Identification of *Klebsiella pneumoniae* complex members using MALDI-TOF mass spectrometry

Carla Rodrigues,[a] Virginie Passet,[a] Sylvain Brisse[a#]

[a] Biodiversity and Epidemiology of Bacterial Pathogens, Institut Pasteur, Paris, France.

Running title: *K. pneumoniae* phylogroups identified by MALDI-TOF MS

Keywords: *Klebsiella pneumoniae*, phylogroups, MALDI-TOF MS, identification, biomarkers

#Corresponding author

Sylvain Brisse

Biodiversity & Epidemiology of Bacterial Pathogens

Institut Pasteur

25 Rue du Docteur Roux

75015 Paris Cedex 15,

France

Telephone: +331 45 68 83 34

E-mail: sylvain.brisse@pasteur.fr

ABSTRACT

Klebsiella pneumoniae (phylogroup Kp1), one of the most problematic pathogens associated with antibiotic resistance worldwide, is phylogenetically closely related to *K. quasipneumoniae* [subsp. *quasipneumoniae* (Kp2) and subsp. *similipneumoniae* (Kp4)], *K. variicola* (Kp3) and two unnamed phylogroups (Kp5 and Kp6). Together, Kp1 to Kp6 make-up the *K. pneumoniae* complex. Currently, the phylogroups can be reliably identified only by gene sequencing. Misidentification using standard methods is common and the clinical significance of *K. pneumoniae* complex members is therefore imprecisely defined. Here, we evaluated the potential of MALDI-TOF mass spectrometry to discriminate *K. pneumoniae* complex members. We report for the first time the existence of mass spectrometry biomarkers associated with the phylogroups, with a sensitivity and specificity ranging between 80-100% and 97-100%, respectively. Strains within phylogroups Kp1, Kp2, Kp4 and Kp5 each shared two specific peaks not observed in other phylogroups. Kp3 strains shared a peak that was only observed otherwise in Kp5. Finally, Kp6 had a diagnostic peak shared only with Kp1. Kp3 and Kp6 could therefore be identified by exclusion criteria (lacking Kp5 and Kp1-specific peaks, respectively). Further, ranked Pearson correlation clustering of spectra grouped strains according to their phylogroup. These results call for incorporation of spectra of all *K. pneumoniae* complex members into reference MALDI-TOF spectra databases, in which they are currently lacking. This advance may allow for simple and precise identification of *K. pneumoniae* and closely related species, opening the way to a better understanding of their epidemiology, ecology and pathogenesis.

INTRODUCTION

Klebsiella pneumoniae is an increasingly challenging human bacterial pathogen, causing hospital or community-acquired infections that are associated with high rates of antibiotic resistance (1, 2). Population diversity studies have shown that *K. pneumoniae* is in fact part of a complex of species, being phylogenetically closely related to *K. quasipneumoniae* (subsp. *quasipneumoniae* and subsp. *similipneumoniae*) and *K. variicola* (3–5). Before recent taxonomic updates (6, 7), *K. pneumoniae* and the other above taxa were designed as *K. pneumoniae* phylogroups Kp1, Kp2, Kp4 and Kp3, respectively (8). Together with two novel phylogroups (Kp5 and Kp6) described recently (5), these taxa constitute the *K. pneumoniae* complex. Although *K. pneumoniae* is numerically the major cause of human infections among members of the complex, the involvement of the other members of the complex in human infections is gaining recognition (4, 8–12). However, the unsuitability of traditional clinical microbiology methods to distinguish species within the complex leads to high rates of misidentifications (most often as *K. pneumoniae*) that are masking the true clinical significance of each phylogroup and their potential epidemiological specificities (8, 9, 12, 13). In fact, the different members of the *K. pneumoniae* complex can be reliably identified only based on gene sequencing (e.g. bla_{LEN}, bla_{OKP}, bla_{SHV}, *rpoB*, *gyrA*, *parC*) (4, 7, 14). Some PCR-based identification methods were developed but they are prone to errors or do not distinguish all phylogroups (8, 15–17). Clearly, there is a need for reliable, cost-effective and fast identification methods able to discriminate members of the *K. pneumoniae* complex.

Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (MS) has revolutionized routine identification of microorganisms, being a fast and cost-effective technique. It now represents a first line identification method in many clinical, environmental and food microbiology laboratories (18). In the case of the *K. pneumoniae* complex, MALDI-TOF MS identification remains largely unsatisfactory given the absence of well characterized, representative members of the complex in spectral databases. Currently, only *K. pneumoniae* and *K. variicola* are included in the Bruker database (https://www.bruker.com/fileadmin/user_upload/1-Products/Separations_MassSpectrometry/MALDI_Biotyper/US_CA_System/MBT_list_of_organisms_10_2017.pdf), and identification of even these two species is imprecise given the lack of reference

spectra of other phylogroups (13, 19). To address this important limitation of currently MALDI-TOF MS technology, we used a collection of well characterized reference strains from the six *K. pneumoniae* complex phylogroups and analyzed them by MALDI-TOF MS in order to define the potential of this method to identify species within the *K. pneumoniae* complex.

MATERIAL AND MEHODS

Bacterial strains. A set of 46 strains previously characterized by whole-genome sequencing or using core gene sequences (5, 7, 20, 21) were analyzed in this study (Table S1). The strains belonged to the taxa *K. pneumoniae* (*sensu stricto*, i.e., Kp1; n=10), *K. quasipneumoniae* subsp. *quasipneumoniae* (Kp2, n=9), *K. quasipneumoniae* subsp. *similipneumoniae* (Kp4, n=7), *K. variicola* (Kp3, n=9), and to two taxonomically undefined lineages named Kp5 (n=6) and Kp6 (n=5). Strains had been stored in brain heart infusion broth containing 25% glycerol at -80°C and were sub-cultivated before use in this study.

Spectra acquisition. An overnight culture on Luria-Bertani agar (37°C, 18h) was used to prepare the samples with the ethanol/formic acid extraction procedure following the manufacturer recommendations (Bruker Daltonics, Bremen, Germany). Samples (1 µL) were spotted onto an MBT Biotarget 96 target plate, air dried and overlaid with 1 µL of a saturated α-cyano-4-hydroxycinnamic acid (HCCA) matrix solution in 50% of acetonitrile and 2.5% of trifluoroacetic acid. Mass spectra were acquired on a Microflex LT mass spectrometer (Bruker Daltonics, Bremen, Germany) using the default parameters (detection in linear positive mode, laser frequency of 60 Hz, ion source voltages of 2.0 and 1.8 kV, lens voltage of 6 kV) within the mass range of 2,000-20,000 Da. For each strain, a total of 24 spectra from 8 independent spots were acquired (3 spectra *per* spot, instrumental replicates). External calibration of the mass spectra was performed using Bruker Bacterial Test Standard (BTS).

Spectra analysis. The spectra were preprocessed by applying the "smoothing" and "baseline subtraction" procedures available in FlexAnalysis software (Bruker Daltonics, Bremen, Germany), exported as peak lists with *m/z* values and signal intensities for each peak in text format, and imported into a dedicated BioNumerics v7.6 (Applied Maths, Ghent, Belgium) database. Peak detection was

performed in BioNumerics using a signal to noise ratio of 20. The instrumental replicates (24 spectra for each strain) were used to generate a mean spectrum for each strain using the following parameters: minimum similarity, 80%; minimum peak detection rate, 60%; constant tolerance, 1; and linear tolerance, 300 ppm. Finally, peak matching was performed to search all distinct peaks (called peak classes in BioNumerics) using as parameters: constant tolerance, 1.9; linear tolerance, 550 ppm; maximum horizontal shift, 1; peak detection rate, 10. The discriminating value of each resulting peak was evaluated by a Mann-Whitney test (22). To allocate proteins associated with peaks, the online tool TagIdent was used (http://web.expasy.org/tagident/). Additionally, a Neighbor Joining tree based on ranked Pearson coefficient was constructed using BioNumerics.

RESULTS AND DISCUSSION

Forty-six reference strains representing the six phylogroups currently known within the *K. pneumoniae* complex were analyzed by MALDI-TOF MS. Based on the MALDI Biotyper Compass database version 4.1.80 (Bruker Daltonics, Bremen, Germany), the 46 strains were identified either as *K. pneumoniae* (31 strains, all belonging to Kp1, Kp2, Kp4 and Kp6) or as *K. variicola* (15 strains, all strains of Kp3 and Kp5). Identification scores ranged between 2.16-2.56 for *K. pneumoniae* and 1.89-2.55 for *K variicola.* Of note, in two cases a replicate was reported in one measure as *K. pneumoniae* and in other as *K. variicola*. These data highlight the need to update the database in order to refine confidence in *K. pneumoniae*/*K. variicola* identification and to enable identification of *K. quasipneumoniae and novel phylogroups*. **Fig. 1** summarizes the peak positions found in each strain. Most (about 97%) of the peaks were concentrated in the region below 10,000 *m/z* and almost no peak was found above this value. The similarity among spectra within the *K. pneumoniae* complex was always above 87% (data not shown), with peaks at 4363, 5379, 6286, 6298, 7241 and 9473 *m/z* being found in all the members of the complex. Interestingly, ten specific biomarkers associated with specific members of the *K. pneumoniae* complex were identified. These peaks were located within the range 3835 - 9553 *m/z*. The specificity and sensitivity of their distribution among phylogroups ranged between 97-100% and 80-100%, respectively (**Fig. 1 and Table 1**). Kp1 (4152 and 8305 *m/z*), Kp2 (4136 and 8271 *m/z*), Kp4 (7670 and 3835 *m/z*) and Kp5 (4777 and 9553 *m/z*) each presented two

specific peaks, which may allow their unambiguous identification. Kp3 strains shared a peak that was only observed otherwise in Kp5 (7768 *m/z*). Finally, Kp6 had a diagnostic peak (5278 *m/z*) shared only with Kp1. Kp3 and Kp6 could therefore be identified by exclusion criteria (lacking Kp5 and Kp1-specific peaks, respectively) (**Fig. 1 and Table 1**). These data reveal the possibility to identify precisely an isolate of the Kp complex based on the specific combination of the above described peaks. To the best of our knowledge, this is the first time that mass spectrometry biomarkers that discriminate the phylogroups of the *K. pneumoniae* complex are described. Furthermore, cluster analysis grouped all strains according to their phylogroup (**Fig. S1**), also showing the potential of whole spectrum comparison for strain identification at the phylogroup level.

About half of the peaks visualized in a bacterial spectrum in the mass range used in this work (2,000-20,000 Da) correspond to ribosomal proteins (18). Here, we were able to presumptively identify two of the specific peaks as ribosomal proteins (S22 and L31, respectively 5278 and 7768 *m/z*), and one as a non-characterized protein specific for Kp4 [7670 *m/z,* locus tag SB30_RS24725 (GenBank Accession number CBZR010000000)]. The specificity of the peaks was supported by the protein alignments obtained from whole-genome sequences (data not shown). The other seven peaks useful for identification could not be associated with a defined protein (**Table 1**).

In conclusion, this work demonstrates the potential of MALDI-TOF MS to identify isolates of the *K. pneumoniae* complex at the phylogroup level. We urge that reference spectra of the various taxa of the *K. pneumoniae* complex be incorporated into reference MALDI-TOF spectra databases, so that the approach could be implemented in microbiology laboratories. Improved identification of *K. pneumoniae* and related taxa will advance our understanding of the epidemiology, ecology and links with pathogenesis of this increasingly important group of pathogens.

Acknowledgments

We thank the Biological Resource Center of Institut Pasteur (Chantal Bizet and Estelle Muhle) for support within the MALDI-TOF MS platform. We acknowledge Andriniaina Rakotondrasoa and Jean-Marc Collard (Institut Pasteur of Madagascar), S. Wesley Long (Houston Methodist Hospital)

and David A. Rasko and J. Kristie Johnson (University of Maryland School of Medicine, Baltimore) for kindly providing strains analyzed in this study.

This work is part of the MedVetKlebs project, a component of the One Health European Joint Programme, which has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.

Conflicts of interests

The authors declare that there are no conflicts of interest.

References

1. European Centre for Disease Prevention and Control (ECDC). 2017. Surveillance of antimicrobial resistance in Europe 2016. Annual report of the European Antimicrobial REsistance Surveillance Network (EARS-Net).
2. Wyres KL, Holt KE. 2016. *Klebsiella pneumoniae* population genomics and antimicrobial-resistant clones. Trends Microbiol 24:944–956.
3. Brisse S, Verhoef J. 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 51:915–924.
4. Holt KE, Wertheim H, Zadoks RN, Baker S, Whitehouse CA, Dance D, Jenney A, Connor TR, Hsu LY, Severin J, Brisse S, Cao H, Wilksch J, Gorrie C, Schultz MB, Edwards DJ, Nguyen K Van, Nguyen TV, Dao TT, Mensink M, Minh V Le, Nhu NTK, Schultsz C, Kuntaman K, Newton PN, Moore CE, Strugnell RA, Thomson NR. 2015. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. Proc Natl Acad Sci U S A 112:E3574-81.
5. Blin C, Passet V, Touchon M, Rocha EPC, Brisse S. 2017. Metabolic diversity of the emerging pathogenic lineages of *Klebsiella pneumoniae*. Environ Microbiol 19:1881–1898.
6. Rosenblueth M, Martínez L, Silva J, Martínez-Romero E. 2004. *Klebsiella variicola*, a novel

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

188 7. Brisse S, Passet V, Grimont PAD. 2014. Description of *Klebsiella quasipneumoniae* sp. nov., a
189 novel species isolated from human infections, with two subspecies *Klebsiella quasipneumoniae*
190 subsp. *quasipneumoniae* subsp. nov. and *Klebsiella quasipneumoniae* subsp. *similipneumoniae*
191 subsp. nov., and demonstration that *Klebsiella singaporensis* is a junior heterotypic synonym
192 of *Klebsiella variicola*. Int J Syst Evol Microbiol 64:3146–3152.

193 8. Brisse S, van Himbergen T, Kusters K, Verhoef J. 2004. Development of a rapid identification
194 method for *Klebsiella pneumoniae* phylogenetic groups and analysis of 420 clinical isolates.
195 Clin Microbiol Infect 10:942–945.

196 9. Seki M, Gotoh K, Nakamura S, Akeda Y, Yoshii T, Miyaguchi S, Inohara H, Horii T, Oishi K,
197 Iida T, Tomono K, Masafumi Seki C. 2013. Fatal sepsis caused by an unusual *Klebsiella*
198 species that was misidentified by an automated identification system. J Med Microbiol 62:801–
199 803.

200 10. Maatallah M, Vading M, Kabir MH, Bakhrouf A, Kalin M, Nauclér P, Brisse S, Giske CG.
201 2014. *Klebsiella variicola* is a frequent cause of bloodstream infection in the Stockholm area,
202 and associated with higher mortality compared to *K. pneumoniae*. PLoS One 9:1–21.

203 11. Breurec S, Melot B, Hoen B, Passet V, Schepers K, Bastian S, Brisse S. 2016. Liver abscess
204 caused by infection with community-acquired *Klebsiella quasipneumoniae* subsp.
205 *quasipneumoniae*. Emerg Infect Dis 22:529–531.

206 12. Becker L, Fuchs S, Pfeifer Y, Semmler T, Eckmanns T, Korr G, Sissolak D, Friedrichs M, Zill
207 E, Tung M-L, Dohle C, Kaase M, Gatermann S, Rüssmann H, Steglich M, Haller S, Werner G.
208 2018. Whole genome sequence analysis of CTX-M-15 Producing *Klebsiella* isolates allowed
209 dissecting a polyclonal outbreak scenario. Front Microbiol 9:322.

210 13. Long SW, Linson SE, Ojeda Saavedra M, Cantu C, Davis JJ, Brettin T, Olsen RJ. 2017.
211 Whole-Genome sequencing of human clinical *Klebsiella pneumoniae* isolates reveals
212 misidentification and misunderstandings of *Klebsiella pneumoniae*, *Klebsiella variicola*, and
213 *Klebsiella quasipneumoniae*. mSphere 2:e00290-17.

214 14. Haeggman S, Lofdahl S, Paauw A, Verhoef J, Brisse S. 2004. Diversity and evolution of the

class A chromosomal beta-lactamase gene in *Klebsiella pneumoniae*. Antimicrob Agents Chemother 48:2400–2408.

15. Fonseca EL, Ramos N da V, Andrade BGN, Morais LLCS, Marin MFA, Vicente ACP. 2017. A one-step multiplex PCR to identify *Klebsiella pneumoniae*, *Klebsiella variicola*, and *Klebsiella quasipneumoniae* in the clinical routine. Diagn Microbiol Infect Dis 87:315–317.

16. Garza-Ramos U, Silva-Sánchez J, Martínez-Romero E, Tinoco P, Pina-Gonzales M, Barrios H, Martínez-Barnetche J, Gómez-Barreto RE, Tellez-Sosa J. 2015. Development of a Multiplex-PCR probe system for the proper identification of *Klebsiella variicola*. BMC Microbiol 15:1–14.

17. Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Nicolas-Chanoine MH, Decré D, Brisse S. 2014. Development of a multiplex PCR assay for identification of *Klebsiella pneumoniae* hypervirulent clones of capsular serotype K2. J Med Microbiol 63:1608–1614.

18. van Belkum A, Welker M, Pincus D, Charrier JP, Girard V. 2017. Matrix-assisted laser desorption ionization time-of-flight mass spectrometry in clinical microbiology: what are the current issues? Ann Lab Med 37:475–483.

19. Berry GJ, Loeffelholz MJ, Williams-Bouyer N. 2015. An investigation into laboratory misidentification of a bloodstream *Klebsiella variicola* infection. J Clin Microbiol 53:2793–4.

20. Bialek-Davenet S, Criscuolo A, Ailloud F, Passet V, Jones L, Delannoy-Vieillard AS, Garin B, Hello S Le, Arlet G, Nicolas-Chanoine MH, Decré D, Brisse S. 2014. Genomic definition of hypervirulent and multidrug-resistant *Klebsiella pneumoniae* clonal groups. Emerg Infect Dis 20:1812–1820.

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

22. Vranckx K, De Bruyne K, Pot B. 2017. Analysis of MALDI-TOF MS Spectra using the BioNumerics Software. MALDI-TOF Tandem MS Clin Microbiol 539–562.

241 **Table 1**. MALDI-TOF mass spectrometry peak biomarkers useful to discriminate *Klebsiella*

242 *pneumoniae* phylogroups.

Present in Kp phylogroup(s):	**Peak Position** (*m/z*)[1]	**Sensitivity**	**Specificity**	**Possible proteins**[2]
Kp1	4152.89	100%	97.3%	-
	8305.17	100%	100%	-
Kp2	4136.09	100%	100%	-
	8271.32	100%	100%	-
Kp3 and **Kp5**	7768.04	100%	100%	Ribosomal protein L31
Kp4	3835.01	100%	100%	-
	7670.21	100%	100%	Uncharacterized protein specific for Kp4
Kp5	4777.43	100%	100%	-
	9553.05	100%	100%	-
Kp1 and **Kp6**	5278.02	80%	100%	Ribosomal protein S22

[1] Position in the spectra using a tolerance of ± 0.05%.
[2] As determined using TagIdent.

243

244

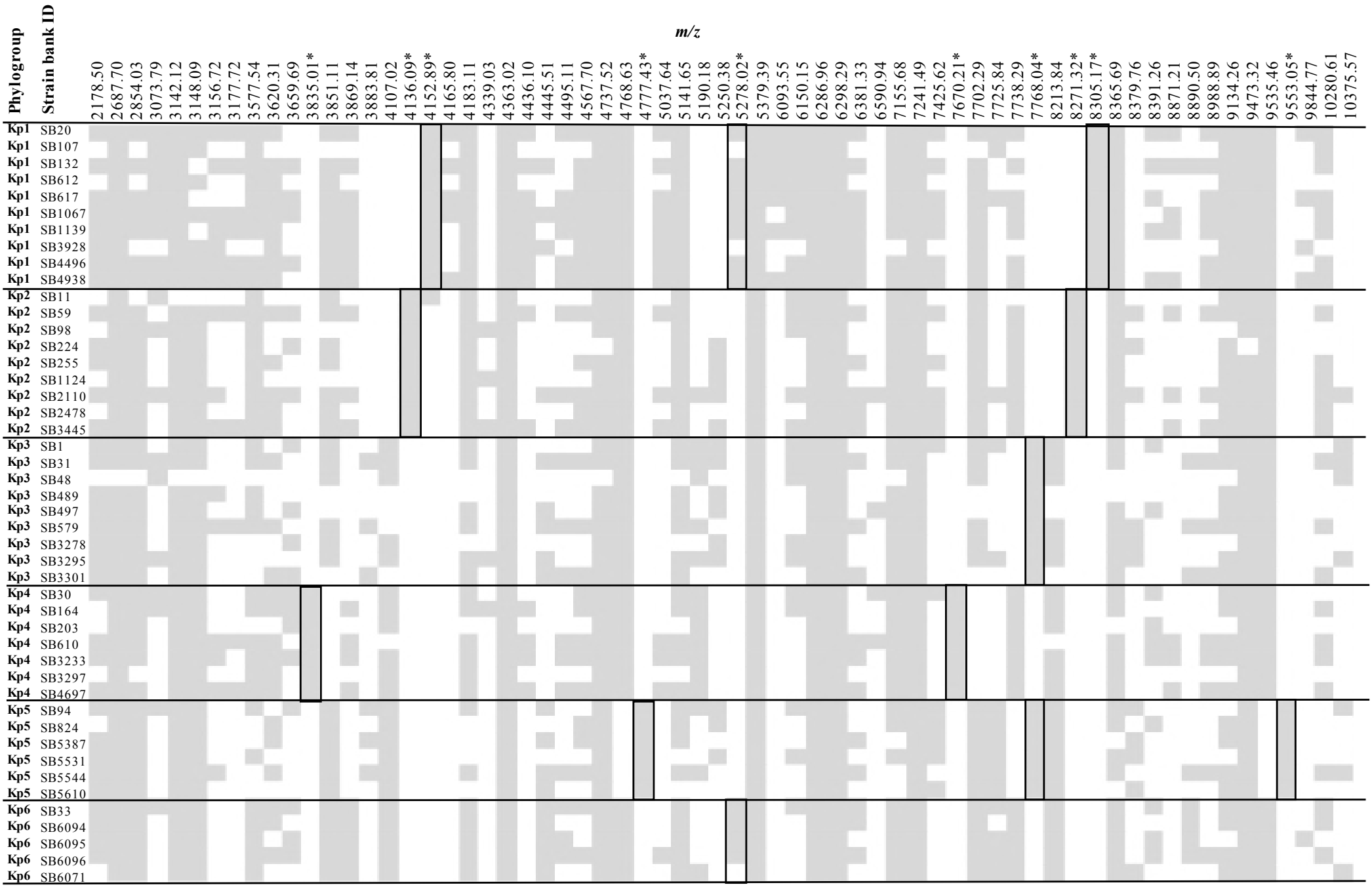
245

246 **Figure 1**. Peak positions (*m/z*) for each of the *K. pneumoniae* complex strains.

247 Star denotes those peaks that are useful for discrimination among phylogroups, as detailed in Table 1.

248

Figure 1. Peak positions (*m/z*) of the *K. pneumoniae* complex isolates.



* Specific biomarkers