

Identification of pathogens from native urine samples by MALDI-TOF/TOF tandem mass spectrometry

Antonio Starcevic (✉ astar@pbf.hr)

Sveuciliste u Zagrebu Prehrambeno-Bioteholoski Fakultet <https://orcid.org/0000-0003-2386-2124>

Damir Oros

Sveuciliste u Zagrebu Prehrambeno-Bioteholoski Fakultet

Marina Ceprnja

Sveuciliste u Zagrebu Farmaceutsko-biokemijski fakultet

Jurica Zucko

Sveuciliste u Zagrebu Prehrambeno-Bioteholoski Fakultet

Mario Cindrić

Institut Ruder Boskovic

Amela Hozic

Institut Ruder Boskovic

Jasenska Skrlin

Klinicka Bolnica Dubrava

Karmela Barisic

Sveuciliste u Zagrebu Farmaceutsko-biokemijski fakultet

Ena Melvan

Macquarie University

Ksenija Uroic

Sveuciliste u Zagrebu Prehrambeno-Bioteholoski Fakultet

Blazenka Kos

Sveuciliste u Zagrebu Prehrambeno-Bioteholoski Fakultet

Research

Keywords: urine, sample preparation, pathogen identification, proteomics, MALDI-TOF/TOF, 16S rRNA sequencing

Posted Date: May 20th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28509/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on June 23rd, 2020. See the published version at <https://doi.org/10.1186/s12014-020-09289-4>.

1 **Identification of pathogens from native urine samples by MALDI-**
2 **TOF/TOF tandem mass spectrometry**

3 **Damir Oros¹, Marina Ceprnja², Jurica Zucko¹, Mario Cindric³, Amela Hozic³, Jasenka**
4 **Skrlin⁴, Karmela Barisic⁵, Ena Melvan⁶, Ksenija Uroic¹, Blazenka Kos¹ and Antonio**
5 **Starcevic¹**

6
7 ¹*Department of Biochemical Engineering, Faculty of Food Technology and Biotechnology,*
8 *Zagreb University, Zagreb, Croatia*

9 ²*Special Hospital Agram, Polyclinic Zagreb, Biochemical Laboratory, Zagreb, Croatia*

10 ³*Division of Molecular Medicine, Ruder Boskovic Institute, Zagreb, Croatia*

11 ⁴*Department for Clinical Microbiology and Hospital Infection, University Hospital Dubrava,*
12 *Zagreb, Croatia*

13 ⁵*Faculty of Pharmacy and Biochemistry, Zagreb University*

14 ⁶*Department of Biological Science, Faculty of Science, Macquarie University, Sydney, Australia*

15
16
17 **Corresponding author: Antonio Starcevic, astar@pbf.hr**

26 **Abstract**

27 **Background:** Reliable high-throughput microbial pathogen identification in human urine
28 samples is crucial for patients with cystitis symptoms. Currently employed methods are time-
29 consuming and could lead to unnecessary or inadequate antibiotic treatment. Purpose of this
30 study was to assess the potential of mass spectrometry for uropathogen identification from a
31 native urine sample.

32 **Methods:** In total, 16 urine samples having more than 10^5 CFU/mL were collected from clinical
33 outpatients. These samples were analysed using standard urine culture methods, followed by 16S
34 rRNA gene sequencing serving as control and here described culture-independent MALDI-
35 TOF/TOF MS method being tested.

36 **Results:** Here we present advantages and disadvantages of bottom-up proteomics, using
37 MALDI-TOF/TOF tandem mass spectrometry, for culture-independent identification of
38 uropathogens (e.g. directly from urine samples). The direct approach provided reliable
39 identification of bacteria at the genus level in monobacterial samples. Taxonomic identifications
40 obtained by proteomics were compared both to standard urine culture test used in clinics and
41 genomic test based on 16S rRNA sequencing.

42 **Conclusions:** Our findings indicate that mass spectrometry has great potential as a reliable high-
43 throughput tool for microbial pathogen identification in human urine samples. In this case, the
44 MALDI-TOF/TOF, was used as an analytical tool for the determination of bacteria in urine
45 samples, and the results obtained emphasize high importance of storage conditions and sample
46 preparation method impacting reliability of MS2 data analysis. The proposed method is simple
47 enough to be utilized in existing clinical settings and is highly suitable for suspected single
48 organism infectious etiologies. Further research is required in order to identify pathogens in
49 polymicrobial urine samples.

50 **Keywords:** urine, sample preparation, pathogen identification, proteomics, MALDI-TOF/TOF,
51 16S rRNA sequencing

52 **Background**

53 Urinary tract infections (UTIs) are the most common form of bacterial infections both in the
54 general population and in hospital patients, attributing to nearly 25% of all infections [1]. UTIs
55 are much more common in females than males. It is estimated that 40-50% of women will
56 develop a UTI during their lives, and approximately 33% of women will have recurrent acute
57 uncomplicated UTI [2]. Common primary bacterial uropathogens are *Escherichia coli*,
58 *Staphylococcus saprophyticus*, *Enterococcus spp*, *Proteus mirabilis*, and *Klebsiella pneumoniae*.
59 While most common secondary uropathogens are *Staphylococcus aureus*, *Klebsiella oxytoca*,
60 *Pseudomonas aeruginosa*, *Streptococcus agalactiae* and fungal pathogen *Candida spp* [3, 4, 5,
61 6]. Approximately 60-80% of all uncomplicated bacterial UTIs are caused by *E. coli*.
62 Researchers have recognized that urine is not sterile and confirmed the importance of resident
63 bacterial flora (urinary microbiota) in the lower urinary tract. Resident urinary microbiota is
64 mostly composed of *Lactobacillus gasseri*, *Corynebacterium coyleae*, *Actinobaculum schaalii*,
65 *Aerococcus urinae*, *Gardnerella vaginalis*, *Streptococcus anginosus*, *Streptococcus epidermis*,
66 *Actinomyces neuui* and *Bifidobacterium spp* [7, 8].

67 In order to identify microorganisms in clinical microbiology laboratories, most used
68 microbiological techniques are still based on cultivation on different culture media [9]. Despite
69 advances in genomics and proteomics, that “standard” urine culture method is still the golden
70 standard for the diagnosis of UTIs. Urine samples containing more than 10^5 CFU/mL of a single
71 microbial species usually indicate clinical relevance. However, there are significant
72 shortcomings to these cultivation-oriented methods. The first limitation is the time required for
73 the cultivation of microorganisms and subsequent identification [10]. Standard incubation times
74 range from 12 to 24 h in order to enable reliable detection of the presence of uropathogens [11].
75 The second limitation is the requirement for fresh urine samples. Some of these limitations may
76 result in overall negative urine cultures in up to 80% of cases, in many microbiology laboratories
77 [12]. Unfortunately, a variety of sampling methods and inappropriate specimen transport are

78 major cause of pre-analytical errors [13].

79 Various methods have been used for detection of microorganisms in clinical microbiology [14,
80 15, 16]. For fast screening of urine samples flow cytometry (such as Sysmex analyser) has been
81 used. However, urine flow cytometer is not able to provide bacteria identification [17, 18].

82 Genomic methods relying on DNA analysis, such as SeptiFast, FilmArray or GeneXpert, are
83 being used, however they are still not approved by the FDA for UTI identification [14]. Usage of
84 real-time PCR methods in the identification of uropathogens has been proven as feasible [19],
85 however it is limited in its scope. Techniques using DNA sequencing regularly show more
86 sensitivity compared to standard urine culture test. For this reason, bacterial identification
87 relying on sequencing of the 16S rRNA genes is becoming a method of choice for detection of
88 uropathogens in urine samples [20, 21].

89 Field of proteomics also offers methods for microbial identification, mass spectrometry (MS)
90 being the most prominent one. MS platforms used include matrix-assisted laser
91 desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) based analysis
92 producing characteristic spectrum called peptide mass fingerprint (PMF), or less frequently used
93 liquid chromatography tandem mass spectrometry (LC-MS/MS) based peptide sequencing. LC-
94 MS/MS depends on initial isolation of bacterial colonies from urine and their subsequent
95 cultivation [22, 23], while MS based analysers claim ability to directly process samples or swabs.

96 Today, MS-based analysers are in routine use, such as the Bruker BioTyper (Bruker Daltonics)
97 and VITEK MS Plus (bioMérieux), both detecting MS1 spectra fingerprint of most abundant
98 proteins present in a wide array of microorganisms [24, 25, 26]. The US Food and Drug
99 Administration (FDA) has issued regulatory approval for using mass spectrometry-based
100 platform MALDI-TOF MS for routine identification of pathogenic microbes from human
101 specimens in clinical microbiology laboratories [27, 23]. The instrument is coupled with
102 dedicated software and database so it can perform a comparison of the recorded MS1 spectra
103 with the mass spectra of known microorganisms stored in the database. However, MALDI-TOF

104 MS has its limitations and does not allow identification of microorganisms at the species level,
105 nor it performs well when more than one species or strain is present in the sample [28, 29, 30].
106 Furthermore, in order to obtain reliable results, samples have to be cultured on selective agar and
107 a single microbial colony is then used to identify an organism. To bypass time-consuming and
108 selective cultivation stage, culture-independent methods have been developed [31, 32, 17, 33,
109 34]. In recent years there has been growing interest in proteomic analyses of urine samples using
110 tandem mass spectrometry [35, 36]. The metaproteomic analysis is able to provide high numbers
111 of strain-specific peptides useful for microbial identification at the genus, species and even
112 strain-level, and it can also be applied to urine samples containing more than one species, or
113 even potential biomarkers for non-invasive monitoring of human diseases [37, 38, 39, 40].
114

115 **Methods**

116 **Urine samples collection and storage**

117 Urine specimens were collected from the Centre for Clinical Microbiology and Hospital
118 Infections, University Hospital Dubrava with only exclusion criteria being antimicrobial therapy.
119 Through the period from October to December 2016 total of 2993 urine specimens were received
120 from patients for which a urinary culture analysis was requested (Supplementary table 1). The
121 samples were collected from patients according to the instructions for collecting the urine by
122 midstream clean-catch technique [42].
123

124 **Urine culture test**

125 The microorganisms were identified by routine microbiology methods [43]. Aliquots made from
126 urine specimens were inoculated onto McConkey agar and blood agar plates using a 1µl
127 calibrated loop and incubated aerobically at 37°C from 18 to 48 h, according to the standard
128 operating procedure at the Centre for Clinical Microbiology and Hospital Infections, University
129 Hospital Dubrava. Single colonies were counted to determine the bacterial concentration.

130 Clinically significant infections were considered those with more than 10^5 CFU/mL.

131

132 **Samples for genomics and proteomics analysis**

133 From samples that were tested positive (total of 1571) on urine culture test, 16 samples were
134 randomly selected matching the following criteria: a.) there were more than 10^5 CFU/mL and b.)
135 sample contained more than 30 ml of urine. All sixteen urine samples (associated with
136 corresponding laboratory reports) were stored at -20°C and used for further analysis.

137

138 **GENOMIC ANALYSIS**

139 **DNA extraction**

140 Frozen samples were thawed at room temperature and homogenised. Bacterial genomic DNA
141 was extracted using the Maxwell 16 Cell DNA Purification Kit on the Maxwell 16 research
142 instrument (Promega, Madison) according to the manufacturer's instructions. The concentration
143 of DNA was determined using a Nano-Drop spectrophotometer (Shimadzu Biotech).

144

145 **16S rRNA sequencing and bioinformatics analysis**

146 Extracted DNA was sent to Next Generation Sequencing Service Provider (MR DNA, Texas,
147 USA). Sequencing was performed on an Illumina MiSeq platform using paired-end sequencing
148 protocol. Amplicons of the 16S rRNA gene were generated using primers targeting V3 and V4
149 variable regions of the ribosomal RNA. A 30 cycle PCR reaction was performed using the
150 HotStarTaq Plus Master Mix Kit (Qiagen, USA). Microbiome bioinformatic analysis was
151 performed using QIIME 2 (Quantitative Insights Into Microbial Ecology) software package
152 version 2018.4 [44]. Paired-end raw sequences were demultiplexed and quality filtered using the
153 `q2-demux` plugin followed by denoising with DADA2 [45]. First 7 bases of forward and reverse
154 read were trimmed, forward read was truncated to 290 bases, and reverse read to 240 bases.
155 Taxonomy was assigned to obtained amplicon sequence variants using the `q2-feature-classifier`
156 [46] `classify-sklearn` naive Bayes taxonomy classifier against the Greengenes 13_8 99% OTUs

157 reference sequences trimmed to variable regions 3 and 4 [47]. Amplicons were analysed using
158 the QIIME 2 (version 2017.4).

159

160 **PROTEOMICS**

161 **Sample preparation**

162 A homogenized aliquot of 10 ml urine sample was centrifuged at 1000 g at room temperature for
163 1 min (Supplementary figure 1). Insoluble sediment was discarded, and supernatant was
164 transferred to a new tube and centrifuged at 16000 g at 4°C for 5 min. The supernatant was
165 discarded, and the bacterial pellet was resuspended in a buffer (25 mM NH₄HCO₃, pH 7.8). The
166 pellet was homogenized on vortex and centrifuged at 16000 g, at 4 °C for 5 min. This procedure
167 was designed to “wash out” mainly excess human cells and it was repeated three times. Proteins
168 were extracted from the bacterial pellet using 100 µL of bacterial protein extraction reagent B-
169 PER (Thermo-Pierce, USA). Following the manufacturer's protocol, sample was incubated at
170 room temperature for 15 min and subsequently heated at 100°C in a water bath for 2 min.
171 Insoluble cellular debris was removed by centrifugation at 16000 g at 4°C for 5 min. Finally,
172 supernatant with soluble proteins contained in B-PER solution was ready for the next step in
173 proteomics sample preparation.

174

175 **In solution digestion**

176 Protein sample contained in B-PER (70 µL) was mixed with 2 µL of trypsin solution (1 mg/mL,
177 Merck, Germany). The in-solution digestion was carried out at 37 °C on a thermoshaker (500
178 rpm) for 18 h (overnight).

179

180 **Peptide fractionation**

181 After 18 hours of trypsin in-solution digestion, fractionation was performed using the Agilent
182 Bravo automated liquid handling platform (96-channel tip head) and AssayMAP SCX cartridges
183 according to the manufacturer's instructions, and fractionation protocol (application note 5991-

184 3602EN), SCX cartridges were primed with 400 mM ammonium formate /1% formic acid/ 25%
185 acetonitrile (ACN), equilibrated with 1% formic acid /25% ACN, loaded with samples, and
186 eluted sequentially using a 40 mM ammonium formate /25% ACN (pH 3.5; 4.0) 40 mM
187 ammonium acetate /25% ACN, (pH 4.5; 5; 5.5) and 100 mM ammonium hydroxide /25% ACN
188 (pH 9.5). From each processed sample, a total of six fractions were collected by chromatography
189 using a pH modulated stepwise elution method.

190

191 **MALDI-TOF/TOF mass spectrometry analysis**

192 For sample analysis, 1 μ l of 5 mg/mL α -CHCA (α -cyano-4-hydroxycinnamic acid) matrix
193 solution was mixed with 1 μ l of each sample fraction (six fractions per sample). From the
194 resulting solution, 1 μ l was spotted onto the Opti-TOF MALDI 384 target plate (AB Sciex).
195 After drying at room temperature, spotted samples were analysed using a 4800 Plus MALDI-
196 TOF/TOF mass spectrometer (Applied Biosystems Inc., Foster City, USA) equipped with a 200
197 Hz, 355 nm Nd: YAG laser. MS spectra were acquired over a mass range of 800 - 4500 m/z.
198 Peptide fragmentation was performed at collision energy (CID) of 1 kV in positive ion reflection
199 mode, using nitrogen as collision gas. For each sample up to 20 most intense peaks of MS
200 spectra were selected for MS/MS spectra analysis. Approximately 1000 single shots were
201 accumulated from different positions for MS analysis, and 2000 shots spectra were recorded for
202 the subsequent fragment ion spectra. Internal calibration using trypsin autolysis fragments was
203 performed. MS and MS/MS spectra were acquired using the 4000 Series Explorer software v
204 3.5.3 (AB Sciex).

205

206 **Analysis of proteomics data**

207 Mascot (version 2.1.; Matrix Science, UK) analysis was carried out to identify peptides and to
208 search for matching proteins in the NCBI nr database (20140312) with taxonomy filter set for
209 *Proteobacteria* (11838333 sequences), *Firmicutes* (5487348 sequences) and *Homo sapiens*

210 (276468 sequences). Search parameters for MS and MS/MS database were as follows: parent ion
211 mass tolerances of 0.3 Da and 0.5 Da fragment ion mass tolerance, trypsin digestion with a
212 maximum of one miscleavage per peptide and methionine oxidation as variable modification.
213 Trypsin specificity was set at C-terminal lysine and arginine unless next residue is proline.
214 Qualitative data analysis was performed with MASCOT using a 95% confidence interval, so the
215 significance threshold was adjusted that the false discovery rate was <5%. In Mascot reports a
216 minimum score of 48 was used.

217

218 **Results and discussion**

219 **Urine culture test**

220 All samples which have undergone proteomics and genomics analyses were benchmarked
221 against standard urine culture test which accompanied all the samples (Supplementary table 2).
222 Among the 16 clinical samples analysed, 13 were classified as monobacterial infections and 3
223 were classified as polymicrobial (at least two identified uropathogens). Thirteen samples showed
224 presence of Gram-negative and only three to Gram-positive bacteria. Regarding taxonomic
225 diversity of the samples analysed, according to standard tests, there were 7 different bacterial
226 species in total, belonging to 4 respective genera (Supplementary table 3).

227

228 **Effect of storage time and temperature on bacteria in urine samples**

229 Guidelines for the collection and storage of urine specimens differ for different diagnostic
230 purposes. This is something we should be aware of. Urine samples should be collected and
231 stored having in mind exact diagnostic procedures to be carried out. In our study, short-term
232 storage (up to 4 weeks) of urines at -20°C showed to be a good choice for the preservation of
233 bacteria in collected samples. Long-term storage (for more than three months) at -80°C led to
234 biomass loss, most likely due to prolonged freezing which caused greater bacterial cell fragility
235 leading to a greater extent of cell disruption during centrifugation (unpublished observations).

236

237 **IDENTIFICATION OF MICROORGANISMS USING GENOMICS**

238 **16S rRNA sequencing results**

239 Identification of bacterial taxa is shown in Table 1. Lowest obtainable taxonomic level for which
 240 assignment was possible is being shown as a result of genomic identification. Table 1 provides
 241 following information: sample number, conventional urine culture result, DNA concentration
 242 and 16S rRNA gene sequencing result.

243
 244 **Table 1** Identification results based on conventional urine culture and 16S rRNA gene
 245 sequencing

N.o.	Urine culture identification	DNA concentration (ng/μL)	16S rRNA sequencing results
UR1	<i>Klebsiella pneumoniae</i> ESBL	3.15	100% Enterobacteriaceae
UR2	<i>Klebsiella oxytoca</i>	9.14	97% Enterobacteriaceae
UR3	<i>Klebsiella pneumoniae</i> ESBL	6.01	90% Enterobacteriaceae; 5.1% Granulicatella; 1.1% Anaerococcus
UR4	<i>Proteus mirabilis</i>	6.96	97.7% <i>Proteus</i> ; 1% Enterobacteriaceae
UR5	<i>Enterococcus faecalis</i>	16.22	21.8% <i>Pseudomonas</i> ; 13.7% <i>Propionibacterium acnes</i> ; 11% <i>Lactobacillus helveticus</i> ; 8.1% <i>Adhaeribacter</i> ; 8% <i>Acinetobacter</i> ; 5.9% <i>Staphylococcus</i> ; 4.9% <i>Stenotrophomonas</i> ; 3.8% <i>Hydrogenophaga</i> ; 3.6% <i>Erysipelotrichaceae</i> ; 3.1% <i>Corynebacterium</i> ; 3.1% <i>Cellulomonas</i> ; 2.3% <i>Aerococcus</i> ; 2% <i>Acidovorax</i> ; 1.9% <i>Lachnospiraceae</i> ; 1.6% <i>Sphingobium</i>
UR6	<i>Enterococcus faecalis</i>	4.93	51.4% <i>Enterococcus</i> ; 46.5% <i>Enterococcaceae</i>
UR7	<i>Enterobacter cloacae</i> ESBL	1.84	98% <i>Enterobacter</i> ; 0.9% <i>Proteus</i>
UR8	<i>Citrobacter koseri</i>	15.94	57.5% <i>Citrobacter koseri</i> ; 4.5% <i>Bacteroides</i> ; 3.7% <i>Dysgonomonas</i> ; 2.7% <i>Bacteroides</i> ; 2.6% <i>Rikenellaceae</i> ; 2.3% <i>Parabacteroides</i> ; 2% <i>Desulfovibrionaceae</i> ; 2% <i>Lachnospiraceae</i> ; 2% <i>Ruminococcaceae</i> ; 1.9% <i>Enterobacteriaceae</i> ; 1.3% <i>Ruminococcus</i> ; 1.3% <i>Erysipelotrichaceae</i> ; 1.2% <i>Enterococcus</i> ; 1.1% <i>Clostridiales</i>
UR9	<i>Proteus mirabilis</i>	0.45	96.7% <i>Proteus</i> ; 2.4% <i>Enterobacteriaceae</i> ; 1.2% <i>Prevotella</i>
UR10	<i>Proteus mirabilis</i>	3.66	97% <i>Proteus</i> ; 1.3% <i>Enterobacteriaceae</i>
UR11	<i>Escherichia coli</i> ; <i>Proteus mirabilis</i> ESBL	9.39	93% <i>Enterobacteriaceae</i> ; 3.5% <i>Proteus</i>
UR12	<i>Proteus mirabilis</i>	1.73	99.2% <i>Proteus</i>
UR13	<i>Enterobacter aerogenes</i>	1.08	75.4% <i>Enterobacteriaceae</i> ; 14% <i>Lactobacillus delbrueckii</i> ; 4.2% <i>Kluyvera</i> ; 4% <i>Enterobacter</i> ; 1% <i>Lactobacillus helveticus</i>
UR14	<i>Enterobacter cloacae</i>	4.14	95.4% <i>Enterobacteriaceae</i> ; 1.2% <i>Clostridium perfringens</i> ; 1% <i>Bifidobacterium pseudolongum</i>
UR15	<i>Enterobacter cloacae</i> ; <i>Enterococcus faecalis</i> ; <i>E coli</i> ; <i>Proteus mirabilis</i>	15.17	86.9% <i>Proteus</i> ; 7.2% <i>Enterobacteriaceae</i> ; 2..2% <i>Enterobacter</i> ; 1% <i>Rhodospirillaceae</i>
UR16	<i>Escherichia coli</i> ; <i>Klebsiella pneumoniae</i>	15..04	91.2% <i>Enterobacteriaceae</i> ; 8.6% <i>Klebsiella</i>

247 What stands out in this table is a disparity in taxonomic identification obtained through 16S
248 rRNA gene sequencing – in the majority of cases bacteria were identified on genus level (44%)
249 and family level (56%), while the identification on species level is usually lacking.
250 It is apparent that *Klebsiella* spp. (UR1-UR3), and *Enterobacter* spp. (UR13-UR14)
251 identifications are difficult to compare due to different levels of taxonomy assignment by the
252 method [48], while there is a significant positive correlation amongst other results for both
253 conventional and genomics methods. A possible explanation for this difficulty might be related
254 to bacterial nomenclature, taxonomy and very high sequence identity. Furthermore, 16S rRNA
255 analysis was not informative at the genus and/or species level in the family *Enterobacteriaceae*
256 [49]. There was a surprising difference of end results between standard test and genomics in
257 sample UR 5. Standard urine culture test indicated *Enterococcus faecalis* as a single uropathogen
258 in this sample, while 16S rRNA indicated polymicrobial mixture without *Enterococcus* genus
259 listed. There are two possible explanations for this disparity, one indicating a urine collection
260 sample contamination [50] which would likely cause a genomics test error, and the other being
261 false-positive result of standard culture-based urine test giving a false positive *Enterococcus*
262 result.

263

264 **METHOD FOR PROTEOMICS-BASED IDENTIFICATION OF UROPATHOGENS**

265 The present study was undertaken to assess the potential of bottom-up proteomics for
266 identification of pathogens directly from the urine samples of patients with UTIs by
267 benchmarking the results obtained against the reference ones (standard urine tests) and using the
268 16S rRNA gene sequencing - genomics for arbitration in cases where proteomics gives results
269 which differ from the standard urine tests.

270

271 **Sample preparation**

272 For the proteomic analysis, a minimum concentration of 10^5 CFU/mL and a volume of 5 mL of

273 fresh urine sample or urine stored in the refrigerator up to 4 weeks. In this preliminary study, we
274 investigated and compared the preparation of samples stored at -20 °C and -80 °C. We based our
275 decision on the optimal storage temperature of samples on visual inspection of pellets during
276 centrifugation. In the case of urine samples stored at -80 °C bacterial cells were lost, and the
277 pellet deemed insufficient for further downstream analysis. On the other hand, samples stored at
278 -20 °C showed abundant biomass, however, this proved to be a challenge to wash. Reason for
279 this could be cell aggregation, probable autoaggregation, especially since blood was present in
280 tested samples [32]. Furthermore, good separation of bacterial cells from other materials such as
281 yeast cells, epithelial cells, leukocytes, erythrocytes, mucus, urinary casts, and different types of
282 crystals that can be present in urine depends on centrifugation speed [32, 50]. Moreover, at high-
283 speed the pellet will likely be abundant with cell debris. Consequently, damaged cells will be
284 washed off during the sample preparation process. Pellet volume was identified as an important
285 element that influenced the success of positive protein identification. Microbial biomass had to
286 be visible to the naked eye after washing steps. The pellet biomass can be seen in supplementary
287 figure 2.

288 Previous studies had considered the impact of ultrasonication on microorganisms to improve
289 sample preparation [51, 32, 52]. In our research protein extraction using B-PER worked for both
290 gram-negative and gram-positive bacteria, so there was no need for additional mechanical
291 methods of cell rupture. In reviewed literature, no data was found on the efficiency of protein
292 digestion in the presence of B-PER. We believe that no other group has reported the use of
293 trypsin in the B-PER solution.

294

295 **Peptide fractionation**

296 During a preliminary study, we found that the amount of data we could get from one sample spot
297 was insufficient. Thus, to overcome this obstacle we used peptide fractionation. We hypothesised
298 that peptide fractionation would help enrich the low-abundance peptides (Supplementary figure

299 4).

300

301 **MALDI-TOF/TOF mass spectrometry results**

302 **Protein identifications and data analysis**

303 While BioTyper and Vitek use reference databases to identify and classify the microorganisms
304 according to their mass spectra fingerprint, we relied on peptide ion fragments from MS/MS
305 scans and MASCOT protein search results which were translated into MASCOT based
306 uropathogen identification ranks. For this purpose, we have combined MASCOT score with a
307 peptide count and made a simple Python script which ranks organisms suspected to be in the
308 sample based on probability of their proteins being detected. First step was protein identification
309 of tryptic peptides conducted using MASCOT search engine [54]. This provided us with both
310 score and number of queries matched for proteins belonging to one or more organisms. The
311 Mascot Score is a statistical score for how well the spectra generated match the database protein
312 sequence [54, 55]. Plainly, a higher score indicates a more confident protein match while the
313 number of queries matched indicates the number of spectra that were matched to this protein.
314 Although it is not unusual for a portion of peptides to be scanned multiple times, overall, the
315 greater the score and the greater the number of queries matched – greater the probability of a true
316 positive match. Therefore, we have combined these two measures into a “summa score”, simply
317 by summing up all individual peptide scores for a given protein match. Proteins and respective
318 taxa were ordered based on this “summa score” in descending order and highest scoring taxa was
319 taken as most likely uropathogen identification. Table 2 compares the results of this analysis
320 with the standard urine culture test. Summarized report on MASCOT identified bacterial proteins
321 is listed in supplementary table 4.

322

323 **Table 2** MALDI-TOF/TOF analysis with MASCOT identification of uropathogens

N.o.	Urine culture identification	MASCOT IDENTIFICATION
UR1	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
UR2	<i>Klebsiella oxytoca</i>	<i>Klebsiella pneumoniae</i>
UR3	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>

UR4	<i>Proteus mirabilis</i>	<i>Proteus mirabilis</i>
UR5	<i>Enterococcus faecalis</i>	<i>Enterococcus faecalis</i>
UR6	<i>Enterococcus faecalis</i>	<i>Enterococcus faecalis</i>
UR7	<i>Enterobacter cloacae</i>	<i>Citrobacter freundii</i>
UR8	<i>Citrobacter koseri</i>	<i>Citrobacter freundii</i>
UR9	<i>Proteus mirabilis</i>	<i>Proteus mirabilis</i>
UR10	<i>Proteus mirabilis</i>	<i>Proteus mirabilis</i>
UR11	<i>Escherichia coli</i> ; <i>Proteus mirabilis ESBL</i>	<i>Escherichia fergusonii</i>
UR12	<i>Proteus mirabilis</i>	<i>Proteus mirabilis</i>
UR13	<i>Enterobacter aerogenes</i>	<i>Enterobacter aerogenes</i>
UR14	<i>Enterobacter cloacae</i>	<i>Klebsiella pneumoniae</i>
UR15	<i>Enterobacter cloacae</i> ; <i>Enterococcus faecalis</i> ; <i>E coli</i> ; <i>Proteus mirabilis</i>	<i>Enterococcus faecalis</i>
UR16	<i>Escherichia coli</i> ; <i>Klebsiella pneumoniae</i>	<i>Pectobacterium atrosepticum</i>

324

325 The proteins ordered by summa score were listed in supplementary table E1. Significant
326 minimum MASCOT summa score obtained for all samples was 53, while maximum reported
327 score was 830. A total number of 382 peptides were reported for all 16 samples. Most of these
328 peptides belong to bacterial proteins (71%). Although we expected the majority of proteins
329 belonging to ribosomes, we identified a rather small percentage of ribosomal proteins (8%). In
330 our case proteins with the highest scores, were membrane proteins including outer membrane
331 porin protein C, peptidoglycan-associated lipoprotein (PAL) and murein lipoprotein (MLP). This
332 interesting result might be associated with the usage of the B-PER [53]. Considering all
333 monobacterial samples, direct identifications provided reliable identification for genus *Klebsiella*
334 (3 samples), *Proteus* (4 samples), *Enterococcus* (2 samples), *Enterobacter* (1 sample) and
335 *Citrobacter* (1 sample). Overall, 87% of correlation with standard urine test was obtained with
336 this simple proteomics approach for monobacterial samples.

337 It is remarkable how pathogenic species were correctly identified at the genus levels considering
338 such a small number of identified bacterial proteins per sample, and the absence of unique
339 peptides. Although our results indicate that proteomics-based identification with a small number
340 of proteins could be carried out, high-throughput setup yielding more spectra and retrieving
341 larger fractions of proteomes would be more favourable.

342

343 **Microbial identification in polymicrobial cultures**

344 To investigate polymicrobial cultures (UR11, UR15 and UR16), we compared the results
345 obtained from the conventional urine culture, 16S rRNA gene sequencing and proteomics
346 (Supplementary table 5). Our previous experience with MALDI-TOF/TOF mass spectrometer
347 indicated that bacterial identification in polymicrobial urine samples using this platform for
348 proteomics has some limitations. As reported previously by other authors, MALDI-TOF MS
349 identification of polymicrobial cultures directly from urine samples did not provide reliable
350 results [50, 17]. Therefore, bacterial identification at the strain-level is still regarded as a
351 challenge. Some of the underlying factors that compromise this method sensitivity in bacterial
352 identification are: sample impurity substances (human proteins), the low abundance of bacterial
353 proteins in the sample [41], insufficient coverage of urinary bacterial species in the databases,
354 shared peptide sequences among proteins from different taxa [38] as well as insufficient data for
355 identification after analysis [39]. Bottom-up tandem MS with growing reference database and
356 data processing through bioinformatic analysis has made significant progress in increasing
357 polymicrobial identification [36, 30] but this is still a field being developed experimentally and
358 far from clinical practice.

359

360 **Human proteins versus contamination**

361 Normal human urine of a healthy individual contains over 2000 proteins [56, 57], while over
362 5000 proteins can be found when the urinary tract is under inflammation [33]. Due to low protein
363 concentration, urine is a difficult proteomic sample to work with [59].

364 We recorded 29% of human proteins in our samples, of which 33% were found to be repetitive
365 (Supplementary table 6). The most abundant of these repeated human proteins were from
366 haemoglobin subunits (alpha and beta-globin), apolipoprotein and uromodulin. We did not find
367 any evidence of epithelial cells from the urinary or vaginal tract, nor any biomarkers.

368 As can be seen from supplementary figure 3, first two fractions cover more than 50% of the total
369 number of proteins. Furthermore, supplementary figure 4 shows a quantitative overview of
370 bacterial and human proteins of each sample. In terms of future work, it would be interesting to
371 consider two-dimensional fractionation to increase bacterial proteome coverage and enhance the
372 ratio of bacterial vs human proteins.

373

374 **Limitations and future direction**

375 With regard to the research method, the major limitation identified of this study is small number
376 of identified proteins per sample. Many proteomic analyses for bacterial identification were
377 limited to monomicrobial specimens with high CFU/mL concentration based on our need to
378 compare results with those of standard urine culture tests which have own inherent drawbacks.
379 This study lays the groundwork for future research. In the future, a possible direction could be
380 dealing with lower abundant proteins to enhance effectiveness in proteome identification.
381 Switching to a more high-throughput platform such as ESI could solve this issue. Furthermore,
382 to increase the number of proteins, a possible solution could be use of peptide double
383 fractionation or FASP (filter-aided sample preparation) method. To improve bacterial
384 identification, we are developing bioinformatics software based on natural language processing.
385 Urine is clinically underutilized and have a much greater potential in development of non-
386 invasive tests and techniques. Proteomics approach and direct sample analysis have potential to
387 provide us with a broader clinical picture of the patient that could bring us closer to precision
388 medicine.

389

390 **Conclusion**

391 The main goal of the current study was to establish a procedure for analysis of uropathogens by
392 proteomics, tested using MALDI-TOF/TOF mass spectrometry directly from urine specimens.
393 This study has shown that identification of bacteria from a native urine sample, without culturing

394 step, depends on storage conditions, sample preparation method, as well as data analysis.
395 Overall, the results of this study demonstrate that mass spectrometry based proteomics can
396 effectively identify different uropathogens from fresh or cold stored, non-cultivated human urine
397 samples. The direct approach was able to provide reliable identification of bacteria at the genus-
398 level in monobacterial samples despite inherent limitations of mass spectrometry platform used.
399 More research is required in order to handle polymicrobial urine samples.

400
401 **Additional file 1:** Supplementary figure 1. Experimental workflow for the identification of
402 uropathogen from a native urine sample. Supplementary figure 2. Images of 16 urine specimens.
403 Supplementary figure 3. Protein content of each fraction as a percentage of the total protein.
404 Supplementary figure 4. Cumulative number of bacterial and human proteins for each sample per
405 fraction.

406 **Additional file 2:** Supplementary table 1. General information about patients. Supplementary
407 table 2. Results of conventional urine culture and urine dipstick analysis for 16 urine samples.
408 Supplementary table 3. Uropathogenic bacteria in urine samples. Supplementary table 4.
409 Summary reports of identified bacterial proteins for each urine sample sorted by “MASCOT
410 summa score”. Supplementary table 5. The comparative view of urine culture, proteomics and
411 genomic results. Supplementary table 6. Identified human proteins ranked by MASCOT score
412 for each urine sample.

413 **Additional file 3:** Supplementary Excel tables.

414

415 **Abbreviations**

416 ACN: Acetonitrile; CFU/mL: Colony-forming units per millilitre; LC: liquid chromatography;
417 MALDI: Matrix-assisted laser desorption/ionization; MS: Mass spectrometry; PMF: Peptide
418 mass fingerprint; TOF: Time of flight; UTI: Urinary tract infection

419

420 **Author contributions**

421 DO, AH and MČ: Carried out the experiments. DO, JŽ, MC, MČ and AS: Performed data
422 analysis. All authors contributed to the design of the work. DO, MČ, JŽ and AS: Provided major
423 contributions to manuscript writing. JŠ and MČ: Management of patients and helped in
424 processing patient samples. All authors read and approved the final manuscript.

425

426 **Funding**

427 This work was funded by the Croatian Science Foundation under research project “Exploring
428 Gut Microbiome Equilibrium” (grant HRZZ_IP_06_2016_3509). All authors report no conflicts
429 of interest relevant to this article.

430

431 **Availability of data and materials**

432 All the datasets analysed in the current study are available upon reasonable request. There is an
433 archive available at: http://proteinreader.bioinfo.pbf.hr/urine/urine_spectra.zip with the mass
434 spectrometry raw data used.

435 **Ethics approval and consent to participate**

436 This study was approved by the Ethics committee of the University Hospital Dubrava, by the
437 Ethics committee of the Faculty of Pharmacy and Biochemistry, and by the Research Ethics
438 committee of the Faculty of Food Technology and Biotechnology. Written informed consent was
439 obtained from all our patients.

440

441 **Consent for publication**

442 The datasets used and/or analysed during the current study are available from the corresponding
443 author on reasonable request.

444 **Competing interests**

445 The authors declare that they have no competing interest.

446

447 **Author details**

448 ¹Department of Biochemical Engineering, Faculty of Food Technology and Biotechnology,
449 Zagreb University, 10000 Zagreb, Croatia. ²Special Hospital Agram, Polyclinic Zagreb,
450 Biochemical Laboratory, 10000 Zagreb, Croatia. ³Division of Molecular Medicine, Ruđer
451 Bošković Institute, Zagreb, Croatia. ⁴Department for Clinical Microbiology and Hospital
452 Infection, University Hospital Dubrava, 10000 Zagreb, Croatia. ⁵Faculty of Pharmacy and
453 Biochemistry, Zagreb University. ⁶Department of Biological Science, Faculty of Science,
454 Macquarie University, Sydney, Australia

455

456 **References**

- 457 1. Foxman B. The epidemiology of urinary tract infection. *Nat Rev Urol.* 2010;7(12):653–
458 60.
- 459 2. Dason S, Dason JT, Kapoor A. Guidelines for the diagnosis and management of recurrent
460 urinary tract infection in women. *J Can Urol Assoc.* 2011;5(5):316–22.
- 461 3. Bouchillon SK, Badal RE, Hoban DJ, Hawser SP. Antimicrobial Susceptibility of
462 Inpatient Urinary Tract Isolates of Gram-Negative Bacilli in the United States: Results
463 from the Study for Monitoring Antimicrobial Resistance Trends (SMART) Program:
464 2009-2011. *Clin Ther.* 2013;35(6):872–7.
- 465 4. Brubaker L, Wolfe AJ. The New World of the Urinary Microbiome in Women Linda. *Am*
466 *J Obstet Gynecol.* 2015;213(5):644–649.
- 467 5. Koves B, Wullt B. The Roles of the Host and the Pathogens in Urinary Tract Infections.
468 *Eur Urol Suppl.* 2016;15(4):88–94.
- 469 6. Schmiemann G, Kniehl E, Gebhardt K, Matejczyk MM, Hummers-Pradier E. The
470 diagnosis of urinary tract infection: a systematic review. *Dtsch Ärzteblatt Int.*
471 2010;107(21):361–7.
- 472 7. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, et al. Urine is
473 not sterile: Use of enhanced urine culture techniques to detect resident bacterial flora in

- 474 the adult female bladder. *J Clin Microbiol.* 2014;52(3):871–6.
- 475 8. Wolfe AJ, Brubaker L. “Sterile Urine” and the Presence of Bacteria. *Eur Urol.*
476 2015;68(2):173–4.
- 477 9. Ferreira L, Sánchez-Juanes F, Muñoz-Bellido JL, González-Buitrago JM. Rapid method
478 for direct identification of bacteria in urine and blood culture samples by matrix-assisted
479 laser desorption ionization time-of-flight mass spectrometry: Intact cell vs. extraction
480 method. *Clin Microbiol Infect.* 2011;17(7):1007–12.
- 481 10. Burd EM, Kehl KS. A critical appraisal of the role of the clinical microbiology laboratory
482 in the diagnosis of urinary tract infections. *J Clin Microbiol.* 2011;49(9):34–8.
- 483 11. Cavagnolo R. Evaluation of incubation times for urine cultures. *J Clin Microbiol.*
484 1995;33(7):1954–6.
- 485 12. Brilha S, Proença H, Cristino JM, Hänscheid T. Use of flow cytometry (Sysmex UF-100)
486 to screen for positive urine cultures: In search for the ideal cut-off. *Clin Chem Lab Med.*
487 2010;48(2):289–92.
- 488 13. Delanghe J, Speeckaert M. Preanalytical requirements of urinalysis. *Biochem Medica.*
489 2014;24(1):89–104.
- 490 14. Davenport M, Mach KE, Shortliffe LMD, Banaei N, Wang H, Liao JC. New and
491 developing diagnostic technologies for urinary tract infections. *Nat Rev Urol.*
492 2017;14(5):296–310.
- 493 15. Franco-Duarte R, Černáková L, Kadam S, Kaushik KS, Salehi B, Bevilacqua A, et al.
494 Advances in chemical and biological methods to identify microorganisms—from past to
495 present. *Microorganisms.* 2019;7(130).
- 496 16. Singhal N, Kumar M, Kanaujia PK, Viridi JS. MALDI-TOF mass spectrometry: an
497 emerging technology for microbial identification and diagnosis. *Front Microbiol.*
498 2015;6(791).
- 499 17. Zboromyrska Y, Rubio E, Alejo I, Vergara A, Mons A, Campo I, et al. Development of a

- 500 new protocol for rapid bacterial identification and susceptibility testing directly from urine
501 samples. *Clin Microbiol Infect.* 2016;22(6):561.e1.
- 502 18. De Rosa R, Grosso S, Lorenzi G, Bruschetta G, Camporese A. Evaluation of the new
503 Sysmex UF-5000 fluorescence flow cytometry analyser for ruling out bacterial urinary
504 tract infection and for prediction of Gram negative bacteria in urine cultures. *Clin Chim*
505 *Acta.* 2018;484:171–8.
- 506 19. Van Der Zee A, Roorda L, Bosman G, Ossewaarde JM. Molecular diagnosis of urinary
507 tract infections by semi-quantitative detection of uropathogens in a routine clinical
508 hospital setting. *PLoS One.* 2016 Mar 1;11(3).
- 509 20. Mignard S, Flandrois JP. 16S rRNA sequencing in routine bacterial identification: A 30-
510 month experiment. *J Microbiol Methods.* 2006;67(3):574–81.
- 511 21. Sabat AJ, Van Zanten E, Akkerboom V, Wisselink G, Van Slochteren K, De Boer RF, et
512 al. Targeted next-generation sequencing of the 16S-23S rRNA region for culture-
513 independent bacterial identification-increased discrimination of closely related species.
514 *Sci Rep.* 2017;7(3434).
- 515 22. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass
516 spectrometry. *Annu Rev Biochem.* 2001;70:437–73.
- 517 23. Cheng K, Chui H, Domish L, Hernandez D, Wang G. Recent development of mass
518 spectrometry and proteomics applications in identification and typing of bacteria.
519 *Proteomics - Clin Appl.* 2016;10(4):346–57.
- 520 24. Lévesque S, Dufresne PJ, Soualhine H, Domingo MC, Bekal S, Lefebvre B, et al. A Side
521 by Side Comparison of Bruker Biotyper and VITEK MS: Utility of MALDI-TOF MS
522 Technology for Microorganism Identification in a Public Health Reference Laboratory.
523 *PLoS One.* 2015;10(12):e0144878.
- 524 25. Jang KS, Kim YH. Rapid and robust MALDI-TOF MS techniques for microbial
525 identification: a brief overview of their diverse applications. *J Microbiol.* 2018;56(4):209–

- 526 16.
- 527 26. Li Y, Shan M, Zhu Z, Mao X, Yan M, Chen Y, et al. Application of MALDI-TOF MS to
528 rapid identification of anaerobic bacteria. *BMC Infect Dis.* 2019;19(941).
- 529 27. Lathrop JT, Jeffery DA, Shea YR, Scholl PF, Chan MM. US food and drug administration
530 perspectives on clinical mass spectrometry. *Clin Chem.* 2016;62(1):41–7.
- 531 28. Patel R. MALDI-TOF MS for the diagnosis of infectious diseases. *Clin Chem.*
532 2015;61(1):100–11.
- 533 29. Henzel WJ, Watanabe C, Stults JT. Protein identification: The origins of peptide mass
534 fingerprinting. *J Am Soc Mass Spectrom.* 2003;14(9):931–42.
- 535 30. Karlsson R, Gonzales-siles L, Boulund F, Svensson-stadler L, Skovbjerg S, Karlsson A, et
536 al. Proteotyping: Proteomic characterization, classification and identification of
537 microorganisms – A prospectus. *Syst Appl Microbiol.* 2015;38(4):246–57.
- 538 31. Veron L, Mailler S, Girard V, Muller BH, L’Hostis G, Ducruix C, et al. Rapid urine
539 preparation prior to identification of uropathogens by MALDI-TOF MS. *Eur J Clin*
540 *Microbiol Infect Dis.* 2015 Sep 24;34(9):1787–95.
- 541 32. Kitagawa K, Shigemura K, Onuma KI, Nishida M, Fujiwara M, Kobayashi S, et al.
542 Improved bacterial identification directly from urine samples with matrix-assisted laser
543 desorption/ionization time-of-flight mass spectrometry. *J Clin Lab Anal.* 2018;32(3).
- 544 33. Yu Y, Sikorski P, Bowman-Gholston C, Cacciabeve N, Nelson KE, Pieper R. Diagnosing
545 inflammation and infection in the urinary system via proteomics. *J Transl Med.*
546 2015;13(111).
- 547 34. Íñigo M, Coello A, Fernández-Rivas G, Rivaya B, Hidalgo J, Quesada MD, et al. Direct
548 Identification of Urinary Tract Pathogens from Urine Samples , Combining Urine
549 Screening Methods and Matrix-Assisted Laser Desorption Ionization – Time of Flight
550 Mass Spectrometry. *J Clin Microbiol.* 2016;54(4):988–93.
- 551 35. Shao W, Zhang M, Lam H, Lau SCK. A peptide identification-free, genome sequence-

- 552 independent shotgun proteomics workflow for strain-level bacterial differentiation. *Sci*
553 *Rep.* 2015;5(14337).
- 554 36. Boulund F, Karlsson R, Gonzales-Siles L, Johnning A, Karami N, AL-Bayati O, et al.
555 Typing and characterization of bacteria using bottom-up tandem mass spectrometry
556 proteomics. *Mol Cell Proteomics.* 2017;16(6):1052–63.
- 557 37. Boulund F, Karlsson R, Gonzales-Siles L, Johnning A, Karami N, AL-Bayati O, et al.
558 Typing and Characterization of Bacteria Using Bottom-up Tandem Mass Spectrometry
559 Proteomics. *Mol Cell Proteomics [Internet].* 2017;16(6):1052–63. Available from:
560 <http://www.mcponline.org/lookup/doi/10.1074/mcp.M116.061721>
- 561 38. Alves G, Wang G, Ogurtsov AY, Drake SK, Gucek M, Sacks DB, et al. Rapid
562 Classification and Identification of Multiple High-Resolution Tandem Mass Spectrometry.
563 *J Am Soc Mass Spectrom.* 2018;29:1721–37.
- 564 39. Karlsson R, Davidson M, Svensson-Stadler L, Anders Karlsson, Olesen K, Carlsohn E, et
565 al. Strain-Level Typing and Identification of Bacteria Using Mass Spectrometry-Based
566 Proteomics. *J Proteome Res.* 2012;11(5):2710–20.
- 567 40. Yu Y, Sikorski P, Smith M, Bowman-Gholston C, Cacciabeve N, Nelson KE, et al.
568 Comprehensive metaproteomic analyses of urine in the presence and absence of
569 neutrophil-associated inflammation in the urinary tract. *Theranostics.* 2017;7(2):238–52.
- 570 41. Harpole M, Davis J, Espina V. Current state of the art for enhancing urine biomarker
571 discovery. *Expert Rev Proteomics.* 2016;13(6):609–26.
- 572 42. Lifshitz E, Kramer L. Outpatient urine culture: does collection technique matter? *Arch*
573 *Intern Med.* 2000;160(16):2537–40.
- 574 43. Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, et al. The
575 clinical urine culture: Enhanced techniques improve detection of clinically relevant
576 microorganisms. *J Clin Microbiol.* 2016;54(5):1216–22.
- 577 44. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al.

- 578 Reproducible, interactive, scalable and extensible microbiome data science using QIIME
579 2. *Nat Biotechnol.* 2019;37(8):852–7.
- 580 45. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2:
581 High-resolution sample inference from Illumina amplicon data. *Nat Methods.*
582 2016;13(7):581–3.
- 583 46. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing
584 taxonomic classification of marker-gene amplicon sequences with QIIME 2’s q2-feature-
585 classifier plugin. *Microbiome.* 2018;6(90).
- 586 47. McDonald D, Price MN, Goodrich J, Nawrocki EP, Desantis TZ, Probst A, et al. An
587 improved Greengenes taxonomy with explicit ranks for ecological and evolutionary
588 analyses of bacteria and archaea. *ISME J.* 2012;6(3):610–8.
- 589 48. Kotaskova I, Obrucova H, Malisova B, Videnska P, Zwinsova B, Peroutkova T, et al.
590 Molecular techniques complement culture-based assessment of bacteria composition in
591 mixed biofilms of urinary tract catheter-related samples. *Front Microbiol.* 2019;10(462).
- 592 49. Naum M, Brown EW, Mason-Gamer RJ. Is 16S rDNA a reliable phylogenetic marker to
593 characterize relationships below the family level in the enterobacteriaceae? *J Mol Evol.*
594 2008;66:630–42.
- 595 50. Lough ME, Shradar E, Hsieh C, Hedlin H. Contamination in Adult Midstream Clean-
596 Catch Urine Cultures in the Emergency Department: A Randomized Controlled Trial. *J*
597 *Emerg Nurs.* 2019;45(5):488–501.
- 598 51. Huang B, Zhang L, Zhang W, Liao K, Zhang S, Zhang Z, et al. Direct detection and
599 identification of bacterial pathogens from urine with optimized specimen processing and
600 enhanced testing algorithm. *J Clin Microbiol.* 2017 May 1;55(5):1488–95.
- 601 52. Sigdel TK, Nicora CD, Hsieh SC, Dai H, Qian WJ, Camp DG, et al. Optimization for
602 peptide sample preparation for urine peptidomics. *Clin Proteomics.* 2014;11(1):1–8.
- 603 53. Zhang X, Li L, Mayne J, Ning Z, Stintzi A, Figeys D. Assessing the impact of protein

- 604 extraction methods for human gut metaproteomics. *J Proteomics* [Internet].
605 2018;180:120–7. Available from: <http://dx.doi.org/10.1016/j.jprot.2017.07.001>
- 606 54. Cottrell JS. Protein identification using MS/MS data. *J Proteomics*. 2011;74(10):1842–51.
- 607 55. Perkins DN, Darryl J. C. Pappin, David M. Creasy, Cottrell JS. Probability-based protein
608 identification by searching sequence databases using mass spectrometry data.
609 *Electrophoresis*. 1999;20(18):3551–67.
- 610 56. Koenig T, Menze BH, Kirchner M, Monigatti F, Parker KC, Patterson T, et al. Robust
611 prediction of the MASCOT score for an improved quality assessment in mass
612 spectrometric proteomics. *J Proteome Res*. 2008;7(9):3708–17.
- 613 57. Adachi J, Kumar C, Zhang Y, Olsen J V., Mann M. The human urinary proteome contains
614 more than 1500 proteins, including a large proportion of membrane proteins. *Genome*
615 *Biol*. 2006;7(9).
- 616 58. Marimuthu A, O’Meally RN, Chaerkady R, Subbannayya Y, Nanjappa V, Kumar P, et al.
617 A comprehensive map of the human urinary proteome. *J Proteome Res*. 2011;10(6):2734–
618 43.
- 619 59. Lin L, Yu Q, Zheng J, Cai Z, Tian R. Fast quantitative urinary proteomic profiling
620 workflow for biomarker discovery in kidney cancer. *Clin Proteomics*. 2018;15(42).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile2.docx](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile1.docx](#)