference sequences of their respective \textit{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 (44). The resulting directional agreement of \textit{K}-type classification once knowing the \textit{wzi} allele was 94.4\% (CI, 91.1 to 97.7\%). This result indicates that the \textit{K} type can be predicted from a \textit{wzi} allele in a large majority of cases. Conversely, knowing the \textit{K} type predicted the \textit{wzi} allele only weakly (adjusted Wallace coefficient, 44.1\%; CI, 33.5 to 54.7\%), as was expected by the higher diversity of \textit{wzi} alleles than the \textit{K} types.

Conclusions. This study reports the development of a rapid method, \textit{wzi} sequencing, to predict the \textit{K} types of \textit{K. pneumoniae} isolates. PCR amplification was negative for \textit{K}-type reference strains of several \textit{Klebsiella} species but was positive for most strains of \textit{K. pneumoniae} and \textit{K. variicola}, which are the most frequently encountered \textit{Klebsiella} species in clinical isolates (29, 40, 45, 46). The \textit{wzi} sequences of most \textit{K}-type reference strains were distinguished, with few exceptions. Besides, \textit{wzi} sequences from multiple strains of important \textit{K} types fell into recognizable clusters. The typeability rate of \textit{wzi} sequencing was 98.3\%, which is much higher than that of allele-specific PCR assays, which specifically target serotypes associated with liver abscesses or other community-acquired infections (14, 17, 20, 21). Based on the present database, the ability to predict the \textit{K} type from the \textit{wzi} sequence was estimated at nearly 95\%. Furthermore, it should be noted that \textit{K}-type reference strains are old and are likely to represent the current population of \textit{K. pneumoniae} clinical isolates poorly. Therefore, it can be expected that typeability will improve by including additional clinical isolates in the reference database. Given that isolates with distinct \textit{K} types can have closely related \textit{wzi} alleles, we recommend that isolates with novel \textit{wzi} alleles be \textit{K} typed and incorporated into the reference database. We are currently developing a BIGSdb genome database (47) into which the \textit{wzi–K} type correspondence will be incorporated and continuously updated. This will allow for the automated extraction of \textit{wzi} allele information and the deduced \textit{K} type from genomic sequences obtained using high-throughput sequencing technologies. Unfortunately, the \textit{K} type cannot always be deduced based on \textit{wzi} given sequence identities of some strains with distinct \textit{K} types. It is expected that sequence information from other loci of the capsular polysaccha-
ride synthesis cluster could be used to resolve these cases. Nevertheless, wzi sequencing represents a clear improvement over the molecular methods that are currently available for approximating the capsular serotypes 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 a better understanding of the epidemiology of groups of strains of particular medical importance, including hypervirulent and multidrug-resistant clonal groups, and will at the same time allow for 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 their helpful advice. Financial support for this work was provided by a grant from Assistance Publique-Hôpitaux de Paris, contract CRC 08007 (KPath) and from the French government’s Investissement d’Avenir program, Laboratoire d’Excellence “Integrative Biology of Emerging Infectious Diseases” (grant ANR-10-LABX-62-IBEID).

REFERENCES


