Journal of Clinical Microbiology

Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples

Henrik Hasman, Dhany Saputra, Thomas Sicheritz-Ponten, Ole Lund, Christina Aaby Svendsen, Niels Frimodt-Møller and Frank M. Aarestrup

Updated information and services can be found at:
http://jcm.asm.org/content/52/1/139

These include:

REFERENCES	This article cites 27 articles, 11 of which can be accessed free at: http://jcm.asm.org/content/52/1/139#ref-list-1
CONTENT ALERTS	Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»

Information about commercial reprint orders: **http://journals.asm.org/site/misc/reprints.xhtml**
To subscribe to to another ASM Journal go to: **http://journals.asm.org/site/subscriptions/**

Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples

Henrik Hasman,[a] Dhany Saputra,[b] Thomas Sicheritz-Ponten,[b] Ole Lund,[b] Christina Aaby Svendsen,[a] Niels Frimodt-Møller,[c] Frank M. Aarestrup[a]

National Food Institute, Technical University of Denmark, Lyngby, Denmark[a]; Systems Biology, Technical University of Denmark, Lyngby, Denmark[b]; Hvidovre Hospital, Hvidovre, Denmark[c]

Whole-genome sequencing (WGS) is becoming available as a routine tool for clinical microbiology. If applied directly on clinical samples, this could further reduce diagnostic times and thereby improve control and treatment. A major bottleneck is the availability of fast and reliable bioinformatic tools. This study was conducted to evaluate the applicability of WGS directly on clinical samples and to develop easy-to-use bioinformatic tools for the analysis of sequencing data. Thirty-five random urine samples from patients with suspected urinary tract infections were examined using conventional microbiology, WGS of isolated bacteria, and direct sequencing on pellets from the urine samples. A rapid method for analyzing the sequence data was developed. Bacteria were cultivated from 19 samples but in pure cultures from only 17 samples. WGS improved the identification of the cultivated bacteria, and almost complete agreement was observed between phenotypic and predicted antimicrobial susceptibilities. Complete agreement was observed between species identification, multilocus sequence typing, and phylogenetic relationships for *Escherichia coli* and *Enterococcus faecalis* isolates when the results of WGS of cultured isolates and urine samples were directly compared. Sequencing directly from the urine enabled bacterial identification in polymicrobial samples. Additional putative pathogenic strains were observed in some culture-negative samples. WGS directly on clinical samples can provide clinically relevant information and drastically reduce diagnostic times. This may prove very useful, but the need for data analysis is still a hurdle to clinical implementation. To overcome this problem, a publicly available bioinformatic tool was developed in this study.

Microbial whole-genome sequencing (WGS) holds great promise for enhancing diagnostic and public health microbiology (1–3). Its great value in describing and improving our understanding of bacterial evolution, outbreaks, and transmission events has been shown in a number of recent studies, including studies of *Staphylococcus aureus* (4–6), *Vibrio cholerae* (7), *Escherichia coli* (8), and *Mycobacterium tuberculosis* (9) and surveillance of antimicrobial resistance (10).

The next natural step is to translate this technology from a research tool into one with clinical utility in routine diagnostic settings. Retrospective use of benchtop sequencing for selected isolates of methicillin-resistant *Staphylococcus aureus* (MRSA) (11, 12) and *Clostridium difficile* (11) has indicated the great potential of the technology for understanding and potentially limiting intrahospital transmission of these important pathogens. The first attempts to use the technology in real or near-real time have recently been published (13). However, so far the focus has been mainly on using whole-genome sequencing for isolated and purified bacterial isolates.

Rapid diagnostic identification and characterization of infectious pathogens are essential to guide therapy, to predict outcomes, and to detect transmission events or treatment failures. Current clinical microbial diagnostic methods are mainly based on conventional culturing of clinical samples on different agar plates, followed by susceptibility testing and further characterization on a case-by-case basis. Depending on the pathogen, this procedure often takes 1 to 2 days for culturing, an additional 1 to 2 days for species identification and susceptibility testing, and weeks for molecular typing. Using whole-genome sequencing directly on isolates can theoretically reduce the processing time to 1 to 2 days for culturing and around 12 h for

sequencing and analysis (2). However, if it was feasible to perform sequencing directly on clinical samples, then this could further reduce time and improve diagnoses.

Several methods for rapid diagnostic testing directly with clinical samples have been developed and evaluated, including PCR (14) and matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) (15). These technologies, however, do not give information beyond species identification.

Obvious targets for using sequencing directly on clinical samples are slowly growing or difficult-to-culture pathogens. Whole-genome amplification followed by sequencing has recently been performed for the sexually transmitted intracellular pathogen *Chlamydia trachomatis* (16, 17). Another successful study used fecal samples from a recent *E. coli* outbreak and identified the outbreak strain from data generated directly from the samples (18). Both studies focused on an *a priori* known pathogen and showed great dependence on advanced molecular technologies and bioinformatic analysis. Especially the availability of easy and fast bioinformatic analysis that can be used in real time is a press-

ing need for the widespread use of next-generation sequencing in clinical microbiology.

Compared to other clinical samples, urine is a less complex matrix, with limited human DNA contamination and relatively high numbers of bacterial cells. Here, we evaluate the use of WGS directly on urine samples using benchtop sequencing technology, and we compare this with conventional bacteriological methods and WGS of cultured bacteria. Furthermore, we have developed a fast bioinformatic tool for data analysis, which reduces the bioinformatic processing time from days/months to a few hours.

MATERIALS AND METHODS

Clinical samples. The clinical microbiological laboratory at Hvidovre Hospital examines approximately 120,000 clinical samples every year, of which approximately 70,000 are urine samples. Urine samples are collected in sterile tubes (Urine Monovette; Sarsted, Nümbrecht, Germany). A total of 35 random urine samples, each with a volume of approximately 10 ml, from two separate days in April and September 2012 were selected for this study. All urine samples received were from patients suspected to have urinary tract infections (UTIs).

Bacterial isolation, identification, and susceptibility testing. Blood agar plates were used for culturing. From the urine samples, a total of 100 μl and 10- and 100-fold serial dilutions were spread on blood agar plates; after overnight incubation under aerobic conditions, the plates were examined for purity and the numbers of colonies were counted. At least one colony of the predominant colony type was subcultured and identified to the species level using microscopy, KOH testing, and subculturing on BBL CHROMagar Orientation medium (BD Diagnostic Systems). Pure cultures were stored for WGS at −80°C in cryotubes containing 30% glycerol. Antimicrobial susceptibility testing was performed as MIC testing using microtiter plates (10).

DNA isolation and sequencing. DNA isolation from pure cultures was performed using the Easy-DNA kit (Invitrogen) with an additional pretreatment step. Initially, the cells were inoculated onto a blood agar plate from the cryotubes described above and were incubated overnight at 37°C. A single colony was then inoculated into 10 ml brain heart infusion (BHI) broth and incubated overnight at 37°C with gentle shaking (75 rpm). The 10-ml overnight culture was centrifuged at $5,000 \times g$ for 10 min and resuspended in 200 μl phosphate-buffered saline. Lysozyme (30 μl of a 10 g/liter suspension, to a final concentration of 1.3 g/liter) was added to this mixture, and the cells were incubated for 20 min at 37°C. After incubation, 30 μl of 10% sodium dodecyl sulfate was added and the tubes were gently mixed. Finally, 15 μl of proteinase K (20 g/liter) was added and the samples were incubated for 20 min at 37°C. DNA was then purified as described in the Easy-DNA protocol. Between 44 and 100 ng of genomic DNA was used for sequencing. The isolates were sequenced on the Ion Torrent PGM system (Life Technologies), following the manufacturer's protocols for 200-bp genomic DNA fragment library preparation (Ion Xpress Plus gDNA and Amplicon Library 98 preparation), template preparation (Ion OneTouch system), and sequencing (Ion PGM 200 sequencing kit). Purified DNA from urine samples was sequenced individually on 316 DNA chips, while DNA from single isolates was bar coded according to the library kit and sequenced in pairs on 316 DNA chips or in fours on 318 DNA chips. For urine, a total of 10 ml urine was initially centrifuged at $2,000 \times g$ for 30 s to precipitate human cells. The bacterial cells were precipitated by centrifugation at $15,000 \times g$ for 5 min, the supernatant was discarded, and DNA was isolated and sequenced as described above.

Analysis of sequencing results. (i) k-mer-based species identification. A total of 1,647 complete bacterial genomes were downloaded from the NCBI database; 16-mers from these sequences were stored in a database. To limit the size of the database, only 16-mers starting with the sequence ATGA was retained. This reduced the database approximately 4^4 (256-fold). This was implemented in the program maketemplatedb.py. Another program, findtemplate.py, was used to search the database. This program finds the unique k-mers in the input file and outputs the number

of times each of the GenBank entries in the database is identical to one of these k-mers. The program was run with the "winner takes all" option, where each k-mer only counts in the specific template containing it that obtained the most hits in the first round of mapping. The significance of the match was calculated by testing the equation $z = (h - e)/\text{sqrt}(h + e)$ in a normal distribution. Here, h is the number of hits in a given sequence and e is the expected number of hits, $e = H \cdot n/N$, where H is the total number of hits in the database and n and N are the numbers of k-mers in the target sequence and the entire database, respectively. The P value obtained was corrected for multiple testing using the Bonferroni method, by multiplying the P value obtained by the number of entries in the database.

(ii) Sequence analysis and species distribution using MG-RAST. Raw sequencing data from the Ion Torrent PGM system was uploaded to the MG-RAST server (http://metagenomics.anl.gov). Data were analyzed using the following (default) pipeline options: removal of dereplication events, removal of host (*Homo sapiens*) DNA (NCBI v36), dynamic trimming values of 15 (phred score) and 5 bases, a length-filtering value of 2.0, and an ambiguous base filtering value of 5. MG-RAST was used to estimate the level of host contamination and the relative distribution of bacterial species.

(iii) Species and microbial consortium identification with chainmapper.py. To identify species and the microbial community profile from direct sequencing, we developed chainmapper.py. Here the raw sequence data were automatically trimmed and subsequently aligned with different reference genomes using BWA software on our high-performance computing installation. After removal of contamination from human tissue by fast-mapping all reads to the human genome (hs.build37.1; 90% coverage and 80% identity), all remaining nonmapped reads were mapped against all complete bacterial genomes (NCBI, 26 August 2012) and bacterial draft genomes (NCBI, 3 September 2012) using 50% identity over 50% coverage. Again, all reads that were not mapped against any reference genome were then mapped against complete and draft fungal genomes (NCBI, 20 September 2012), sequences from the Human Microbiome Project and MetaHIT (NCBI, 28 September 2012), and complete and draft protozoan and viral genomes (NCBI, 27 September 2012). The final remaining reads that did not match anything were then mapped against the complete nucleotide database using Bowtie. The organism composition summary was created by the number of reads mapped to each distinct organism, and the community profile and abundance estimation graph were produced as a .pdf file giving information on the number and percentage of reads mapping to each database/bacterial species. A threshold of 50,000 reads and a minimum of 1% of all reads were set for identification of a bacterial species.

(iv) Multilocus sequence typing and determination of resistance genes. The multilocus sequence type was determined from WGS sequencing data for all samples and isolates for which a multilocus sequence typing (MLST) scheme is available, as described previously (19). The presence of known acquired resistance genes was determined by mapping the data from all samples and isolates to an online database of almost 2,000 resistance gene variants (20).

(v) Phylogenetic analysis. Based on results for isolated bacteria and data obtained from direct sequencing, a phylogenetic tree was constructed for the most commonly identified bacterial species, using a previously reported online method (21).

RESULTS

Conventional identification. For 19 of the 35 samples, bacterial colonies were growing on blood agar plates after overnight incubation at 37°C. In two cases, the cultures were mixed to such an extent that it was not possible to differentiate specific colony types; in two cases, two different types of colonies were identified. A total of 19 different isolates were selected for species identification and antimicrobial susceptibility testing (Table 1). Using conventional identification, nine isolates from eight samples were identified as

TABLE 1 Isolation and identification of organisms from urinary samples based on conventional methods, whole-genome sequencing of isolated bacteria, and direct sequencing

Sample no.	Culture result (CFU)	Conventional identification	WGS-based identification, strain	Direct sequencing identification		Chainmapper identification (%)[a]	
				k-mer	MG-RAST (%)	Species	Genus
1				*Clostridium* sp.	*Lactobacillus* (42)	*L. iners* (3.5)	*Lactobacillus* (4.8)
3	≥10^5	*Enterococcus* spp.	*E. faecalis*, ST40	*E. faecalis*	*Enterococcus* (50)	*E. faecalis* (28)	*Enterococcus* (28)
4	10^4	Gram-positive rods	*Clostridium*	*Clostridium* sp.	*Lactobacillus* (78)	*L. iners* (33), *Lactobacillus* sp. (11.6)	*Lactobacillus* (45.8)
6	≥10^5	*E. coli*	*E. coli*, ST14	*E. coli*	*E. coli* (52)	*E. coli* (60), *Escherichia* sp. (10.8), *Bifidobacterium breve* (1.5)	*Escherichia* (71.3), *Bifidobacterium* (1.7), *Shigella* (1.2)
7				*Clostridium* sp.	*G. vaginalis* (15), *Bifidobacterium* (15)	*G. vaginalis* (3.78)	*Gardnerella* (3.78)
8				*Clostridium* sp.	*Lactobacillus* (53)	*L. iners* (6), *Lactobacillus* sp. (2.4)	*Lactobacillus* (8.7)
10	≥10^5	*E. coli*	*E. coli*, ST409	*E. coli*	*E. coli* (50)	*E. coli* (44), *Escherichia* sp. (15), *Citrobacter freundii* (5.5), *Citrobacter* sp. (5.2), *Shigella* sp. (3.3)	*Escherichia* (60), *Citrobacter* (12.2), *Shigella* (5.2)
12	≥10^5	*E. coli*	*E. coli*, ST95	*E. coli*	*E. coli* (38)	*E. coli* (23), *Escherichia* sp. (7), *Bifidobacterium bifidum* (1.6)	*Escherichia* (30), *Bifidobacterium* (1.6)
13				*Clostridium* sp.	*Prevotella* (22)	*Prevotella timonensis* (2)	*Prevotella* (4)
16	NC[b]	*Proteus* sp.	*Proteus mirabilis*	*Clostridium* sp.			
19	≥10^5	*E. coli*	*E. coli*, ST127	*P. mirabilis*	*Proteus* (13)	*P. mirabilis* (2.9)	*P. mirabilis* (3)
20	≥10^5	*E. coli*	*E. coli*, ST1193	*E. coli*	*E. coli* (63)	*E. coli* (78), *Escherichia* sp. (13)	*Escherichia* (91)
21	NC	*Proteus* sp.	*P. mirabilis*	*E. coli*	*E. coli* (54)	*E. coli* (43), *Escherichia* sp. (10)	*Escherichia* (54)
	10^4	*E. coli*	*E. coli*, ST998				
24	≥10^5	*E. coli*	*E. coli*, ST227	*P. mirabilis*	*Proteus* (18), *E. coli* (11)	*P. mirabilis* (26), *Aerococcus urinae* (19.45), *E. coli* (7.5)	*Proteus* (26), *Aerococcus* (19.47), *Escherichia* (8.8)
		E. coli	*E. coli*, ST227	*E. coli*			
25	10^3	*Enterococcus* sp.	*E. faecalis*, ST16	*E. coli*	*E. coli* (57)	*E. coli* (51), *Escherichia* sp. (20), *Shigella* sp. (3)	*Escherichia* (73), *Shigella* (4.5)
26	10^3	*E. coli*	*E. coli*, ST597	*E. faecalis*	*Enterococcus* (48)	*E. faecalis* (25)	*Enterococcus* (26)
27	≥10^5	*Staphylococcus* sp.	*S. lugdunensis*	*E. coli*	*E. coli* (19)	*E. coli* (4)	*Escherichia* (5)
28	10^4	Mixed culture	ND[c]	*S. lugdunensis*	*Staphylococcus* (83)	*S. lugdunensis* (59)	*Staphylococcus* (60)
29	≥10^5	*Enterococcus* sp.	*E. faecalis*, ST19	*E. coli*	*E. coli* (26)	*E. coli* (23)	*Escherichia* (27.6)
31	≥10^5	*Enterococcus* sp.	*Clostridium*	*E. faecalis*	*Enterococcus* (48)	*E. faecalis* (17.4)	*Enterococcus* (17.5)
32	≥10^5	*Enterococcus* sp.	*E. faecalis*, ST41	*Clostridium* sp.	*Enterococcus* (29)	*E. faecium* (5.9)	*Enterococcus* (6.2)
33	≥10^5	*Enterococcus* sp.	*E. faecalis*, ST40	*E. faecalis*	*Enterococcus* (65)	*E. faecalis* (44)	*Enterococcus* (44)
34	≥10^5	Mixed culture	ND	*E. faecalis*	*Enterococcus* (40)	*E. faecalis* (13), *E. coli* (1.2)	*Enterococcus* (13), *Escherichia* (1.7)
35[d]				*E. coli*, *E. faecalis*	*Enterococcus* (12), *E. coli* (9)	*E. faecalis* (3), *E. coli* (2)	*Enterococcus* (3), *Escherichia* (2.6)

[a] Percentages of the sequencing reads mapping to a given species when using Chainmapper are included in parentheses.
[b] NC, not countable.
[c] ND, not determined.
[d] Polymicrobial sample.

Escherichia coli, six as *Enterococcus* spp., two as *Proteus* spp., and one as a *Staphylococcus* sp. One isolate could not be identified (Table 1).

Sequencing of cultured isolates. Whole-genome sequencing of the 19 isolates obtained by cultivation confirmed the results from the conventional identification in 17 cases (Table 1). In one case, an isolate that could not be identified to the genus level using the simple conventional scheme was identified by WGS as *Clostridium* sp. The WGS approach further led to species identification of the six isolates conventionally identified as *Enterococcus* species, five as *Enterococcus faecalis*, and one as *Enterococcus faecium*. A single *Staphylococcus* isolate was further identified as *Staphylococcus lugdunensis*.

An MLST type was obtained for all eight *E. coli* isolates and all five *E. faecalis* isolates. Except for two *E. faecalis* isolates that both belonged to sequence type 40 (ST40), all isolates belonged to different sequence types (Table 1). The *E. faecium* isolate could not be assigned to a known MLST type. Antimicrobial resistance genes were observed in 11 of the 17 culture-positive samples, and the predicted susceptibility pattern was equal to that observed using phenotypic testing except for samples 21 and 28, which were phenotypically resistant to nalidixic acid and sulfonamides, respectively.

Sequencing directly on clinical samples. Sufficient amounts of DNA to perform WGS on the Ion Torrent PGM system were isolated from 23 of the 35 urine samples, including all 19 culture-positive samples. MG-RAST and the newly developed Chainmapper program gave almost the same results with regard to species identification, including percentage distributions (Table 1). In our hands, it took approximately 2 days to obtain a result using

TABLE 2 Phenotypically measured antimicrobial resistance and resistance genes present and predicted resistance among organisms from urinary samples based on conventional methods, whole-genome sequencing of isolated bacteria, and direct sequencing

| Sample no. | Resistance pattern[a] | | | | |
| | | WGS (single isolates) | | Direct sequencing | |
	Conventional	Resistance gene(s)	Predicted resistance	Resistance genes (no. of reads)	Predicted resistance
1				lsa(C) (11), tet(M) (3), tet(Q) (3), $bla_{CTX-M-101}$ (1), $catA1$ (1), $strB$ (1)	S
3	TET	tet(M)	TET	tet(M) (56), lsa(A) (21)	TET
4	S	None	S	lsa(C) (77), tet(M) (1)	S
6	AMP, STR, TET, SMX, TMP	$strA$, $strB$, bla_{TEM-1}, $sul1$, $sul2$, tet(A), $dfrA7$	AMP, STR, TET, SMX, TMP	bla_{TEM-1} (55), $dfrA7$ (43), $qacE$ (34), $strA$ (56), $strB$ (51), $sul1$ (29), $sul2$ (31), tet(A) (52), $aadA1$ (2)	AMP, STR, TET, SMX, TMP
7				tet(O) (48), erm(F) (2), erm(A) (1), $strB$ (1)	TET
8				lsa(C) (16), $dfrC$ (1)	S
10	S	None	S	bla_{CMY-41} (42), $qacE$ (31)	ESC
12	S	None	S	$qacE$ (25)	S
13				tet(Q) (43), $cfxA$ (15), tet(M) (2), erm(A) (5), $cfxA6$ (9), $cfxA5$ (2), $cfxA2$ (7)	AMP, TET
19	CST, TET	tet(J)	CST,[b] TET	tet(J) (5), aac(6′)-aph(2″) (2), msr(C) (1)	CST,[b] TET
20	S	None	S	$qacE$ (32), $sul2$ (1)	S
21	AMP, CIP, GEN, NAL	aac(3)-IId, bla_{TEM-1}, tet(A)	AMP, GEN, TET	aac(3)-IId (76), bla_{TEM-1} (34), tet(A) (62), $qacE$ (12), $atA1$ (1), $sul2$ (1)	AMP, GEN, TET
24	CST, CHL, TET	cat, tet(J)	CST,[b] CHL, TET	cat (10), $qacE$ (7), tet(J) (25), bla_{CEPA} (1), bla_{TEM-1} (1), $strB$ (1), tet(A) (1), tet(Q) (1)	CST,[b] CHL, TET
25	AMP, CHL, STR, TET, SMX, TMP	$strA$, $strB$, bla_{TEM-1}, $catA1$, $sul2$, tet(A), $dfrA14$	AMP, CHL, STR, TET, SMX, TMP	bla_{TEM-1} (123), $catA1$ (29), $dfrA14$ (98), $qacE$ (15), $strA$ (185), $strB$ (214), $sul2$ (126), $tetA$ (426), tet(O) (2)	AMP, CHL, STR, TET, SMX, TMP
26	S	lsa(A)	S	lsa(A) (32), bla_{TEM-15} (1), $sul2$ (1)	S
27	S	None	S	tet(A) (10), bla_{TEM-1} (1), $bla_{TEM-122}$ (1), bla_{TEM-15} (1), $bla_{TEM-171}$ (1), $catA1$ (3), $cfiA14$ (1), $cfxA3$ (1), $cfxA6$ (2), erm(B) (1), lsa(A) (3), $qacE$ (3) $strA$ (4), $strB$ (1), $sul2$ (1), tet(40) (1), tet(K) (3), tet(O) (8), tet(Q) (4), tet(W) (5)	TET
28	PEN, SMX	$blaZ$	PEN	$blaZ$ (138), $bla_{TEM-148}$ (1), $bla_{TEM-171}$ (1), $bla_{TEM-190}$ (1), $dfrA14$ (2), $strA$ (1), $sul2$ (1), tet(A) (3), tet(K) (1), tet(M) (1)	PEN
29				$qacE$ (11), aac(6′)-aph(2″) (1), $blaZ$ (1), lnu(B) (1)	S
31	S	tet(M)	S	lsa(A) (38), tet(M) (106), aac(6′)-aph(2″) (1), ant(6) (1)	S
32	ERY, GEN, TET	aac(6′)-aph(2″), ant(6), erm(B), lnu(B), msr(C), tet(M)	ERY, GEN, TET	aac(6′)-aph(2″) (16), ant(6) (19), erm(B) (29), lnu(B) (25), msr(C) (16), tet(M) (15), aac(6′)-Ii (6), aph(3′)-III (10), msr(D) (1), tet(K) (2), tet(Q) (2)	ERY, GEN, TET
33	TET	tet(M)	TET	lsa(A) (146), tet(M) (259), tet(K) (3)	TET
34	TET	tet(M)	TET	lsa(A) (36), $strA$ (18), $strB$ (10), tet(M) (73), aac(3)-IIa (3), aac(6′)-$Ib-cr$ (2), $aadA5$ (5), bla_{CEPA} (2), $bla_{CEPA-29}$ (1), $bla_{CEPA-44}$ (1), $bla_{CTX-M-101}$ (1), $bla_{CTX-M-108}$ (1), $bla_{CTX-M-80}$ (3), bla_{OXA-30} (3), bla_{OXA-31} (1), $catB3$ (1), $cfxA$ (1), $cfxA2$ (2), $cfxA6$ (2), $dfrA17$ (2), mph(A) (2), $qacE$ (2), $sul1$ (4), $sul2$ (8), tet(B) (9), tet(Q) (6)	TET, STR
35				lsa(A) (20), tet(M) (35), aac(3)-IIe (1), aac(6′)-Iz (5), aph(3′)-IIc (2), bla_{L1} (2), bla_{TEM-1} (8), $cfxA6$ (1), $dfrA14$ (1), $qacE$ (2), $qnr-S1$ (8), sph (3), $strB$ (4), $sul2$ (1), $sul3$ (1), tet(A) (8), tet(B) (1), tet(K) (1), tet(O) (2)	TET

[a] S, sensitive; PEN, penicillin; AMP, ampicillin; CIP, ciprofloxacin; CST, colistin; NAL, nalidixic acid; STR, streptomycin; SMX, sulfamethoxazole; TET, tetracycline; TMP, trimethoprim; CHL, chloramphenicol; GEN, gentamicin; ESC, extended-spectrum cephalosporinase; ERY, erythromycin.
[b] Based on species identification.

MG-RAST, whereas Chainmapper gave a species identification, a microbial community profile, and an abundance estimation in approximately 40 min, as well as indicating the presence of all resistance genes within 3 min.

In all 17 cases in which it was possible to isolate a pure culture isolate, the use of WGS directly on the samples yielded the same species identification and MLST type as WGS performed on pure isolates. In addition, the direct sequencing approach enabled identification of *E. coli* and a mixture of *E. coli* and *E. faecalis* in the

two samples that were contaminated using the culturing approach. The remaining four samples all contained *Lactobacillus*, *Prevotella*, *Gardnerella*, or *Bifidobacterium* (Table 1). Direct sequencing identified *Aerococcus urinae* in sample 24, in addition to the mixture of *Proteus* and *E. coli* which was observed using culturing.

Direct sequencing performed on the urine pellet resulted in some cases in the detection of an increased number of resistance genes, compared to those observed in the cultured iso-

lates (Table 2). When only the abundant resistance genes were included, however, in almost all cases the same resistance genes, with the same predicted susceptibility profiles, were obtained using direct sequencing and sequencing of single isolates (Table 2). Additional genes that were not observed in the cultured isolates were detected in samples 10 and 27. Furthermore, resistance genes were detected in one of the samples with a mixed culture and in two of the four culture-negative samples. Compared with sequencing of culture isolates, no resistance genes were missed by direct sequencing of the samples.

Comparative phylogenetic analysis and SNP trees. Single-nucleotide polymorphism (SNP)-based phylogenetic trees were generated for all *E. coli* and *E. faecalis* data obtained using direct sequencing or single-isolate sequencing (Fig. 1). Even though some SNP differences were observed, almost perfect phylogenetic matches between WGS data obtained from the pure isolates and directly from the samples were observed. In addition, it was possible to include data from the samples in the phylogenetic tree when cultures were contaminated.

DISCUSSION

Rapid diagnostic testing is important to detect and to control outbreaks, to initiate the correct treatment, and to determine the progress of infections. UTIs are one of the most common causes of infections in humans and account for more than one-half of all microbiological examinations at hospitals in Denmark.

Using whole-genome sequencing on cultured isolates from the 17 samples with pure cultures, we were able to obtain rapidly precise species and clonal information, as well as predicted antimicrobial susceptibility profiles equal to those obtained by phenotypic methods. Direct sequencing of the urine samples yielded the same bacterial species identification, clonal identification, and identification of resistance genes as observed for the cultured isolates.

Importantly, direct sequencing on the clinical samples also yielded information on the presence of bacteria that were not detected using conventional (aerobic) culturing. Thus, *Lactobacillus iners*, *Gardnerella vaginalis*, *Prevotella*, and *A. urinae* have all been implicated in UTIs (22–24), even though their precise roles as pathogens and normal colonizers of the genital tract have not been firmly established. It is noteworthy that *A. urinae* is a rarely reported pathogen that is usually misclassified as *Streptococcus*, *Enterococcus*, or *Staphylococcus* (25). Previous studies using 16S rRNA gene-based classification of urinary samples have identified a large number of different fastidious bacterial species in culture-negative samples (26, 27). In the future, more-widespread use of whole-genome sequencing could potentially lead to increased detection of fastidious urinary tract pathogens and polymicrobial infections. This could lead to improved understanding of infectious diseases and novel ways of defining pathogens.

We found a larger number of resistance genes in the data obtained directly from the urine samples than in the data obtained from pure cultured isolates. This is not surprising, since urine most likely contains small numbers of other bacterial species originating from the natural flora present in the urethra. This could potentially lead to overestimation of the occurrence of resistance in a patient sample and perhaps even treatment with broader-spectrum antibiotics than necessary. However, filtering genes with low coverage removed almost all of the resistance genes not observed in the cultured isolates and, even though direct sequencing might give a slight overestimate of the resistance, it is noteworthy that this procedure did not miss any genes, compared to sequencing of the purified isolates.

The current conventional procedures for clinical diagnostic testing often include the use of multiple cultivation and incubation steps followed by species-specific identification, susceptibility testing, and typing (Fig. 2). As suggested in a recent review by Didelot et al. (2), the recent availability of new benchtop sequencing systems constitutes an important step toward both simplifying and improving clinical diagnostic testing. In this study, starting from either a urine sample or an overnight culture of pure isolates, it took us approximately 18 h to purify DNA, prepare the DNA libraries, and sequence the samples or isolates. After establishment of the bioinformatic pipeline, the analysis could be performed in less than 6 h. Thus, compared to conventional bacteriology, where the time needed for identification and susceptibility testing would be 48 to 72 h, sequencing of pure isolates would give results within 48 h and sequencing directly on the clinical samples could yield results in less than 24 h. In addition, the genomic approach would give complete strain information, allowing immediate identification of transmission or recurrent infection. A comparison of the approaches is depicted in Fig. 2.

A number of other technologies are also available for direct detection of pathogens in clinical samples. These include PCR-based methods and matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) (14, 15). Both methods are cheap and rapid and can yield reliable species identification, as well as detection of specific resistance genes for the PCR-based methods. The methods are limited, however, in the sense that they do not yield clonal information, only yield information regarding a limited number of species and genes, and cannot be easily compared between laboratories.

The findings of our study indicate that there could be major value in performing whole-genome sequencing in real time directly on clinical samples as an integral part of routine diagnostic testing and surveillance in the hospital setting. The features of this technology include rapid turnaround, affordability, and the provision of clinically relevant information to health care personnel that can be interpreted without specialist knowledge of whole-genome sequencing. To facilitate widespread clinical utilization, for this study we developed a rapid method for analyzing whole-community sequence data produced from clinical samples. This method can identify both species and resistance genes in a sample and additionally give information on the presence of other DNA, including human and fungal DNA, all within a clinically relevant time frame. Chainmapper can already be used to obtain most clinically relevant information and, in combination with tools for clonal analysis, will also give epidemiologically important information.

Whole-genome sequencing may still be too expensive for routine use in most clinical microbial laboratories. However, given the competition between current and emerging sequencing platforms, the price and turnaround time will most likely fall. Once data interpretation is fully automated, we predict that whole-genome sequencing will become a standard tool for infection detection and control and will provide the ability to monitor the spread and evolution of major pathogens in real time, both within and outside hospitals.
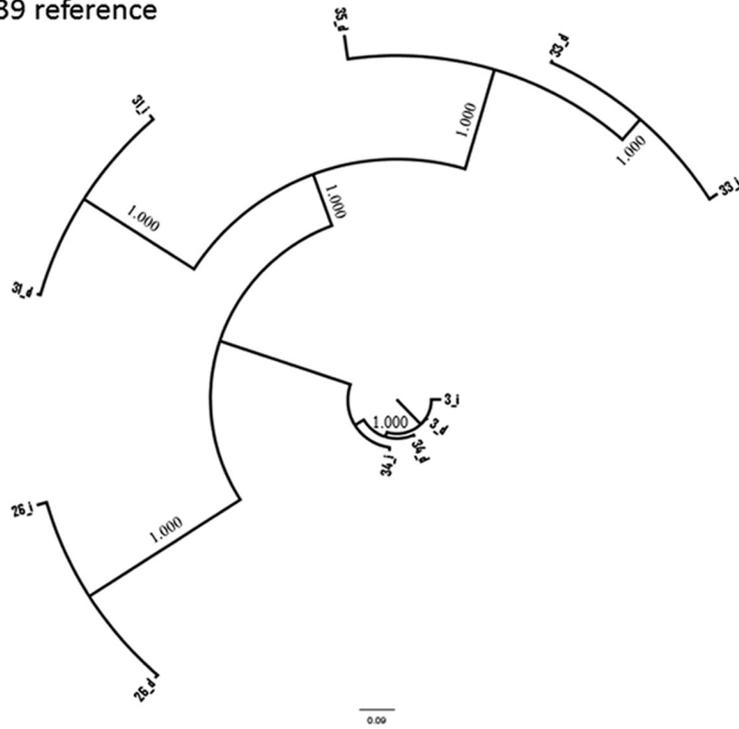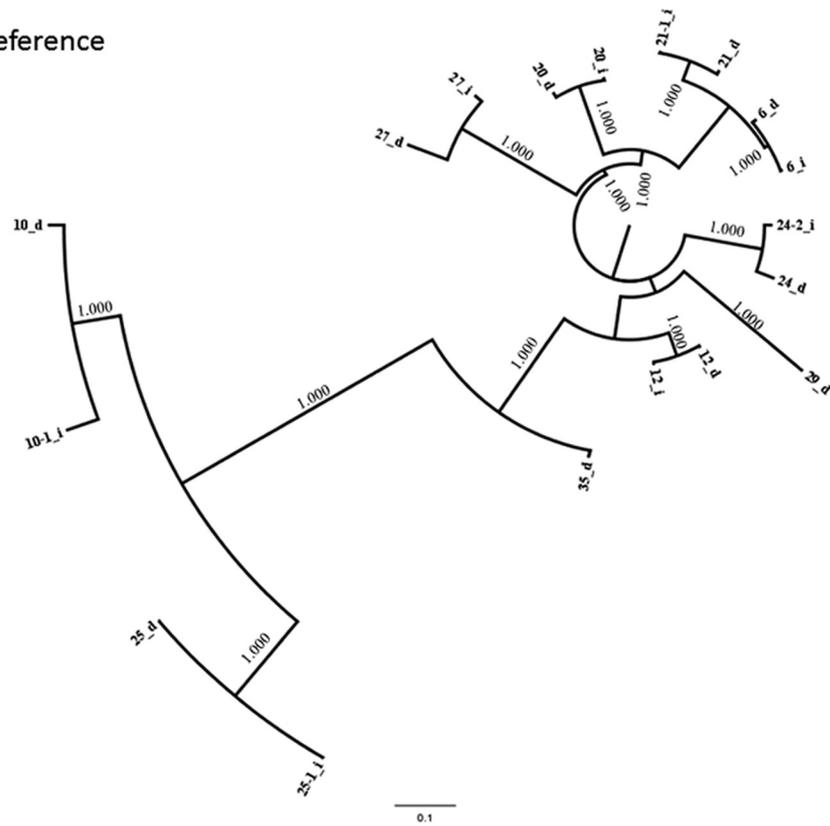
FIG 1 Phylogenetic relationships among *Escherichia coli* (A) and *Enterococcus faecalis* (B) strains identified using whole-genome sequencing of purified single isolates (labeled with sample numbers followed by "_i") and direct sequencing of urine samples (labeled with sample numbers followed by "_d"). SNP trees were constructed using an online application (21). Data were obtained from single isolates and directly from samples clustered together, and it was also possible to place data in a phylogenetic context when it was not possible to culture single isolates. Branch length in the snpTree output indicates number of substitutions per site.
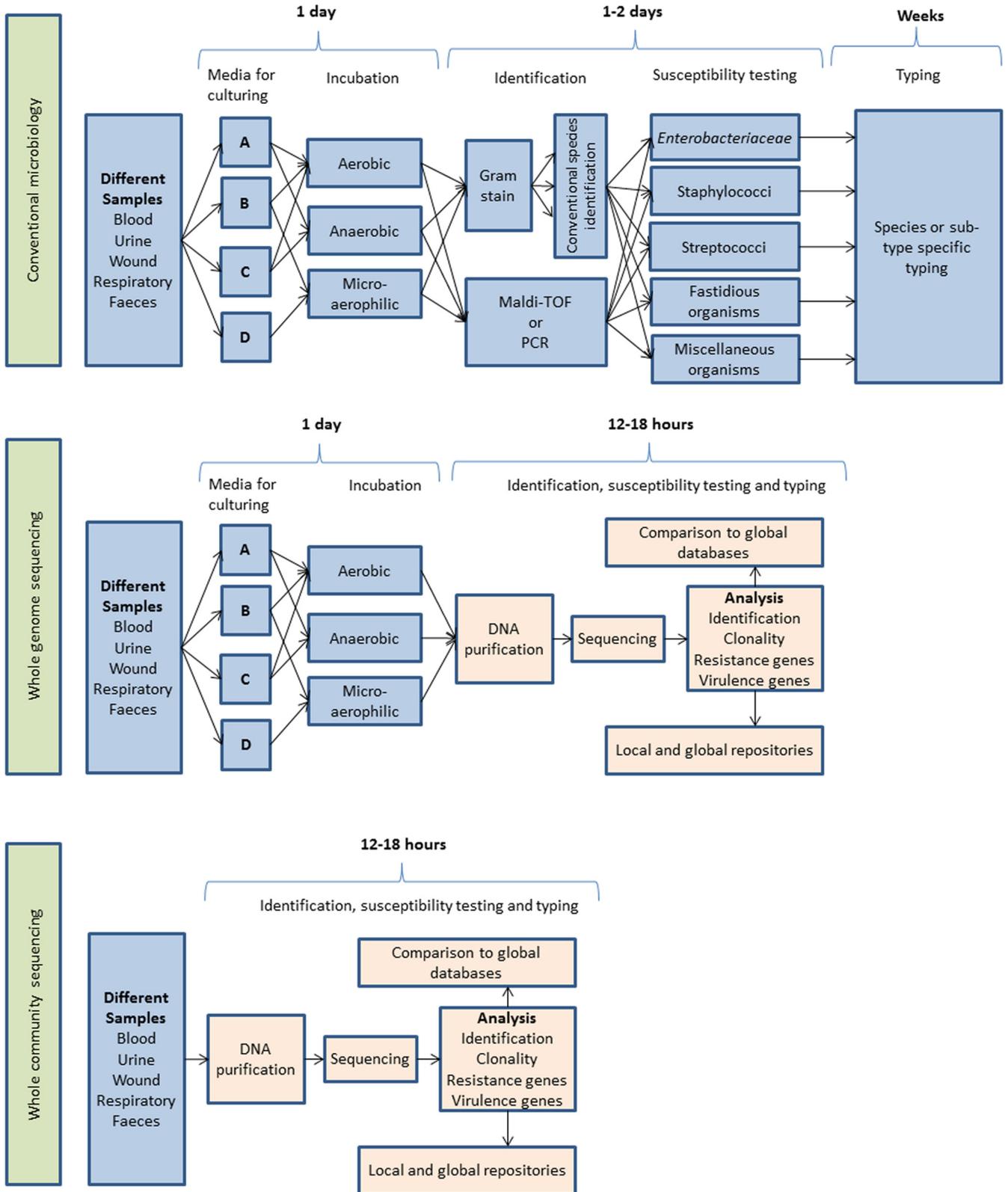
FIG 2 Schematic representation of the workflow anticipated after adoption of whole-genome sequencing used either on cultured isolates or directly on the clinical samples, with an expected time scale.

ACKNOWLEDGMENTS

REFERENCES

1. **Aarestrup FM, Brown EW, Detter C, Gerner-Smidt P, Gilmour MW, Harmsen D, Hendriksen RS, Hewson R, Heymann DL, Johansson K, Ijaz K, Keim PS, Koopmans M, Kroneman A, Lo Fo Wong D, Lund O, Palm D, Sawanpanyalert P, Sobel J, Schlundt J.** 2012. Integrating genome-based informatics to modernize global disease monitoring, information sharing, and response. Emerg. Infect. Dis. **18:**e1. http://dx.doi.org /10.3201/eid.1811.120453.

2. **Didelot X, Bowden R, Wilson DJ, Peto TE, Crook DW.** 2012. Transforming clinical microbiology with bacterial genome sequencing. Nat. Rev. Genet. **13:**601–612. http://dx.doi.org/10.1038/nrg3226.

3. **Köser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, Holden MT, Dougan G, Bentley SD, Parkhill J, Peacock SJ.** 2012. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS Pathog. **8:**e1002824. http://dx.doi .org/10.1371/journal.ppat.1002824.

4. **Harris SR, Cartwright EJ, Török ME, Holden MT, Brown NM, Ogilvy-Stuart AL, Ellington MJ, Quail MA, Bentley SD, Parkhill J, Peacock SJ.** 2013. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. Lancet Infect. Dis. **13:**130–136. http://dx.doi.org/10.1016/S1473-3099(12)70268-2.

5. **Price LB, Stegger M, Hasman H, Aziz M, Larsen J, Andersen PS, Pearson T, Waters AE, Foster JT, Schupp J, Gillece J, Driebe E, Liu CM, Springer B, Zdovc I, Battisti A, Franco A, Zmudzki J, Schwarz S, Butaye P, Jouy E, Pomba C, Porrero MC, Ruimy R, Smith TC, Robinson DA, Weese JS, Arriola CS, Yu F, Laurent F, Keim P, Skov R, Aarestrup FM.** 2013. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. mBio **4:**e00520–12. http://dx.doi.org /10.1128/mBio.00520-12.

6. **Young BC, Golubchik T, Batty EM, Fung R, Larner-Svensson H, Votintseva AA, Miller RR, Godwin H, Knox K, Everitt RG, Iqbal Z, Rimmer AJ, Cule M, Ip CL, Didelot X, Harding RM, Donnelly P, Peto TE, Crook DW, Bowden R, Wilson DJ.** 2012. Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. Proc. Natl. Acad. Sci. U. S. A. **109:**4550–4555. http://dx.doi.org/10.1073/pnas .1113219109.

7. **Hendriksen RS, Price LB, Schupp JM, Gillece JD, Kaas RS, Engelthaler DM, Bortolaia V, Pearson T, Waters AE, Upadhyay BP, Shrestha SD, Adhikari S, Shakya G, Keim PS, Aarestrup FM.** 2011. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. mBio **2:**e00157–11. http://dx.doi.org/10.1128/mBio .00157-11.

8. **Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther S, Rothberg JM, Karch H.** 2011. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One **6:**e22751. http://dx.doi.org/10.1371/journal.pone.0022751.

9. **Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE.** 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect. Dis. **13:**137–146. http://dx.doi .org/10.1016/S1473-3099(12)70277-3.

10. **Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agersø Y, Lund O, Larsen MV, Aarestrup FM.** 2013. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. J. Antimicrob. Chemother. **68:**771–777. http://dx .doi.org/10.1093/jac/dks496.

11. **Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, Ip CL, Wilson DJ, Didelot X, O'Connor L, Lay R, Buck D, Kearns AM, Shaw A, Paul J, Wilcox MH, Donnelly PJ, Peto TE, Walker AS, Crook DW.** 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus*

12. **Köser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ.** 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N. Engl. J. Med. **366:**2267–2275. http://dx.doi.org/10.1056 /NEJMoa1109910.

13. **Török ME, Reuter S, Bryant J, Köser CU, Stinchcombe SV, Nazareth B, Ellington MJ, Bentley SD, Smith GP, Parkhill J, Peacock SJ.** 2013. Rapid whole-genome sequencing for the investigation of a suspected tuberculosis outbreak. J. Clin. Microbiol. **51:**611–614. http://dx.doi.org/10.1128 /JCM.02279-12.

14. **Cunningham SA, Sloan LM, Nyre LM, Vetter EA, Mandrekar J, Patel R.** 2010. Three-hour molecular detection of *Campylobacter*, *Salmonella*, *Yersinia*, and *Shigella* species in feces with accuracy as high as that of culture. J. Clin. Microbiol. **48:**2929–2933. http://dx.doi.org/10.1128/JCM.00339-10.

15. **Croxatto A, Prod'hom G, Greub G.** 2012. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. FEMS Microbiol. Rev. **36:**380–407. http://dx.doi.org/10.1128/JCM.00339-10.

16. **Seth-Smith HM, Harris SR, Skilton RJ, Radebe FM, Golparian D, Shipitsyna E, Duy PT, Scott P, Cutcliffe LT, O'Neill C, Parmar S, Pitt R, Baker S, Ison CA, Marsh P, Jalal H, Lewis DA, Unemo M, Clarke IN, Parkhill J, Thomson NR.** 2013. Whole-genome sequences of *Chlamydia trachomatis* directly from clinical samples without culture. Genome Res. **23:**855–866. http://dx.doi.org/10.1101/gr.150037.112.

17. **Andersson P, Klein M, Lilliebridge RA, Giffard PM.** 2013. Sequences of multiple bacterial genomes and a *Chlamydia trachomatis* genotype from direct sequencing of DNA derived from a vaginal swab diagnostic specimen. Clin. Microbiol. Infect. **19:**e405–e408. http://dx.doi.org/10.1111 /1469-0691.12237.

18. **Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ, Quick J, Weir JC, Quince C, Smith GP, Betley JR, Aepfelbacher M, Pallen MJ.** 2013. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104: H4. JAMA **309:**1502–1510. http://dx.doi.org/10.1001/jama.2013.3231.

19. **Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O.** 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J. Clin. Microbiol. **50:**1355–1361. http://dx.doi.org/10.1128/JCM.06094-11.

20. **Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV.** 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. **67:**2640–2644. http: //dx.doi.org/10.1093/jac/dks261.

21. **Leekitcharoenphon P, Kaas RS, Thomsen MC, Friis C, Rasmussen S, Aarestrup FM.** 2012. snpTree: a web-server to identify and construct SNP trees from whole genome sequence data. BMC Genomics **13**(Suppl 7):S6. http://dx.doi.org/10.1186/1471-2164-13-S7-S6.

22. **Lam MH, Birch DF, Fairley KF.** 1988. Prevalence of *Gardnerella vaginalis* in the urinary tract. J. Clin. Microbiol. **26:**1130–1133.

23. **Domann E, Hong G, Imirzalioglu C, Turschner S, Kühle J, Watzel C, Hain T, Hossain H, Chakraborty T.** 2003. Culture-independent identification of pathogenic bacteria and polymicrobial infections in the genitourinary tract of renal transplant recipients. J. Clin. Microbiol. **41:**5500–5510. http://dx.doi.org/10.1128/JCM.41.12.5500-5510.2003.

24. **Sierra-Hoffman M, Watkins K, Jinadatha C, Fader R, Carpenter JL.** 2005. Clinical significance of *Aerococcus urinae*: a retrospective review. Diagn. Microbiol. Infect. Dis. **53:**289–292. http://dx.doi.org/10.1016/j .diagmicrobio.2005.06.021.

25. **Zhang Q, Kwoh C, Attorri S, Clarridge JE, III.** 2000. *Aerococcus urinae* in urinary tract infections. J. Clin. Microbiol. **38:**1703–1705.

26. **Imirzalioglu C, Hain T, Chakraborty T, Domann E.** 2008. Hidden pathogens uncovered: metagenomic analysis of urinary tract infections. Andrologia **40:**66–71. http://dx.doi.org/10.1111/j.1439-0272.2007.00830.x.

27. **Siddiqui H, Nederbragt AJ, Lagesen K, Jeansson SL, Jakobsen KS.** 2011. Assessing diversity of the female urine microbiota by high throughput sequencing of 16S rDNA amplicons. BMC Microbiol. **11:**244. http://dx .doi.org/10.1186/1471-2180-11-244.