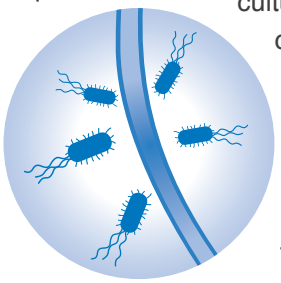
Next-Generation Sequencing (NGS) for Culture-Independent, Metagenomics-Based Detection of Uropathogens

INTRODUCTION

Urinary tract infections (UTIs)

UTIs are the most common outpatient infections in the US and one of the most common healthcare-acquired infections.[1] UTIs affect 150 million people worldwide and in the US lead to 10 million office visits and 1 million hospitalizations.[2] These hospitalizations resulted in charges of $9.7 billion and a real total cost of $2.8 billion. The mean real cost per case has increased by 90.8% between 2001 and 2011.[3] More than half of all women (50-60%) will have at least one UTI in their life; men are affected at lower frequency.[1]



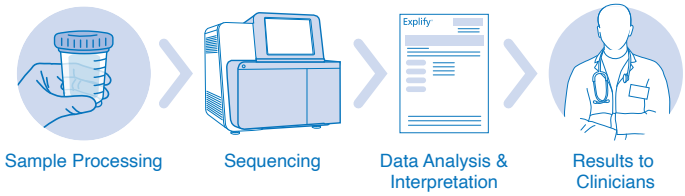
The majority of UTIs (65-75%) are caused by *uropathogenic Escherichia coli* (UPEC). Other causative agents include Gram-positive and Gram-negative bacteria such as *Klebsiella pneumoniae*, *Proteus mirabilis*, *Pseudomonas aeruginosa*, *Enterococcus faecalis, Enterococcus faecium*, *Staphylococcus aureus*, *Staphylococcus saprophyticus*, *Streptococcus agalactiae*, and some fungi, including Candida spp.[2]

Accurate identification of uropathogens is critical to manage UTIs and to limit the development and spread of antimicrobial resistance.[4] Routine testing for UTIs includes urinalysis, which provides indirect evidence of infection by measuring the urine's appearance and composition (e.g. nitrites, leukocyte esterase produced by leukocytes) and urine culture to identify viable bacteria. As urine may get contaminated with flora from the urethra, the ideal urine specimen is obtained through suprapubic aspiration. However, this invasive procedure is reserved for unusual cases. Instead, clean-voided midstream urine ('clean-catch') is collected to minimize contamination and only bacterial concentrations of $\geq 10^5$ colony forming units per milliliter of urine (CFU/mL) are considered significant in most symptomatic patients.[5] Antimicrobial susceptibility patterns of isolated bacteria can be assessed by culturing bacteria in the presence of varying concentrations of antibiotics (phenotypic susceptibility testing). Cutoffs are applied to the lowest concentration of an antibiotic that inhibits bacterial growth *in vitro* to predict effectiveness *in vivo*.

Timely and accurate identification of uropathogens and assessment of antimicrobial susceptibility is central to the management of UTIs.[4] However, approximately 20% of women with symptoms of UTI have a negative urine culture.[6]

Clinical Metagenomics

Diagnostic testing using metagenomics (clinical metagenomics) is a powerful method for unbiased, culture-independent detection, quantification, and genetic profiling of pathogen using next-generation sequencing. DNA extracted from urine is sequenced in an agnostic ('shotgun sequencing') approach and a large number of common and rare bacterial and fungal



pathogens can be identified and profiled via their DNA sequences. Clinical metagenomics using shotgun sequencing can provide complete bacterial genome information, which allows for detection of genetic markers for antimicrobial resistance (AMR). Antibiotic resistance can be predicted based on the detected AMR markers.

Clinical metagenomics can increase the diagnostic yield over culture-based pathogen detection for a number of infectious diseases, including UTI.[7] Here, we provide a streamlined specimen-to-report workflow for detection, identification, and quantification of 97 bacterial and fungal uropathogens directly from urine. This workflow also tests for 371 AMR markers that may confer resistance to 12 antibiotics widely used for treatment of the 8 most common uropathogens. Testing is based on an optimized laboratory protocol for DNA extraction, the Nextera DNA Flex Library Prep Kit, Illumina sequencing and IDbyDNA's Explify Platform for sequencing data analysis and result reporting.

METHODS

Urine specimens and reference method

Residual urine specimens (Table 1) were collected after routine urine bacterial culture was completed by automated methods on a WASPLab system (Copan DIAGNOSTICS Inc.). Residual specimens were stored at 4°C for 8-11 days prior to extraction of DNA for clinical metagenomics.

Table 1. Urine culture results for the specimens used in this study.

Cultured Bacteria	$<10^5$ CFU/mL[a]	$\geq 10^5$ CFU/mL	Total
Enterococcus faecalis		7	7
Enterococcus faecium		1	1
Escherichia coli	29	34	63
Klebsiella pneumoniae		10	10
Proteus mirabilis		7	7
Pseudomonas aeruginosa		3	3
Staphylococcus aureus	1	2	3
Staphylococcus saprophyticus		1	1
Total	**30**	**65**	**115[b]**

[a] CFU/mL, colony forming units per mL estimates calculated by manual plate counting
[b] including 20 specimens with no pathogen detected

Quality control

For each batch of specimens, one positive and one negative external control was used. The ZymoBIOMICS Microbial Community Standard (MCS, Zymo Research, Catalog number D6300) was used as a positive control. This positive control contains a total of ~1.4 x 10^{10} cells/mL in stabilizing buffer (DNA/RNA Shield), composed of three Gram-negative bacteria (*Escherichia coli* (8.5%), *Pseudomonas aeruginosa* (6.1%) and *Salmonella enterica* (8.8%)), five Gram-positive bacteria (*Bacillus subtilis* (10.3%), *Enterococcus faecalis* (14.6%), *Lactobacillus fermentum* (21.9%), *Listeria monocytogenes* (13.9%) and *Staphylococcus aureus* (15.3%)) and two yeasts (*Cryptococcus neoformans* (0.18%) and *Saccharomyces cerevisiae* (0.29%)). Synthetic urine (Sigmatrix Urine Diluent, Sigma Aldrich, catalog number SAE0074-1L) was used as a negative control. Ultra-pure bacteriophage was spiked as an internal control (process control, source: e.g. ready-to-use ultra-pure T7 or T4 DNA bacteriophage from Microbiologics, stock concentration of 10^{11} plaque-forming units per milliliter, PFU/mL).

DNA extraction from urine specimens and removal of human DNA (de-hosting)

For DNA extraction, 1.8 ml urine was centrifuged at 8000xg for 10 minutes. The supernatant was discarded, and the pellet was resuspended in 320 µl 1X DNA/RNA shield (Zymo Research, catalog number R1100-250). The resuspended pellet was spiked with $2x10^7$ PFU ultra-pure Bacteriophage T7 (0.2 µl of the stock per specimen), mixed with 10 µl of 1% saponin (Alfa Aesar, catalog number AAA18820-14) and incubated at room temperature for 5 minutes. To deplete human DNA, 81 µl of DNase I mix (40 µl DNase I + 41 µl 10X buffer with Mg^{2+}, ThermoFisher, catalog number EN0525) was added to each specimen and incubated at 37°C for 30 minutes followed by 95°C for 5 minutes to inactivate DNase I. Subsequently, specimens were mixed with 400 µl bashing bead buffer (Zymo Research, catalog number D6001-3-150), transferred to ZR BashingBead Lysis Tubes (0.1 & 0.5 mm, Zymo Research, catalog number S6012-50) and homogenized on a Fastprep 24 (MP Biomedicals; 6 m/s for 60 seconds per round, 4 rounds with 5 minute pauses between rounds) as part of the protocol for Tough-to-Lyse Specimens of the Quick-DNA Urine Kit (Zymo Research, catalog number D3061). DNA was eluted in 70 µl of molecular biological grade water (VWR, catalog number 46-000-CI/CM) and stored frozen until library preparation.

Library preparation and next-generation sequencing[8]

DNA libraries were prepared with the Nextera DNA Flex kit (Illumina, catalog number 20018705) and IDT for Illumina Nextera DNA UD Indexes Set A (Illumina, catalog number 20027213) per package insert. Sequencing libraries were quantified, pooled, and sequenced to a median depth of 8.5 x 10^6 reads/specimen on a NextSeq 550 instrument using either a high-output (Illumina, catalog number 20024907) or mid-output v2.5 (catalog number 20024904) sequencing kit with 150 cycles.

Sequencing data analysis

Sequencing data was demultiplexed, quality filtered, and analyzed with the Explify UTI Data Analysis Solution (IDbyDNA) that uses an interactive Result Review Portal (Figure 1) for automated interpretation of QC together with validated detection thresholds to detect 97 of the most relevant uropathogens (Table 2). Sequencing data were analyzed in 6 formats for each specimen: 150 bp reads as generated (median of ~8.5 million per specimen), down-sampled to 5 million and 2 million reads per specimen; reads were also trimmed to 100 bp length before analyzing data at a median of ~8.5 million, 5 million, and 2 million reads per specimen.

Figure 1. Screen shot of the Explify UTI Data Analysis Solution showing an example for detection of *Escherichia coli* from urine by clinical metagenomics.

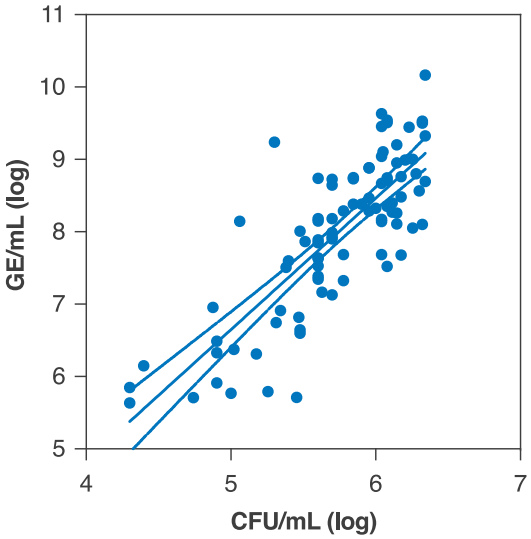
Bacteria — 1 organisms (Filter applied: 1) — Show All — Hide

MD	Organism Name	Evidence	Type	% Coverage	ANI	Median Depth	Reads	Quantity	Reference Length	Details
○	*Escherichia coli*	9,883	RNA	100.0%	100.0%	512	4,386	> 10^8 GE/mL	1,251	Show
			DNA	93.0%	98.7%	46	94,644		272,681	

Table 2. Most relevant uropathogens detected by the Explify UTI Data Analysis Solution test.

BACTERIA			FUNGI
Acinetobacter baumannii	*Haemophilus ducreyi*	*Pseudomonas aeruginosa*	*Aspergillus flavus*
Acinetobacter calcoaceticus	*Haemophilus influenzae*	*Pseudomonas putida*	*Candida albicans*
Actinotignum schaalii (Actinobaculum schaalii)	*Haemophilus parainfluenzae*	*Salmonella enterica*	*Candida dubliniensis*
Aerococcus sanguinicola	*Hafnia alvei*	*Serratia marcescens*	*Candida fabianii (Cyberlindnera fabianii)*
Aerococcus urinae	*Klebsiella aerogenes (Enterobacter aerogenes)*	*Shigella flexneri*	*Candida glabrata (Nakaseomyces glabrata)*
Aerococcus viridans	*Klebsiella oxytoca*	*Shigella sonnei*	*Candida krusei (Pichia kudriavzevii)*
Aeromonas hydrophila	*Klebsiella pneumoniae*	*Staphylococcus aureus*	*Candida lusitaniae (Clavispora lusitaniae)*
Bacteroides fragilis	*Kluyvera ascorbata*	*Staphylococcus epidermidis*	*Candida parapsilosis*
Burkholderia cepacia	*Listeria monocytogenes*	*Staphylococcus haemolyticus*	*Candida tropicalis*
Campylobacter jejuni	*Moraxella catarrhalis*	*Staphylococcus hominis*	*Cryptococcus neoformans*
Chlamydia trachomatis	*Morganella morganii*	*Staphylococcus intermedius*	*Trichosporon asahii*
Chromobacterium violaceum	*Mycobacterium avium complex*	*Staphylococcus lugdunensis*	*Trichosporon beigelii*
Chryseobacterium indologenes	*Mycobacterium gordonae*	*Staphylococcus saprophyticus*	
Citrobacter freundii	*Mycobacterium kansasii*	*Staphylococcus simulans*	
Citrobacter koseri	*Mycobacterium marinum*	*Staphylococcus warneri*	
Corynebacterium genitalium	*Mycobacterium scrofulaceum*	*Stenotrophomonas maltophilia*	
Corynebacterium glucuronolyticum	*Mycobacterium simiae*	*Streptococcus agalactiae*	
Corynebacterium jeikeium	*Mycobacterium tuberculosis*	*Streptococcus anginosus*	
Corynebacterium renale	*Mycobacterium xenopi*	*Streptococcus constellatus*	
Corynebacterium riegelii	*Mycobacteroides chelonae (Mycobacterium chelonae)*	*Streptococcus intermedius*	
Corynebacterium urealyticum	*Mycoplasma genitalium*	*Streptococcus pneumoniae*	
Enterobacter cloacae complex	*Mycoplasma hominis*	*Treponema pallidum*	
Enterococcus faecalis	*Neisseria gonorrhoeae*	*Ureaplasma parvum*	
Enterococcus faecium	*Oligella urethralis*	*Ureaplasma urealyticum*	
Enterococcus gallinarum	*Pasteurella multocida*	*Vibrio cholerae*	
Escherichia coli	*Proteus mirabilis*	*Yersinia enterocolitica*	
Fusobacterium necrophorum	*Proteus vulgaris*	*Yersinia pseudotuberculosis*	
Fusobacterium nucleatum	*Providencia rettgeri*		
Gardnerella vaginalis	*Providencia stuartii*		

Figure 2. Correlation between culture-based (CFU/mL) and metagenomics-based (GE/mL) quantification.



The Explify UTI Data Analysis Solution is able to quantify the number of bacterial and fungal genome equivalents (GE) or genomic copies based on a standard curve generated with bacterial and fungal suspensions of known concentration. Quantification results are reported in GE/mL. Comparison of GE/mL with CFU/mL estimates calculated by manual plate counting of paired specimen cultures tested by quantitative urine culture has shown a high correlation (correlation coefficient of 0.81, *n*=89, Figure 2). Based on these results, GE/mL ranges as shown in Table 3 can be used to deduce the bacterial loads CFU/mL loads in urine specimens.

Table 3. Bacterial quantification using culture and clinical metagenomics.

Range	Approximate Quantities Observed by Urine Culture	General Clinical Interpretation for Culture
$<10^7$ GE/mL[a]	$<10^5$ CFU/mL[b]	Insignificant
10^7 - 10^8 GE/mL	≥10^5 CFU/mL	Significant
>10^8 GE/mL		

[a] GE, genome equivalents
[b] CFU, colony forming units

Results

Table 4. Results for the ZymoBIOMICS Microbial Community Standard (MCS).

Microorganisms[a] in MCS		GE/mL[b]	Pos		Pos, 5M		Pos, 2M	
			150 bp	100 bp	150 bp	100 bp	150 bp	100 bp
Escherichia coli	GNR[c]	1.2x10^9	3/3	3/3	3/3	3/3	3/3	0/3
Pseudomonas aeruginosa	GNR	8.5x10^8	3/3	3/3	3/3	3/3	3/3	0/3
Salmonella enterica	GNR	1.2x10^9	3/3	3/3	3/3	3/3	3/3	3/3
Enterococcus faecalis	GPC[d]	2.0x10^9	3/3	3/3	3/3	3/3	3/3	3/3
Listeria monocytogenes	GPR[e]	1.9x10^9	3/3	3/3	3/3	3/3	3/3	3/3
Staphylococcus aureus	GPC	2.1x10^9	3/3	3/3	3/3	3/3	3/3	3/3
Cryptococcus neoformans	Yeast	5.2x10^7	3/3	3/3	3/3	3/3	3/3	0/3

[a] The MCS also contains *Bacillus subtilis*, *Lactobacillus fermentum* and *Saccharomyces cerevisiae*, which are not uropathogens and thus not reported by the Explify UTI Data Analysis Solution
[b] Concentrations per product information for the MCS
[c] Gram-negative rods
[d] Gram-positive cocci
[e] Gram-positive rods

Detection of uropathogens by clinical metagenomics compared to culture

Sequencing data was analyzed with the Explify UTI Data Analysis Solution, which starts with the raw sequencing run directory and includes specimen and data quality control (QC). Data QC for the sequencing run (run-level), positive and negative controls (batch-level), and each specimen (specimen-level) were interpreted based on pre-defined criteria. Identification of uropathogens is performed per pre-defined and validated detection thresholds and results are visualized in the Explify Result Review Portal (Figure 1). Table 5 shows the overall results for all specimens. Tables 6 and 7 show the clinical metagenomics results for urine specimens with bacterial loads of ≥1x10^5 or <1x10^5 CFU/mL, respectively. There was a high correlation between the culture and clinical metagenomics results for specimens. Table 8 shows that clinical metagenomics detected established uropathogens that failed to grow or were not identified by culture (e.g. *Enterococcus faecalis* in 5 specimens). In addition, clinical metagenomics also detected additional pathogens that are not routinely reported by urine culture. One additional potential uropathogen was detected in 16 specimens, 2 additional potential uropathogens in 10 specimens, 3 additional potential uropathogens in 3 specimens and 4 additional potential uropathogens in one specimen.

Table 5. Detection of uropathogens in culture-negative and culture-positive urine.

Metagenomics Results	Culture Result		
	Negative (*n*=20)	<10^5 CFU/mL (*n*=30)	≥10^5 CFU/mL (*n*=65)
Positive, count	1[a]	14	62
Positive, percent	5%	47%	95%

[a] Enterococcus faecalis was detected at >10^8 GE/mL in one culture-negative specimen

Table 6. Detection of uropathogens in culture-positive urine specimens (≥1x10^5 CFU/mL) by cultured bacterium.

Cultured Uropathogen	Pos (%)[a]		Pos, 5M (%)[b]		Pos, 2M (%)[c]	
	150 bp	100 bp	150 bp	100 bp	150 bp	100 bp
Enterococcus faecalis	7/7 (100%)	7/7 (100%)	7/7 (100%)	7/7 (100%)	7/7 (100%)	7/7 (100%)
Enterococcus faecium	1/1 (100%)	1/1 (100%)	1/1 (100%)	1/1 (100%)	1/1 (100%)	1/1 (100%)
Escherichia coli	32/34 (94%)	32/34 (94%)	30/34 (88%)	30/34 (88%)	29/34 (85%)	29/34 (85%)
Klebsiella pneumoniae	9/10 (90%)	9/10 (90%)	9/10 (90%)	7/10 (70%)	6/10 (60%)	6/10 (60%)
Proteus mirabilis	7/7 (100%)	7/7 (100%)	7/7 (100%)	7/7 (100%)	7/7 (100%)	7/7 (100%)
Pseudomonas aeruginosa	3/3 (100%)	3/3 (100%)	3/3 (100%)	3/3 (100%)	3/3 (100%)	2/3 (67%)
Staphylococcus aureus	2/2 (100%)	2/2 (100%)	2/2 (100%)	2/2 (100%)	2/2 (100%)	2/2 (100%)
Staphylococcus saprophyticus	1/1 (100%)	1/1 (100%)	1/1 (100%)	1/1 (100%)	1/1 (100%)	1/1 (100%)
Total	**62/65 (95%)**	**62/65 (95%)**	**60/65 (92%)**	**58/65 (89%)**	**56/65 (86%)**	**55/65 (85%)**

[a] Data was analyzed as generated (median of 8.5 million reads/specimen)
[b] Data was down-sampled to 5 million reads per specimen
[c] Data was down-sampled to 2 million reads per specimen

Table 7. Detection of uropathogens in culture-positive urine specimens (<1x10^5 CFU/mL) by cultured bacterium.

Cultured Uropathogen	Pos (%)[a]		Pos, 5M (%)[b]		Pos, 2M (%)[c]	
	150 bp	100 bp	150 bp	100 bp	150 bp	100 bp
Escherichia coli	14/29 (48%)	14/29 (48%)	14/29 (48%)	14/29 (48%)	13/29 (45%)	12/29 (41%)
Staphylococcus aureus	0/1 (0%)	0/1 (0%)	0/1 (0%)	0/1 (0%)	0/1 (0%)	0/1 (0%)
Total	**14/30 (47%)**	**14/30 (47%)**	**14/30 (47%)**	**14/30 (47%)**	**13/30 (43%)**	**12/30 (40%)**

[a] Data was analyzed as generated (median of 8.5 million reads/specimen)
[b] Data was down-sampled to 5 million reads per specimen
[c] Data was down-sampled to 2 million reads per specimen

Table 8. Additional potential uropathogens detected by clinical metagenomics but not reported by culture.

Potential Uropathogens	Culture-negative (*n*=20)	<1x10^5 CFU/mL (*n*=30)	≥1x10^5 CFU/mL (*n*=65)	Total Detected
Actinotignum schaalii[a]	0	2	2	4
Aerococcus sanguinicola	0	0	2	2
Aerococcus urinae	0	2	1	3
Bacteroides fragilis	0	0	1	1
Corynebacterium jeikeium	0	1	1	2
Corynebacterium riegelii	0	1	1	2
Corynebacterium urealyticum	0	0	1	1
Corynebacterium glucuronolyticum	1	0	0	1
Enterococcus faecalis	1	2	2	5
Fusobacterium nucleatum	0	1	0	1
Gardnerella vaginalis	1	5	6	12
Mycoplasma hominis	0	2	0	2
Oligella urethralis	0	0	1	1
Staphylococcus epidermidis	0	2	3	5
Staphylococcus hominis	1	0	0	1
Streptococcus agalactiae	0	0	2	2
Streptococcus anginosus	1	7	2	10
Total Detected	**5**	**25**	**25**	**55**

[a] Formerly *Actinobaculum schaalii*

Conclusion

This study evaluated the concordance of clinical metagenomics and urine culture. Clinical metagenomics leverages the latest advances in next-generation sequencing to detect a wide range of uropathogens at genome-level resolution. Results indicate high concordance of culture-independent clinical metagenomics with routine urine culture to detect and quantify common and unusual causes of urinary tract infections. These results help laboratories choose the preferred sequencing depth and read length.

References

1. Medina M, Castillo-Pino E. An introduction to the epidemiology and burden of urinary tract infections. Ther Adv Urol. 2019;11:1756287219832172.
2. Flores-Mireles AL, Walker JN, Caparon M, Hultgren SJ. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol. 2015;13(5):269-84.
3. Simmering JE, Tang F, Cavanaugh JE, Polgreen LA, Polgreen PM. The Increase in Hospitalizations for Urinary Tract Infections and the Associated Costs in the United States, 1998-2011. Open Forum Infect Dis. 2017;4(1):ofw281.
4. Davenport M, Mach KE, Shortliffe LMD, Banaei N, Wang TH, Liao JC. New and developing diagnostic technologies for urinary tract infections. Nat Rev Urol. 2017;14(5):296-310.
5. Dason S, Dason JT, Kapoor A. Guidelines for the diagnosis and management of recurrent urinary tract infection in women. Can Urol Assoc J. 2011;5(5):316-22.
6. Foxman B. The epidemiology of urinary tract infection. Nat Rev Urol. 2010;7(12):653-60.
7. Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, Garcia-Cobos S, et al. Application of next generation sequencing in clinical microbiology and infection prevention. J Biotechnol. 2017;243:16-24.
8. For Research Use Only. Not for use in diagnostic procedures.

Appendix A – Specimen Report

Explify® UTI

PATIENT IDENTIFIER:
REPORT DATE:

DATE OF BIRTH:
SEX:
IDbyDNA ACCESSION:

MEDICAL FACILITY:
ACCOUNT ID:
ORDERING PHYSICIAN:

SPECIMEN RECEIVED:
DATE OF COLLECTION:
SPECIMEN TYPE:

Test Ordered: UTI - Data Analysis Solution
Medical Director: Robert Schlaberg, M.D.

RESULTS:

POTENTIAL UROPATHOGENS DETECTED

Predicted Resistance: Amoxicillin, Cefalexin, Ceftriaxone, Imipenem

Analytical Sensitivity: Adequate

The listed antibiotics are limited to those associated with the genetic markers reported below. Potential uropathogen(s) may be intrinsically resistant to antibiotics not listed below; see Test Information at the back of this report for intrinsic resistance information.

POTENTIAL UROPATHOGENS

The detected microorganisms reported below are known uropathogens or have been linked to urinary tract infection (UTI) based on published studies. The microorganisms listed in Test Scope and not reported below are not detected in this specimen. Clinical correlation is recommended.

BACTERIA	QUANTITY
Escherichia coli	> 10^8 GE/mL

FUNGI

None Detected

GENETIC MARKERS FOR ANTIBIOTIC RESISTANCE

This test targets 371 genetic markers for antibiotic resistance. Detected marker(s) are reported when one or more of eight associated uropathogens is also detected (see Test Information). Detection of genetic markers does not always predict phenotypic resistance; lack of detection does not indicate susceptibility.

GENETIC MARKER	PREDICTED RESISTANCE
ampC	Amoxicillin, Cefalexin, Ceftriaxone, Imipenem