

1 Title: Metagenomic sequencing to replace semi-quantitative urine culture for detection of urinary  
2 tract infections: a proof of concept.

3  
4 Victoria A. Janes<sup>1</sup>#, Sébastien Matamoros<sup>1</sup>, Niels Willemse<sup>1</sup>, Caroline E. Visser<sup>1</sup>, Bob de  
5 Wever<sup>1</sup>, Marja E. Jakobs<sup>1</sup>, Poorani Subramanian<sup>2</sup>, Nur A. Hasan<sup>2</sup>, Rita R. Colwell<sup>2</sup>, Menno D. de  
6 Jong<sup>1</sup>, Constance Schultsz<sup>1,3</sup>.

7  
8 <sup>1</sup> Department of Medical Microbiology, Academic Medical Center, University of Amsterdam,  
9 Amsterdam, The Netherlands.

10 <sup>2</sup> CosmosID, Rockville, Maryland, United States of America.

11 <sup>3</sup> Department of Global Health – Amsterdam Institute for Global Health and Development,  
12 Amsterdam, The Netherlands.

13

14 #Corresponding author: Victoria A. Janes, v.a.janes@amc.nl

15 **ABSTRACT**

16           Semi-quantitative bacterial culture is the standard method to diagnose urinary tract  
17 infections (UTI), but bacterial growth rate limits diagnostic speed and it is unreliable when  
18 patients have been pre-treated with antibiotics. Metagenomics could increase diagnostic speed  
19 and accuracy by sequencing the microbiome and resistome directly from urine samples,  
20 bypassing culture. However, a semi-quantitative approach – as needed for diagnosing UTIs – has  
21 not been established.

22           Metagenomics was deployed to identify and semi-quantify bacterial presence indicative of  
23 UTI, predict antimicrobial susceptibility (AMR), and results were compared to semi-quantitative  
24 culture. Whole genome sequencing of the corresponding uropathogens was done for comparison.  
25 Analysis time and cost were tracked.

26           Forty-one consecutive urine samples underwent metagenomic analysis. All culture  
27 positive samples contained >200ng of DNA, suggestive of a threshold below which UTI could  
28 be ruled out solely based on DNA quantity. A semi-quantitative Diagnostic Index (DI) was  
29 created by multiplying the total DNA quantity by the relative abundance of uropathogens per  
30 urine sample. The DI allowed discrimination of UTI from non-UTI samples in all but 1 case.  
31 Metagenomic detection of AMR determinants correctly predicted the phenotype of uropathogens  
32 in 20 of 32 cases. The metagenomic work-flow was 31h and cost €116 per sample, but could be  
33 reduced to 4.5h and €5 for low-DNA-yield non-UTI samples.

34           The genomic determinants of AMR and their distribution across uropathogens need to be  
35 better understood for prediction of AMR phenotypes by metagenomics. The introduction of the

36 DI demonstrates the potential of semi-quantitative metagenomics to replace culture as rapid  
37 diagnostic method for UTI.

38

## 39 **INTRODUCTION**

40 Urinary tract infections (UTIs) are amongst the most common infections that require  
41 antibiotic prescription, and in catheterized and hospitalized patients UTIs are the most common  
42 nosocomial infections(1-3). The high incidence of UTIs, paired with increasing resistance to first  
43 and second line antibiotics for common uropathogens, such as *Escherichia coli*, stress the need  
44 for rapid diagnostics in aid of appropriate antimicrobial treatment(4). However, the current  
45 standard for diagnosing UTIs is time-consuming as it is based on (semi-)quantitative culture of  
46 urine, followed by identification and antimicrobial susceptibility testing (AST) of isolated  
47 pathogens, which usually takes 1-3 days and for some pathogens up to 7 days. In addition,  
48 growth of clinically relevant bacteria may be hampered by prior antimicrobial treatment, which is  
49 common practice worldwide, resulting in diagnostic challenges and risks of misdiagnoses.

50 Dependence on culture for UTI diagnosis can be bypassed by quantitative molecular  
51 detection of uropathogens in urine specimens. Multiplex real-time PCR-based methods have  
52 shown promise for this purpose but are hampered by exclusive detection of selected micro-  
53 organisms and lack of, or incomplete information on antimicrobial resistance (AMR) patterns(5,  
54 6). Direct metagenomic sequencing of microbial communities in urine provides opportunities for  
55 unbiased detection of uropathogens, including the presence of AMR genes, while current next  
56 generation sequencing (NGS) methods and bio-informatic tools facilitate generation of such  
57 results in a timely fashion. Metagenomic approaches have indeed shown promise in identifying

58 uropathogens and their antimicrobial susceptibility patterns in urine samples(7). However, semi-  
59 quantitative detection of uropathogens is essential in order to replace culture-based diagnostics by  
60 metagenomic approaches since international clinical guidelines use bacterial loads (e.g.  $< 10^3$  or  
61  $10^5$  CFU/mL) to define evidence of UTI in different patient populations(8, 9). To facilitate such  
62 semi-quantitative metagenomic diagnostics for UTI, we developed and tested a metagenomic  
63 sequencing-based algorithm for determination of clinically relevant levels of bacterial DNA,  
64 indicative of urinary tract infections. Results were compared to the reference test semi-  
65 quantitative urine culture followed by species identification by MALDI-TOF, and automated  
66 AST. Whole genome sequencing (WGS) of isolated uropathogens served as benchmark for  
67 expected AMR gene content in metagenomes. We recorded the time to result and direct costs.  
68 The aim of this study was to provide proof of concept for the use of metagenomics as a tool for  
69 detecting UTI and predicting antimicrobial susceptibility patterns.

70

## 71 **METHODS**

### 72 *Urine Samples*

73 During two days, all urine samples obtained from patients with suspected UTIs that were  
74 assessed at the clinical microbiology laboratory for culture, identification, and AST, were  
75 included in the study. We collected up to a maximum of 20ml of surplus urine. Urine samples,  
76 corresponding routine culture plates and AST results were collected and anonymized, according  
77 to local and national ethical requirements for diagnostic studies using surplus routine care patient  
78 samples.

79

80 *Urine Culture*

81 Urine was cultured according to local standard operating procedures (see supplementary  
82 data for summarized SOP). Bacterial isolates were identified morphologically and by MALDI-  
83 TOF MS (MALDI Biotyper, Bruker, Karlsruhe, Germany) and were reported in a semi-  
84 quantified way as the number of colony forming units (CFU) per ml.

85 A culture-positive urine was defined as growth of clinically relevant bacteria at numbers  
86  $>10^3$  CFU/ml and for the purpose of this study all culture-positive urine samples were assigned to  
87 the UTI group. Growth of  $10^3$  CFU/ml or less was reported as “no significant growth”, while  
88 mixed growth of commensal vaginal, rectal, or mucocutaneous bacteria in any numbers or  
89 absence of any growth were reported as “commensal flora” or “no growth”, respectively. The  
90 latter three culture results together formed the non-UTI group. The clinically relevant bacteria  
91 from culture-positive urines were subcultured on Columbia agar plates with 5% sheep blood  
92 (Biomérieux, Marcy-l'Étoile, France) prior to DNA extraction and WGS.

93

94 *Antimicrobial susceptibility testing*

95 AST was done directly on urine by disc diffusion if Gram staining of the urine revealed  
96 the presence of abundant bacteria of a single morphology, or on identified bacterial isolates using  
97 VITEK 2 (version 06.01; Biomérieux), both according to EUCAST guidelines and  
98 breakpoints(see Supplementary data, Summary of SOPs)(10).

99

100

101 *DNA extraction and sequencing*

102 DNA was extracted from 20 ml of urine. If <20 ml of urine was available, the sample was  
103 supplemented with PBS. The urine was first centrifuged for removal of human cells (2000 g, 30  
104 s), followed by centrifugation of the supernatant for pelleting of bacterial cells (8000 g, 10 min).  
105 The bacterial pellet was pre-lysed with an in-house enzyme cocktail comprising  
106 achromopeptidase, mutanolysine, lysostaphine (1000:100:3) and lysozyme (1mg/ml) (Sigma-  
107 Aldrich, St. Louis, MO, USA) in TE buffer. This was followed by lysis using proteinase K and an  
108 in-house lysis buffer (Sodium-docecyl-Sulphate (1 %), Tween-20 (0.5 %) and Sarkosyl (0.5 %)  
109 in TE-buffer), after which automated DNA extraction was performed immediately using the  
110 NucliSENS easyMag platform (Biomérieux) following manufacturer's instructions. DNA from  
111 the cultured bacteria was extracted with the Wizard Genomic DNA Purification Kit (Promega,  
112 Madison, Wisconsin, USA). The Qbit dsDNA HS Assay Kit (ThermoFisher, Waltham,  
113 Massachusetts, USA) was used to measure DNA concentration. The Ion Xpress™ Plus  
114 Fragment Library Kit (Thermo Fisher Scientific) was used for PCR-free, manual library  
115 preparation according to manufacturer's specifications. Library quantification was performed  
116 with the Ion Library TaqMan Quantification Kit (Thermo Fisher Scientific). DNA was  
117 sequenced on the Ion Torrent Proton platform, set to produce an average of 1 million 200 base-  
118 pair length single-end reads per sample. Trimmomatic was used to remove low quality reads  
119 (settings WGS: headcrop 15, sliding window 4:20, Phred score 15; settings metagenomics:  
120 headcrop 15, crop 270, sliding window 4:20, Phred score 10, minlen 50)(11).

121

122

123 *Metagenomic analysis*

124 Human sequences in the metagenomic datasets were removed with Deconseq version  
125 0.4.3, human genome version 37(12). The remaining reads were analyzed for identification of  
126 microbial DNA at subspecies level and determination of the organism's relative abundance using  
127 the CosmosID bioinformatics software package (CosmosID Inc., Rockville, MD)(13-16). A  
128 cloud version is accessible at <https://app.cosmosid.com>. The relative abundance of each bacterial  
129 organism per sample was expressed as a percentage of the total number of bacterial reads  
130 belonging to that organism, normalized for organism-specific genome length. Reads identified as  
131 eukaryote, viral, or archaeal were excluded.

132 All bacteria identified by metagenomics were classified as commensal or uropathogen.  
133 Commensal bacteria comprised members of the genito-urinary tract and skin microbiota  
134 generally not considered as urinary pathogenic. Uropathogens comprised all *Enterobacteriaceae*,  
135 other species known as common causative agents of UTI and putative uropathogens that have  
136 been described to cause UTI in rare cases(17). A full list of bacterial species detected and their  
137 designation is provided in the Supplementary data, table 1.

138 Finally, the resistome i.e., the pool of resistance genes present in the microbial  
139 community, was characterized from the metagenome with the CosmosID (CosmosID Inc.,  
140 Rockville, MD) metagenomic software package using CosmosID's curated antibiotic resistance  
141 gene database. Resistome predictions were compared with results of AST and WGS of the  
142 corresponding isolated bacteria.

143

144

145 *Whole genome sequencing*

146 SPAdes/3.6.0 was used for read-assembly (settings: --iontorrent, -k 21,33,55 --  
147 careful)(18). KmerFinder version 2.1 identified the bacterial species(19). If KmerFinder was  
148 inconclusive, a BLAST search of the NCBI database was done, using the assembled genome(20).

149 ResFinder version 2.1 was used with default settings (90% identity match, 60% coverage)  
150 for detection of acquired antimicrobial resistance (AMR) genes  
151 (<https://cge.cbs.dtu.dk/services/ResFinder/> analysis date: 14<sup>th</sup> June 2017)(21). Additional  
152 acquired and chromosomal genes and mutations were detected from the assembled reads using  
153 the web-based “Resistance Gene Identifier” tool for searching the Comprehensive Antibiotic  
154 Resistance Database (CARD) using “strict” and “perfect” matches only with identity match of  
155  $\geq 99\%$  (<https://card.mcmaster.ca/analyze/rgi> analysis date: 14<sup>th</sup> June 2017)(22). AMR genes  
156 identified by WGS served as benchmark of expected AMR gene content in metagenomic  
157 sequence data.

158

159 *Metagenomic quantification*

160 In order to integrate both DNA quantity and relative abundance in a single diagnostic  
161 measure, thus emulating the qualitative and quantitative properties of the reference test culture,  
162 we created the diagnostic index (DI). The DI is the product of the total quantity of extracted DNA  
163 in nanograms (D) and the relative abundance of uropathogens (RA) for each urine sample:  $DI =$   
164  $D * RA$ . Semi-quantitative culture and metagenomics results were compared one on one per urine  
165 sample for species identification and quantification. The median and interquartile ranges (IQR)  
166 for DNA yield, RA and DI for the groups UTI and non-UTI were compared. Thus, we assessed



167 which metagenomic measure had best discriminatory power, displaying the least overlapping  
168 values between groups UTI and non-UTI.

169

## 170 *Statistics*

171 Differences in relative abundance between uropathogens, as well as differences in total  
172 DNA yield between specimens were analyzed using the Mann-Whitney-U test (RStudio version  
173 0.99.902). A significant difference was defined as a p-value <0.05. Graphs were designed with  
174 RStudio version 0.99.902. Sensitivity, specificity, and 95% confidence intervals (95% CI) were  
175 calculated using MEDCALC ([https://www.medcalc.org/calc/diagnostic\\_test.php](https://www.medcalc.org/calc/diagnostic_test.php)).

176

## 177 **RESULTS**

### 178 *Culture and WGS of isolates*

179 Forty-six consecutive urine samples collected for routine culture were included, results of  
180 which were reported as no-, commensal- or no significant ( $\leq 10^3$  CFU/ml) growth for respectively  
181 12, 19 and 4 urines, hence classified as the non-UTI group. Eleven samples were classified as  
182 culture-positive and constituted the UTI-group. From these 11 specimens, 15 uropathogens were  
183 identified in total: 2 *Enterobacter cloacae*, 6 *Escherichia coli*, 2 *Klebsiella pneumoniae*, 2  
184 *Staphylococcus aureus* and 1 each of *Proteus mirabilis*, *Morganella morganii*, and *Pseudomonas*  
185 *aeruginosa*. Of these, 14 were available for WGS. Species identification by WGS confirmed  
186 identification by MALDI-TOF in all 14 cases.

187

188 *Metagenomics of urine samples*

189 Metagenomic sequencing and analysis was performed for 41 of the 46 included urine  
190 samples. In the UTI group, one sample did not contain surplus urine for metagenomic analysis,  
191 and DNA extraction was unsuccessful for another due to technical difficulties. In the non-UTI  
192 group, zero DNA was extracted from three samples for which culture showed no growth for one  
193 and commensal flora in  $10^2$ CFU/ml for two samples. Thus, 9 UTI and 32 non-UTI urine  
194 samples, containing 5.7-1210.0 ng of DNA, were used for metagenomic sequencing and analysis.

195 Non-UTI group

196 The 11 sequenced urine samples yielding no growth in culture, showed at least one  
197 bacterial species to be present by metagenomics. In these samples, the most frequently detected  
198 genera were *Gardnerella*, *Bifidobacterium*, *Enterococcus*, and *Lactobacillus* (Supplementary  
199 data table 2-C). Of the 19 urine samples that grew commensal flora in culture, 17 were analyzed  
200 by metagenomics and the most frequently observed bacterial taxa were *Bifidobacterium*,  
201 *Lactobacillus*, *Prevotella*, *E. coli*, *Staphylococcus epidermidis*, and *Gardnerella vaginalis*  
202 (Supplementary data table 2-B). All four samples with no significant growth in culture,  
203 respectively showed *E. coli*, *Bifidobacterium* and *Lactobacillus* as most abundant taxa  
204 (Supplementary data table 2-D).

205 UTI group

206 In 9 of 11 urine samples from the UTI group that underwent metagenomic sequencing and  
207 analysis, 12 of the 13 cultured uropathogens were identified amongst the most abundant species  
208 by metagenomics, but some differences were observed (Table 1). Where *S. aureus* (sample 30),  
209 present in mixed growth with *E. coli*, was cultured at a concentration of  $10^4$ CFU/ml, it was not

210 found in the metagenomic dataset. Sample 19 showed *Klebsiella pneumoniae* dominance (90.4%  
211 relative abundance) with only 0.9% relative abundance of *E. coli*, whilst culture suggested equal  
212 growth of  $>10^5$ CFU/ml of both species. Similarly, culture of sample 49 suggested equal growth  
213 of *Enterobacter cloacae* and *Pseudomonas aeruginosa*, whilst metagenomics revealed *E. cloacae*  
214 dominance (relative abundance 53.9%) with *P. aeruginosa* present at 7.8% relative abundance.  
215 Additionally *E. coli* was detected at a RA of 29.4%. Only *S. aureus* ( $10^4$  CFU/ml) was cultured  
216 from sample 8 whilst *Klebsiella species* was additionally detected by metagenomics (Table 1) in  
217 a patient pre-treated with trimethoprim. Metagenomics identified additional taxa compared to  
218 culture in 6 other UTI samples including *Bifidobacterium species*, *Shigella species*, *Lactobacillus*  
219 *species*, *Bacillus cereus*, *Anaerococcus lactolyticus*, *Peptoniphilus harei* and *Atopobium vaginae*.

220

#### 221 *Metagenomic quantification*

222 The median relative abundance (RA) of uropathogens in urine samples from the UTI-  
223 group (93.6%; IQR 7.4) was significantly different ( $p<0.001$ ) from that of the non-UTI group  
224 (0.3%; IQR 11.2) (Supplementary data, figure 1). The median DNA yield from urine samples of  
225 the UTI-group (642.4ng; IQR 456.5; range 203.5 – 1210.0ng) was and significantly different  
226 ( $p<0.001$ ) from the non-UTI group (127.3ng; IQR 295.3; range 0.0-734.8 ng)(Supplementary  
227 data, Figure 2). Where the non-UTI samples had a variable DNA yield, all urine samples from the  
228 UTI group, yielded  $> 203$  ng of DNA, forming a clear threshold below which no UTIs were  
229 found. However, both RA and DNA yield lacked discriminatory power as individual sample  
230 values were overlapping (Supplementary Data, figures 1 and 2).

231 Integrating RA and DNA yield per urine sample in the diagnostic index (DI)  
232 discriminated all 9 culture positive urine samples of the UTI-group from the remaining 32 non-  
233 UTI samples except for one. This sample corresponded to a culture showing no growth whilst  
234 metagenomics identified *Aerococcus urinae* as being present (RA 88%, DNA yield 374 ng)  
235 (Figure 1 and Supplementary Data, Figure 3. The median difference in DI for groups UTI and  
236 non-UTI was significant ( $p < 0.0001$ ), with only 1 overlapping value between groups. The  
237 sensitivity and specificity of the DI for correctly allocating a urine sample to the UTI or non-UTI  
238 group were respectively 100% (95%-CI: 66.4-100%) and 96.7% (95%-CI: 83.8-99.9%)(23).

239

#### 240 *Antimicrobial susceptibility*

241 Phenotypic AST of the isolated uropathogens showed resistance against 0 to 6 drug  
242 classes per isolate, with resistance against an average of 3 drug classes per isolate, cumulating to  
243 38 cases of AMR against a drug class for all 15 uropathogens together (Table 2, Supplementary  
244 data Table 3). Combined Resfinder and CARD analyses of WGS gave genotype-phenotype  
245 agreement in 22 of 38 resistant phenotypes.

246 Nine of 11 urine samples from the UTI group were available for metagenomic resistome  
247 analysis. These 9 samples contained 13 of the 15 cultured uropathogens, comprising 32 cases of  
248 resistance against an antimicrobial drug class, as per phenotypic AST. Metagenomic resistome  
249 analysis and phenotypic AST were concordant in 20 out of 32 cases (Table 2).

250 Five isolates (2 *E. cloacae*, 1 *K. pneumoniae*, 1 *M. morgani* and 1 *P. mirabilis*) derived  
251 from individual urine samples showed a nitrofurantoin resistance phenotype, but a nitrofurantoin  
252 resistance conferring gene was not detected, neither by metagenomic resistome nor by WGS

253 analysis. Metagenomics did not detect a matching AMR gene to explain seven additional  
254 resistance phenotypes in five samples (Table 2) whilst AMR genes corresponding to these  
255 phenotypes were detected in the WGS assembly of the isolates.

256 In four samples, different AMR genes were identified by metagenomics and WGS of the  
257 corresponding isolate, although both could explain the phenotype (Table 2). In 2 of these samples  
258 (*bla<sub>SHV</sub>* and *bla<sub>TEM</sub>* genes detected by WGS in sample 19 and *bla<sub>DHA</sub>* genes in sample 45),  
259 coverage of the metagenomic datasets was too low to distinguish the highly homologous gene  
260 variants (5-10 SNP differences between identified genes) (Supplementary data, Figure 4). The  
261 third, a *bla<sub>ACT-28</sub>* gene detected by WGS in sample 7 was absent in the CosmosID database and  
262 was instead identified as the *bla<sub>ACT-9</sub>* gene (a close variant with 5 SNPs difference) in the  
263 corresponding metagenomic sample. Finally, in sample 30, containing *E. coli* phenotypically  
264 resistant to 3<sup>rd</sup> generation cephalosporins, metagenomic analysis identified an *ampC* gene whilst  
265 the *bla<sub>CTX-M-27</sub>* ESBL-gene was identified in the WGS of the isolate.

266 Metagenomics outperformed WGS in three cases. In sample 28, containing *E. coli*  
267 phenotypically resistant to 3<sup>rd</sup> generation cephalosporins, metagenomics identified an *ampC* gene,  
268 whilst WGS detected *bla<sub>TEM-1B</sub>*, which encodes resistance against penicillins, but not 3<sup>rd</sup>  
269 generation cephalosporins. In two instances a resistance encoding genotype was identified by  
270 metagenomics but an equivalent coding gene was not detected in the corresponding WGS  
271 analysis. Both genotypes matched to a resistant phenotype: *oqxA* and *oqxB* coding for norfloxacin  
272 resistance in sample 45 and *ampC* for amoxicillin resistance in sample 23.

273

274

275 *Cost and timing*

276           The time from urine arrival at the laboratory to obtaining results by culture was  
277 approximately 24h for non-UTI samples and 48-72h for culture and AST for UTI samples. The  
278 cost of this analysis per sample was approximately €4 for a non-UTI and €10 for a urine sample  
279 from the UTI group, excluding cost of staffing.

280           Total time required for metagenomic analysis was approximately 31h per sample, which  
281 included 4.5h for sample centrifugation, DNA extraction, and DNA shearing. A total of 15.5h  
282 were required for library preparation. Four hours were needed for sequencing. The computation  
283 time for CosmosID analysis for all 41 datasets was 3.88 min per sample using a Linux analysis  
284 server with two physical processors, a total of 12 cores (non-multithreaded) and 64GB of  
285 memory. The cost of library preparation and sequencing was €116 per sample, excluding cost of  
286 staffing. Should a screening be made based on quantity of extracted DNA per sample, non-UTI  
287 samples would have a processing time of approximately 4h and a cost of €5, namely the cost of  
288 DNA extraction only.

289

290 **DISCUSSION**

291           This study demonstrates the potential of metagenomics for rapid detection of clinically  
292 relevant levels of uropathogens, thus distinguishing UTI from non-UTI urine samples.

293           Whilst several reports were previously published on the microbial community  
294 composition of urine(24, 25), successful metagenomic identification of uropathogens present in  
295 clinical samples has only been reported once to the best of our knowledge(7). However, that

296 study did not include the semi-quantitative analysis required to establish a diagnosis of UTI(2, 3,  
297 26). If metagenomics is to replace culture, a semi-quantitative analysis of clinically relevant  
298 bacteria present within the urine microbial community is critical to obtain a diagnostic test result,  
299 as has been done in this study.

300 Whilst all samples of the UTI group yielded >203.5 ng of DNA, 12 of 21 samples above  
301 this threshold were non-UTI urine samples, reflective of the fact that the procedure of urine  
302 sample collection is highly prone to bacterial contamination and that urine itself is not sterile(27).  
303 However, for none of the UTI samples was a low amount of DNA (<200 ng) associated with  
304 clinically relevant bacterial growth in culture, independent of the volume of urine from which  
305 DNA was extracted which ranged between 2 and 20 ml (data not shown). This observation  
306 suggests that establishing an unambiguous threshold of DNA extraction yield per urine volume is  
307 feasible due to the fact that in the presence of a UTI the bacterial DNA load in the urine increases  
308 exponentially. Thus, DNA quantity alone could serve as an initial screening step for a diagnostic  
309 algorithm, directing sequencing only those samples with a high DNA yield.

310 The total amount of time needed to detect and identify uropathogens and AMR genes in  
311 the UTI samples was significantly less (31h vs. 48-72hs) for metagenomics than for standard  
312 culture whilst metagenomics provided greater detail. Automated library preparation will further  
313 reduce processing time of UTI samples by several hours. Metagenomic costs were €116 vs. €10  
314 for culture per sample. Although metagenomic sequencing is more expensive at the present time,  
315 sequencing costs are continuing to decline(28). At present €4 and 24h are required for ruling out  
316 a UTI by culture. Should a validated metagenomic diagnostic algorithm be in place, a UTI could  
317 be ruled out in a low yield DNA sample within 4.5 hours and costs would be reduced to the cost  
318 of DNA extraction – approximately €5 – making metagenomics highly competitive.

319 Overlapping values for the UTI and non-UTI samples were observed for both DNA  
320 quantity and relative abundance of uropathogens, indicating these parameters are unsuitable for  
321 identifying urine samples from patients with UTI. However, the DI discriminated UTI and non-  
322 UTI samples extremely well. The single outlier was a sample containing *Aerococcus urinae* in  
323 high relative abundance and high DNA yield, that had been reported to be culture negative, even  
324 though Gram-staining revealed Gram-positive cocci. *A. urinae* can be overlooked in cultures  
325 because of its morphological similarity to coagulase negative staphylococci(29), leading to false  
326 negative culture results.

327 Metagenomic analysis was successfully employed to identify uropathogens in all but two  
328 of the culture positive urine samples (Table 1). In the initial analysis *Morganella morganii* in  
329 sample 45 was not identified. The species was successfully identified after addition of the  
330 genome to CosmosID's reference database, highlighting the importance of well-curated and  
331 complete databases. In sample 30, *S. aureus* was cultured at a concentration of  $>10^4$  CFU/ml. It is  
332 unlikely that insufficient DNA extraction due to poor lysis of Gram-positive bacteria can explain  
333 the apparent lack of *S. aureus* DNA, since an extensive pre-lysis protocol was used, although this  
334 cannot be ruled out(30). Alternatively, contamination of the agar plates may have occurred.

335 CosmosID metagenomic bioinformatics system identified AMR encoding genes that were  
336 confirmed to be present by pure isolate WGS analysis. However for a selected number of cases,  
337 metagenomic and WGS analyses identified a highly similar, but non-identical AMR gene. An  
338 explanation for this discrepancy can be the different databases used in the metagenomic and  
339 WGS analysis. For WGS, it has been shown that different bioinformatic AMR identification tools  
340 produce different results and that databases are often incomplete(31). In addition, low coverage  
341 of gene segments by metagenomic reads resulting in less accurate gene calling could explain



342 differences between metagenomics and WGS. Interestingly, even though metagenomic and WGS  
343 methods identified different gene variants in those few instances, the variants were predicted to  
344 encode the same phenotype. However, accurate phenotype prediction on the basis of genomic  
345 information, whether WGS or metagenomics, remains a challenge. Phenotype clearly depends on  
346 more than the presence or absence of a resistance encoding gene. Whether antibiotic resistance  
347 mechanisms are expressed, depends on complex interplay between accumulating mutations or  
348 acquisition of resistance encoding genes and as yet unknown regulators, as well as environmental  
349 factors(32).

350 In conclusion, we provided a proof of concept of semi-quantitative metagenomic  
351 diagnostics for UTI, including the development of a “diagnostic index” based on uropathogen  
352 relative abundance and total DNA yield per sample, to facilitate rapid classification of urine  
353 samples to UTI and non-UTI groups.. In future studies, clinical characteristics of patients and  
354 prior antibiotic use should be included to determine the sensitivity, specificity, and the positive  
355 and negative predictive value of this metagenomics analysis approach applied to urine samples  
356 using culture as the reference test.

357

## 358 **ACKNOWLEDGEMENTS**

359 The authors would like to thank Cosimo Cristella for assistance in using RStudio and  
360 Frank Baas for critical input in sequencing planning and experiment design.

361 This study was supported by the COMPARE Consortium, which has received funding  
362 from the European Union’s Horizon 2020 research and innovation program under grant

363 agreement No. 64347. The funders had no role in study design, data collection and interpretation,  
364 or the decision to submit the work for publication.

365 Rita R. Colwell is Founder and Chairman of the Board of CosmosID®, a bioinformatics  
366 company, and some of the other authors are employees of the company. Affiliation with  
367 CosmosID does not alter the authors' adherence to all JCM policies as detailed in the online  
368 instructions for authors. Authors declare no conflict of interest with regards to this manuscript.

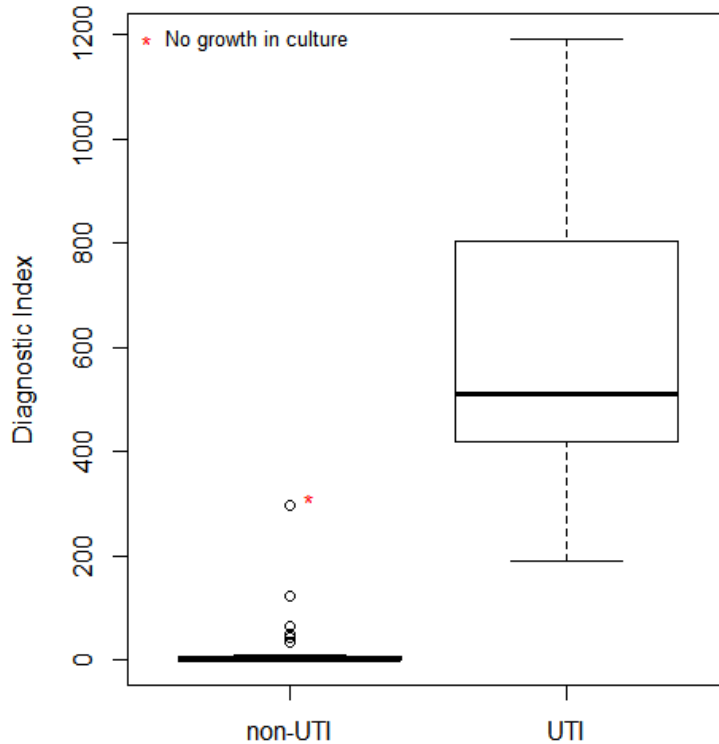
369

## 370 **AUTHORS CONTRIBUTION**

371 V. A. J. wrote the manuscript. V. A. J., S. M. and M. E. J. performed the experiments. V.  
372 A. J., S. M., N. W., P. S. and N. A. H. performed the bio-informatics analysis. C. E. V.  
373 supervised the clinical diagnostic analysis. S. M., V. A. J., C. E. V., B. d W., R. R. C., M. D. d J.,  
374 and C. S. designed the study. All authors provided critical review of the data and manuscript.

375

376 Figure 1. Diagnostic Index (DI) for urine samples in the groups UTI and non-UTI



377  
378 Legend Figure 1. The “Diagnostic Index” (DI), the product of the relative abundance of  
379 uropathogens, and total DNA quantity per sample was computed for the groups UTI and non-  
380 UTI. The median DI for the UTI group was 512.0 (IQR 382.8) versus 0.11 (IQR 5.4) for the  
381 non-UTI group, which was a significant difference ( $p < 0.00001$ , Mann Whitney-U test). The red  
382 asterisk represents urine sample 44 showing *Aerococcus urinae* in metagenomic analysis (88%  
383 relative abundance, 325.6ng DNA extracted from that sample), which was classified as a  
384 uropathogen in this study. Culture showed no growth.

385

386 Table 1. Comparing species identification by semi-quantitative culture and metagenomics'  
 387 relative abundance (Rel. ab.) of organisms per urine sample of the UTI group.

388  
 389

Sample ID	Culture		Metagenomics	
	Species	CFU/ml	Species	Rel. ab. (%)
7	<i>Enterobacter cloacae</i>	10 <sup>4</sup>	<i>Enterobacter cloacae</i>	58.8
			<i>Enterobacter species</i>	34.8
			<i>Bacillus cereus</i>	6.4
8	<i>Staphylococcus aureus</i>	>10 <sup>4</sup>	<i>Staphylococcus aureus</i>	76.3
			<i>Klebsiella species</i>	23.7
18	<i>Escherichia coli and skin flora</i>	>10 <sup>4</sup>	<i>Escherichia coli</i>	55.8
			<i>Bifidobacterium pseudocatenulatum</i>	25.0
			<i>Bifidobacterium longum</i>	11.4
			<i>Shigella flexneri</i>	3.1
			Other	4.7
19	<i>Escherichia coli</i>	>10 <sup>5</sup>	<i>Klebsiella species</i>	90.4
			<i>Klebsiella pneumoniae</i>	8.1
	<i>Klebsiella pneumoniae</i>	>10 <sup>5</sup>	<i>Escherichia coli</i>	0.9
			Other	0.7
23	<i>Escherichia coli</i>	10 <sup>4</sup>	<i>Escherichia coli</i>	100.0
28	<i>Escherichia coli</i>	10 <sup>4</sup>	<i>Escherichia coli</i>	86.6
			<i>Shigella sonnei</i>	6.6
			Other	6.8
30	<i>Escherichia coli</i>	>10 <sup>5</sup>	<i>Escherichia coli</i>	80.4
			<i>Anaerococcus</i>	
	<i>Staphylococcus aureus</i>	10 <sup>4</sup>	<i>lactolyticus</i>	9.3
			<i>Escherichia species</i>	5.3
			<i>Peptoniphilus harei</i>	3.6
			<i>Staphylococcus aureus</i>	0.0
			Other	1.5
45	<i>Proteus mirabilis</i>	>10 <sup>5</sup>	<i>Proteus mirabilis</i>	66.4
			<i>Morganella morganii</i>	26.9
	<i>Morganella morganii</i>	>10 <sup>5</sup>	<i>Klebsiella variicola</i>	2.1
			<i>Klebsiella pneumoniae</i>	1.7
			Other	2.8
49	<i>Enterobacter cloacae</i>	>10 <sup>4</sup>	<i>Enterobacter cloacae</i>	53.9
			<i>Pseudomonas aeruginosa</i>	
	<i>Pseudomonas aeruginosa</i>	>10 <sup>4</sup>	<i>Escherichia coli</i>	29.4
			<i>Atopobium vaginae</i>	8.9
			<i>Pseudomonas aeruginosa</i>	7.8

390 Table 2. Phenotypic AST per drug class compared to genotypic AMR gene detection using the  
 391 detection tools ResFinder and CARD (WGS), and CosmosID (metagenomics).

392

Sample ID	Species	CULTURE	WHOLE GENOME SEQUENCING		METAGENOMICS
		Antimicrobial drug class resistance phenotype	ResFinder <sup>1</sup>	CARD <sup>2</sup>	CosmosID
7	<i>E. cloacae</i>	2 <sup>nd</sup> gen. cephalosporin Pip/taz <sup>3</sup> Fosfomycin Nitrofurantoin	fosA	blaACT-28 blaACT-28	blaACT-9, fosA
8	<i>S. aureus</i>	None	NA <sup>4</sup>	NA <sup>4</sup>	None
9	<i>E. coli</i>	Folic acid synthesis inhibitors	sul2-like	sul3	NA <sup>4</sup>
17	<i>K. pneumoniae</i>	Penicillins 3 <sup>rd</sup> gen. cephalosporin Fluoroquinolones Folic acid synthesis inhibitors Nitrofurantoin	blaCTX-M-15, blaSHV-26, blaTEM-1B blaCTX-M-15 oqxA-like, oqxB-like, QnrB66-like dfrA14-like	blaCTX-M-15, blaSHV-26, blaTEM-1 blaCTX-M-15 QnrB1 dfrA14	NA <sup>4</sup>
18	<i>E. coli</i>	Folic acid synthesis inhibitors		sul3	None
19	<i>K. pneumoniae</i>	Penicillins Nitrofurantoin	blaSHV-1	blaSHV-1	
	<i>E. coli</i>	Penicillins (amoxicillin) Fluoroquinolones Folic acid synthesis inhibitors	blaTEM-1B dfrA1, drfA14-like, sul1, sul2	blaTEM-206 parC, gyrA, mfd dfrA1	blaTEM-128, blaSHV-69, AmpC, oqxA, oqxB, drfA14
23	<i>E. coli</i>	Penicillins (amoxicillin)			AmpC
28	<i>E. coli</i>	Penicillins 3 <sup>rd</sup> gen. cephalosporin Folic acid synthesis inhibitors	blaTEM-1B dfrA5, Sul2	blaTEM-1 sul2	AmpC, dfrA5, strB, aadA1
30	<i>E. coli</i>	Penicillins 3 <sup>rd</sup> gen. cephalosporin Pip/taz <sup>3</sup> Fluoroquinolones	blaCTX-M-27-like blaCTX-M-27-like blaCTX-M-27-like	gyrA	AmpC
	<i>S. aureus</i>	Penicillins (penicillin)	blaZ-like		
45	<i>M. morgani</i>	Penicillins 2 <sup>nd</sup> gen. cephalosporins Fluoroquinolones Fosfomycine Tobramycine Nitrofurantoin	blaDHA-1-like blaDHA-1 QnrD-like	blaDHA-14 blaDHA-14	blaDHA-2, QnrD, oqxA, oqxB, FosA
	<i>P. mirabilis</i>	Penicillins (amoxicillin) Nitrofurantoin			
49	<i>E. cloacae</i>	2 <sup>nd</sup> gen. cephalosporin Nitrofurantoin	blaACT-15-like	blaACT-15	
	<i>P. aeruginosa</i>	3 <sup>rd</sup> gen. cephalosporins Carbapenems (imipenem) Fosfomycin	blaOXA-50-like, blaPAO-like blaOXA-50-like, blaPAO-like fosA-like	PDC-7 PDC-7 fosA	blaACT-15

393 Legend Table 2. If only 1 antibiotic was tested per antimicrobial drug class, the tested drug was  
394 noted between brackets. <sup>1</sup> Database contains acquired AMR determining genes only. <sup>2</sup> Database  
395 contains chromosomal and acquired AMR determining genes and mutations.  
396 <sup>3</sup>Piperacillin/tazobactam. <sup>4</sup>Not available.

397

398

## 399 REFERENCES

- 400 1. Linder JA, Huang ES, Steinman MA, Gonzales R, Stafford RS. 2005. Fluoroquinolone prescribing in  
401 the United States: 1995 to 2002. *Am J Med* 118:259-68.
- 402 2. Hooton TM, Bradley SF, Cardenas DD, Colgan R, Geerlings SE, Rice JC, Saint S, Schaeffer AJ,  
403 Tambayh PA, Tenke P, Nicolle LE, Infectious Diseases Society of A. 2010. Diagnosis, prevention,  
404 and treatment of catheter-associated urinary tract infection in adults: 2009 International Clinical  
405 Practice Guidelines from the Infectious Diseases Society of America. *Clin Infect Dis* 50:625-63.
- 406 3. Gupta K, Hooton TM, Naber KG, Wullt B, Colgan R, Miller LG, Moran GJ, Nicolle LE, Raz R,  
407 Schaeffer AJ, Soper DE, Infectious Diseases Society of A, European Society for M, Infectious D.  
408 2011. International clinical practice guidelines for the treatment of acute uncomplicated cystitis  
409 and pyelonephritis in women: A 2010 update by the Infectious Diseases Society of America and  
410 the European Society for Microbiology and Infectious Diseases. *Clin Infect Dis* 52:e103-20.
- 411 4. Cagnacci S, Gualco L, Debbia E, Schito GC, Marchese A. 2008. European emergence of  
412 ciprofloxacin-resistant *Escherichia coli* clonal groups O25:H4-ST 131 and O15:K52:H1 causing  
413 community-acquired uncomplicated cystitis. *J Clin Microbiol* 46:2605-12.
- 414 5. van der Zee A, Roorda L, Bosman G, Ossewaarde JM. 2016. Molecular Diagnosis of Urinary Tract  
415 Infections by Semi-Quantitative Detection of Uropathogens in a Routine Clinical Hospital Setting.  
416 *PLoS One* 11:e0150755.
- 417 6. Hansen WL, van der Donk CF, Bruggeman CA, Stobberingh EE, Wolffs PF. 2013. A real-time PCR-  
418 based semi-quantitative breakpoint to aid in molecular identification of urinary tract infections.  
419 *PLoS One* 8:e61439.
- 420 7. Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Moller N, Aarestrup FM.  
421 2014. Rapid whole-genome sequencing for detection and characterization of microorganisms  
422 directly from clinical samples. *J Clin Microbiol* 52:139-46.
- 423 8. Hooton TM, Bradley SF, Cardenas DD, Colgan R, Geerlings SE, Rice JC, Saint S, Schaeffer AJ,  
424 Tambayh PA, Tenke P, Nicolle LE. 2010. Diagnosis, prevention, and treatment of catheter-  
425 associated urinary tract infection in adults: 2009 International Clinical Practice Guidelines from  
426 the Infectious Diseases Society of America. *Clin Infect Dis* 50:625-63.
- 427 9. Nicolle LE, Bradley S, Colgan R, Rice JC, Schaeffer A, Hooton TM. 2005. Infectious Diseases  
428 Society of America guidelines for the diagnosis and treatment of asymptomatic bacteriuria in  
429 adults. *Clin Infect Dis* 40:643-54.

- 430 10. EUCAST. Breakpoint tables for interpretation of MICs and zone diameters. Version 7.0, 2017. .  
431 <http://www.eucast.org> Accessed 05-01-2017.
- 432 11. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence  
433 data. *Bioinformatics* 30:2114-20.
- 434 12. Schmieder R, Edwards R. 2011. Fast identification and removal of sequence contamination from  
435 genomic and metagenomic datasets. *PLoS One* 6:e17288.
- 436 13. Ottesen A, Ramachandran P, Reed E, White JR, Hasan N, Subramanian P, Ryan G, Jarvis K, Grim C,  
437 Daquigan N, Hanes D, Allard M, Colwell R, Brown E, Chen Y. 2016. Enrichment dynamics of  
438 *Listeria monocytogenes* and the associated microbiome from naturally contaminated ice cream  
439 linked to a listeriosis outbreak. *BMC Microbiol* 16:275.
- 440 14. Ponnusamy D, Kozlova EV, Sha J, Erova TE, Azar SR, Fitts EC, Kirtley ML, Tiner BL, Andersson JA,  
441 Grim CJ, Isom RP, Hasan NA, Colwell RR, Chopra AK. 2016. Cross-talk among flesh-eating  
442 *Aeromonas hydrophila* strains in mixed infection leading to necrotizing fasciitis. *Proc Natl Acad*  
443 *Sci U S A* 113:722-7.
- 444 15. Hasan NA, Young BA, Minard-Smith AT, Saeed K, Li H, Heizer EM, McMillan NJ, Isom R, Abdullah  
445 AS, Bornman DM, Faith SA, Choi SY, Dickens ML, Cebula TA, Colwell RR. 2014. Microbial  
446 community profiling of human saliva using shotgun metagenomic sequencing. *PLoS One*  
447 9:e97699.
- 448 16. Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, Gibbons SM, Larsen P,  
449 Shogan BD, Weiss S, Metcalf JL, Ursell LK, Vazquez-Baeza Y, Van Treuren W, Hasan NA, Gibson  
450 MK, Colwell R, Dantas G, Knight R, Gilbert JA. 2014. Longitudinal analysis of microbial interaction  
451 between humans and the indoor environment. *Science* 345:1048-52.
- 452 17. Foxman B. 2010. The epidemiology of urinary tract infection. *Nat Rev Urol* 7:653-60.
- 453 18. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,  
454 Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.  
455 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J*  
456 *Comput Biol* 19:455-77.
- 457 19. Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H, Sicheritz-Ponten T,  
458 Aarestrup FM, Ussery DW, Lund O. 2014. Benchmarking of methods for genomic taxonomy. *J*  
459 *Clin Microbiol* 52:1529-39.
- 460 20. Mount DW. 2007. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc* 2007:pdb  
461 top17.
- 462 21. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen  
463 MV. 2012. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother*  
464 67:2640-4.
- 465 22. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S,  
466 Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL,  
467 Pawlowski AC, Johnson TA, Brinkman FS, Wright GD, McArthur AG. 2017. CARD 2017: expansion  
468 and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids*  
469 *Res* 45:D566-D573.
- 470 23. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. 1981. Selection and interpretation of  
471 diagnostic tests and procedures. *Principles and applications. Ann Intern Med* 94:557-92.
- 472 24. Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ,  
473 Schreckenberger PC. 2014. Urine Is Not Sterile: Use of Enhanced Urine Culture Techniques To  
474 Detect Resident Bacterial Flora in the Adult Female Bladder. *Journal of Clinical Microbiology*  
475 52:871-876.
- 476 25. Siddiqui H, Lagesen K, Nederbragt AJ, Jeansson SL, Jakobsen KS. 2012. Alterations of microbiota  
477 in urine from women with interstitial cystitis. *BMC Microbiology* 12:205.

- 478 26. Burd EM, Kehl KS. 2011. A Critical Appraisal of the Role of the Clinical Microbiology Laboratory in  
479 the Diagnosis of Urinary Tract Infections. *Journal of Clinical Microbiology* 49:S34-S38.
- 480 27. Thomas-White K, Brady M, Wolfe AJ, Mueller ER. 2016. The bladder is not sterile: History and  
481 current discoveries on the urinary microbiome. *Curr Bladder Dysfunct Rep* 11:18-24.
- 482 28. NIH NHGRI. July 6, 2016 2016. <https://www.genome.gov/sequencingcosts/>. Accessed June 8,  
483 2017.
- 484 29. Grude N, Jenkins A, Tveten Y, Kristiansen BE. 2003. Identification of *Aerococcus urinae* in urine  
485 samples. *Clinical Microbiology and Infection* 9:976-979.
- 486 30. Ezaki T, Suzuki S. 1982. Achromopeptidase for lysis of anaerobic gram-positive cocci. *Journal of*  
487 *Clinical Microbiology* 16:844-846.
- 488 31. Xavier BB, Das AJ, Cochrane G, De Ganck S, Kumar-Singh S, Aarestrup FM, Goossens H, Malhotra-  
489 Kumar S. 2016. Consolidating and exploring antibiotic resistance gene data resources. *Journal of*  
490 *Clinical Microbiology* doi:10.1128/jcm.02717-15.
- 491 32. Hughes D, Andersson DI. 2017. Environmental and genetic modulation of the phenotypic  
492 expression of antibiotic resistance. *FEMS Microbiology Reviews* 41:374-391.

493