

Analytical validation of a highly accurate and reliable next-generation sequencing-based urine assay

Mara Couto-Rodriguez,¹ David C. Danko,¹ Heather L. Wells,¹ Sol Rey,¹ Xavier Jirau Serrano,¹ Gabor Fidler,¹ John Papciak,¹ P. Ford Combs,¹ Anna Plourde,² Michael Augenbraun,² Christopher E. Mason,^{1,3,4,5,6} Caitlin Otto,¹ Niamh B. O'Hara,^{1,7} Dorottya Nagy-Szkal^{1,7}

AUTHOR AFFILIATIONS See affiliation list on p. 13.

ABSTRACT Urinary tract infections (UTIs) are diagnosed based on symptoms and confirmed by urine culture, despite its limitations in sensitivity. False-negative cultures can lead to inappropriate antimicrobial use or urosepsis in high-risk patients. Next-generation sequencing (NGS)-based metagenomics offers a comprehensive and precise alternative but is rarely applied clinically. We developed and validated BIOTIA-ID, a clinical-grade NGS-based diagnostic pipeline for pathogen detection in urine. Remnant clinical and spiked urine samples underwent extraction, metagenomic library preparation, and Illumina NextSeq 550 sequencing. We trained and applied a bioinformatic pipeline that uses machine learning to identify pathogens and resistance markers. BIOTIA-DX was intentionally designed and trained to increase stringency and reduce false positive detection of urogenital commensals or opportunistic microbes present at colonization levels. Internal controls ensured standardized, high-stringency results. The assay was validated on 1,470 urine specimens evaluating over 14.5k analytes. The clinical validation achieved a 97.2% sensitivity and 99.6% specificity with a limit of detection (LoD) of <15,000 CFU/mL for most bacterial species and <5,000 CFU/mL for fungal species. Discordant results were reconciled by target-specific qPCR or 16S Sanger sequencing, and 87% of the NGS results were concordant with the comparator. A subset of 332 clinical specimens was tested and validated for antimicrobial resistance (AMR). *sul* and *blaSHV* genes were commonly associated with *Escherichia coli* and *Klebsiella pneumoniae*, while *cfxA* was found in *Prevotella* and *Pseudomonas* spp. detected by BIOTIA-ID. Overall, these data demonstrate that BIOTIA-ID is a comprehensive, highly accurate end-to-end diagnostic assay with notable advantages over current culture-based diagnostics.

IMPORTANCE Urinary tract infections (UTIs) are among the most common infections, yet current diagnostic methods, including urine culture, often fail to detect pathogens accurately, leading to delayed treatment and inappropriate antimicrobial use. Clinical metagenomics offers a powerful alternative, especially in complicated cases. BIOTIA-ID is a validated, clinical-grade next-generation sequencing (NGS)-based assay that provides highly accurate pathogen identification and antimicrobial resistance profiling. By incorporating machine learning and stringent quality controls, BIOTIA-ID minimizes false positives and enhances diagnostic precision. Our study demonstrates its superior performance over culture, with potential to improve UTI diagnostics, guide targeted therapy, and support antimicrobial stewardship. The implementation of urine metagenomic diagnostics could support recurrent and complicated UTI patient management, providing a more reliable alternative to traditional methods.

Editor Rup Lal, University of Delhi, Delhi, India

Address correspondence to Dorottya Nagy-Szkal, dorottya.nagy-szkal@downstate.edu.

M.C.-R., D.C.D., H.L.W., S.R., X.J.S., J.P., P.F.C., G.F., C.E.M., C.O., N.B.O., and D.N.-S. are employees at Biotia, Inc. The other authors do not report any conflict of interest.

Received 3 July 2025

Accepted 19 February 2026

Published 21 April 2026

Copyright © 2026 Couto-Rodriguez et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

KEYWORDS clinical metagenomics, next-generation sequencing, urinary tract infection, analytical validation, antimicrobial stewardship, machine learning, infectious disease, urogenital pathogens, clinical diagnostics, precision medicine

Annually, 11 million people in the United States and 404 million people worldwide are diagnosed with a urinary tract infection (UTI) (1, 2). Immunocompromised patients and patients with urological conditions are at higher risk to develop complicated and/or recurrent UTI (cUTI/rUTI), as well as to progress to urosepsis, resulting in higher morbidity and mortality and increased cost of care (3–6). In hospitalized patients, UTIs are associated with 2.3% of the mortality rate and an estimated annual cost of \$340 to \$450 million in the United States alone (1). Approximately 30% of patients with sepsis have an infection originating from the urogenital tract, with a multinational study revealing that 12% of nosocomial UTI patients progressed to urosepsis (5, 6). Rapidly identifying UTI pathogens in high-risk patients is crucial to administering an appropriate treatment and minimizing unnecessary broad-spectrum antibiotic use.

Urine culture is the standard of care (SOC) for the detection and identification of pathogens causing UTIs. However, culture-dependent approaches are focused on an Enterobacterales-centric paradigm, which is based on assumptions that urine is sterile (7, 8), disregards mixed infections as contamination (9), and is based on clinical studies employing heterogeneous definitions and thresholds for a UTI (10). Culture has several limitations, including (i) an inability to identify hard-to-grow microbial organisms; (ii) challenges in identifying rare pathogens or organisms not routinely cultured; and (iii) risk of false negative results for patients being treated with antibiotics (11–13). Long diagnostic turnaround time (TAT) or inconclusive results could lead to the empiric treatment of suspected infections with broad-spectrum antibiotics that are often inappropriate and may contribute to increased rates of drug resistance (14).

The emergence of new molecular and culture-based techniques has underscored the drawbacks of traditional culture and challenged our understanding of UTIs, raising questions around what is considered a urogenital pathogen and the role of mixed bacterial communities in UTIs (9, 11, 15–18). Genomics-based assays enable direct specimen analysis without the need for culturing by sequencing genetic material, which is analyzed for the presence of pathogenic organisms through the detection of their DNA. By comparing detected genetic sequences to a comprehensive microbial genome database, genomics-based tests can accurately and rapidly identify pathogens in cUTI samples, which may have atypical pathogens, multiple co-infecting organisms, and complex drug resistance profiles. Concerns have been raised about the sensitivity of molecular approaches to differentiate between opportunistic and commensal microbes detection and differentiation of colonization versus infection (16, 19, 20). However, careful development of genomics-based assays that center on extensive validation leveraging culture, molecular, and bioinformatics techniques can increase the utility of this methodology in routine clinical practice, thus offering a promising supplemental testing method to the current gold standard UTI testing.

Here, we developed, optimized, and clinically validated a next-generation sequencing (NGS)-based urine assay, BIOTIA-ID, that has high diagnostic sensitivity and specificity and provides comprehensive detection of urogenital pathogens within a single test.

MATERIALS AND METHODS

Clinical validation strategy

All specimens were processed with our version-controlled BIOTIA-ID Urine NGS assay laboratory protocol, analytical pipeline, and reference microbial database in Biotia's CLIA-certified laboratory. The assay and clinical validation were designed based on New York State Department of Health validation guidelines for submission of a Laboratory Developed Test for bacteriology and mycology nucleic acid amplification assays (Fig. 1).

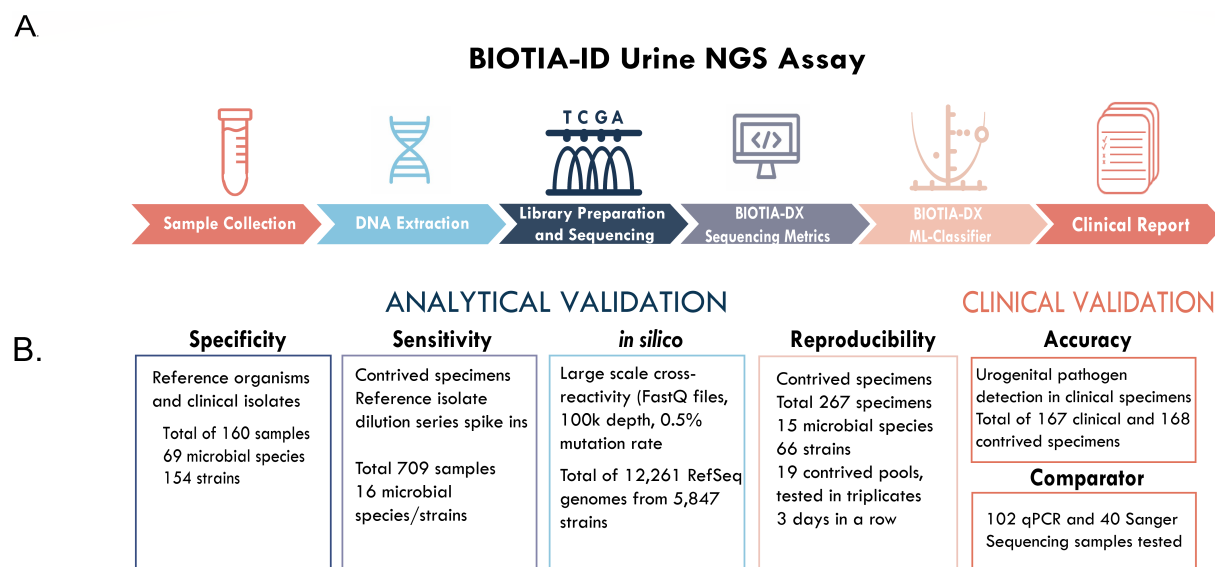


FIG 1 (A) An overview of the BIOTIA-ID Urine NGS assay workflow, consisting of sample collection (urine specimens in standard urine transport tubes [UTT]), nucleic acid extraction, library preparation, and sequencing using the Illumina platform, BIOTIA-DX metagenomic analysis, multiple decision tree-based machine learning modeling, and generation of a clinical report with urogenital pathogens detected. (B) Analytical validation experiments were conducted with clinical and contrived specimens to assess specificity, sensitivity, and BIOTIA-DX *in silico* performance. The clinical validation established the accuracy of urogenital pathogen detection in clinical specimens with additional comparator studies using qPCR and Sanger sequencing in cases where culture and NGS results disagreed. Furthermore, we completed an inter/intra-assay reproducibility study to show the validity and consistency of our results.

Clinical specimens and reference materials

Deidentified urine specimens in urine transport tubes (UTT) were collected as residual samples after routine clinical testing from different microbiology reference laboratories and were provided with the culture results for each specimen following their standard operating procedures to report pathogens. Samples were collected and processed under an institutional review board (Advarra Pro00038083). A collection of clinical and reference isolates (ATCC, Zeptomatrix) was also used for the clinical validation, including bacteria, fungi, viruses, and parasites (listed in Table 1). Reference and clinical microbial isolates used in the contrived specimens were inoculated and grown in Tryptic Soy Agar with 5% Sheep Blood at 37°C for 24–48 h.

Clinical-grade metagenomic sequencing for infectious disease diagnostics

BIOTIA-ID urine NGS assay

Genomic DNA (gDNA) extraction from 2 mL of urine was performed with QIAcube-MDx, and yields were quantified with the Qubit Flex. Normalized gDNA was spiked with 5% internal positive control (IPC) and processed with the Illumina DNA prep library preparation kit. Libraries were quality checked for size and concentration with TapeStation 4200 and Qubit Flex, respectively. Libraries were pooled in equimolar concentrations and sequenced on an Illumina NextSeq 550 platform.

BIOTIA-DX software

After sequencing, the FastQ files were first subjected to a filtering step, which removed low-quality reads and sequencing adapters. Reads with >40% of bases with phred scores <15, reads with >5 N's, or reads with complexity <30% were discarded using *fastp* v0.24.0. Detection of the internal positive control was performed for each sample as a confirmation of overall sequencing quality. Human sequence data were then removed from downstream analysis by removing reads which mapped to either the human or chimpanzee genomes. The remaining reads were first compared against a large

TABLE 1 Reference organisms and clinical isolates tested and validated on the specificity study^b

Gram-negative enterobacterales	Gram-negative non-enterobacterales	Gram-positive	Fungi
<i>Citrobacter koseri</i>	<i>Acinetobacter calcoaceticus-baumannii-complex</i>	Other <i>Staphylococcus</i> spp.	<i>Candida albicans</i>
<i>Citrobacter freundii</i>	<i>Acinetobacter lwoffii</i>	Anginosus group <i>Streptococci</i>	<i>Candida auris</i>
<i>Enterobacter cloacae-complex</i>	<i>Pseudomonas aeruginosa</i>	<i>Streptococcus agalactiae</i>	<i>Candida dubliniensis</i>
<i>Escherichia coli</i>	<i>Stenotrophomonas maltophilia</i>	<i>Streptococcus mitis</i>	<i>Candida glabrata</i>
<i>Klebsiella aerogenes</i>	Anaerobic bacteria^c	Other bacteria	<i>Candida guilliermondii</i>
<i>Klebsiella oxytoca</i>	<i>Anaerococcus</i> spp.	<i>Chlamydia trachomatis</i> ^a	<i>Candida kefyr</i>
<i>Klebsiella pneumoniae-complex</i>	<i>Bacteroides fragilis</i>	<i>Gardnerella vaginalis</i>	<i>Candida krusei</i>
<i>Klebsiella variicola</i>	<i>Prevotella</i> spp.	<i>Mycoplasma genitalium</i> ^a	<i>Candida lusitanae</i>
<i>Morganella morganii</i>	Gram-positive	<i>Mycoplasma hominis</i> ^a	<i>Candida parapsilosis</i>
<i>Proteus mirabilis</i>	<i>Aerococcus</i> spp.	<i>Neisseria gonorrhoeae</i> ^a	<i>Candida tropicalis</i>
<i>Proteus vulgaris</i>	<i>Corynebacterium urealyticum</i>	<i>Treponema pallidum</i> ^a	<i>Cryptococcus neoformans</i> ^a
<i>Providencia rettgeri</i>	<i>Enterococcus faecalis</i>	<i>Ureaplasma</i> spp. ^a	<i>Rhodotulula mucilaginosa</i> ^a
<i>Providencia stuartii</i>	<i>Enterococcus faecium</i>	Viruses	<i>Saccharomyces cerevisiae</i> ^a
<i>Raoultella ornithinolytica</i>	<i>Staphylococcus aureus</i>	CMV ^a	Parasites^a
<i>Salmonella Typhimurium</i> ^a	<i>Staphylococcus epidermidis</i>	HPV ^a	<i>Cryptosporidium</i> spp. ^a
<i>Serratia marcescens</i>	<i>Staphylococcus lugdunensis</i>	HSV1 ^a	<i>Entamoeba</i> spp. ^a
<i>Shigella flexneri</i> ^a	<i>Staphylococcus saprophyticus</i>	HSV2 ^a	<i>Giardia lamblia</i> ^a
			<i>Trichomonas vaginalis</i> ^a

^aOrganisms tested with BIOTIA-ID assay but not reported as a key urogenital pathogen.^bWhole organisms of urogenital pathogens detected by the assay, related organisms, and other organisms that can be present in urine specimens were tested by spiking microbial cells into a negative urine matrix (50,000 CFU/mL for bacteria and 25,000 CFU/mL for fungi). A total of 69 microbial species (bacteria, fungi, viruses, and parasites) and 154 strains were evaluated.^cGray shading and bolded text indicate the headers for the listed pathogens.

database of microbial genomes in a coarse classification step using a commonly used metagenomics tool, *kraken2* (21). We use a custom *kraken2* microbial database within BIOTIA-DX that includes curated, high-quality genetic sequences from bacterial, fungal, viral, and non-fungal eukaryotic parasitic organisms. In total, more than 7,000 organisms are represented in the database, with a diverse representation of microbial species and strains.

Despite any degree of database curation, metagenomics tools often result in spurious, false-positive hits, which have historically made it difficult to use NGS in a clinical diagnostic setting. To distinguish between clinically relevant infections, off-target spurious hits by the metagenomics software, or low levels of cross-contamination, we designed a classifier using ML that estimates the probability that a microbial organism is present in a sample. The ML classifier uses a human-validated training set where known pathogens and clinically relevant infections have been verified. These “ground truth” samples were used to train the ML classifier to distinguish between pathogenic infections and non-pathogenic identifications of other organisms. In this way, we can ensure that the ML model is highly tuned to organisms causing clinically significant infections and is not detecting commensal organisms, non-pathogens, or spurious noise in the metagenomics. Microbial targets exceeding predefined *kraken2* thresholds in the coarse classification step are tested using this ML classifier in a fine classification step. Finally, organisms that exceeded a pre-defined probability threshold from the classifier were called present in the sample.

Analytical validation

Specificity

Reference organisms and clinical isolates were tested and validated to establish the specificity of the assay. Whole organisms of urogenital pathogens detected by the assay, genetically related organisms, and other organisms that can be present in urine specimens were tested by spiking microbial cells into a negative urine matrix. A total of 69 microbial species (bacteria, fungi, viruses, and parasites) and 154 strains were

evaluated (Table 1). Microbial isolates were spiked into a negative urine matrix at concentrations of 25,000–50,000 CFU/mL and processed with the assay.

Analytical sensitivity

The analytical sensitivity was assessed in 16 of the most prevalent urogenital pathogens by spiking reference whole organisms into a negative urine matrix. Serial dilutions of each pathogen were spiked in triplicate, followed by gDNA extraction. The limit of detection was defined as the concentration of the target organism detected in at least 95% of the samples.

Clinical validation

Accuracy

We tested a combination of urine clinical specimens and contrived samples (whole organism microbial reference strains and clinical isolates spiked into negative clinical matrix) to assess accuracy. Deidentified, remnant clinical urine samples were obtained from a reference laboratory and stored at -80°C until processing with the assay. Urine clinical samples were collected based on the pathogens diagnosed by culture. At least 30 specimens, each with the most common urogenital pathogens (*Escherichia coli*, *Enterococcus faecalis*, *Klebsiella pneumoniae*, *Proteus mirabilis*, and *Staphylococcus aureus*), were collected. When fewer than 30 samples for a given pathogen were available, contrived specimens were used to bring the total number of samples for that target to 30 (a total of 335 samples; 167 clinical specimens; and 168 contrived). Contrived samples were used for *S. aureus* ($n = 30$), *P. mirabilis* ($n = 9$), *K. pneumoniae* ($n = 8$), *C. albicans* ($n = 33$), *C. auris* ($n = 30$), *C. glabrata* ($n = 30$), *C. krusei* ($n = 33$), *C. parapsilosis* ($n = 34$), and *C. tropicalis* ($n = 30$), with a spike in concentration close to the limit of detection (LoD) ($\sim 2.5\text{--}5\times$ LoD). In addition, 35 specimens negative by culture were tested.

See Supplemental materials for additional methodology details on the assay controls, software training set, *in silico* analytical specificity, inter/intra-reproducibility, and comparator testing.

Antimicrobial resistance gene marker identification

The BIOTIA-DX antimicrobial resistance (AMR) module is designed to detect a comprehensive set of 12 beta-lactamase genes (*blaOXA*, *blaNDM*, *blaCTX-M*, *blaTEM*, *blaKPC*, *blaCMY*, *blaADC*, *blaSHV*, *blaVIM*, *blaPDC*, *blaZ*, *cfxA*), four *mec* gene classes (A–D), four *sul* genes (*sul1–4*), and two *van* genes (*vanA–B*). Members within these gene families often share a high degree of sequence homology. To address this, gene sequences were aggregated into consensus sequences to enable the general detection of these target gene families. Sequences were collected from the CARD and AMRFinder databases and clustered using CD-HIT with an 80% identity threshold. Gene families with higher levels of sequence diversity (e.g., *blaOXA* or *blaCTX-M*) were represented with multiple clusters covering all variants. The resulting cluster sequences were aligned using MAFFT to construct a base frequency matrix. Consensus sequences for each cluster were defined based on the most frequent nucleotide at each position. This database of gene family consensus sequences was then used as the reference for read mapping in the AMR detection process. We used a subset ($n = 332$) of culture-positive and culture-negative clinical specimens to test for AMR gene markers.

RESULTS

We developed an end-to-end urine clinical diagnostic metagenomics assay to detect urogenital pathogens and validated its performance on 1,470 samples sequenced in 65 sequencing runs (Fig. 1). The sequencing performance results (microbial read counts and control metrics) are summarized in Fig. 2.

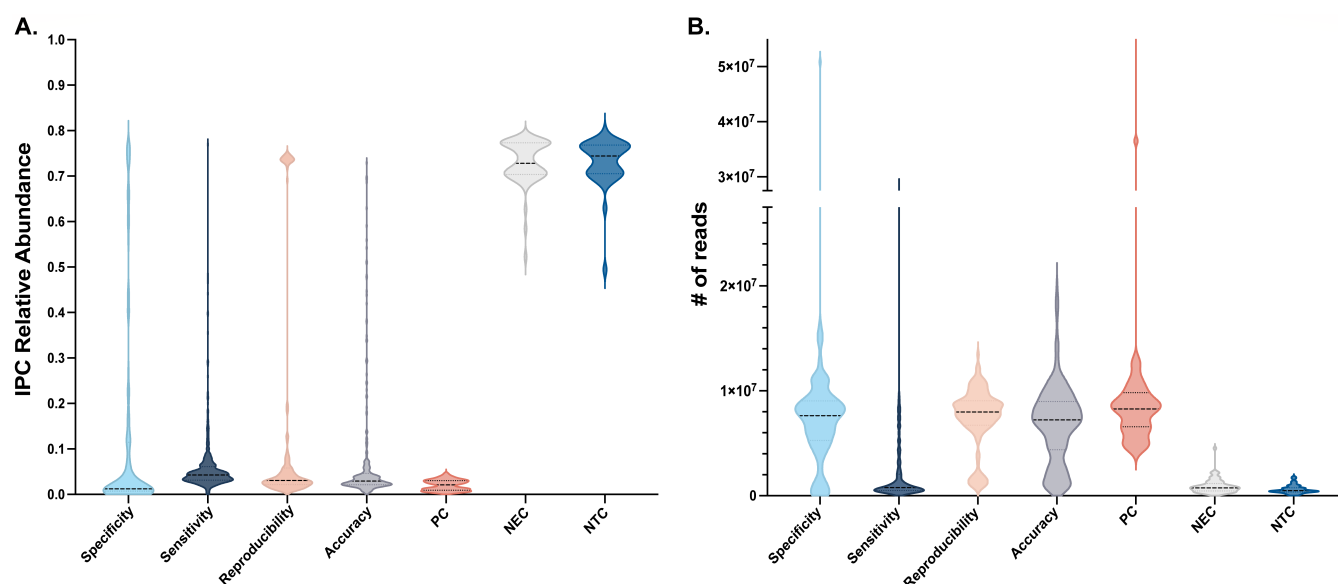


FIG 2 The assay performance of specimens processed in the analytical (specificity, sensitivity, and reproducibility) and clinical validation (accuracy) studies was defined by evaluating the IPC and the number of microbial reads. (A) The IPC was detected in all specimens and positive controls (PC) at an expected range (1%–5% abundance). The IPC reads represented the majority of the reads detected in negative extraction controls (NEC) and negative template controls (NTC). (B) Violin plots depicting the number of microbial reads obtained in each study and controls. As expected, the clinical and contrived specimens processed through the studies and the PCs generated similar microbial read depth (average 7.9M of microbial reads), while the NEC and NTC had microbial reads below 750k. Violin plots show IPC abundance and microbial read distribution for all samples and controls processed in each study: specificity (light blue, $n = 160$); sensitivity (navy blue, $n = 709$); reproducibility (light pink, $n = 267$); accuracy (gray, $n = 325$), PC (dark pink, $n = 65$), NEC (light gray, $n = 130$), NTC (cerulean blue, $n = 130$).

Key performance characteristics of BIOTIA-ID

Specificity

The similarity among different microbial genomes, the diversity of clinical isolates from reference genomes, and contaminating DNA fragments impact the analytical sensitivity of a metagenomic test. Urogenital pathogens detected by the assay and genetically related organisms were processed with BIOTIA-ID (Table 1), yielding a sensitivity and specificity of 100% and 99.96% for bacterial targets and 100% sensitivity and specificity for fungal targets. During assay development, several taxa not extensively represented in the clinical training set generated false positive and false negative results, leading us to process 64 additional species in 80 contrived urine samples to feed more training data to the model. This approach, coupled with hyperparameter tuning of the model, ensured the increased sensitivity and specificity reported above.

Sensitivity

The analytical sensitivity studies established the lowest concentration of pathogen reliably detected by our assay. Ten bacterial and six fungal species were tested in six replicates per concentration, per analyte evaluated (Fig. 3). A total of 709 contrived specimens were processed. Specifically, the LoDs determined for each species were 1,000 CFU/mL for *C. glabrata*, *C. krusei*, *C. parapsilosis*, and *C. tropicalis*; 5,000 CFU/mL for *C. albicans* and *C. auris*; 7,500 CFU/mL for *E. coli*, *G. vaginalis*, *K. pneumoniae*, and *P. mirabilis*; 10,000 CFU/mL for *E. faecalis*; 12,500 CFU/mL for *Prevotella* spp.; 15,000 CFU/mL for *A. baumannii* and *S. aureus*; 25,000 CFU/mL for *P. aeruginosa* and *B. fragilis*.

Accuracy

The accuracy verification study evaluated the precision of the assay in detecting uropathogens in clinical urine specimens compared to the current gold standard

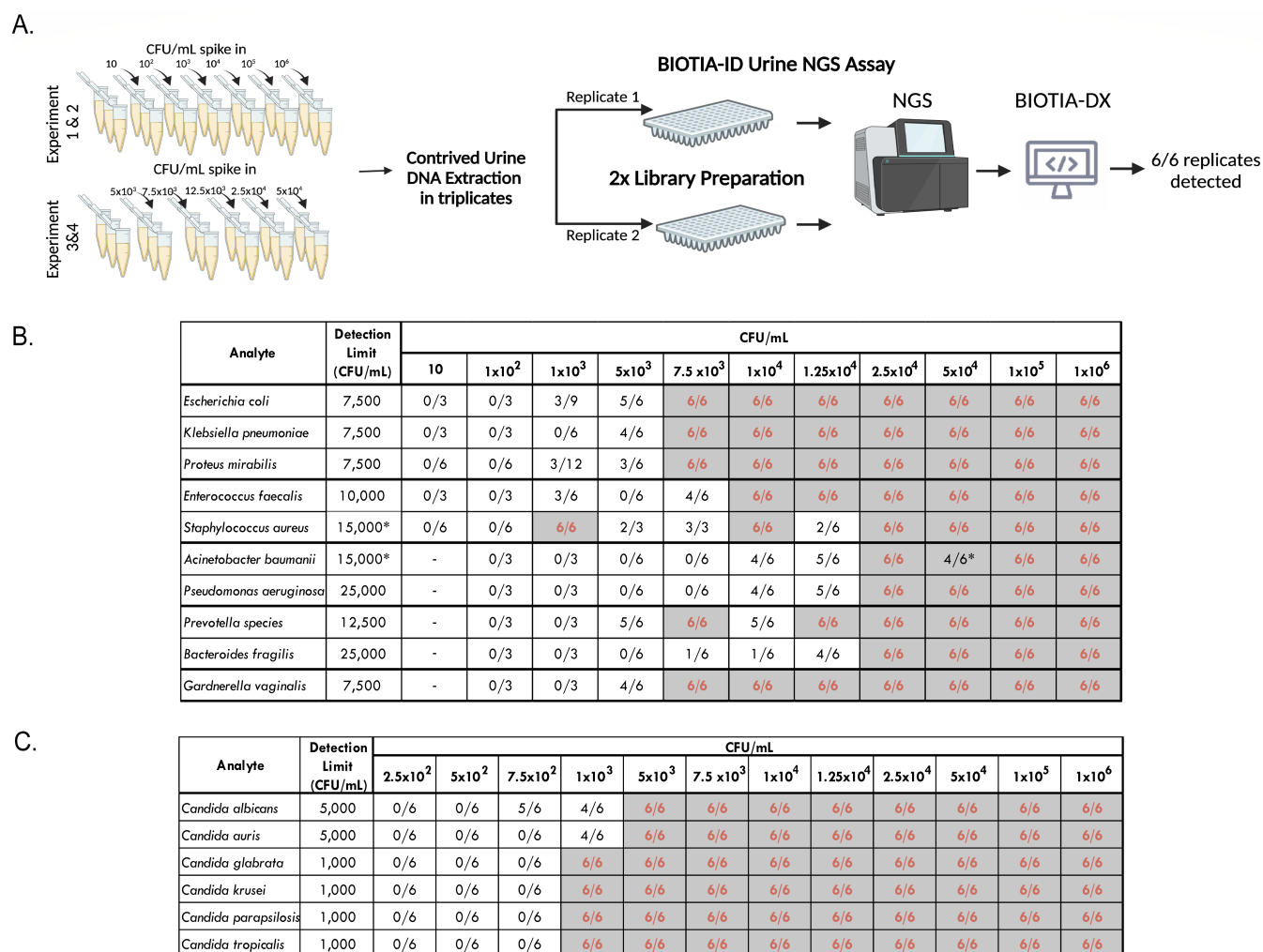


FIG 3 The analytical sensitivity was assessed in the most frequently found urogenital pathogens (10 bacteria and 6 fungal species) by spiking reference whole organisms into a negative urine matrix. (A) Each pathogen was spiked in triplicate per concentration (CFU/mL), followed by genomic DNA extraction. Experiments 1 and 2 included six replicates for each concentration of the 10-fold dilution spike-ins, ranging from 10 to 1 million CFU/mL. Experiments 3 and 4 were used to fine-tune the LoD by testing six replicates per concentration, ranging from 5×10^3 to 5×10^4 CFU/mL. (B and C) Summary of replicates detected per concentration tested for each urogenital pathogen. The LoD was reproducibly verified and determined based on a 100% positivity rate, resulting in an overall LoD of <25,000 CFU/mL for bacteria and <5,000 CFU/mL for fungi. Specifically, the LoDs determined for each species were 1,000 CFU/mL for *C. glabrata*, *C. krusei*, *C. parapsilosis*, and *C. tropicalis*; 5,000 CFU/mL for *C. albicans* and *C. auris*; 7,500 CFU/mL for *E. coli*, *G. vaginalis*, *K. pneumoniae*, and *P. mirabilis*; 10,000 CFU/mL for *E. faecalis*; 12,500 CFU/mL for *Prevotella* spp.; 15,000 CFU/mL for *A. baumannii* and *S. aureus*; 25,000 CFU/mL for *P. aeruginosa* and *B. fragilis*.

diagnostic (culture). A total of 335 samples (167 clinical and 168 contrived) were tested, including 35 culture-negative specimens.

BIOTIA-ID detected 100% of the spiked microorganisms in contrived specimens and identified a total of 250 urogenital pathogens in the clinical specimens, classifying 83.2% as positive for UTI (Fig. 4A). A single organism was detected in 43.1% of the specimens, while more than one organism was detected in 40.1% (of those 19.8% polymicrobial) (Fig. 4B). Conversely, 72.5% of specimens were observed as single organisms by culture. As depicted in Fig. 4C, BIOTIA-ID detected 26 different urogenital pathogen species with higher frequencies of the key urogenital pathogens used throughout this validation (a total 71.8% of *E. coli*, *E. faecalis*, *K. pneumoniae*, *P. mirabilis*, and *S. aureus*). The remaining 28.2% was uropathogens commonly reportable by culture (e.g., *S. agalactiae*, *Staphylococcus* spp., gram-negative Enterobacterales species, and *Aerococcus* spp.), and organisms not usually culturable with the SOC (e.g., *G. vaginalis* and *Prevotella* spp.). NGS detected *Candida* species in six specimens, while culture reported yeast in only two.

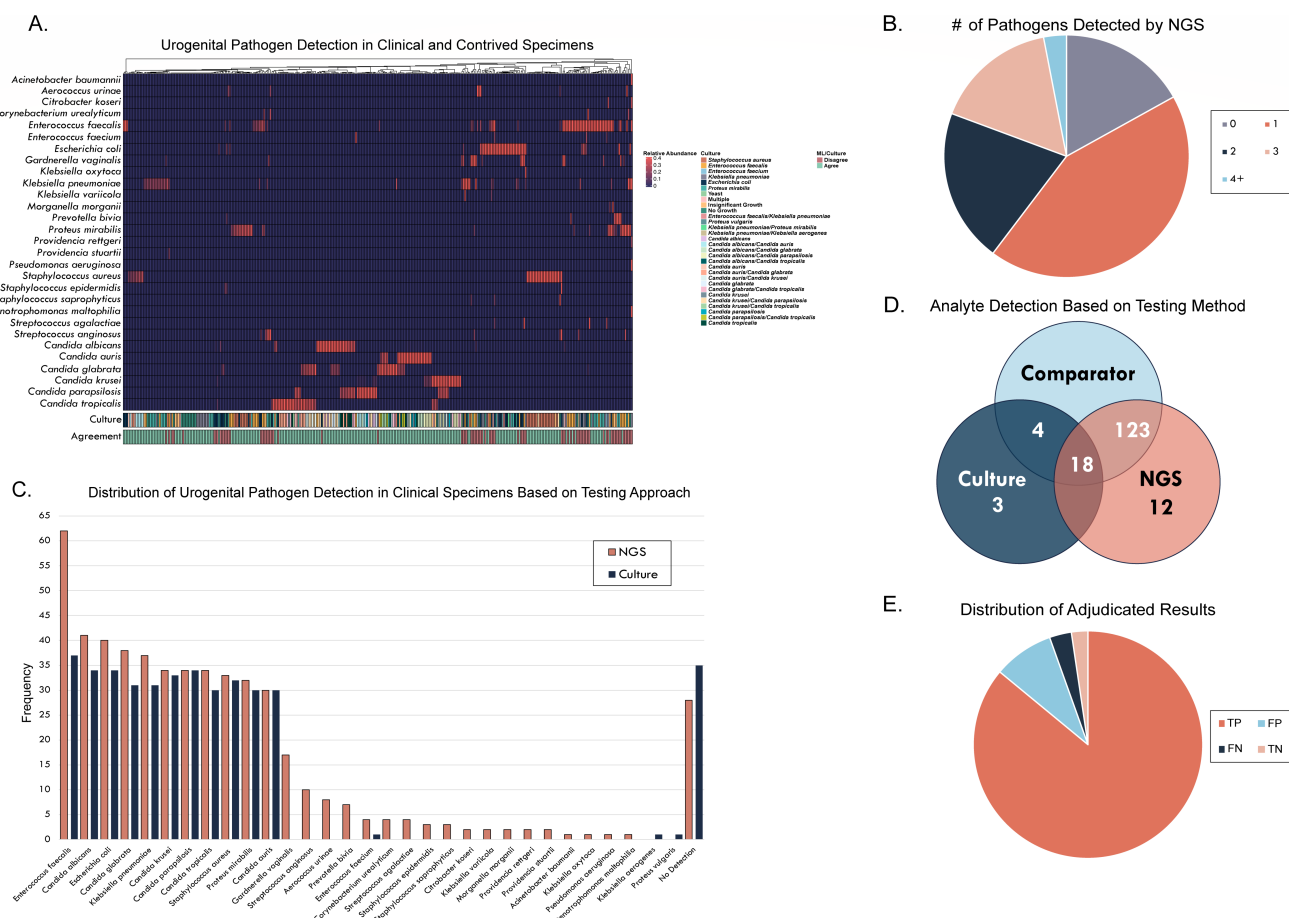


FIG 4 The BIOTIA-ID Urine NGS assay accurately detects and classifies urogenital pathogens in clinical ($n = 167$) and contrived ($n = 168$) specimens in the accuracy study. (A) A heatmap depicts the relative abundance of taxa detected by NGS and is annotated with culture diagnosis, and agreement (green) or disagreement (purple) between BIOTIA-DX ML and culture. (B) Pie chart showing the distribution of the number of urogenital pathogens detected in clinical urine specimens tested with BIOTIA-ID. (C) Bar chart depicts the distribution of microbial analytes detected with culture (navy blue) and with NGS (pink). (D) A Venn diagram illustrates the distribution and overlap of the analytes tested and detected with comparator testing for adjudicating the discrepancies found between NGS and SOC. Apparent false positive and false negative results from 19 analytes, where BIOTIA-ID differed from the original culture result (50.2% of clinical samples), were tested by qPCR or Sanger sequencing ($n = 123$), and (E) 87% of these NGS results were found to be correct ($n = 107$) as shown on pie chart showing the distribution of adjudicated results. TP: true positive ($n = 104$, dark pink), TN: true negative ($n = 3$, light pink), FP: false positive ($n = 12$, dark blue), FN: false negative ($n = 4$, dark gray). *E. faecalis*, *K. pneumoniae*, *E. coli*, *Candida* spp., *Aerococcus* spp., and *G. vaginalis* represented the highest frequency of disagreements (65%) between the SOC and NGS. Comparator reconciled results by taxa tested are shown in Fig. S1. Moreover, analytes adjudicated as false positives by NGS mostly belong to *G. vaginalis* and Anginosus group streptococci, two common opportunistic pathogens of the urogenital tract. However, false negative samples with analytes *E. faecalis*, *K. pneumoniae*, and *E. coli* yielded below threshold BIOTIA-ID prediction probabilities, demonstrating BIOTIA-DX's low statistical confidence in defining the presence of the pathogen reported by the SOC.

Of the clinical specimens tested, 50.9% revealed a disagreement between NGS and the gold standard culture (11.3% completely disagreed on the species detected). Due to cultural limitations and the high sensitivity of NGS, comparator testing of 123 specimens with 19 analytes (qPCR $n = 82$, Sanger sequencing $n = 39$) was conducted to reconcile discrepancies between findings from both methods. A subset of the 19 concordant analytes was used as clinical positive controls, and 100% of the analytes detected agreed across the three methods (Fig. 4D; Table S5). Apparent false positives and false negatives were then tested, and 87% of these NGS results were concordant with the comparator (Fig. 4E).

Following comparator adjudication, the accuracy study achieved a 97.2% sensitivity and 99.6% specificity (Table 2). Overall, BIOTIA-ID accomplished >99.9% sensitivity and specificity across all samples and analytes processed, exceeding ~14.5k analytes

TABLE 2 BIOTIA-ID clinical and analytical validation performance characteristics summary by analyte category (bacteriology, mycology, overall)^a

Category	Method	TP	TN	FN	FP	Total	Sensitivity (%)	Specificity (%)
Bacteriology	Clinical	210	8,323	7	35	8,575	96.77	99.58
	Contrived	1,099	43,122	0	14	44,048	100.00	99.97
	<i>In silico</i>	8,119	272,775	3	147	281,044	99.96	99.95
	Total	9,431	332,483	10	196	341,933	99.89	99.94
Mycology	Clinical	30	530	0	0	560	100.00	100.00
	Contrived	925	9,240	1	2	10,168	99.89	99.98
	<i>in silico</i>	4,142	41,420	0	0	45,562	100.00	100.00
	Total	5,097	51,190	1	2	56,290	99.98	100.00
Overall	Clinical	240	8,853	7	35	9,135	97.17	99.61
	Contrived	2,024	52,362	1	16	54,216	99.95	99.97
	<i>In silico</i>	12,261	314,195	3	147	326,606	99.98	99.95
	Total	14,528	383,673	11	198	398,223	99.92	99.95

^aTP, true positive; TN, true negative; FN, false negative; FP, false positive.

evaluated (Table 2). Contrived specimens yielded >99.9% sensitivity and specificity, and analyte breakdown is presented in Table S6.

Resistance marker detection using BIOTIA-ID

A total of 332 clinical specimens were tested for AMR markers, of which 141 (42%) tested positive for at least one target. The culture-positive ($n = 131$) and culture-negative samples ($n = 201$) exhibited similar proportions of target genes, with 68 and 73 positive markers detected, respectively. Both groups showed comparable AMR profiles, with notable differences in the prevalence of *blaSHV* (21.1% vs 7.6%) and *mecA* (5.7% vs 13.4%) (Fig. 5A). Overall, the most frequently detected markers were *cfxA* ($n = 54$), *sul* ($n = 47$), and *blaSHV* ($n = 34$). The *sul* and *blaSHV* genes were commonly associated with *E. coli* and *K. pneumoniae*, while *cfxA* was found in *Prevotella* and *Pseudomonas* species detected by BIOTIA-ID. The predictions were validated using AMRFinder and confirmed by qPCR testing. Both tools demonstrated 100% accuracy for *sul*, *blaTEM*, *blaOXA*, *mecA*, *van*, and *blaKPC*. However, the accuracy for *cfxA* was lower, with BIOTIA-DX achieving 86% and AMRFinder 82%. For three *cfxA* samples incorrectly predicted as negative, PCR confirmation revealed higher cycle threshold (Ct) values (>34) compared to the average Ct of 25 for all *cfxA* samples. Additionally, AMRFinder exhibited reduced accuracy for *blaSHV* (87%) and *blaCTX-M* (94%) compared to BIOTIA-DX (Fig. 5B).

DISCUSSION

In this study, we developed and validated a clinical-grade NGS assay, BIOTIA-ID, for accurate pathogen identification in urine specimens using ML-classification. BIOTIA-ID overcomes challenges specific to urine molecular diagnostics, including sample collection and transport, low gDNA yields, presence of inhibitors, high human background, and taxonomic bias, due to suboptimal extraction (22–26). The laboratory process is compatible with UTT, the preservative of choice for standard urine cultures in a clinical setting. DNA extraction was optimized for efficient lysis of gram-positive bacteria and fungi, ensuring sufficient gDNA yields from 2 mL of urine, specifically from low biomass samples, such as those collected from males, from catheterized patients, or from patients with prior antibiotic use. Three external controls and one internal control were implemented to validate the process (PC), monitor cross-contamination coming from the process or reagents (negative extraction controls [NEC], negative template controls [NTC]), and ensure the integrity and validity of each sample, minimizing false negatives due to inhibition (IPC). BIOTIA-DX has a clinically curated pathogen database, and the ML-classifier was trained with a combination of urine clinical specimens, clinical and

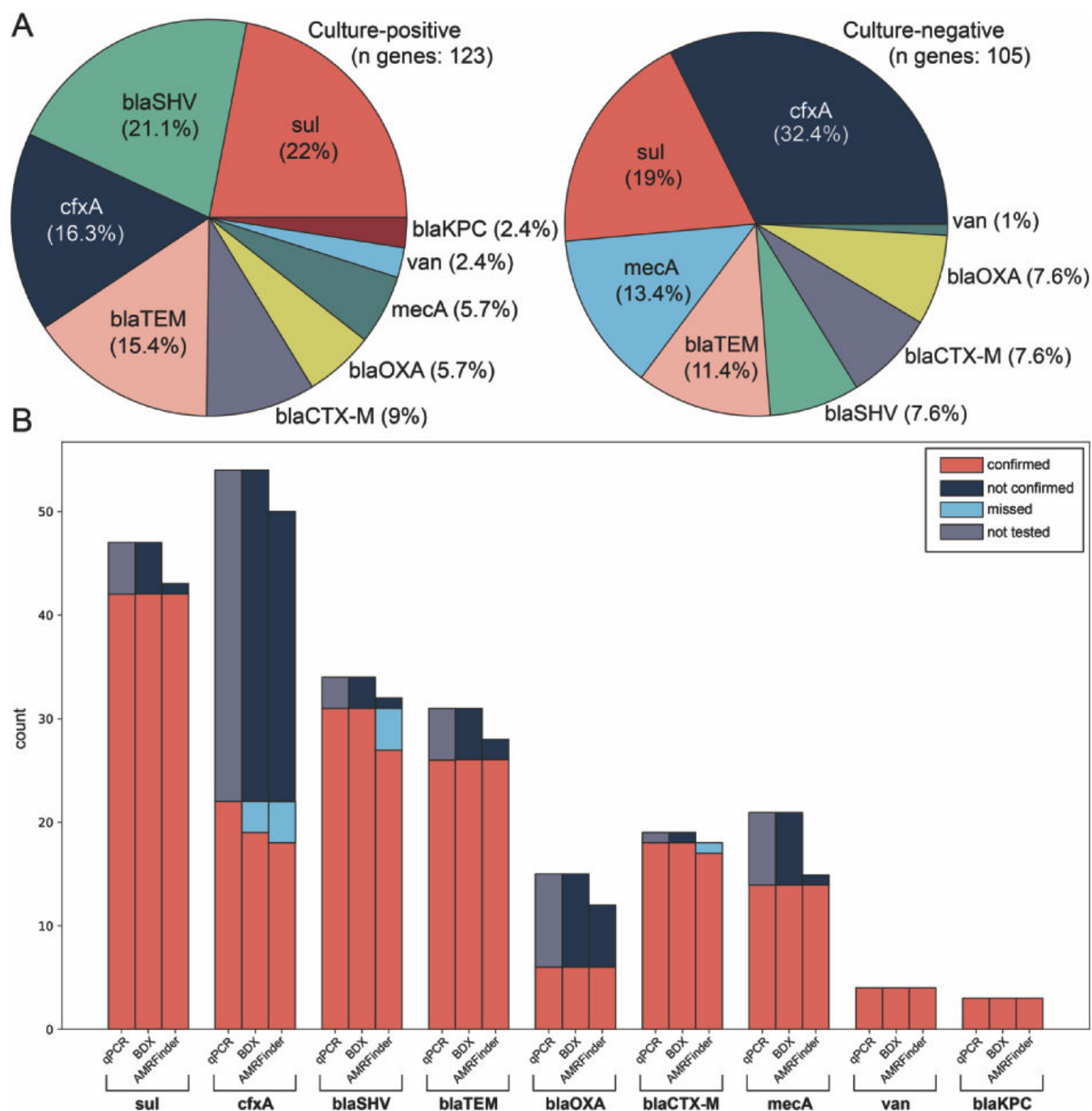


FIG 5 Prevalence of BIOTIA-DX AMR marker detection in both culture-positive and culture-negative clinical specimens where at least one of the target markers was present. (A) In culture-positive samples, the most frequently detected markers were *sul*, *blaSHV*, and *cfxA*, whereas in culture-negative samples, the most prevalent markers were *cfxA*, *sul*, and *mecA*. Prediction results were compared with AMRfinder, and gene presence was validated with qPCR testing. (B) The bar charts display the number of qPCR-confirmed markers (shown in orange). Additionally, the light blue portions represent the false-negative ratios for the two prediction tools. Areas where gene presence was not tested by qPCR are colored in gray, and the dark blue portions indicate the predictions that were not confirmed.

reference isolates, and synthetic metagenomic specimens. Comparator validation of discrepant results, along with a comprehensive training set, increased the specificity and robustness of pathogen prediction. Analytical performance was assessed on 20 prevalent urogenital pathogens in 1,305 contrived samples, 12,250+ *in silico* simulations, and 167 clinical specimens. BIOTIA-ID demonstrated high sensitivity (99.92%) and specificity

(99.95%) in predicting uropathogens, making it a promising alternative testing to culture-based diagnostics.

Although urine culture is highly reliable, with reported ~90% sensitivity and 86% specificity in healthy outpatient women, its specificity is exceptionally variable in chronically ill patients (76%–95%) and in patients with indwelling catheters, where it drops close to 0% (27). BIOTIA-ID's ability to precisely detect pathogens at low concentrations enhances its clinical utility, particularly in the context of managing high-risk immunocompromised patients, rUTI and cUTIs, and/or catheter-associated UTIs. UTI diagnostics currently rely on culture, and while effective, 20%–30% of cases go undiagnosed (13), with this number higher in complicated cases. Culture-based diagnostics fail due to numerous issues, including prior antibiotic exposure, collection or storage errors, cross-contamination, selectivity of culture conditions, and inadequate identification of fastidious, anaerobic, and fungal infections (12, 13), many of which can be addressed with NGS-based diagnostics.

In this study, half of the NGS tested clinical specimens revealed a disagreement with culture findings, with most cases detecting more than one urogenital pathogen, including the species reported by culture. About 11.3% of specimens had different species detected with NGS, and about 15% exhibited clinically relevant discrepancies, like detecting a gram-positive (*E. faecalis*) instead of a gram-negative pathogen (*E. coli*). Despite being highly abundant by both NGS and qPCR testing, *E. faecalis* was the top missed organism by culture, reported as negative. Notably, close to 43% of the clinical specimens which culture diagnosed as *P. mirabilis* showed *E. faecalis* as the dominant pathogen with an NGS-based approach. *E. faecalis* is frequently under-reported by standard culture, with 50% detection rates when compared to Enhanced Quantitative Urine Culture (11). Furthermore, *Candida* species were missed ($n = 17$) or only reported by culture as yeast ($n = 2$), representing 11.9% of the clinical specimens tested. While *Candida* can be isolated with the SOC, sensitivity is rather poor. A study compared *Candida* detection in urine samples with confirmed yeast presence on urinalysis, finding that standard urine culture identified 37% of *Candida* infections, whereas using a fungal-specific media (Sabouraud Dextrose Agar) increased detection to 98% (28). Studies have shown that standard urine culture does not provide optimal growth conditions for all urogenital pathogens, causing organisms, such as Enterobacteriales, to outcompete other microbes like *Enterococcus* and *Candida* (11, 15). Failure to detect the correct pathogen results in physicians prescribing inadequate antibiotics, thus exacerbating the emergence of antimicrobial resistance and increasing patient suffering (29, 30).

Like other molecular-based studies, BIOTIA-ID showed improved detection of key urogenital pathogens, atypical microbes, and fastidious bacteria, organisms not reported with culture-based diagnostics (16, 31–34). Although the rise of molecular-based diagnostics and clinical studies using these technologies has revealed that gram-positive, atypical, or fastidious pathogens account for a larger percentage of cUTI infections, more interventional and clinical utilization studies are needed to understand the pathogenesis and epidemiology of these understudied microbes (11, 35). Despite specimen collection bias, BIOTIA-ID identified a significant number of atypical pathogens in the tested clinical specimens, leading us to hypothesize that the detection frequency would likely rise in clinical studies that focus on culture-negative cUTI cases, a population that could greatly benefit from improved diagnostics.

Polymicrobial infections with more than two pathogens are usually considered contamination under the current SOC, and reporting of these results as infection is a topic of debate (10). In accordance with other UTI studies employing molecular approaches, about 40% of clinical specimens tested in this validation showed more than one pathogen detected. Most clinical algorithms designed for uropathogen detection provide optimal growth conditions for a limited number of microbes and are based on a threshold of 10^5 CFU/mL. Understandably, many guidelines aim to prevent over-prescription of antibiotics. Nonetheless, the development of improved diagnostics can help

better guide antimicrobial stewardship efforts while still providing a comprehensive report. Regarding concerns about the potential oversensitivity of an NGS-based assay, BIOTIA-DX was intentionally designed and trained to increase stringency and reduce false positive detection of urogenital commensals or opportunistic microbes present at colonization levels often a weakness of many molecular diagnostic tools and clinical studies (18, 19, 35). Relative abundance of microbial species, percentage of human reads, IPC detection, and complexity of the microbial profiles are features implemented to enhance the BIOTIA-DX's ability to correctly predict an infection. This study highlights how ML, coupled with extensive laboratory and bioinformatic validation, could improve infectious disease diagnostics by addressing the failures of culture-based diagnostics. Preliminary data suggested a strong linear correlation between spiked CFU/mL and relative abundance of LoD contrived samples (data not shown). The complexity of a clinical urine specimen requires the consideration and development of additional pipeline features for the implementation of a quantification metric that could be relevant and useful to physicians. Additional clinical studies are required to generate a robust data set, facilitating the creation of an updated, user-friendly NGS-based metric that aligns more closely with current clinical and antibiotic stewardship guidelines for UTI management.

Further prospective clinical studies and real-world evaluations are warranted to validate the clinical impact, cost-effectiveness, and health-system benefits of BIOTIA-ID as a UTI diagnostic tool. The observed performance characteristics suggest potential clinical applications where the advantages of accuracy and comprehensive testing outweigh some of the limitations associated with sequencing cost and TAT. Some use cases that would most benefit from this approach include symptomatic patients with negative culture, cUTI and rUTI, and immunocompromised patients with nonspecific infection symptoms who are at high risk of developing sepsis. The current end-to-end turnaround time of approximately 36 h and the expected cost—projected to align with advanced multiplex PCR panels—highlight the emerging feasibility of sequencing-based diagnostics in clinical settings.

BIOTIA-ID represents a significant advancement in UTI diagnostics, offering notable advantages over culture-based methods. Its comprehensive nature, high accuracy, low LoD, and potential to guide targeted antimicrobial therapy support the goals of antimicrobial stewardship. Implementation of BIOTIA-ID in clinical settings with the right diagnostic stewardship has the potential to improve patient outcomes, reduce the misuse of antibiotics, and contribute to the global fight against antimicrobial resistance.

ACKNOWLEDGMENTS

We thank Zaineb Bello for clinical specimen collection and laboratory operation support; Tiara Rivera and Tamara Goncharuk for sequencing support; Yehudah Gruenstein and LabQ Team for clinical specimen collection; Adam Nagyhazi-Horvath, Eszter Szollosy, Patrik Blik, the Abesse Team, Pierre Davidoff, Cory Mason for IT support; Taylor Paisie for bioinformatic support; Talisa Vega-Kline for logistical support.

The BIOTIA-DX software pipeline work used Cromwell On Azure; the Microsoft Genomics supported implementation of the Broad Institute's Cromwell workflow engine on Azure. The development and validation work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant (ACI-1548562). Specifically, it used the Bridges system, which is supported by an NSF award (ACI-1445606), at the Pittsburgh Supercomputing Center (PSC).

M.C.-R., D.C.D., C.E.M., C.O., N.B.O., and D.N.-S. designed and conceptualized the validation study; A.P. and M.A. provided a subset of the clinical specimens (resources); M.C.-R., D.C.D., H.L.W., S.R., G.F., J.P., and P.F.C. clinical and genomic data curation, software development, data analysis, and visualizations (data curation and formal analysis, visualization); M.C.-R., S.R., X.J.S., and H.L.W. performed the investigation and analytical validation (data curation and formal analysis, visualization); M.C.-R., C.O., and D.N.-S.

wrote, reviewed, and edited the manuscript with input from all authors; D.C.D., H.L.W., G.F., P.F.C., A.P., M.A., and N.B.O. edited the manuscript; D.N.-S., N.B.O., and C.E.M. provided funding and resource allocation.

AUTHOR AFFILIATIONS

¹Biotia Inc., New York, New York, USA

²The Department of Medicine, SUNY Downstate Health Sciences University, New York, New York, USA

³Tri-Institutional Computational Biology & Medicine Program, Weill Cornell Medicine of Cornell University, New York, New York, USA

⁴Weill Cornell Medicine, The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, New York, New York, USA

⁵Weill Cornell Medicine, The WorldQuant Initiative for Quantitative Prediction, New York, New York, USA

⁶Weill Cornell Medicine, The Feil Family Brain and Mind Research Institute, New York, New York, USA

⁷The Department Cell Biology/College of Medicine, SUNY Downstate Health Sciences University, New York, New York, USA

AUTHOR ORCIDs

Anna Plourde  <http://orcid.org/0000-0002-2824-1821>

Dorottya Nagy-Szakal  <http://orcid.org/0009-0003-5972-6973>

AUTHOR CONTRIBUTIONS

Mara Couto-Rodriguez, Conceptualization, Data curation, Formal analysis, Project administration, Supervision, Visualization, Writing – original draft, Writing – review and editing | David C. Danko, Conceptualization, Data curation, Formal analysis, Software, Visualization, Writing – review and editing | Heather L. Wells, Data curation, Formal analysis, Software, Visualization, Writing – review and editing | Sol Rey, Data curation, Formal analysis, Writing – review and editing | Xavier Jirau Serrano, Data curation, Formal analysis | Gabor Fidler, Data curation, Formal analysis, Software, Visualization, Writing – review and editing | John Papciak, Data curation, Formal analysis, Software | P. Ford Combs, Data curation, Formal analysis, Software, Visualization, Writing – review and editing | Anna Plourde, Resources, Writing – review and editing | Michael Augenbraun, Resources, Writing – review and editing | Christopher E. Mason, Conceptualization, Funding acquisition, Resources, Software | Caitlin Otto, Conceptualization, Supervision, Writing – original draft | Niamh B. O'Hara, Conceptualization, Funding acquisition, Resources, Supervision, Writing – review and editing | Dorottya Nagy-Szakal, Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The data supporting the study findings are available from the corresponding author on request. Sequencing data that support the findings of this study (with human reads removed) have been deposited in GeoSeeq (<https://portal.geoseeq.com/projects/bid-urine-analytical-data>), a publicly available metagenomic platform, and are available upon request. The BIOTIA-DX software used is described in the Materials and Methods. The software leverages the following openly available tools: fastp v0.23.3, bowtie2 v2.5.1, minimap2 v2.22, SNAP aligner 2.0.3, samtools v1.6, and kraken2 v2.1.3. The ML classifier is a proprietary portion of the code.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (Spectrum02026-25-s0001.pdf). Supplemental methods and results; Fig. S1 and S2; Tables S1 to S6.

REFERENCES

- Yang X, Chen H, Zheng Y, Qu S, Wang H, Yi F. 2022. Disease burden and long-term trends of urinary tract infections: a worldwide report. *Front Public Health* 10. <https://doi.org/10.3389/fpubh.2022.888205>
- Zeng Z, Zhan J, Zhang K, Chen H, Cheng S. 2022. Global, regional, and national burden of urinary tract infections from 1990 to 2019: an analysis of the global burden of disease study 2019. *World J Urol* 40:755–763. <https://doi.org/10.1007/s00345-021-03913-0>
- Brumwell A, Sutton G, Lantos PM, Hoffman K, Ruffin F, Brinkac L, Clarke TH, Adams MD, Fowler VG, Fouts DE, Thaden JT. 2023. *Escherichia coli* ST131 associated with increased mortality in bloodstream infections from urinary tract source. *J Clin Microbiol* 61:e0019923. <https://doi.org/10.1128/jcm.00199-23>
- Tocut M, Zohar I, Schwartz O, Yossepowitch O, Maor Y. 2022. Short- and long-term mortality in patients with urosepsis caused by *Escherichia coli* susceptible and resistant to 3rd generation cephalosporins. *BMC Infect Dis* 22:571. <https://doi.org/10.1186/s12879-022-07538-5>
- Wagenlehner FME, Pilatz A, Weidner W, Naber KG. 2015. Urosepsis: overview of the diagnostic and treatment challenges. *Microbiol Spectr* 3. <https://doi.org/10.1128/microbiolspec.UTI-0003-2012>
- Bjerkklund Johansen TE, Cek M, Naber K, Stratchounski L, Svendsen MV, Tenke P, PEP and PEAP study investigators, European Society of Infections in Urology. 2007. Prevalence of hospital-acquired urinary tract infections in urology departments. *Eur Urol* 51:1100–1111; <https://doi.org/10.1016/j.eururo.2006.08.012>
- Brubaker L, Chai TC, Horsley H, Khasriya R, Moreland RB, Wolfe AJ. 2023. Tarnished gold—the “standard” urine culture: reassessing the characteristics of a criterion standard for detecting urinary microbes. *Front Urol* 3:1206046. <https://doi.org/10.3389/fruro.2023.1206046>
- Wolfe AJ, Brubaker L. 2015. “Sterile urine” and the presence of bacteria. *Eur Urol* 68:173–174. <https://doi.org/10.1016/j.eururo.2015.02.041>
- LaRocco MT, Franek J, Leibach EK, Weissfeld AS, Kraft CS, Sautter RL, Baselski V, Rodahl D, Peterson EJ, Cornish NE. 2016. Effectiveness of preanalytic practices on contamination and diagnostic accuracy of urine cultures: a laboratory medicine best practices systematic review and meta-analysis. *Clin Microbiol Rev* 29:105–147. <https://doi.org/10.1128/CMR.00030-15>
- Bilsen MP, Jongeneel RMH, Schneeberger C, Platteel TN, van Nieuwkoop C, Mody L, Caterino JM, Geerlings SE, Köves B, Wagenlehner F, Conroy SP, Visser LG, Lambregts MMC. 2023. Definitions of urinary tract infection in current research: a systematic review. *Open Forum Infect Dis* 10:ofad332. <https://doi.org/10.1093/ofid/ofad332>
- Price TK, Dune T, Hilt EE, Thomas-White KJ, Kliethermes S, Brincat C, Brubaker L, Wolfe AJ, Mueller ER, Schreckenberger PC. 2016. The clinical urine culture: enhanced techniques improve detection of clinically relevant microorganisms. *J Clin Microbiol* 54:1216–1222. <https://doi.org/10.1128/JCM.00044-16>
- Ishihara T, Watanabe N, Inoue S, Aoki H, Tsuji T, Yamamoto B, Yanagi H, Oki M, Kryukov K, Nakagawa S, Inokuchi S, Ozawa H, Imanishi T. 2020. Usefulness of next-generation DNA sequencing for the diagnosis of urinary tract infection. *Drug Discov Ther* 14:42–49. <https://doi.org/10.5582/ddt.2020.01000>
- Heytens S, De Sutter A, Coorevits L, Cools P, Boelens J, Van Simaey L, Christiaens T, Vanechoutte M, Claeys G. 2017. Women with symptoms of a urinary tract infection but a negative urine culture: PCR-based quantification of *Escherichia coli* suggests infection in most cases. *Clin Microbiol Infect* 23:647–652. <https://doi.org/10.1016/j.cmi.2017.04.004>
- Murray CJL, Ikuta KS, Sharara F, Swetschinski L, Robles Aguilar G, Gray A, Han C, Bisignano C, Rao P, Wool E, et al. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* 399:629–655. [https://doi.org/10.1016/S0140-6736\(21\)02724-0](https://doi.org/10.1016/S0140-6736(21)02724-0)
- Hilt EE, McKinley K, Pearce MM, Rosenfeld AB, Zilliox MJ, Mueller ER, Brubaker L, Gai X, Wolfe AJ, Schreckenberger PC. 2014. Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *J Clin Microbiol* 52:871–876. <https://doi.org/10.1128/JCM.02876-13>
- Szlachta-McGinn A, Douglass KM, Chung UYR, Jackson NJ, Nickel JC, Ackerman AL. 2022. Molecular diagnostic methods versus conventional urine culture for diagnosis and treatment of urinary tract infection: a systematic review and meta-analysis. *Eur Urol Open Sci* 44:113–124. <https://doi.org/10.1016/j.euro.2022.08.009>
- Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, Khan AM, Woodford N, Saunders NJ, Wain J, O’Grady J, Livermore DM. 2017. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. *J Antimicrob Chemother* 72:104–114. <https://doi.org/10.1093/jac/dkw397>
- Lagier J-C, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, Levasseur A, Rolain J-M, Fournier P-E, Raoult D. 2018. Culturing the human microbiota and culturomics. *Nat Rev Microbiol* 16:540–550. <https://doi.org/10.1038/s41579-018-0041-0>
- Eberly AR, Beebout CJ, Carmen Tong CM, Horn GT, Green HD, Fitzgerald MJ, De S, Apple EK, Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA, Clayton DB, Stratton CW, Schmitz JE, Hadjifrangiskou M. 2020. Defining a molecular signature for uropathogenic versus urocolonizing *Escherichia coli*: the status of the field and new clinical opportunities. *J Mol Biol* 432:786–804. <https://doi.org/10.1016/j.jmb.2019.11.008>
- Minich JJ, Sanders JG, Amir A, Humphrey G, Gilbert JA, Knight R. 2019. Quantifying and understanding well-to-well contamination in microbiome research. *mSystems* 4:e00186-19. <https://doi.org/10.1128/mSystems.00186-19>
- Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
- Karstens L, Asquith M, Caruso V, Rosenbaum JT, Fair DA, Braun J, Gregory WT, Nardos R, McWeeney SK. 2018. Community profiling of the urinary microbiota: considerations for low-biomass samples. *Nat Rev Urol* 15:735–749. <https://doi.org/10.1038/s41585-018-0104-z>
- Ackerman AL, Anger JT, Khalique MU, Ackerman JE, Tang J, Kim J, Underhill DM, Freeman MR, NIH Multidisciplinary Approach to the Study of Chronic Pelvic Pain (MAPP). 2019. Optimization of DNA extraction from human urinary samples for microbiome community profiling. *PLoS One* 14:e0210306. <https://doi.org/10.1371/journal.pone.0210306>
- Munch MM, Chambers LC, Manhart LE, Domogala D, Lopez A, Fredricks DN, Srinivasan S. 2019. Optimizing bacterial DNA extraction in urine. *PLoS One* 14:e0222962. <https://doi.org/10.1371/journal.pone.0222962>
- Nearing JT, Comeau AM, Langille MGI. 2021. Identifying biases and their potential solutions in human microbiome studies. *Microbiome* 9:113. <https://doi.org/10.1186/s40168-021-01059-0>
- Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, Braun J, McWeeney S. 2019. Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *mSystems* 4:e00290-19. <https://doi.org/10.1128/mSystems.00290-19>
- Chan-Tack KM, Trautner BW, Morgan DJ. 2020. The varying specificity of urine cultures in different populations. *Infect Control Hosp Epidemiol* 41:489–491. <https://doi.org/10.1017/ice.2020.16>
- Helbig S, Achkar JM, Jain N, Wang X, Gialanella P, Levi M, Fries BC. 2013. Diagnosis and inflammatory response of patients with candiduria. *Mycoses* 56:61–69. <https://doi.org/10.1111/j.1439-0507.2012.02201.x>
- Durkin MJ, Keller M, Butler AM, Kwon JH, Dubberke ER, Miller AC, Polgreen PM, Olsen MA. 2018. An assessment of inappropriate antibiotic

- use and guideline adherence for uncomplicated urinary tract infections. *Open Forum Infect Dis* 5:ofy198. <https://doi.org/10.1093/ofid/ofy198>
30. Shafrin J, Marijam A, Joshi AV, Mitrani-Gold FS, Everson K, Tuly R, Rosenquist P, Gillam M, Ruiz ME. 2022. Impact of suboptimal or inappropriate treatment on healthcare resource use and cost among patients with uncomplicated urinary tract infection: an analysis of integrated delivery network electronic health records. *Antimicrob Resist Infect Control* 11:133. <https://doi.org/10.1186/s13756-022-01170-3>
31. Festa RA, Luke N, Mathur M, Parnell L, Wang D, Zhao X, Magallon J, Remedios-Chan M, Nguyen J, Cho T, Ngo A, Murphy M, Baunoch D. 2023. A test combining multiplex-PCR with pooled antibiotic susceptibility testing has high correlation with expanded urine culture for detection of live bacteria in urine samples of suspected UTI patients. *Diagn Microbiol Infect Dis* 107:116015. <https://doi.org/10.1016/j.diagmicrobio.2023.116015>
32. Jia K, Huang S, Shen C, Li H, Zhang Z, Wang L, Zhao G, Wu Z, Lin Y, Xia H, Tang M, Yang H, Hu H. 2023. Enhancing urinary tract infection diagnosis for negative culture patients with metagenomic next-generation sequencing (mNGS). *Front Cell Infect Microbiol* 13:1119020. <https://doi.org/10.3389/fcimb.2023.1119020>
33. Moustafa A, Li W, Singh H, Moncera KJ, Torralba MG, Yu Y, Manuel O, Biggs W, Venter JC, Nelson KE, Pieper R, Telenti A. 2018. Microbial metagenome of urinary tract infection. *Sci Rep* 8:4333. <https://doi.org/10.1038/s41598-018-22660-8>
34. McDonald M, Kameh D, Johnson ME, Johansen TEB, Albala D, Mouraviev V. 2017. A head-to-head comparative phase II study of standard urine culture and sensitivity versus DNA next-generation sequencing testing for urinary tract infections. *Rev Urol* 19:213–220. <https://doi.org/10.3909/riu0780>
35. Neugent ML, Hulyalkar NV, Nguyen VH, Zimmern PE, De Nisco NJ. 2020. Advances in understanding the human urinary microbiome and its potential role in urinary tract infection. *mBio* 11:e00218-20. <https://doi.org/10.1128/mBio.00218-20>