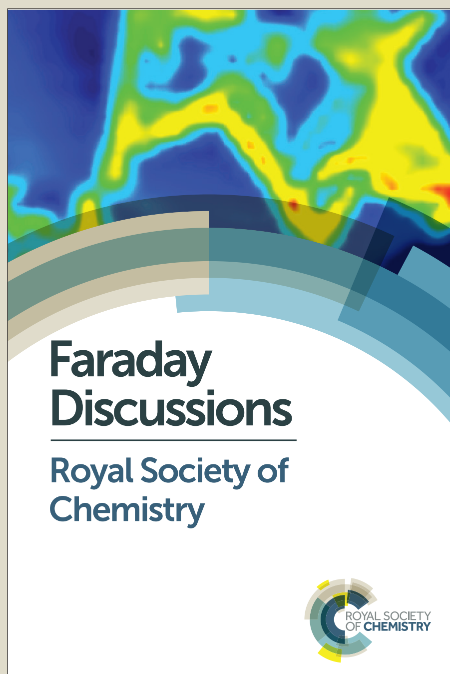
Faraday Discussions

Accepted Manuscript

This manuscript will be presented and discussed at a forthcoming Faraday Discussion meeting. All delegates can contribute to the discussion which will be included in the final volume.

Register now to attend! Full details of all upcoming meetings: http://rsc.li/fd-upcoming-meetings



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

Rapid Characterisation of *Klebsiella oxytoca* Isolates from Contaminated Liquid Hand Soap Using Mass Spectrometry, FTIR and Raman Spectroscopy

Ralf Dieckmann [1*], Jens Andre Hammerl [1], Hartmut Hahmann [4], Amal Wicke [4], Sylvia Kleta [1], Piotr Wojciech Dabrowski [5], Andreas Nitsche [3], Maren Stämmler [2], Sascha Al Dahouk [1], Peter Lasch [2]

[1] Federal Institute for Risk Assessment, Department of Biological Safety, Max-Dohrn-Str. 8-10, D-10589 Berlin, Germany
[2] Robert Koch-Institut, ZBS 6 - *Proteomics and Spectroscopy*, Seestraße 10, D-13353 Berlin, Germany
[3] Robert Koch-Institut, ZBS 1 - *Highly Pathogenic Viruses*, Seestraße 10, D-13353 Berlin, Germany
[4] Landesamt für Verbraucherschutz Sachsen-Anhalt, Fachbereich Lebensmittelsicherheit, Freiimfelder Str. 68, D-06112 Halle, Germany
[5] Robert Koch-Institut, NG4 – *Bioinformatics*, Seestraße 10, D-13353 Berlin, Germany

Running Title: Characterisation of K. oxytoca isolates from contaminated hand soap

* Corresponding author. Mailing address: Federal Institute for Risk Assessment, Department of Biological Safety, Max-Dohrn-Str. 8-10, D-10589 Berlin, Germany. Phone: +49 30 18412-2182. Fax: +49 30 18412-2791. E-Mail: ralf.dieckmann@bfr.bund.de

Abbreviations: AAHC, antibiotic-associated haemorrhagic colitis; CFU, colony forming unit; DTGS, deuterated triglycine sulphate detector; FT, Fourier transform; FTIR, Fourier transform infrared; HCCA, α-cyano-4-hydroxycinnamic acid; IR, infrared; MALDI TOF, matrix-assisted laser desorption/ionization time-of-flight; MLST, multi locus sequence typing; MS, mass spectrometry; NIR, near-infrared; NJ, neighbour joining; PFGE, pulsed-field gel electrophoresis; SNP, single nucleotide polymorphism; SNR, signal-to-noise ratio; ST, sequence type; TFA, trifluoroacetic acid; UHCA, unsupervised hierarchical cluster analysis; WGS, whole genome sequencing

Faraday Discussions Accepted Manuscript

ABSTRACT

Microbiological monitoring of consumer products and the efficiency of early warning systems and outbreak investigations depend on the rapid identification and strain characterisation of pathogens posing risks to the health and safety of consumers. This study evaluated the potential of three rapid analytical techniques for identification and subtyping of bacterial isolates obtained from a liquid hand soap product, which has been recalled and notified through the EU RAPEX system due to its severe bacterial contamination. Ten isolates recovered from two bottles of the product were identified as *Klebsiella oxytoca* and subtyped using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI TOF MS), near-infrared Fourier transform (NIR FT) Raman spectroscopy and Fourier transform infrared (FTIR) spectroscopy. The comparison of the classification results obtained by these phenotype-based techniques with outcomes of the DNA-based methods pulsed-field gel electrophoresis (PFGE), multi locus sequence typing (MLST) and single nucleotide polymorphism (SNP) analysis of whole genome sequencing (WGS) data revealed a high level of concordance. In conclusion, a set of analytical techniques might be useful for rapid, reliable and cost-effective microbial typing to ensure safe consumer products and allow source tracking.

Faraday Discussions Accepted Manuscript

INTRODUCTION

Bacteria belonging to the genus *Klebsiella* are a frequent cause of human nosocomial and community-acquired infections leading to a wide range of diseases such as pneumonia, urinary tract infections, antibiotic-associated haemorrhagic colitis (AAHC), septicaemia, and skin and soft tissue infections. Klebsiellae are nonmotile, rod-shaped, Gram-negative bacteria with a prominent capsular polysaccharide. In particular, *Klebsiella (K.) pneumoniae* and *K. oxytoca* are associated with nosocomial infections often affecting immunocompromised patients and those requiring intensive care. They are both opportunistic pathogens found in the environment and in mammalian mucosal surfaces. Hand sanitizer and liquid soap dispensers have frequently been with the source of hospital-acquired infections (1-7) and multiple nosocomial outbreaks have been reported as a result from inadequate antisepsis or disinfection (8). Personal care products with a high water content such as hand soaps or shampoos as well as disinfectants used in private or public areas are at risk of being contaminated by bacteria leading to product deterioration and health risks to consumers (9-12). Intrinsic (occurring during manufacture) as well as extrinsic (occurring during use) contamination of personal care products or disinfectant solutions have been implicated in human infections (8).

Bacterial typing is necessary to identify sources and routes of product contamination and is a prerequisite for targeted control measures. The term subtyping refers to the characterisation beyond the species or subspecies level allowing the determination of clonal relationships and phylogenetic relatedness of bacterial strains. Nowadays, classical typing methods such as lysotyping or serotyping are more and more replaced by molecular biological approaches. Pulsed-field gel electrophoresis (PFGE) is the molecular typing method of first choice for several bacterial agents because of its high discriminatory power, epidemiological relevance and standardisation (13-15). It is based upon the variable migration of large DNA restriction fragments in an electrical field of alternating polarity and has been widely applied to characterise numerous bacterial species including *K. oxytoca* (16-19), however, it is technically demanding, time-consuming (approx. 1 week) and costly. Multi locus sequence typing (MLST) is based on the allelic differences among housekeeping genes and provides unambiguous data that are suitable for global epidemiology and evolutionary studies (20) and databases for many bacterial species are publicly available (http://pubmlst.org/databases/). Recently, a MLST method was developed for *K. oxytoca*, and analysis of nosocomial isolates showed that MLST can discriminate between isolates from patients suffering from nosocomial pneumonia and from antibiotic-associated haemorrhagic colitis (21). In recent years, whole genome sequencing (WGS) has developed rapidly and has become

Faraday Discussions Accepted Manuscript

a feasible typing method for microbiology laboratories in recent years. It provides ultimate discriminatory power and the potential to gather additional information such as the presence of antibiotic resistance genes, virulence genes or extrachromosomal elements (22, 23). For single nucleotide polymorphism (SNP) analysis whole-genome sequence data can be mapped against a reference genome and proved to be suitable for epidemiological typing and outbreak investigations (24-28).

Various analytical techniques often referred to as whole-organism fingerprinting methods are currently under investigation for rapid, objective and reliable bacterial identification at the species level and for bacterial typing. Such techniques not only reduce costs and efforts via automation and high-throughput solutions, but also generate results more rapidly. Particularly, matrix-assisted laser desorption/ionisation time-of-flight (MALDI TOF) mass spectrometry (MS), Fourier-transform infrared (FTIR) spectroscopy and Raman spectroscopy are promising tools because these methods are rapid, simple and cost-effective. Furthermore, the requirements for sample preparation are modest allowing automation and high-throughput analysis. In contrast to PFGE, MLST, or WGS which target the bacterial genome, these "phenotypic" techniques are based on the detection of large numbers of spectral features originating from various molecular building blocks (proteins, lipids, polysaccharides and nucleic acids). Since different microorganisms differ in their overall molecular composition, the spectra generated can be interpreted as highly complex taxon-specific fingerprints. MALDI TOF mass spectrometry mainly targets highly abundant ribosomal and nucleic acid-binding proteins (29, 30). Vibrational spectroscopic techniques, i.e. infrared (IR) and Raman spectroscopy are based on the specific vibrational motions of small functional groups: while IR spectroscopy measures the absorption of infrared light, Raman spectroscopy uses the samples' inelastic scattering properties following excitation by a monochromatic light source (laser). All three phenotypic methods allow for label-free analysis of microbial samples and therefore might be suitable for monitoring purposes and outbreak investigations without prior knowledge of the bacterial targets. Advantageously, they do not require prior amplification of nucleic acids like PCR-based methods, but detect biomolecules that are constitutively expressed by each individual cell.

In the present study ten bacterial isolates of *K. oxytoca* were analysed along with a selection of clinical and reference strains by a panel of phenotypic (MALDI TOF MS, FTIR and Raman spectroscopy) and genotypic (PFGE, MLST, WGS) characterisation techniques. Bacterial isolates were obtained from two lots of an intrinsically contaminated liquid hand soap product. This sanitary product was notified in 2013 through the EU RAPEX system (Rapid Alert System) for non-food dangerous products and recalled because it posed a microbiological risk due to the severe contamination

Faraday Discussions Accepted Manuscript

with *K. oxytoca* with an average concentration of 2.5 × 10^4 cfu/ml (cfu: colony forming unit).

The main objectives of the study were: (1) to examine the phylogenetic relatedness of the *K. oxytoca* hand soap isolates, (2) to determine the potential utility of three analytical techniques i.e. MALDI TOF MS, FTIR and FT Raman spectroscopy for subtyping *Klebsiella* sp. in terms of reproducibility, accuracy and taxonomic resolution, (3) to assess the degree of correlation between the results provided by the phenotype-based techniques and those attained by DNA-based approaches. In summary, a combined analysis of classification data obtained by distinct but complementary microbial characterisation techniques should be established to ensure product hygiene measures and allow tracing back a contamination site.

MATERIALS AND METHODS

Bacterial strains and cultivation.

Ten isolates originating from two bottles (lot I and lot II) of an intrinsically contaminated liquid hand soap produced in Greece, four clinical isolates, among them three from different infants with enteritis in 1995, and three reference strains were included in this study (Table 1). The liquid soap strains were isolated within routine product quality monitoring using standardised microbiological methods and a standard protocol for germ count. *K. oxytoca* was confirmed by standard biochemical methods and a commercial test kit (BBL CRYSTAL E/NF, Becton Dickinson, Sparks, USA). The product was recalled because it posed a microbiological risk due to a high concentration of *K. oxytoca* and was notified through the EU Rapid Alert System for non-food dangerous products (EU RAPEX, A12/1308/13). The average number of bacteria detected in the soap was 2.5 × 10^4 cfu/ml. Bacterial samples were prepared by growing each strain in three independent batches by two passages under aerobic conditions. Cells were cultured for nucleic acid-based characterisation on tryptic soy agar with neutralising agents Histidine, Lecithin, Tween-80 (Caso-HLT-Agar, Merck, Darmstadt, Germany) for 72 h at 30°C. For phenotypic characterisation cultivation was carried out on Caso agar plates (Merck, Darmstadt, Germany) for 24 h at 37°C.

Multi locus sequence typing (MLST).

Genomic DNA (gDNA) of the isolates was prepared using the QIAamp DNA mini kit (Qiagen, Hilden, Germany) as recommended by the manufacturers. All isolates were analysed by MLST typing of seven housekeeping genes (*gapA*, *infB*, *mdh*, *pgi*, *phoE*, *rpoB*, *tonB*) as described by Herzog et al. (21). PCR products were purified with the QIAquick PCR purification kit (Qiagen) and sequenced by MWG Eurofins (Ebersberg,

Faraday Discussions Accepted Manuscript

Germany). Sequence analysis was performed using the software packages of DNAstar (Lasergene) and Accelrys DS Gene. The sequences of the genetic loci investigated were subjected to the PubMLST database (http://pubmlst.org/koxytoca/) to assign sequence types (ST). For confirmation of MLST data, sequences of the housekeeping genes were extracted from WGS data of the *Klebsiella* sp. isolates using MLST 1.8 at https://cge.cbs.dtu.dk/services/MLST/ (31). Phylogenetic analysis of MLST sequences was performed using the Mega 6.0 software (32). Neighbour joining (NJ) trees for concatenated sequences were constructed using the Kimura-2 parameter model to estimate the genetic distances (33, 34).

Pulsed-field gel electrophoresis (PFGE).

Whole DNA from each *Klebsiella* sp. isolate was analysed by macrorestriction profiling using XbaI, and resolved by PFGE analysis in accordance with the CDC (Centers for Disease Control and Prevention) Pulse Net protocol (www.cdc.gov/pulse.net) as described by Hunter *et al*. (35), using a CHEF-DRIII SYS220/240 (Bio-Rad Laboratories GmbH, Munich, Germany) system. PFGE patterns were analysed using BioNumerics software package (version 6.6, Applied Maths, Sint-Martens-Latem, Belgium). Similarity analysis was performed using Dice coefficients with a band matching tolerance of 0.5 % and 1 % optimisation. Ward's method was used as the clustering method.

Analysis of single nucleotide polymorphisms (SNP) of whole genome sequencing (WGS) data.

Isolation of genomic DNA (gDNA) was conducted as described previously (36) using the QIAamp DNA mini kit (Qiagen, Hilden, Germany). A total of 1 ng of gDNA was subjected to library preparation using the Illumina Nextera® XT DNA Sample Preparation Kit. Samples were tagged, pooled and paired-end 2x300bp sequencing of the DNA was performed using an Illumina MiSeq which yielded 47.2 million reads with a total of 10.5 billion bases. SNPs differentiating the strains from each other were identified by mapping the reads to the sequence of *K. oxytoca* strain M1 (CP008841.1) using bowtie2 (37). A total of 38,214 discriminating SNPs were identified using an in-house pipeline based on GATK 3.4 (38). Tree calculation was conducted with PhyML 3.0 under the GTR+γ model (39).

FTIR spectroscopy.

For IR absorbance measurements, bacterial cells were harvested and prepared as previously described (40-42). Briefly, small amounts of late-exponential-phase cells (~10 to 60 µg dry weight) were carefully removed with a platinum loop from regions of confluent colony growth and suspended in 80 µl of distilled and sterilised water by

Faraday Discussions Accepted Manuscript

vortexing. An aliquot of 35 µl of the suspension was transferred to a ZnSe optical plate in a multisampling cuvette and dried in a desiccator over a drying agent (Sicapent®, Merck) with the application of a moderate vacuum (approx. 8 kPa) to form a transparent film suitable for FTIR measurements. Prior to spectral measurements the sample holder was sealed with a KBr cover plate to control the humidity and to protect the instrument from contamination. Spectra were recorded in the region between 500 and 4,000 cm^{-1} using an IFS 28/B FTIR spectrometer (Bruker Optics, Ettlingen, Germany). The instrument was specially designed for the measurement of microorganisms and was equipped with a deuterated triglycine sulphate (DTGS) detector. For each FTIR spectrum 64 scans were co-added and averaged. Spectra were recorded at a nominal spectral resolution of 6 cm^{-1}. Fourier transformation was done using a Blackmann-Harris 3-term apodisation function and a zerofilling factor of 4 giving a point spacing of approximately 1 cm^{-1}. The spectrometer was continuously purged by dry air to reduce interference with atmospheric water vapour and CO_2. IR spectra were acquired from three independent cultivations (biological replicates) of each individual bacterial strain.

Near-infrared (NIR) FT Raman spectroscopy.

Approximately 5-10 wire loops corresponding to 4-8 mg dry weight of bacterial material were transferred into cavities of custom-made steel cups produced by Körber & Körber (Birkenwerder, Germany). These steel cups have a diameter of 10 mm and a height of 4 mm. A centred hole of 2 mm in diameter and a depth of 0.5 mm allowed deposition of the bacterial biomass. Samples were dried in a desiccator over Sicapent® (Merck) with the application of a moderate vacuum (8-10 kPa). The cups were covered with custom-designed CaF_2 plates (Korth Kristalle, Altenholz, Kiel, Germany) to prevent contamination of the instrument. FT Raman measurements were performed using a MultiRam NIR FT Raman spectrometer (Bruker Optics) equipped with a Nd:YAG laser operating at 1064 nm and a liquid nitrogen cooled germanium (Ge) detector. The laser power was adjusted to 300-550 mW. FT Raman spectra were obtained with a spectral resolution of 6 cm^{-1}. Fourier transformation was done using a Blackmann-Harris 3-term apodisation function and a zero filling factor of 4. For each spectrum 250 individual scans were co-added. To further increase the signal-to-noise-ratio (SNR) three technical replicate spectra were consecutively collected from each of the three biological replicates (individual cultures) of each strain. For further analyses (cluster analysis) spectra from technical replicates were averaged.

MALDI TOF mass spectrometry.

For MS sample preparation the modified trifluoroacetic acid (TFA) inactivation protocol for highly pathogenic microorganisms was used (43). Briefly, 3-5 loops from bacterial

Faraday Discussions Accepted Manuscript

cultures (growth conditions as described above) were suspended in 20 µl HPLC grade water (Mallinckrodt Baker B.V., Deventer, Netherlands) before 80 µl of pure TFA (Uvasol®, Merck) were added. After an incubation period of 30 min, the solutions were diluted ten-fold with HPLC grade water (Mallinckrodt Baker B.V.). A volume of 2 µl of the microbial dilutions was mixed with an equal volume of a 12 mg/mL α-cyano-4-hydroxycinnamic acid (HCCA) solution (Bruker Daltonics, Bremen, Germany). The HCCA solution was prepared by dissolving HCCA in a 2:1 (v/v) mixture of 100% acetonitrile and 0.3% TFA. 2 µl of each mixture were spotted onto ground steel sample targets from Bruker Daltonics. Microbial mass spectra were acquired using an Autoflex I mass spectrometer (Bruker Daltonics) controlled by the FlexControl 3.0 data acquisition software. The mass spectrometer was equipped with a nitrogen laser (λ=337 nm) operating at a pulse rate of 10 Hz (manual mode of spectra acquisition). Pulse ion extraction time was 200 ns and the sampling rate was set to 0.5 GHz. Mass spectral measurements were carried out in the linear mode using an acceleration voltage of 20.00 (ion source 1) and 18.25 (ion source 2) kV. Lens voltage was 6.50 kV. Spectra were generated in the range between m/z 2,000 and 20,000. For calibration we used *E. coli* K12 strain DSM 3871 as an external standard. At least 300 individual laser shots were co-added for each MALDI TOF mass spectrum. Two technical replicate spectra were collected from each of the three biological replicates of each strain. For further analyses spectra from technical replicates were averaged.

Analysis of FTIR, Raman and MALDI mass spectra.

Recording of spectra, data storage and analysis of FTIR and NIR FT Raman spectra were performed by means of the OPUS software package provided by Bruker Optics. Spectra were pre-processed by applying a second-derivative Savitzky-Golay smoothing/derivative filter with nine smoothing points (FTIR only) and vector-normalisation (FTIR, FT Raman). Pre-processed spectra were subjected to unsupervised agglomerative hierarchical cluster analysis (UHCA) using the information contained in the spectral windows 750-900 cm^{-1}, 900-1200 cm^{-1}, 1300-1500 cm^{-1} and 2800-3000 cm^{-1} for FTIR or 350-720 cm^{-1}, 860-1000 cm^{-1} and 1030-1670 cm^{-1} for FT Raman, respectively. Dendrograms were obtained by means of the OPUS options *scaling to first range (*=D-values) (44) (distance method) and *average linkage* (linkage method). For processing microbial mass spectra, including calculation of mean spectra and UHCA, the MicrobeMS software (45) was utilised. For UHCA, mass spectra were smoothed, baseline corrected and normalised (46, 47) Peak detection was carried out by setting the number of peaks for each individual spectrum to 30. These peak tables were converted to bar-code spectra in which the peak amplitude information (intensity) has been omitted. Bar code spectra served as inputs for UHCA and dendrograms were

Faraday Discussions Accepted Manuscript

constructed with the same distance and linkage method used for FTIR and FT Raman spectra (see above).

Combined cluster analysis of phenotype data.

Cluster membership functions obtained on the basis of UHCA of microbial MALDI TOF mass, FTIR and FT Raman spectra served as inputs for a combined phenotypic characterisation method. To this end, vectors containing the individual cluster membership values were compiled for each spectrum of a biological replicate after fusion steps 1-6 (MALDI and FTIR), or 1-7 (Raman). These vectors were in the following subjected to a top-level unsupervised hierarchical cluster analysis carried out by means of the *pdist* (Euclidean distance) and *linkage* (average linkage) functions of Matlab's statistics toolbox (Matlab version R2014a, The Mathworks, Natick, MA, USA). The dendrogram was generated using the *dendrogram* function of Matlab.

RESULTS & DISCUSSION

Genotyping

The genetic relatedness of the isolates was determined by MLST, PFGE and SNP analysis of WGS data. Allelic profiles at seven loci based on the MLST scheme established by Herzog et al. (21) were assigned according to the PubMLST database (see supplemental Table S1). The ten hand soap isolates of *K. oxytoca* shared a single MLST profile (3;5;21;13; 20 (^{62}C$\rightarrow$T);6;12). The most similar sequence type found in the PubMLST database was ST151 (3;5;21;13;20;6;12), which displayed only a single nucleotide polymorphism. The corresponding strain (no. M936) was isolated from human faeces in France, 2009 (see supplemental information, Figure S1). Sequence data were concatenated in the order of loci used to define the allelic profile leading to a single sequence of 3,003 bp for each strain. The resulting neighbour-joining phylogeny is depicted in Figure 1A. In addition, a combined analysis with all *K. oxytoca* isolates present in the PubMLST database was undertaken (see Figure S1). According to the nomenclature defined by Herzog *et al*. (21), the ten hand soap isolates were found in a branch of cluster B1 together with strains which were mainly isolated from respiratory cultures. Within our study, *K. oxytoca* CB 4072 was most closely related to the hand soap isolates and clustered in a subbranch of cluster B1. Isolate CB 4063 was found in a branch of cluster B2, while all other isolates were found in cluster A except strain *K. oxytoca* CCUG 15788. According to MLST, the assignment of this reference strain to the species *K. oxytoca* seems questionable. As pointed out by Herzog et al., the majority of AAHC (antibiotic-associated haemorrhagic colitis) isolates of *K. oxytoca* belong to cluster A, while respiratory isolates are almost exclusively found in subcluster B1

Faraday Discussions Accepted Manuscript

suggesting a pathogenic potential of the soap isolates that is associated with pneumonia rather than gastrointestinal tract infections (21).
All isolates included in this study were examined by PFGE using XbaI macrorestriction enzyme. Between 10 and 20 bands per isolate were obtained. Their restriction patterns and a dendrogram created using Dice similarity coefficient and Ward's method is shown in Figure 1B. In congruency to the MLST phylogeny, the hand soap isolates of *K. oxytoca* formed a separate PFGE cluster and belonged to a single pattern. Six profiles contained an additional band at about 72 kb which was missing in the remaining four patterns and presumably represented a plasmid fragment (see red arrow in Figure 1B). The other analysed clinical isolates and reference strains had unique genotypes and exhibited no genetic relatedness based on PFGE data.
Whole-genome based SNP analysis was used to infer the relationships of the isolates and a phylogenetic tree was reconstructed using SNPs differentiating these genomes (Figure 1C). A SNP was defined as a position in the genome sequence at which at least two strains displayed different bases. Two outlier strains (*K. pneumoniae* ssp. *ozaenae* ATCC 25926 and *K. oxytoca* CCUG15788) and one hand soap isolate (PHS-891) were omitted from this analysis due to the low coverage of their genome sequences. The SNP tree was computed from 38,214 concatenated qualified SNPs using the complete genome sequence of strain *K. oxytoca* M1 (Acc. No. CP008841) as the reference. Whole-genome based SNP analysis clustered all hand soap isolates into one single cluster differentiating them accurately from the reference strains and other clinical isolates. The inset of Fig. 1C shows a detailed view of the hand soap isolate clade differentiating isolates from the two hand soap bottles, lot I and lot II, based on 134 of the 38,214 SNPs.

Phenotyping.

The degree of correlation of the similarity trees (dendrograms) produced by the different phenotype-based analytical techniques, MALDI TOF mass spectrometry, FTIR and FT Raman spectroscopy was compared with the phylogenetic trees obtained by the different genotype-based methods, i.e. MLST, PFGE and WGS. Phenotype-based analytical techniques are known to produce taxon-specific spectral fingerprints suitable for bacterial identification at the species or subspecies level (MALDI), or even at the strain level (FTIR, FT Raman). Figure 2A-C presents typical examples of MALDI TOF mass, FTIR and NIR FT Raman spectra from two exemplarily chosen isolates of *K. oxytoca*. Since in phenotype-based analyses some variation is common between spectra of technical and/or biological replicates, each of the *Klebsiella* sp. isolates or strains given in Table 1 was characterised from three individual microbial cultures.

Faraday Discussions Accepted Manuscript

MALDI TOF mass spectrometry produced about 50 to 100 mass peaks in the m/z range from 2,000 to 13,000 that mainly originate from highly abundant ribosomal proteins and from nucleic acid binding proteins (29, 30). Lipids, carbohydrates and nucleic acids are usually not detectable by MALDI TOF mass spectrometry using common experimental protocols for microbial identification. Due to the high level of standardisation and the manual mode of MS measurements, the SNR and the spectral reproducibility were particularly high (cf. spectra of *K. oxytoca* hand soap isolates in Figure 2A).
In contrast, the two vibrational spectroscopic methods, FTIR and NIR FT Raman spectroscopy, can be used to record signatures carrying signals from all main building blocks of microbial cells (e.g. proteins, lipids, nucleic acids, carbohydrates). Both methods are known to be suitable for differentiation, identification and classification of diverse microbial species down to the strain level (44, 48-51). In IR and Raman spectra of microbial samples the useful information is generally distributed over broad spectral regions; in combination with various methods of spectral pre-processing the spectral information can be extracted from the typically broad and complex spectral contours by multivariate pattern recognition methods such as cluster analysis, or analysis by neural networks (44, 52-55). As shown by Figures 2B and C, raw FTIR and FT Raman spectra of *K. oxytoca* isolates PHS-890 PHS-891 exhibit a remarkable degree of similarity. The latter aspect is quite typical: in vibrational spectroscopy, the usually very small, but highly characteristic differences between spectra of closely related microbial strains can often be visualised only by application of advanced methods of resolution enhancement (derivation, or Fourier self-deconvolution) and normalisation (52).
Whole-cell MALDI TOF MS was initially used to confirm species identification by the MALDI Biotyper system from Bruker Daltonics (56). All isolates were identified at the species level with log score values above 2.2 (data not shown). Unsupervised hierarchical cluster analysis (UHCA) of the mass peak tables revealed some degree of intraspecies diversity and the dendrogram (see Figure 3A) demonstrated a good correlation with phylogenetic data based on molecular typing methods indicating the discriminatory power of the method. Spectra from the liquid soap isolates formed a homogeneous subcluster clearly distinct from all other *Klebsiella* sp. isolates (cf. cluster 1 in the dendrogram of Figure 3A). The degree of heterogeneity within this cluster is comparable with the heterogeneity found in the remaining clusters 2-7. In the dendrograms of Figure 3 the horizontal red lines are arbitrarily set cutoff values used for group definition and differentiation of the *Klebsiella* sp. strains investigated.

Sets of discriminating biomarker peaks were tentatively assigned to protein identities by database searches using experimental peak positions of the MALDI TOF MS spectra. Major biomarker ions discriminating the hand soap isolates from the reference and control strains included the 30S ribosomal subunit protein S15 (MW of the $[M\text{-}Met+H]^+$

Faraday Discussions Accepted Manuscript

ion of 10,079 and 10,095 Da corresponding to an A→S amino acid exchange), integration host factor subunit alpha (m/z [M-Met+H]$^+$ of 11,237 and 11,251 Da, D→E) and 50S ribosomal protein L25 (m/z [M+H]+ of 10,650 and 10,677 (L→V; N→T; L→I) and m/z [M+H]+ 10,637 (unknown sequence).

Dendrograms obtained by UHCA of FTIR and NIR FT Raman spectra from the *Klebsiella* sp. data set are given in Figure 3B and Figure 3C, respectively. The dendrogram of the FTIR data (Figure 3B) demonstrates an acceptable level of separation between spectra from *K. pneumoniae* and *K. oxytoca* and - with the exception of *K. oxytoca* strains ATCC 13182 and CB4063 which form a mixed cluster - the presence of strain-specific clusters. Outliers are not included in this dendrogram and there is a high degree of similarity among spectra of the hand soap isolates: these isolates form a large and homogenous cluster, cluster 1, which interestingly does not show any signs of internal substructure, neither in terms of subclusters composed by replicate spectra of individual isolates, nor in terms of isolates from the different sample lots (see Table 1). Analysis of the NIR FT Raman spectra by UHCA indicates a somewhat reduced level of separation: although all spectra of the hand soap isolates also form one large cluster, we observe outliers (*K. oxytoca* CB4074.3 and CB4063.2) and the formation of two mixed clusters, cluster 2 which is formed by spectra from *K. oxytoca* strains Oman 61 and CCUG 15788 and cluster 3 composed by spectra from *K. oxytoca* strains CB4063 and ATCC 13182. Furthermore, the dendrogram of Figure 3C was obtained by omitting the spectral information contained in the CH-stretching region (2800-3050 cm^{-1}). This region is known to be dominated by contributions of lipids. Cluster analysis attempts of FT Raman spectra that included the CH-stretching region revealed a tendency towards a less clear separation between the individual strains and more outliers (not shown). Although the main structure of the dendrogram of Figure 3C is similar to the dendrogram structures of Figures 3A and 3B, this example suggests, that FT-Raman spectroscopy provides less taxonomic information than the two other phenotypic characterisation methods. Whether this observation represents a fundamental limitation of NIR FT Raman, or whether this is merely only a consequence of a reduced SNR – UHCA with derivative spectra does not improve the results - cannot be decided on the basis of the relatively small spectral data set investigated in this study.

Figure 4 shows the results of a combined cluster analysis of data acquired by the different microbial characterisation techniques MALDI TOF mass spectrometry, FTIR and FT Raman spectroscopy. Spectra obtained from the different biological replicates of the hand soap isolates fall in one single cluster (cluster 1). The dendrogram of Figure 4 indicates that replicates of each strain compose strain-specific clusters with a maximum intra-strain distance that is significantly lower than any distance between strains. This

Faraday Discussions Accepted Manuscript

and the absence of outliers suggest that all *K. oxytoca* hand soap isolates are phenotypically indistinguishable.

Comparison of techniques

Rapid and reliable bacterial typing systems are key tools to identify routes of transmission and infection, so that targeted control measures can be timely initiated. The potential of rapid phenotypic techniques for discriminating *K. oxytoca* isolates below the species level has been examined in this work and compared with the respective performance of DNA-based methods. Clustering of spectra from the *Klebsiella* isolate collection which included ten *K. oxytoca* strains from two different lots of a contaminated liquid hand soap product revealed a high level of concordance for all six microbial typing techniques used. MALDI TOF mass spectrometry and the vibrational spectroscopic methods of FTIR and NIR FT Raman spectroscopy are all physico-chemical techniques capable of rapid and non-subjective fingerprinting of bacteria. All three phenotype-based characterisation techniques have in common that spectral analysis is based on pattern recognition in which spectral fingerprints of unknowns are compared against libraries of bacterial reference spectra. Phenotypic characterisation techniques can be universally applied without prior knowledge of genus or species identity and are easily adaptable to routine analysis. Measurements can be automated under high throughput conditions providing results within minutes. The techniques have the advantages of low costs and of low requirements in terms of sample amount and sample preparation efforts. For the above reasons phenotypic methods have been suggested in the past as valuable alternatives for classifying microorganisms below the species level and for epidemiological typing (FTIR (57-63); Raman (64-67); MALDI TOF MS (68-73).

The depth of taxonomic resolution attainable by the different phenotypic characterisation techniques has been the subject of intense research. While MALDI TOF mass spectrometry is particularly robust for diagnostics at the genus and species level with some potential for strain level identification (71), FTIR spectroscopy suffers from the fact that a high level of intraspecies diversity of the IR spectral fingerprints may lead to reduced accuracy of species identification (48). On the other hand, the discriminative potential of FTIR spectroscopy for microbial typing is well documented (49, 50, 74, 75). This discrepancy and the proposed higher strain typing capability of FTIR spectroscopy have been explained by the particular information content of the IR spectra carrying information from several classes of biomolecules, including lipids, nucleic acids and carbohydrates (lipopolysaccharides, LPS). The sensitivity of IR spectroscopy towards carbohydrates and LPS present in outer membrane structures of Gram-negative bacteria, including *Klebsiella*, may constitute the compositional basis for its ability to discriminate bacterial strains. In contrast, MALDI TOF MS detects mainly conserved

Faraday Discussions Accepted Manuscript

ribosomal and to a lesser extend nucleic acid-binding proteins which exhibit only limited strain specificity. Whether the aforementioned considerations are valid also for the complementary vibrational spectroscopic method of NIR FT Raman spectroscopy, is difficult to judge on the basis of the available data. The dendrograms produced from the experimental IR and Raman spectra suggest, however, a reduced taxonomic resolution of NIR FT Raman compared to FTIR spectroscopy (see outliers in Figure 3C). This and the finding of a generally reduced separability of the individual microbial strains in the NIR FT Raman setup are likely a result of an only moderate SNR. It must be stated however, that the number of strains used in this study is rather limited and does not allow to draw general conclusions. Furthermore, the taxonomic resolution of all three phenotypic characterisation techniques is certainly taxon-dependent (76).

Vibrational spectral fingerprints are dependent on the metabolic state of the cells under study. Growth and nutrient conditions may exert a notable influence on the spectral patterns. Therefore, microbial characterisation by FTIR and NIR FT Raman spectroscopy requires a high level of standardisation to assure adequate spectral reproducibility. Interestingly, the standardisation requirements are less strict in a MALDI-TOF MS-based workflow (see above) whereas DNA-based techniques are considered to be invariant to the cell's metabolic state. The presence of spectral databases suitable for exchange between different laboratories is a further requirement of a phenotype-based microbial identification methodology. While such libraries are not available for IR or Raman-based applications, MALDI TOF MS has been already implemented in the routine practice of microbial identification, mainly because of the fact that commercial spectral databases are today readily available (77). It was certainly the higher robustness of the mass spectral fingerprints and – compared to IR and Raman spectroscopy – the lower requirements in terms of standardisation (growth conditions, type of sample preparation) which have enforced the development of commercial MALDI TOF MS databases. This is to our opinion the main reason why MS and not IR or Raman typing has found its way in the microbiological routine.

Among DNA-based typing techniques PFGE is currently the gold standard for epidemiological studies and outbreak detection (78). PFGE has a very high discriminative power but is labour-intensive and offers only low sample throughput. Using PFGE for source-tracking purposes it should be kept in mind, that differences in PFGE profiles of isogenic strains may be attributed to differences in the contents of mobile genetic elements (e.g. plasmids or prophages) (79). In the highly related liquid hand soap isolates, PFGE revealed the presence of additional single bands in some isolates, which are likely due to the carriage of large plasmids. MLST also has a high discriminatory power and has the advantages to provide highly standardised data readily transportable between laboratories. However, because genetic mutations accumulate

Faraday Discussions Accepted Manuscript

slowly in housekeeping genes, classical MLST schemes may be more suited to study long-term global epidemiology rather than local epidemiology. Like PFGE, MLST is known as a labour-intensive, time-consuming and costly technique. Higher resolution methods utilizing whole genome sequencing (WGS) will clearly improve surveillance and tracing pathogens and are likely to replace other typing methods in the near future. WGS has already been used for back-tracking outbreaks including *K. pneumonia* (80). SNP analysis using WGS shows great promise for use as an epidemiological typing tool. In our study it provided the highest discriminatory power and allowed differentiation even of highly related isogenic strains. In fact, WGS was the only method that was able to separate the liquid hand soap isolates into distinct lot-specific clades (see inset of Figure 1C). WGS is often perceived as a relatively complex and time-consuming technique. However, this is not always true, as time-consuming cultivation can be sometimes avoided. In cases where a sufficient bacterial load is available and DNA from other sources is absent sequencing can be performed directly from the isolated sample DNA. In such cases, WGS has a major speed advantage over methods that require cultivation. It is important to emphasise that – as in the present study – WGS analysis from isolated sample DNA is not always feasible.

In conclusion, the present study suggests that phenotype-based screening could assist in dereplication of *K. oxytoca* isolates with different profiles. Phenotype methods may be applied as rapid alternatives to PFGE and MLST for initial classification, or source-tracking of bacterial pathogens before results from DNA-based techniques are available. While MALDI-TOF MS will be the method of choice due to its availability in many microbiological laboratories, FTIR spectroscopy appears to be particularly well-suited among the phenotype-based techniques because of its high discriminatory power, simplicity and the advantage of low costs. The results of NIR FT Raman measurements suggest that under the given experimental conditions (NIR excitation at λ=1064 nm, Ge-detector, FT setup) this method is not optimally suitable for microbial typing.

Whole genome sequencing is certainly the most powerful technique which still could benefit from a combination with rapid phenotyping methods. Indeed, techniques like MALDI-TOF MS or FTIR spectroscopy could assist and guide time-consuming data analysis procedures of WGS investigations. The combination of genotyping and phenotyping techniques would be helpful to overcome the inherently slowness of a purely genotype-based identification system. For further pathogen characterisation e.g. regarding the presence of virulence or antibiotic resistance genes and extrachromosomal elements and unambiguous confirmation of the typing results obtained by phenotypic analyses WGS will certainly play a key role in the future.

Faraday Discussions Accepted Manuscript

ACKNOWLEDGEMENTS

The authors are thankful for the excellent technical assistance of Marett Splett, Jane Kowall and Anna Louisa Hauffe at the BfR Berlin and of Andy Schneider and Stefanie Herfort from the Proteomics and Spectroscopy Unit (ZBS6) at the RKI Berlin. Furthermore, the authors are grateful to Dr. Lothar Beutin (BfR Berlin) and Dr. Daniela Jacob (RKI Berlin, Highly Pathogenic Microorganisms/ZBS2) for providing selected strains of *K. oxytoca*. The authors acknowledge intramural funding at the German Federal Institute for Risk Assessment (BfR), Grants 47-001 and 1322-621.

Faraday Discussions Accepted Manuscript

FIGURE LEGENDS

Figure 1. Genotyping results of *K. oxytoca* hand soap isolates, clinical isolates and reference strains.

A. Dendrogram produced by multi-locus sequence typing (MLST) as described by Herzog et al. (21). For MLST seven housekeeping genes (*gapA, infB, mdh, pgi, phoE, rpoB, tonB*) were used.

B. Analysis of pulsed-field gel electrophoresis (PFGE) patterns (XbaI restriction) of *K. oxytoca* and *K. pneumoniae*.

C. SNP analysis of whole genome sequencing (WGS) data of *K. oxytoca*. Inset: Detail view of the clade containing only the data of the hand soap isolates.

Figure 2. Spectra acquired by means of phenotyping methods (MALDI TOF MS, FTIR and FT Raman spectroscopy) from cultures of *K. oxytoca* isolates PHS-890 and PHS-891 (hand soap isolates). Each spectral trace was generated by averaging the available spectra from three biological replicates. For more clarity spectra were min-max normalised and shifted along the y-axis.

A. MALDI TOF mass spectra of the *K. oxytoca* isolates PHS-890 (upper trace) and PHS-891 (lower trace).

B. FTIR spectra in the mid-IR spectral region ranging from 500 to 4000 cm^{-1}.

C. NIR FT Raman spectra of *K. oxytoca* isolates. The wavelength λ of NIR Raman laser was 1064 nm.

Figure 3. Dendrograms obtained by unsupervised hierarchical cluster analysis (UHCA) of MALDI TOF mass, FTIR and NIR FT Raman spectra from *Klebsiella* spp. strains. Each microbial strain was characterised in triplicate. Horizontal lines (red, dotted) indicate the similarity cutoff value used arbitrarily for group definition and differentiation of *Klebsiella* sp. strains below the species level.

A. Dendrogram produced from MALDI TOF mass spectra of *Klebsiella* spp. strains

A. Dendrogram obtained by clustering FTIR spectra

C. Dendrogram obtained by clustering NIR FT Raman spectra

Figure 4. Dendrogram of a combined analysis of microbial phenotypes. Spectral data from *Klebsiella* spp. cultures were obtained by MALDI TOF mass spectrometry, FTIR and FT Raman spectroscopy (triplicate measurements of individual microbial cultures by each technique). Microbial spectra collected by the different phenotyping techniques were subsequently individually analysed by unsupervised hierarchical clustering (UHCA). Cluster membership functions were established for each individual culture by

Faraday Discussions Accepted Manuscript

setting the number of clusters to 7 (MALDI, FTIR), or 8 (Raman). The membership functions served as inputs for a top-level UHCA. The dendrogram of the combined analysis suggests that all *K. oxytoca* hand soap isolates are indistinguishable (see text for details).

Supplemental Figure S1.
Radial Dendrogram produced by MLST on the basis of seven housekeeping genes (*gapA, infB, mdh, pgi, phoE, rpoB, tonB*). Allele sequences were concatenated to generate a 3,003 bp concatemer for each strain. A phylogenetic tree was constructed in MEGA ver. 6 using UPGMA. Distances were computed using Maximum Composite Likelihood and the reliability of the tree was assessed by bootstrap analysis with 1,000 replicates. Sequences of 113 isolates present in the PubMLST database were included in the analysis. The isolates analysed are displayed in bold.

Faraday Discussions Accepted Manuscript

TABLES

Table 1. Overview on strains of *Klebsiella* spp. isolates included in this study.

Strain	Country of origin and year of isolation	Source
K. oxytoca PHS-890	Greece, 2013	Hand soap, lot I
K. oxytoca PHS-891	Greece, 2013	Hand soap, lot I
K. oxytoca PHS-892	Greece, 2013	Hand soap, lot I
K. oxytoca PHS-893	Greece, 2013	Hand soap, lot I
K. oxytoca PHS-894	Greece, 2013	Hand soap, lot I
K. oxytoca PHS-895	Greece, 2013	Hand soap, lot II
K. oxytoca PHS-896	Greece, 2013	Hand soap, lot II
K. oxytoca PHS-897	Greece, 2013	Hand soap, lot II
K. oxytoca PHS-898	Greece, 2013	Hand soap, lot II
K. oxytoca PHS-899	Greece, 2013	Hand soap, lot II
K. oxytoca CB4063	Germany, 1995	Child, Enteritis
K. oxytoca CB4074	Germany, 1995	Child, Enteritis
K. oxytoca CB4072	Germany, 1995	Child, Enteritis
K. oxytoca CCUG 15788	Sweden	Environmental
K. oxytoca Oman 61	Oman, 2011	Clinical isolate
K. oxytoca ATCC 13182	USA	Pharyngeal tonsil
K. pneumoniae subsp. ozaenae ATCC 25926	Belgium	Human blood

Faraday Discussions Accepted Manuscript

Table S1. Allelic profiles of *Klebsiella* sp. isolates and reference strains analysed by MLST

	gapA	*infB*	*mdh*	*pgI*	*phoE*	*rpoB*	*tonB*	*cluster*
K. oxytoca PHS-890	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca PHS-891	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca PHS-892	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca PHS-893	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca PHS-894	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca PHS-895	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca PHS-896	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca PHS-897	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca PHS-898	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca PHS-899	3	5	21	13	20 (^{62}C→T)	6	12	B1
K. oxytoca CB4063	5	6	19	10	46	5	9 (^{75}A→G ^{78}G→A)	B2
K. oxytoca CB4074	1	2	2	1	4	1	2	A
K. oxytoca CB4072	3	4	15 (^{330}A →T)	8	25	3	4	B1
K. oxytoca Oman 61	1	3	13	24	2 (^{321}G →A)	10	10	A
K. oxytoca ATCC 13182	2	2	2	3	19	2	2	A
K. oxytoca CCUG 15788	16 (16 diff.)	25 (20 diff.)	39 (49 diff.)	23 (20 diff.)	15 (35 diff.)	24 (17 diff.)	34 (55 diff.)	out

Faraday Discussions Accepted Manuscript

REFERENCES

1. Eiref SD, Leitman IM, Riley W. Hand sanitizer dispensers and associated hospital-acquired infections: Friend or fomite? Surgical Infections. 2012;13(3):137-40.
2. Brooks SE, Walczak MA, Hameed R, Coonan P. Chlorhexidine resistance in antibiotic-resistant bacteria isolated from the surfaces of dispensers of soap containing chlorhexidine. Infection Control and Hospital Epidemiology. 2002;23(11):692-5.
3. Brooks SE, Walczak MA, Malcolm S, Hameed R. Intrinsic *Klebsiella pneumoniae* contamination of liquid germicidal hand soap containing chlorhexidine. Infection Control and Hospital Epidemiology. 2004;25(10):883-5.
4. Drancourt M. Nosocomial *Serratia marcescens* infections associated with extrinsic contamination of a liquid nonmedicated soap. Infection Control and Hospital Epidemiology. 2000;21(3):196-9.
5. Buffet-Bataillon S, Rabier V, Bétrémieux P, Beuchée A, Bauer M, Pladys P, et al. Outbreak of *Serratia marcescens* in a neonatal intensive care unit: contaminated unmedicated liquid soap and risk factors. Journal of Hospital Infection. 2009;72(1):17-22.
6. Lanini S, D'Arezzo S, Puro V, Martini L, Imperi F, Piselli P, et al. Molecular epidemiology of a *Pseudomonas aeruginosa* hospital outbreak driven by a contaminated disinfectant-soap dispenser. PLoS ONE. 2011;6(2).
7. Aktaş E, Taşpinar E, Alay D, Ögedey EDK, Külah C, Cömert F. Extrinsic contamination of liquid soap with various gram-negative bacteria in a hospital in Turkey. Infection Control and Hospital Epidemiology. 2010;31(11):1199-201.
8. Weber DJ, Rutala WA, Sickbert-Bennett EE. Outbreaks associated with contaminated antiseptics and disinfectants. Antimicrobial agents and chemotherapy. 2007;51(12):4217-24.
9. Lorenz LA, Ramsay BD, Goeres DM, Fields MW, Zapka CA, Macinga DR. Evaluation and remediation of bulk soap dispensers for biofilm. Biofouling. 2012;28(1):99-109.
10. Chattman M, Maxwell SL, Gerba CP. Occurrence of heterotrophic and coliform bacteria in liquid hand soaps from bulk refillable dispensers in public facilities. Journal of Environmental Health. 2011;73(7):26-9.
11. Zapka CA, Campbell EJ, Maxwell SL, Gerba CP, Dolan MJ, Arbogast JW, et al. Bacterial hand contamination and transfer after use of contaminated bulk-soap-refillable dispensers. Applied and environmental microbiology. 2011;77(9):2898-904.
12. Madani TA, Alsaedi S, James L, Eldeek BS, Jiman-Fatani AA, Alawi MM, et al. *Serratia marcescens*-contaminated baby shampoo causing an outbreak among newborns at King Abdulaziz University Hospital, Jeddah, Saudi Arabia. Journal of Hospital Infection. 2011;78(1):16-9.
13. Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerging infectious diseases. 2001;7(3):382-9.
14. Swaminathan B, Gerner-Smidt P, Ng LK, Lukinmaa S, Kam KM, Rolando S, et al. Building PulseNet International: An interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. Foodborne Pathogens and Disease. 2006;3(1):36-50.
15. Arbeit RD, Arthur M, Dunn R, Kim C, Selander RK, Goldstein R. Resolution of recent evolutionary divergence among *Escherichia coli* from related lineages: The application of pulsed field electrophoresis to molecular epidemiology. Journal of Infectious Diseases. 1990;161(2):230-5.
16. Hansen DS, Skov R, Benedí JV, Sperling V, Kolmos HJ. *Klebsiella* typing: Pulsed-field gel electrophoresis (PFGE) in comparison with O:K-serotyping. Clinical Microbiology and Infection. 2002;8(7):397-404.

Faraday Discussions Accepted Manuscript

17. Decré D, Burghoffer B, Gautier V, Petit JC, Arlet G. Outbreak of multi-resistant *Klebsiella oxytoca* involving strains with extended-spectrum β-lactamases and strains with extended-spectrum activity of the chromosomal β-lactamase. Journal of Antimicrobial Chemotherapy. 2004;54(5):881-8.
18. Cheng VCC, Yam WC, Tsang LL, Yau MCY, Siu GKH, Wong SCY, et al. Epidemiology of *Klebsiella oxytoca*-associated diarrhea detected by simmons citrate agar supplemented with inositol, tryptophan, and bile salts. Journal of Clinical Microbiology. 2012;50(5):1571-9.
19. Tsakris A, Poulou A, Markou F, Pitiriga V, Piperaki ET, Kristo I, et al. Dissemination of clinical isolates of *Klebsiella oxytoca* harboring CMY-31, VIM-1, and a new OXY-2-type variant in the community. Antimicrobial agents and chemotherapy. 2011;55(7):3164-8.
20. Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proceedings of the National Academy of Sciences of the United States of America. 1998;95(6):3140-5.
21. Herzog KAT, Schneditz G, Leitner E, Feierl G, Hoffmann KM, Zollner-Schwetz I, et al. Genotypes of *Klebsiella oxytoca* isolates from patients with nosocomial pneumonia are distinct from those of isolates from patients with antibiotic-associated hemorrhagic colitis. Journal of Clinical Microbiology. 2014;52(5):1607-16.
22. Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. Bacterial whole-genome sequencing revisited: Portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. Journal of Clinical Microbiology. 2014;52(7):2365-70.
23. Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agersø Y, Lund O, et al. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. Journal of Antimicrobial Chemotherapy. 2013;68(4):771-7.
24. Hammerum AM, Hansen F, Skov MN, Stegger M, Andersen PS, Holm A, et al. Investigation of a possible outbreak of carbapenem-resistant *Acinetobacter baumannii* in Odense, Denmark using PFGE, MLST and whole-genome-based SNPs. Journal of Antimicrobial Chemotherapy. 2014;70(7):1965-8.
25. Leekitcharoenphon P, Kaas RS, Thomsen MC, Friis C, Rasmussen S, Aarestrup FM. snpTree--a web-server to identify and construct SNP trees from whole genome sequence data. BMC genomics. 2012;13 Suppl 7.
26. Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O, Aarestrup FM. Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. PLoS ONE. 2014;9(2).
27. Stucki D, Ballif M, Bodmer T, Coscolla M, Maurer AM, Droz S, et al. Tracking a tuberculosis outbreak over 21 years: Strain-specific single-nucleotide polymorphism typing combined with targeted whole-genome sequencing. Journal of Infectious Diseases. 2015;211(8):1306-16.
28. Taylor AJ, Lappi V, Wolfgang WJ, Lapierre P, Palumbo MJ, Medus C, et al. Characterization of foodborne outbreaks of *Salmonella enterica* serovar enteritidis with whole-genome sequencing single nucleotide polymorphism-based analysis for surveillance and outbreak detection. Journal of Clinical Microbiology. 2015;53(10):3334-40.
29. Ryzhov V, Fenselau C. Characterization of the protein subset desorbed by MALDI from whole bacterial cells. Analytical Chemistry. 2001;73(4):746-50.
30. Pineda FJ, Antoine MD, Demirev PA, Feldman AB, Jackman J, Longenecker M, et al. Microorganism identification by matrix-assisted laser/dissorption ionization mass spectrometry and model-derived ribosomal protein biomarkers. Analytical Chemistry. 2003;75(15):3817-22.

Faraday Discussions Accepted Manuscript

31. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total-genome-sequenced bacteria. Journal of Clinical Microbiology. 2012;50(4):1355-61.
32. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9.
33. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980;16(2):111-20.
34. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406-25.
35. Hunter SB, Vauterin P, Lambert-Fair MA, Van Duyne MS, Kubota K, Graves L, et al. Establishment of a universal size standard strain for use with the pulsenet standardized pulsed-field gel electrophoresis protocols: Converting the national databases to the new size standard. Journal of Clinical Microbiology. 2005;43(3):1045-50.
36. Hammerl JA, Lasch P, Nitsche A, Dabrowski PW, Hahmann H, Wicke A, et al. Draft Genome Sequences of *Klebsiella oxytoca* Isolates Originating from a Highly Contaminated Liquid Hand Soap Product. Genome Announc. 2015;3(4).
37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.
38. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-303.
39. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010;59(3):307-21.
40. Helm D, Labischinski H, Naumann D. Elaboration of a procedure for identification of bacteria using Fourier-Transform IR spectral libraries: A stepwise correlation approach. Journal of Microbiological Methods. 1991;14(2):127-42.
41. Helm D, Labischinski H, Schallehn G, Naumann D. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. Journal of General Microbiology. 1991;137(1):69-79.
42. Naumann D, Helm D, Labischinski H. Microbiological characterizations by FT-IR spectroscopy. Nature. 1991;351(6321):81-2.
43. Lasch P, Nattermann H, Erhard M, Stämmler M, Grunow R, Bannert N, et al. MALDI-TOF mass spectrometry compatible inactivation method for highly pathogenic microbial cells and spores. Analytical Chemistry. 2008;80(6):2026-34.
44. Lasch P, Naumann D. Infrared Spectroscopy in Microbiology. Encyclopedia of Analytical Chemistry. 2015.
45. Lasch P. MicrobeMS: A Matlab Toolbox for Analysis of Microbial MALDI-TOF Mass Spectra. http://www.microbe-ms.com. 2015.
46. Lasch P, Wahab T, Weil S, Palyi B, Tomaso H, Zange S, et al. Identification of Highly Pathogenic Microorganisms by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: Results of an Interlaboratory Ring Trial. J Clin Microbiol. 2015;53(8):2632-40.
47. Lasch P, Fleige C, Stämmler M, Layer F, Nübel U, Witte W, et al. Insufficient discriminatory power of MALDI-TOF mass spectrometry for typing of *Enterococcus faecium* and *Staphylococcus aureus* isolates. Journal of Microbiological Methods. 2014;100(1):58-69.
48. Wenning M, Breitenwieser F, Konrad R, Huber I, Busch U, Scherer S. Identification and differentiation of food-related bacteria: a comparison of FTIR spectroscopy and MALDI-TOF mass spectrometry. J Microbiol Methods. 2014.
49. Beutin L, Wang Q, Naumann D, Han W, Krause G, Leomil L, et al. Relationship between O-antigen subtypes, bacterial surface structures and O-antigen gene clusters in *Escherichia*

Faraday Discussions Accepted Manuscript

coli O123 strains carrying genes for Shiga toxins and intimin. Journal of medical microbiology. 2007;56(PART 2):177-84.
50. Grunert T, Wenning M, Barbagelata MS, Fricker M, Sordelli DO, Buzzola FR, et al. Rapid and reliable identification of *Staphylococcus aureus* capsular serotypes by means of artificial neural network-assisted fourier transform infrared spectroscopy. Journal of Clinical Microbiology. 2013;51(7):2261-6.
51. Rebuffo-Scheer CA, Kirschner C, Staemmler M, Naumann D. Rapid species and strain differentiation of non-tubercoulous mycobacteria by Fourier-Transform Infrared microspectroscopy. Journal of Microbiological Methods. 2007;68(2):282-90.
52. Lasch P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging. Chemom Intell Lab Syst. 2012;117:100-14.
53. Helm D, Labischinski H, Schallehn G, Naumann D. Classification and identification of bacteria by Fourier-transform infrared spectroscopy. Journal of general microbiology. 1991;137(1):69-79.
54. Schmitt J, Udelhoven T, Löchte T, Flemming HC, Naumann D. Artificial Neural Network applied to FTIR and FT-Raman spectra in biomedical applications. AIP Conf Proc. 1998;430:260-3.
55. Naumann D. FT-IR spectroscopy of microorganisms at the Robert Koch-Institute: Experiences gained during a successful project. Proc SPIE. 2008;6853.
56. Maier T, Klepel S, Renner Z, Kostrzewa M. Fast and reliable MALDI-TOF MS-based microorganism identification. Nature Methods. 2006;3:324-34.
57. Rebuffo-Scheer CA, Schmitt J, Scherer S. Differentiation of *Listeria monocytogenes* serovars by using artificial neural network analysis of fourier-transformed infrared spectra. Applied and environmental microbiology. 2007;73(3):1036-40.
58. Kim S, Reuhs BL, Mauer LJ. Use of Fourier transform infrared spectra of crude bacterial lipopolysaccharides and chemometrics for differentiation of *Salmonella enterica* serotypes. Journal of applied microbiology. 2005;99(2):411-7.
59. Kuhm AE, Suter D, Felleisen R, Rau J. Identification of *Yersinia enterocolitica* at the species and subspecies levels by fourier transform infrared spectroscopy. Applied and environmental microbiology. 2009;75(18):5809-13.
60. Patrícia Coutinho C, Sá-Correia I, Almeida Lopes J. Use of Fourier transform infrared spectroscopy and chemometrics to discriminate clinical isolates of bacteria of the *Burkholderia cepacia* complex from different species and ribopatterns. Analytical and Bioanalytical Chemistry. 2009;394(8):2161-71.
61. Vaz M, Meirinhos-Soares L, Sousa CCS, Ramirez M, Melo-Cristino J, Lopes JA. Serotype discrimination of encapsulated *Streptococcus pneumoniae* strains by Fourier-transform infrared spectroscopy and chemometrics. Journal of Microbiological Methods. 2013;93(2):102-7.
62. Davis R, Mauer LJ. Subtyping of *Listeria monocytogenes* at the haplotype level by Fourier transform infrared (FT-IR) spectroscopy and multivariate statistical analysis. International Journal of Food Microbiology. 2011;150(2-3):140-9.
63. Preisner O, Guiomar R, Machado J, Menezes JC, Lopes JA. Application of fourier transform infrared spectroscopy and chemometrics for differentiation of *Salmonella enterica* serovar enteritidis phage types. Applied and environmental microbiology. 2010;76(11):3538-44.
64. Guyot K, Biran V, Doit C, Moissenet D, Guillard T, Brasme L, et al. Raman spectroscopic analysis of the clonal and horizontal spread of CTX-M-15-producing *Klebsiella pneumoniae* in a neonatal intensive care unit. European Journal of Clinical Microbiology and Infectious Diseases. 2012;31(10):2827-34.
65. Willemse-Erix D, Bakker-Schut T, Slagboom-Bax F, Jachtenberg JW, Lemmens-Den Toom N, Papagiannitsis CC, et al. Rapid typing of extended-spectrum β-lactamase- and

Faraday Discussions Accepted Manuscript

carbapenemase-producing *Escherichia coli* and *Klebsiella pneumoniae* isolates by use of spectracell RA. Journal of Clinical Microbiology. 2012;50(4):1370-5.
66. Willemse-Erix DFM, Scholtes-Timmerman MJ, Jachtenberg JW, Van Leeuwen WB, Horst-Kreft D, Schut TCB, et al. Optical fingerprinting in bacterial epidemiology: Raman spectroscopy as a real-time typing method. Journal of Clinical Microbiology. 2009;47(3):652-9.
67. Maquelin K, Dijkshoorn L, Van Der Reijden TJK, Puppels GJ. Rapid epidemiological analysis of *Acinetobacter* strains by Raman spectroscopy. Journal of Microbiological Methods. 2006;64(1):126-31.
68. Treviño M, Navarro D, Barbeito G, García-Riestra C, Crespo C, Regueiro BJ. Molecular and epidemiological analysis of nosocomial carbapenem-resistant *Klebsiella* spp. using repetitive extragenic palindromic-polymerase chain reaction and matrix-assisted laser desorption/ionization-time of flight. Microbial Drug Resistance. 2011;17(3):433-42.
69. Berrazeg M, Diene SM, Drissi M, Kempf M, Richet H, Landraud L, et al. Biotyping of Multidrug-Resistant *Klebsiella pneumoniae* Clinical Isolates from France and Algeria Using MALDI-TOF MS. PLoS ONE. 2013;8(4).
70. Mencacci A, Monari C, Leli C, Merlini L, De Carolis E, Vella A, et al. Typing of nosocomial outbreaks of *Acinetobacter baumannii* by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. Journal of Clinical Microbiology. 2013;51(2):603-6.
71. Sandrin TR, Goldstein JE, Schumaker S. MALDI TOF MS profiling of bacteria at the strain level: A review. Mass Spectrometry Reviews. 2013;32(3):188-217.
72. Dieckmann R, Malorny B. Rapid screening of epidemiologically important *Salmonella enterica* subsp. *enterica* serovars by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Applied and environmental microbiology. 2011;77(12):4136-46.
73. Dieckmann R, Helmuth R, Erhard M, Malorny B. Rapid classification and identification of *salmonellae* at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Applied and environmental microbiology. 2008;74(24):7767-78.
74. Sockalingum GD, Bouhedja W, Pina P, Allouch P, Mandray C, Labia R, et al. ATR-FTIR spectroscopic investigation of imipenem-susceptible and -resistant *Pseudomonas aeruginosa* isogenic strains. Biochemical and biophysical research communications. 1997;232(1):240-6.
75. Wenning M, Buchl NR, Scherer S. Species and strain identification of lactic acid bacteria using FTIR spectroscopy and artificial neural networks. J Biophotonics. 2010;3(8-9):493-505.
76. Ghyselinck J, Van Hoorde K, Hoste B, Heylen K, De Vos P. Evaluation of MALDI-TOF MS as a tool for high-throughput dereplication. J Microbiol Methods. 2011;86(3):327-36.
77. Faron ML, Buchan BW, Hyke J, Madisen N, Lillie JL, Granato PA, et al. Multicenter Evaluation of the Bruker MALDI Biotyper CA System for the Identification of Clinical Aerobic Gram-Negative Bacterial Isolates. PLoS One. 2015;10(11):e0141350.
78. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl J, Laurent F, et al. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill. 2013;18(4):20380.
79. Soto SM, Rodríguez I, Rodicio MR, Vila J, Mendoza MC. Detection of virulence determinants in clinical strains of Salmonella enterica serovar Enteritidis and mapping on macrorestriction profiles. Journal of medical microbiology. 2006;55(4):365-73.
80. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Henderson DK, Palmore TN, et al. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. Sci Transl Med. 2012;4(148).

Faraday Discussions Accepted Manuscript

Faraday Discussions Accepted Manuscript