

Klebsiella MALDI TypeR: a web-based tool for Klebsiella identification based on MALDI-TOF mass spectrometry

Sébastien Bridel, Stephen C Watts, Louise M Judd, Taylor Harshegyi, Virginie Passet, Carla Rodrigues, Kathryn E Holt, Sylvain Brisse

▶ To cite this version:

Sébastien Bridel, Stephen C Watts, Louise M Judd, Taylor Harshegyi, Virginie Passet, et al.. Klebsiella MALDI TypeR: a web-based tool for Klebsiella identification based on MALDI-TOF mass spectrometry. Research in Microbiology, 2021, 172 (4-5), pp.103835. 10.1016/j.resmic.2021.103835. pasteur-03230671

HAL Id: pasteur-03230671 https://pasteur.hal.science/pasteur-03230671

Submitted on 20 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1	Klebsiella MALDI TypeR: a web-based tool for Klebsiella				
2	identification based on MALDI-TOF mass spectrometry				
3					
4	Running title: Web-based MALDI-TOF identification of Klebsiella				
5					
6	Sébastien Bridel ¹ , Stephen C. Watts ² , Louise M. Judd ² , Taylor Harshegyi ² , Virginie Passet ¹				
7	Carla Rodrigues ¹ , Kathryn E. Holt ^{2,3} , and Sylvain Brisse ¹ *				
8	¹ Biodiversity and Epidemiology of Bacterial Pathogens, Institut Pasteur, Paris, France				
9	² Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne,				
10	Victoria 3004, Australia				
11	³ London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK				
12					
13	* Corresponding author: S. Brisse. Biodiversity and Epidemiology of Bacterial Pathogens,				
14	Institut Pasteur, 25 rue du Docteur Roux, F-75724 Paris, France. E-mail:				
15	sylvain.brisse@pasteur.fr; Phone +33 1 45 68 83 34				

16 Abstract

Klebsiella pathogens affect human and animal health and are widely distributed in the 17 18 environment. Among these, the Klebsiella pneumoniae species complex, which includes 19 seven phylogroups, is an important cause of community and hospital infections. The 20 Klebsiella oxytoca species complex also causes hospital infections and antibiotic-associated 21 haemorrhagic colitis. The unsuitability of currently used clinical microbiology methods to 22 distinguish species within each of these species complexes leads to high rates of misidentifications that are masking the true clinical significance and potential epidemiological 23 24 specificities of individual species.

We developed a web-based tool, *Klebsiella* MALDI TypeR, a platform-independent and user-friendly application that enables uploading MALDI-TOF mass spectrometry data in order to identify *Klebsiella* isolates at the species complex and phylogroup levels. The tool, available at <u>https://maldityper.pasteur.fr</u>, leverages a database of previously identified biomarkers that are specific for species complexes, individual phylogroups, or related phylogroups. We obtained 84% to 100% identification accuracy depending on phylogroup. Identification results are obtained in a few seconds from batches of uploaded spectral data.

Klebsiella MALDI TypeR enables fast and reliable identification of *Klebsiella* strains
 that are often misidentified with standard microbiological methods. This web-based
 identification tool may be extended in the future to other human bacterial pathogens.

35 Keywords

2

Klebsiella, mass spectrometry, software

37 Introduction

38 Bacteria of the genus Klebsiella are human and animal pathogens with a wide 39 ecological distribution [1]. The genus *Klebsiella*, a member of the family *Enterobacteriaceae*, 40 encompasses a huge diversity in terms of phylogenetic lineages, genomic content, pathogenic 41 properties, and ecological distribution. Two major subdivisions of Klebsiella are the Klebsiella pneumoniae species complex (KpSC) and the Klebsiella oxytoca species complex 42 43 (KoSC) [2]. The KpSC currently includes five taxonomic species, two of which comprise two 44 subspecies. These taxa were initially defined as seven distinct phylogroups, Kp1 to Kp7 [3]. 45 K. pneumoniae (phylogroup Kp1) is one of the most problematic pathogens associated with 46 antibiotic resistance worldwide. This species is phylogenetically closely related to K. 47 quasipneumoniae [subsp. quasipneumoniae (Kp2) and subsp. similipneumoniae (Kp4)], K. 48 variicola [subsp. variicola (Kp3) and subsp. tropica (Kp5)], K. quasivariicola (Kp6) and K. africana (Kp7). In turn, the KoSC comprises 6 species and 7 phylogroups [4]: K. 49 50 michiganensis (Ko1), K. oxytoca (phylogroup Ko2), K. spallanzanii (Ko3), K. pasteurii 51 (Ko4), K. grimontii (Ko6) and K. huaxiensis (Ko8). K. oxytoca is well known as a cause of opportunistic human infections and post-antibiotic hemorrhagic diarrhea [4,5]. 52

Recent literature has brought into focus the difficulties that microbiology laboratories encounter in the identification of individual members of the *K. pneumoniae* and *K. oxytoca* species complexes [5–9]. Indeed, *Klebsiella* species are genetically similar and have few biochemical discriminatory biomarkers. As a consequence, the clinical significance of individual *Klebsiella* species and their potential epidemiological specificities are still not well defined. While Matrix Assisted Laser Desorption Ionization - Time of Flight (MALDI-TOF) 59 mass spectrometry can identify *Klebsiella* species through peak discrimination [4,5,10], the 60 current software implementation is limited to identifying *K. pneumoniae*, *K. variicola* and *K.* 61 *oxytoca* [9–11]. Here we present *Klebsiella* MALDI TypeR, a tool that leverages novel 62 discriminant biomarkers in MALDI-TOF spectra to expand the number of identifiable 63 *Klebsiella* species from MALDI-TOF data and improve identification accuracy, using a 64 previously-published method [12].

65 Material and methods

Bacterial strains and reference identification by genomic sequencing. A total of 66 67 341 strains were included. A first set consisted of 78 clinical isolates (50 belonging to the Klebsiella pneumoniae species complex, 28 to the Klebsiella oxytoca species complex), 68 69 selected to be representative of the phylogenetic diversity of these taxa. These isolates have been used in previous studies [4,5]. Three and 24 spectra per isolate were recorded for the 70 71 KpSC and for the KoSC, respectively, as described previously [4,5]. A second set of 262 72 isolates belonging to the KpSC was collected in Australia from humans and animals. Isolate 73 genomes were sequenced using Illumina paired-end technology and assembled using SPAdes. 74 Species identification was performed with Kleborate [13,14], which compares each genome 75 to a curated set of species references using Mash distances; these species calls were used as 76 reference in subsequent analyses.

77 Preparation of bacterial samples and MALDI-TOF MS analysis. An overnight 78 culture on Luria-Bertani agar (37°C, 18 h) was used to prepare the samples with the ethanol/formic acid extraction procedure following the manufacturer recommendations 79 80 (Bruker Daltonics, Bremen, Germany). Samples (1 µL) were spotted onto an MBT Biotarget 96 target plate, air dried and overlaid with 1 μ L of a saturated α -cyano-4-hydroxycinnamic 81 82 acid matrix solution in 50% of acetonitrile and 2.5% of trifluoroacetic acid. External 83 calibration of the mass spectra was performed using Bruker Bacterial Test Standard. For the 84 assessment of the direct colony picking protocol, a single bacterial colony was picked with a 85 sterile toothpick and directly deposited onto a 96-spot polished steel target and processed as the sample spot previously mentioned. In addition, the same two protocols were followed 86

using blood agar medium (Columbia agar plus 5% sheep blood). A MALDI Biotyper
Microflex LT controlled by Compass flexControl software (version 3.4; Bruker Daltonics,
Billerica, Massachusetts, USA) was used to generate mass spectra for all isolates. Mass
spectra were acquired using automatic mode and default settings (2000 to 20 0000 Da; linear
positive mode; 240 laser shots).

92 Identification of phylogroups based on MALDI-TOF data in Klebsiella MALDI TypeR. The KpSC and the KoSC encompass species and subspecies that are defined as 93 phylogroups. As such, we aimed to distinguish them on the basis of MALDI-TOF data, using 94 95 information from the spectra, defined as peak shifts [15]. Identification of phylogroups is therefore based on a small selection of biomarkers for which one (or more) protein weight 96 97 modifications have been previously reported [4,5]. In fact, each phylogroup could be seen as 98 comprising MALDI-Types, as described by Bridel et al. 2020 [12]: a specific combination of 99 protein m/z values of these biomarkers. *Klebsiella* MALDI TypeR was run on a 4 CPUs 16-100 GB RAM CentOS Linux 7 (Core) virtual machine.

101 **Results and Discussion**

102 Development of *Klebsiella* MALDI TypeR and its web interface. *Klebsiella* 103 MALDI TypeR was developed as a web application with R (version 3.4.0) and R Shiny 104 (version 1.4). A user can upload either a single spectrum or multiple spectra (a maximum 105 number of spectra is not defined, but we recommend working with batches of up to 100 106 spectra per analysis). The uploaded spectra must be in zip archive format. The zip archive 107 must contain the raw data from any spectrometer. This includes the tree structure, which must

108 include spectra in the expected format (txt, csv, fid, ciphergen, mzXML, mzML, imzML, 109 analyze, cdf or msd - see MALDIquant Foreign documentation for further details). This 110 application has been tested with spectra (raw data folders) from Bruker spectrometer only but 111 could theoretically handle any proper archive containing one of the previously mentioned file 112 formats. Supplementary Figure 1 describes each step to reach one raw data folder. Each 113 archive may correspond either to a single raw spectrum or several raw spectra of the same 114 isolate (technical replicates). For better results, we recommend using at least 3 technical 115 replicates for each isolate.

116 The spectra can be visually inspected after upload (**Figure 1**). A panel shows each 117 uploaded spectrum for quick visual inspection (*e.g.*, intensity checking) and a table highlights 118 the presence (if any) of empty spectra.

119 For identification, the tool was designed with a 2-step process. The first step (Figure 120 2, panel B1) identifies isolates at the species complex level using two sets of biomarkers, 121 each set corresponding to reference biomarkers for (i) the KpSC (29 biomarkers, Table S1) or 122 (ii) the KoSC (57 biomarkers, **Table S1**). This step can be performed, based on user choice, 123 with one of two alternative peak detection algorithms. The first one is based on the R package 124 MALDIquant [16] and is set by default, as it runs faster. The second one uses the 125 MassSpecWavelet package [17] and generally detects more significant peaks, but is slower. 126 Species complex identification is based on the ratio of common peaks between the sample 127 peak list and a reference peak list, to the number of peaks in the corresponding reference (this 128 ratio is called "Identification Value"). The input sample is assigned to the species complex 129 with the highest Identification Value.

130 The second step consists in the identification of the specific phylogroup 131 (corresponding to a species or subspecies) within the defined species complex (Figure 2, 132 panel B2). The raw spectrum is processed again by the Multi Peak Shift Typing (MPST) 133 method [12], but this time using the set of biomarkers corresponding to the assigned species 134 complex, which allows recalibration of m/z values. The biomarkers used for the two MPST 135 schemes (one for each species complex, see **Tables 1 to 4**) are based on previous work [4,5]. 136 Briefly, an MPST scheme is analogous to a multi-locus sequence typing (MLST) scheme 137 where an allele type (AT) corresponds to an isoform (IF) for one protein (each IF is defined 138 by a specific mass – or a m/z value on the spectrum) and a MALDI-Type (MT) corresponds to 139 a sequence type (ST). Each MT was observed within a single phylogroup. The tool attempts 140 to define the IF for each biomarker, and then to attribute the spectrum to the phylogroup 141 corresponding to the relevant MT (Figure 2, panel B3). The only exception to this process 142 was introduced for the discrimination of K. oxytoca phylogroups Ko1, Ko4 and Ko6 (K. 143 oxytoca, K. pasteurii and K. grimontii, respectively). These three species form a phylogenetic 144 branch of closely related species within the K. oxytoca species complex [4], and their spectra 145 are very similar such that they share the same MT. In order to distinguish among these three 146 phylogroups, we added 3 previously described biomarkers [4], for which the tool checks for 147 presence/absence within the spectra (i.e., these 3 biomarkers do not correspond to peak 148 shifts). If any of the above biomarkers (peak shifts or presence/absence) is not found, 149 automatic identification is not considered possible, and the result is reported to the user as NT 150 (not typable). In this case, manual analysis or data re-acquisition are needed.

151 To help resolve these cases and allow users to further explore the data, we included a 152 feature enabling to interactively explore the spectra. The plot window (**Figure 3**) shows the 153 sample spectra overlaid on the reference spectra (1 to 3 spectra for each phylogroup).
154 Predefined windows (one for each biomarker) allow the user to directly investigate specific
155 spectral regions. This feature could also enable the detection of new isoforms or biomarkers.
156 It also contains a customizable option that allows the user to select any region of the spectra.

157 The *Klebsiella* MALDI TypeR tool was made available at <u>https://maldityper.pasteur.fr</u> 158 . A single isolate analysis required approximately 10 seconds to complete and the analysis of 159 10 isolates takes about 35 seconds. The tool does not have a limit to the number of isolates 160 that can be analyzed simultaneously, and it performed well with batches of 100 isolates that 161 were uploaded at once.

162 Evaluation of the identification tool. We first tested fifty KpSC isolates grown overnight on lysogeny broth (LB) agar (37°C, 18h) and subjected to ethanol/formic acid 163 164 extraction following the manufacturer recommendations (Bruker Daltonics, Bremen, 165 Germany). For each strain, a total of 3 spectra from the same spot were acquired (Bruker Microflex LT mass spectrometer, Paris). This dataset yielded correct assignments to species 166 167 complex for 50/50 (100%) isolates, 47 (94%) of which had an identification value >50% 168 (Tables S2 and S3). The same samples prepared using different experimental conditions (two 169 media: LB and Columbia agar plus 5% sheep blood; each with both extraction and direct 170 application in the MALDI-TOF target) also yielded correct complex assignments for all 171 isolates, albeit with lower confidence (Table S2). Regarding phylogroup assignment, using 172 LB-extraction 90% of samples were correctly identified at phylogroup level (Table S3), while 173 the other sample preparation methods yielded better results overall (Table S4). To validate 174 robustness of the tool to instrument and user variation, we tested an additional dataset comprising 248 KpSC isolates prepared using the LB-extraction method (176 isolates with a
single spectrum, 26 in duplicate and 68 in triplicate); these spectra were generated on a
different instrument (Bruker MALDI Biotyper microflex LT-SH) in a different laboratory
(Melbourne, Australia). 244/248 isolates (98%) were assigned to the correct species complex,
and 219/244 (90%) were correctly identified at the phylogroup level (84% to 100% for each
phylogroup, see Table S5).

The KoSC assignments were tested using 28 isolates prepared using LB-extraction and triplicate spectra generated on the Bruker Microflex LT mass spectrometer instrument (Paris). All isolates were correctly assigned to complex level and most (89%) were assigned to the correct MALDI-Type (**Tables S6 and S7**). However, discrimination of MALDI-Type MT1 (n=16, 57%) into phylogroups Ko1, Ko4 and Ko6 was poor (**Table S8**) because the three additional biomarkers intended to discriminate between these phylogroups were captured only rarely.

188 Conclusions

We developed the *Klebsiella* MALDI TypeR web tool for identification of *Klebsiella* species based on MALDI-TOF MS. Our tool allows the fast analysis of single or multiple spectra, including their visualization. We show that MALDI-TOF MS leads to accurate identification for *Klebsiella* phylogroups of the KpSC, the major *Klebsiella* species complex found in clinical samples. *Klebsiella* MALDI TypeR therefore represents an important resource for the identification of the KpSC members, which are currently difficult to identify in routine microbiology practice. For the KoSC, we found promising data, but further 196 validation based on a larger dataset would be required. The current biomarker approach does 197 not appear to be suitable for accurate discrimination of Ko1, Ko4 and Ko6. In the future, the 198 website may be updated to integrate additional methods for identification and typing, such as 199 supervised learning.

200 **Declarations**

	11	•	-	
202	Not applicable			
203				
204	Consent for publication			
205	Not applicable			

Ethics approval and consent to participate

206

201

207 Availability of data and materials

208 The datasets used and/or analyzed during the current study are available from the

209 corresponding author on reasonable request.

210

211 Competing interests

212 The authors declare that they have no competing interests.

213

214 Funding

215 This work was financially supported by the MedVetKlebs project, a component of European

216 Joint Programme One Health EJP, which has received funding from the European Union's

217 Horizon 2020 Research and Innovation Programme under Grant Agreement No. 773830.

218 Authors' contributions

- LMJ, TH, CR and VP performed the MALDI-TOF data acquisition. SBridel, SCW and CR
 analyzed the MALDI-TOF data. SBridel designed and programmed the initial version of the *Klebsiella* MALDI TypeR tool and website. SCW, CR, KEH and SBrisse provided input in
 the choice of data processing algorithms and web site design. SBridel wrote the original draft.
 All authors contributed to and approved the final version of the manuscript.
- 224

225 Acknowledgements

We acknowledge the help of Stevenn Volant and Amine Ghozlane from the Bioinformatics and Biostatistics HUB of Institut Pasteur for help with R Shiny deployment of the web tool on the web server.

229 **References**

- 230 [1] Brisse S, Grimont F, Grimont PAD. The genus Klebsiella. In: Dworkin M, Falkow S,
- Rosenberg E, Schleifer K-H, Stackebrandt E, editors. Prokaryotes · Handb. Biol. Bact., vol. 6:
 Proteobacteria: Gamma Subclass. 3rd edition, New York: Springer; 2006.
- 233 [2] Wyres KL, Lam MMC, Holt KE. Population genomics of Klebsiella pneumoniae. Nat 234 Rev Microbiol 2020;18:344–59. https://doi.org/10.1038/s41579-019-0315-1.
- Rev Microbiol 2020, 18:344–39. https://doi.org/10.1038/s41379-019-0313-1.
 Rodrigues C, Passet V, Rakotondrasoa A, Diallo TA, Criscuolo A, Brisse S.
- 236 Description of Klebsiella africanensis sp. nov., Klebsiella variicola subsp. tropicalensis subsp.
- nov. and Klebsiella variicola subsp. variicola subsp. nov. Res Microbiol 2019;170:165–70.
- 238 https://doi.org/10.1016/j.resmic.2019.02.003.
- 239 [4] Merla C, Rodrigues C, Passet V, Corbella M, Thorpe HA, Kallonen TVS, et al.
- Description of Klebsiella spallanzanii sp. nov. and of Klebsiella pasteurii sp. nov. Front
 Microbiol 2019;10. https://doi.org/10.3389/fmicb.2019.02360.
- 242 [5] Rodrigues C, Passet V, Rakotondrasoa A, Brisse S. Identification of Klebsiella
- 243 pneumoniae, Klebsiella quasipneumoniae, Klebsiella variicola and Related Phylogroups by
- 244 MALDI-TOF Mass Spectrometry. Front Microbiol 2018;9.
- 245 https://doi.org/10.3389/fmicb.2018.03000.
- 246 [6] Brisse S, van Himbergen T, Kusters K, Verhoef J. Development of a rapid
- identification method for Klebsiella pneumoniae phylogenetic groups and analysis of 420
 clinical isolates. Clin Microbiol Infect 2004;10:942–5. https://doi.org/10.1111/j.1469-
- 248 clinical isolates. Clin Microbiol Infect 2004;10:942–5. https://doi.org/10.1111/j.1469-
- 249 0691.2004.00973.x.
- [7] Seki M, Gotoh K, Nakamura S, Akeda Y, Yoshii T, Miyaguchi S, et al. Fatal sepsis
 caused by an unusual Klebsiella species that was misidentified by an automated identification
 system. J Med Microbiol 2013;62:801–3. https://doi.org/10.1099/jmm.0.051334-0.
- 253 [8] Fonseca EL, Ramos N da V, Andrade BGN, Morais LLCS, Marin MFA, Vicente
- ACP. A one-step multiplex PCR to identify Klebsiella pneumoniae, Klebsiella variicola, and
 Klebsiella quasipneumoniae in the clinical routine. Diagn Microbiol Infect Dis 2017;87:315–
 7. https://doi.org/10.1016/j.diagmicrobio.2017.01.005.
- 257 [9] Long SW, Linson SE, Saavedra MO, Cantu C, Davis JJ, Brettin T, et al. Whole-
- 258 Genome Sequencing of Human Clinical Klebsiella pneumoniae Isolates Reveals
- 259 Misidentification and Misunderstandings of Klebsiella pneumoniae, Klebsiella variicola, and
- Klebsiella quasipneumoniae. MSphere 2017;2. https://doi.org/10.1128/mSphereDirect.0029017.
- 262 [10] Dinkelacker AG, Vogt S, Oberhettinger P, Mauder N, Rau J, Kostrzewa M, et al.
- 263 Typing and Species Identification of Clinical Klebsiella Isolates by Fourier Transform
- Infrared Spectroscopy and Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass
 Spectrometry. J Clin Microbiol 2018;56. https://doi.org/10.1128/JCM.00843-18.
- 266 [11] Berry GJ, Loeffelholz MJ, Williams-Bouyer N. An Investigation into Laboratory
- 267 Misidentification of a Bloodstream Klebsiella variicola Infection. J Clin Microbiol
- 268 2015;53:2793–4. https://doi.org/10.1128/JCM.00841-15.
- 269 [12] Bridel S, Bourgeon F, Marie A, Saulnier D, Pasek S, Nicolas P, et al. Genetic diversity
- and population structure of Tenacibaculum maritimum, a serious bacterial pathogen of marine
- fish: from genome comparisons to high throughput MALDI-TOF typing. Vet Res 2020;51:60.

- 272 https://doi.org/10.1186/s13567-020-00782-0.
- 273 [13] Holt K. katholt/Kleborate. 2020.
- 274 [14] Lam MMC, Wick RR, Wyres KL, Holt KE. Genomic surveillance framework and
- 275 global population structure for Klebsiella pneumoniae. BioRxiv 2020:2020.12.14.422303.
- 276 https://doi.org/10.1101/2020.12.14.422303.
- 277 [15] Sauget M, Valot B, Bertrand X, Hocquet D. Can MALDI-TOF Mass Spectrometry
- 278 Reasonably Type Bacteria? Trends Microbiol 2017;25:447–55.
- 279 https://doi.org/10.1016/j.tim.2016.12.006.
- 280 [16] Gibb S, Strimmer K. MALDIquant: a versatile R package for the analysis of mass
- 281 spectrometry data. Bioinformatics 2012;28:2270–1.
- 282 https://doi.org/10.1093/bioinformatics/bts447.
- 283 [17] Du P, Kibbe WA, Lin SM. Improved peak detection in mass spectrum by
- 284 incorporating continuous wavelet transform-based pattern matching. Bioinforma Oxf Engl
- 285 2006;22:2059–65. https://doi.org/10.1093/bioinformatics/btl355.

286