Fast identification of *Escherichia coli* in urinary tract infections using a virulence gene based PCR approach in a novel thermal cycler

Jolanda K. Brons\(^a\)*, Stefanie N. Vink\(^a\), Marjon G.J. de Vos\(^a\), Stefan Reuter\(^b\), Ulrich Dobrindt\(^c\), Jan Dirk van Elsas\(^a,\)*

\(^a\) Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands
\(^b\) Department of Medicine D, Division of General Internal Medicine, Nephrology and Rheumatology, University Hospital of Münster, Münster, Germany
\(^c\) Institute of Hygiene, University of Münster, Münster, Germany

**ARTICLE INFO**

**Keywords:**

Uropathogenic *Escherichia coli* (UPEC)

(Fast) PCR (identification)

Virulence genes

Urinary tract infection (UTI)

Novel fast thermal cycler (Nextgen PCR thermal cycler)

**ABSTRACT**

Uropathogenic *Escherichia coli* (UPEC) is the most common causal agent of urinary tract infections (UTIs) in humans. Currently, clinical detection methods take hours (dipsticks) to days (culturing methods), limiting rapid intervention. As an alternative, the use of molecular methods could improve speed and accuracy, but their applicability is complicated by high genomic variability within UPEC strains. Here, we describe a novel PCR-based method for the identification of *E. coli* in urine. Based on *in silico* screening of UPEC genomes, we selected three UPEC-specific genes predicted to be involved in pathogenesis (*c3509, c3686, yrbH*) and (*chuA*), and one *E. coli* specific marker gene (*uidA*). We validated the method on 128 clinical (UTI) strains. Despite differential occurrences of these genes in uropathogenic *E. coli*, the method, when using multi-gene combinations, specifically detected the target organism across all samples. The lower detection limit, assessed with model UPEC strains, was approximately 10⁴ CFU/ml. Additionally, the use of this method in a novel ultrafast PCR thermal cycler (Nextgen PCR) allowed a detection time from urine sampling to identification of only 52 min. This is the first study that uses such defined sets of marker genes for the detection of *E. coli* in UTIs. In addition, we are the first to demonstrate the potential of the Nextgen thermal cycler. Our *E. coli* identification method has the potential to be a rapid, reliable and inexpensive alternative for traditional methods.

1. Introduction

Urinary tract infections (UTIs) affect up to 150 million people worldwide each year (Harding and Ronald, 1994). UTIs are caused by both Gram-negative and Gram-positive bacteria, as well as yeasts such as *Candida* spp. Uropathogenic *Escherichia coli* (UPEC) is the most common etiologic agent, which is currently responsible for approximately 90% of all globally occurring UTIs (Zhang and Foxman, 2003). Clinically, a UTI is defined as a bacteriuria with ≥10⁵ bacterial (and/or yeast) colony forming units per ml (CFU/ml) midstream urine, in combination with clinical symptoms (Sheffield and Cunningham, 2005). UTIs are classified as either “uncomplicated” or “complicated”. Uncomplicated UTIs affect individuals (mostly females) that have no structural or neurological urinary tract (UT) abnormalities, while complicated UTIs affect individuals with underlying problems, such as UT abnormalities, kidney transplantation and/or catheterization (Flores-Mireles et al., 2015). In terms of clinical presentation, UTIs are classified as either cystitis (CY, infection of the bladder/lower urinary tract), pyelonephritis (PY, infection of the kidney/upper urinary tract) and/or urosepsis (US, UTI with sepsis). Asymptomatic bacteriuria (ASB) is not considered as an infection, but represents a risk factor in certain circumstances, e.g. in pregnancy and before traumatizing interventions of the urinary tract (Johansen et al., 2011).

Commensal *E. coli* strains are part of the normal gut microbiome and rarely cause disease in the gut or in the UT. However, particular *E. coli* strains have evolved, presumably via horizontal gene transfer (HGT), into pathogens that cause various diseases, including UTIs (Jacobsen et al., 2008). Based on its ecology, *E. coli* can be classified into three groups: (i) commensal (encompassing beneficial colonizers of the gastrointestinal tract), (ii) intestinal pathogenic (enteric disease or diarrheagenic) and (iii) extra-intestinal pathogenic (ExPEC, among them UPEC) strains (Russo and Johnson, 2000). Eight phylogenetic groups of *E. coli* are recognized, seven (A, B1, B2, C, D, E, F) belonging to *E. coli sensu stricto*, and the eighth constituting the so-called *Escherichia* cryptic clade I (Clermont et al., 2013). Most UPEC strains are members of the B2 group, although uropathogens have also been found within the other
phylogenetic groups. Human commensal strains mainly belong to *E. coli* phylogenetic group A (Johnson and Stell, 2000).

Genes involved in the colonization by *E. coli* of the UT, and subsequent pathogenesis, may provide the key to detect UPECs. Pathogenesis is thought to commence with colonization of the urethra, followed by ascension of bacteria into the bladder and growth in the urine. UPEC cells then increasingly adhere to the bladder surface, interacting with the epithelial defense system. Eventually, biofilms are formed. Then, invasion (through replication and formation of intracellular bacterial communities) occurs and UPEC cells may travel upwards to colonize the kidneys, where they may cause tissue damage with increased risk of sepsis (Terlizzi et al., 2017).

The infective process is known to depend on a suite of bacterial traits ("virulence factors"), which - together - determine the sequence of events leading to progressive disease. Most of these traits have been found to be encoded on pathogenicity islands (PAIs) within the *E. coli* chromosome or, alternatively, on plasmids. Given their continuous acquisition (and occasional loss) of genetic material, the genomes of UPECs are generally larger than those of commensal *E. coli* strains (Hacker et al., 1997; Ahmed et al., 2008; Touchon et al., 2009). For example, the genomes of three well-known UPEC strains (CFT073, 536 and UTI89, all belonging to phylogenetic group B2), that constitute currently accepted reference strains for UT infections, contain 8–22% more open reading frames (and are, consequently, 6–13% larger) than the genome of the (commensal) K-12 reference strain MG1655 (Welch et al., 2002; Bruszukiewicz et al., 2006; Chen et al., 2006).

The distribution of genes for virulence traits in the genomes of UPEC strains is a key issue that will define the criteria for developing a system for identification that is based on their roles in disease development (Ahmed et al., 2008). UPECs exhibit a high degree of genetic diversity (primarily caused by the differential presence or absence of specific genes on PAIs) (Oelschlaeger et al., 2002). Key candidate genes for an identification system can be divided into two groups: (1) genes encoding factors associated with the bacterial cell surface, and (2) genes encoding factors that are secreted and transported to the site of interaction with the host cell (Emody et al., 2003). Both groups include genes encoding: (i) type-1 and -P, next to -S/F1C, fimbrial adhesins, (ii) toxins, i.e. cytotoxic necrotizing factor 1, α-hemolysin and secreted autotransporter toxins, (iii) host defense avoidance mechanisms, i.e. capsule or O-specific antigen, and (iv) iron acquisition systems, i.e. aerobactin, enterobactin, salmochelin and yersiniabactin (Lloyd et al., 2007).

The current "gold standard" for diagnosis of a UTI is detection of the pathogen concomitantly with the presence of clinical symptoms. Dipsticks that detect leukocyte esterase (LE) activity as an indicator of pyuria (condition of urine containing white blood cells or pus) and urinary nitrite (NIT) production as an indicator of bacteriuria are frequently used for presumptive diagnosis of UTIs (Semeniu and Church, 1999). When the dipstick test is positive and clinical symptoms are present, broad-spectrum antibiotics active against the presumed pathogen are prescribed until culture results are available. Pathogen identification typically takes 18–30 h, with antimicrobial susceptibility testing (AST) adding another 24–48 h. AST is performed as a phenotypic assay that measures bacterial growth in the presence of specific antimicrobial agents (Davenport et al., 2017). Treatments with broad-spectrum antibiotics lead to alterations of the gastrointestinal tract microbiome, and can promote the undesirable development of antibiotic-resistant microorganisms (Kostakiot et al., 2012). Thus, a rapid method for detection of UTI causal agents is required.

In a recent overview, Davenport et al. (2017) described current (new and developing) diagnostic technologies for UTIs, i.e. urinalysis and microscopy, MALDI-TOF mass spectrometry, fluorescent in situ hybridization (FISH), microfluidics, PCR, immunology-based and forward light scattering assays. However, most of these approaches have limiting factors such as: no pathogen identification and/or antimicrobial susceptibility testing, expensive, poor data from low concentrations of cells and poor sensitivities and specificities (Davenport et al., 2017).

PCR enables the detection of target genes with great speed, specificity and sensitivity (Nissen and Sluots, 2002). A number of (multiplex) PCR methods have thus far been described for the identification of UPECs. However, these approaches, based on genes such as *fimH* (Johnson and Stell, 2000; Padmavathy et al., 2012; Ren et al., 2016) and *rfbH* (Van der Zee et al., 2016), are not specific for UPECs, as commensal strains also harbour these genes. Other studies, based on the genes *ecp, fyuA, sfa*/ *focDE* (Blackburn et al., 2009; López-Banda et al., 2014) were found to suffer from similar drawbacks. Hence, identification of UPECs on the basis of PCR with genetic markers clearly requires improvement. To identify such UPEC-specific genetic markers, we examined a suite of currently available *E. coli* genomes for (1) the presence of virulence genes in UPECs and (2) the absence of these genes in commensal *E. coli* and other uropathogenic organisms. Based on this screening, we developed a PCR assay for the rapid identification of UPECs. We selected three UPEC-specific genes as targets, in addition to a generic *E. coli* marker gene.

2. Material and methods

2.1. Bacterial strains and reagents

Three reference UPEC strains, i.e. XPKO359, CAS874 (obtained from Prof. dr. J. Degener, University Medical Center Groningen, the Netherlands) and CFT073 (DSMZ 103538), and one commensal *E. coli* K-12 strain, i.e. MG1655 (DSMZ 18039) were used. All strains were grown in 20 ml Luria-Bertani (LB) broth (tryptone 10 g, yeast extract 5 g, NaCl 5 g, distilled water 1 l; pH 7.2 at 37 °C for 18–24 h). Following growth, cultures were centrifuged and the resulting pellets washed twice with Artificial Urine Medium (AUM) (Brooks and Keesil, 1997), after which the final pellets were dissolved in 1 ml AUM. In addition, AUM (19 ml) was spiked separately with 1 ml of the overnight-grown UPEC strains (10⁶ CFU/ml) XPKO359, CAS874 and CFT073.

2.2. In silico selection of UPEC/ *E. coli* marker genes and primer design

To select UPEC/ *E. coli* marker genes, we determined the presence of a suite of 131 genes potentially involved in virulence (exclusively present in UPEC strain CFT073 compared to commensal and fecal strains (Lloyd et al., 2007), in the genomes of 11 UPEC strains. The latter included seven group B2 strains, as well as one strain each of groups D, F, B1 and A, as follows: CFT073 (phylogenetic group B2), 536 (B2), UTI89 (B2), 26–1 (B2), NU14 (B2), ST131 strain EC958 (B2), NA114 (B2), UMN026 (D), IA139 (F), C15 (B1) and VR50 (A) (Table 1). Blastn was used, with a 95% identity cut-off (Altschul et al., 1990). To ensure *E. coli* specificity, and using the same Blastn method, we checked the occurrence of the same genes in three commensal *E. coli* strains (HS (A), K-12 MG1655 (A), K-12 W3110(A)) (Table 1), as well as in the uropathogens *Klebsiella pneumoniae* (taxid:573), *Staphylococcus saprophyticus* (taxid:29385), *Enterococcus* (taxid:1350), group B *Streptococcus* (taxid:1319), *Proteus mirabilis* (taxid:584), *Pseudomonas aeruginosa* (taxid:287), *Staphylococcus aureus* (taxid:1280), and *Candida* (taxid:1535326).

Genes that occurred in only one or two of the UPEC phylogenetic group(s) and in either commensal or non-*E. coli* species were excluded from further analyses. From the overall screening, three genes (potentially involved in the ecology of virulence) were chosen, that met the criteria to serve as targets in the detection system: c3509, c3686 (yrhH) and chuA. The *E. coli* specific gene, uidA, was added as a control/reference gene (Kiel et al., 2018). Thus, a four-gene detection method was created in silico. We designed primers based on all available genetic information for each gene in Clone Manager 6, using as criteria: target region of 190–260 bp, primer length of 19/ 20 bp, and calculated annealing temperature of 60 °C (Table 2). This ensured specificity for
UPECs (based on the genomes of the reference UPEC strains), as well as amplification at the selected annealing temperature.

2.3. E. coli virulence gene PCR optimization

PCR mixtures were composed as follows: 8 μl of 5 × KAPA2G Buffer A (KAPA Biosystems, Wilmington, United States), 60 nmol MgCl₂ (KAPA Biosystems), 12 nmol of each deoxyribonucleoside triphosphate (Sigma-Aldrich), 40 pmol of each primer, 10% dimethylsulfoxide (DMSO) and 1 U of (5 U/μL) KAPA2G Fast HotStart DNA polymerase (KAPA Biosystems), were combined with Ambion nuclease free water (Thermo Fisher Scientific, Waltham, United States) to 40 μl in a 0.2-ml microfuge tube. The DNA template was 1 μl of either: the spiked AUM samples, DNA extracted from these spiked AUM samples or the clinical UPEC DNA. The mixture was split in two (20ul each), one mixture was incubated in a Mastercycler Nexus PCR thermal cycler (Eppendorf, Hamburg, Germany), and the other in a Nextgen PCR thermal cycler (Molecular Biology Systems, Goes, the Netherlands), programmed for the four marker genes as follows: initial denaturation of double-stranded DNA for 30 s. at 95 °C; 30 cycles consisting of 7 s. at 95 °C, 12 s. at 60 °C, and 12 s. at 75 °C; and extension for 30 s. at 75 °C. Cycling conditions were similar for the two thermal cyclers. The PCR program in the regular thermal cycler takes 51 min, while the Nextgen PCR thermal cycler can perform this PCR in 16 min. All amplification products were analyzed by electrophoresis in 1.0% (wt/vol) agarose gels, followed by ethidium bromide staining (1.2 mg/l ethidium bromide in 1× Tris-acetate-EDTA) (Mullis, 1990; Sambrook and Fritsch, 1989), gels, followed by ethidium bromide staining (1.2 mg/l ethidium bromide in 1× Tris-acetate-EDTA) (Mullis, 1990; Sambrook and Fritsch, 1989), and the pellet dissolved in 100 μl sterile dH₂O; then, 500 μl of PB buffer was added and the mixture was vortexed for 30 s. The total volume was transferred to a column and centrifuged for 30 s. at 10,000 x g. The flow through was discarded and the column was washed with 600 μl PE buffer. The DNA was eluted in 50 μl EB buffer (10,000 x g for 30 s.). The total time of extraction was 4.5 min.

2.6. Adaptation of the fast DNA extraction method

Strain CFT073 was grown overnight in LB medium at 37 °C, after which the cells were centrifuged and washed two times with AUM to remove any LB residue. Dilution series were made with AUM, after which PCR (using primer set c3686) was performed on samples processed as follows: (i) Deelman DNA extraction method (3 min), (ii) modified Deelman DNA extraction method (4.5 min), (iii) Qiagen DNeasy ultraclean microbial kit (21 min), and (iv) no DNA extraction (cells in urine, 0 min). Dilution plating was performed to establish the number of CFU/ml in the systems (Hoben and Somasegaran, 1982).

2.7. Determining the lower detection limit of target genes in reference strain samples

The overnight LB-grown E. coli reference strains (XPko359, CASC874, CFT073) were centrifuged and washed two times with AUM to remove any LB medium residue, after which dilution series (10⁰–10⁴ CFU/ml) were made in sterile AUM. Dilution plating was performed to determine the CFU densities in the starting culture (Hoben and Somasegaran, 1982).
and Somasegaran, 1982). PCR was performed with DNA extracted from these dilutions (10^9–10^1 CFU/ml) using the fast DNA extraction method, in both the regular PCR thermal cycler and the fast PCR thermal cycler (Nextgent PCR thermal cycler). Here, all four primer sets (c3509, c3686, chuA, uidA) were examined to determine the lower detection limits of the E. coli PCR.

2.8. Origin of clinical UPEC samples

The novel E. coli PCR-based detection system was validated on genomic DNA isolated from clinical UPEC isolates collected at the University Hospital of Münster, Münster, Germany. These clinical samples were subcultured and confirmed as E. coli by MALDI-TOF (Caprioli et al., 1997). The DNA was isolated, from pure cultures grown overnight in LB medium, by using the MagAttract HMW kit from Qiagen according to the manufacturer’s recommendations (performed in Münster, Germany). A total of 128 UPEC strains isolated from female patients (age 19–89 years, median age 54.4) with an uncomplicated UTI (n = 70) and patients with a complicated UTI (kidney transplantation, n = 58) was tested.

2.9. Novel fast PCR thermal cycler (Nextgent PCR thermal cycler)

The detection method was tested in both a regular PCR thermal cycler (Mastercycler Nexus PCR thermal cycler, Eppendorf, Germany) and the novel fast PCR thermal cycler (Nextgent PCR thermal cycler, Molecular Biology Systems (MBS), Goes, the Netherlands). The Nextgent PCR thermal cycler operates with six aluminum heat blocks, in three zones, and is programmable to the desired temperature zone for denaturing, annealing and extension. A thin (50 μM) polypropylene PCR plate (96 or 384 wells) is transferred to the desired temperature zone in 0.1 s, and clamped between the two aluminum blocks in that zone. Due to this set-up, there is no ramping time and thus the time it takes to complete a PCR is strongly reduced.

3. Results

3.1. In silico selection of UPEC/ E. coli marker genes

The genomes of 11 selected UPEC strains were examined with respect to their distinction from those of three commensal E. coli strains. We specifically screened for the presence/absence of 131 genes that had been reported to be exclusively present in a set of UPEC strains as compared to commensal and fecal E. coli strains (Lloyd et al., 2007). In addition, we screened for the absence of these 131 genes in the other causal agents of UTIs: Klebsiella pneumoniae, Proteus mirabilis, Pseudomonas aeruginosa, Enterococcus sp., group B Streptococcus (GBS), Staphylococcus aureus, S. saprophyticus and Candida sp. (Fig. 1).

We used three criteria to select for UPEC-specific genes from the set of 131 genes identified by Lloyd et al. (2007): 1) they should not occur in commensal E. coli strains and/or non-E. coli uropathogens, 2) they should occur in phylogenetic group B2 and in at least two out of four additional phylogenetic groups (among groups D, F, B1 and A), and 3) the genes should not be co-localized in the same genomic region. This third criterion was used in order to prevent covariant behavior of the marker genes. Thirty-four genes were discarded based on the first criterion, and an additional 80 based on criterion 2. Thus, 17 genes were identified that were present in at least three phylogenetic groups, including group B2. These 17 genes were found to be distributed over eight different regions within the CFT073 reference genome. Based on this scattered occurrence, three of these 17 genes, i.e. c3509, c3686 (yrbH) and chuA (Lloyd et al., 2007) were selected to serve as target genes for the UPEC identification system. By predicted function, all three genes could be tentatively linked to processes associated with bacterial virulence, as outlined hereunder.

Gene c3509, located on pathogenicity island PAI-CFT073-metV, was predicted to encode a putative ATP-binding protein of an ABC transport system. It was present in the genomes of all UPECs from the phylogenetic groups screened, i.e. B2, D, F, B1 and A. The second gene, c3686, located on pathogenicity island PAI-CFT073-pheV, was predicted to encode a D-arabinose 5-phosphate isomerase (API). It was present in UPECs of phylogenetic groups B2, D, F and A, but absent from the B1 group E. coli CI5 strain. The third gene, chuA, was predicted to encode an outer membrane heme/ hemoglobin receptor. It was present in the genomes of UPECs from phylogenetic groups B2, D and F. However, it was absent from E. coli strains CI5 (group B1) and VR50 (group A). Based on these findings, we hypothesized that these three genes, in combination, would cover key virulence determinants across a large majority of clinically relevant UPECs. An E. coli-specific marker gene, uidA, was included in order to confirm the identity of the UTI agent, as well as to serve as an internal control (Kiel et al., 2018).

3.2. Development of a PCR-based system for identification of uropathogenic E. coli in UTIs

Primers were designed to amplify the target genes in E. coli of all phylogenetic groups where these were present. Table 2 gives an overview of the selected primers and their amplification specificities. All selected primer combinations were tested, and PCR conditions optimized, using DNA of reference B2 group UPEC strain CFT073. As a negative control, DNA from the commensal E. coli K-12 strain MG1655 (phylogenetic group A) was used. On the basis of DNA from strain CFT073, the PCR amplicons were of the expected sizes for all four primer sets used. Thus, amplicons of sizes of, respectively, 208 (c3509), 259 (c3686), 221 (chuA) and 259 bp (uidA) were detected, and no side products were observed (Supplementary Fig. S1A). The E. coli K-12 strain MG1655 was negative for all three UPEC-specific virulence genes, and, as intended, did yield amplicons using the uidA primers.

3.3. Adaptation of a DNA extraction method from urine

Theoretically, a PCR is productive if at least a single copy of a target gene region is present, provided this region is sufficiently available for primers to anneal and kick-start the chain reaction. Since the area and/or other compounds that are present in urine can inhibit polymerases (Schrader et al., 2012), we tested the PCR in the presence of AUM (Brooks and Keevil, 1997). We used primer set c3686 on DNA from strain CFT073 obtained using three different DNA extraction/purification methods (Deelman DNA extraction method, modified Deelman DNA extraction method, and reference Qiagen DNeasy ultraclean microbial kit), versus direct detection on a cell culture in AUM (Brooks and Keevil, 1997). The Deelman DNA extraction method enabled detection of the target from preparations down to 1.1 × 10^5 CFU/ml (Fig. 2A), whereas the modified Deelman DNA extraction method was more sensitive, giving a detection limit of 1.1 × 10^3 CFU/ml (Fig. 2B). The reference Qiagen DNeasy ultraclean microbial kit required an input of 1.1 × 10^4 CFU/ml for positive detection (Fig. 2C). In contrast, PCR amplification directly from cells (no DNA extraction) only allowed detection of cell numbers of 1.1 × 10^5 CFU/ml and up (Fig. 2D). The modification of the Deelman DNA extraction method thus enhanced the sensitivity approximately 100-fold, as compared to detection from cells. An additional advantage of the modification was that the extraction could be performed in 4.5 min (see Materials and Methods).

3.4. Lower detection limits of target genes in the genomes of reference strains

To examine the breadth of the multi-gene E. coli identification system, we tested the four primer sets (c3509, c3686, chuA, and uidA) on DNA from dilution series of three selected UPEC strains CASC874, CFT073 and XPKO359 in AUM (10^2–10^5 CFU/ml), using the modified Deelman DNA extraction method. All PCR systems consistently yielded...
Fig. 1. Presence/absence analysis and potential function of 131 selected genes across E. coli and non-E. coli (uropathogenic) bacterial strains. Black: gene present, white: gene absent. Strains were: 3 commensal E. coli (HS, K-12 MG1655, K-12 W3110), 11 UPEC (CFT073, 536, UTI89, 26–1, NU14, ST131 strain EC958, NA114, UMN026, IAI39, CI5, and VR50) and 8 non-E. coli uropathogenic strains (Klebsiella pneumoniae, Staphylococcus saprophyticus, Enterococcus sp., group B Streptococcus (GBS), Proteus mirabilis, Pseudomonas aeruginosa, Staphylococcus aureus, and Candida sp.). Genes were detected using Blastn with a 95% identity cut-off.
amplicons of the expected sizes, as evidenced from gel electrophoresis. The sole exception was the chuA gene in strain XPKO359, which yielded no amplicons. Across all four target gene sources, the c3509 primer set had a lower detection limit in the 10^{4} CFU/ml range, whereas the other three primer sets had lower detection limits ranging from about 1.1 × 10^{3} to 7.7 × 10^{4} CFU/ml (Fig. 3A and Supplementary Table 1A).

3.5. Validation on clinical strains

To validate the identification method, we screened the DNA produced from 128 recently isolated E. coli clinical strains from female patients with diagnosed UTIs for the presence of the selected genes. All strains were classified as belonging to the E. coli phylogenetic groups A, B1, B2, D, F and clade V. Specifically, 47.7% (61/128) belonged to group B2, 21.9% (28/128) to A, 11.7% (15/128) to D, 10.2% (13/128) to B1, 7.0% (9/128) to F and 1.6% (2/128) to clade V. Moreover, the strains originated from 54.7% (70/128) uncomplicated versus 45.3% (58/128) complicated UTIs (kidney transplantation). The clinical pictures were diverse, i.e. ASB (39.8%, 51/128), CY (45.3%, 58/128), PY (12.5%, 16/128), and US (2.3%, 3/128) (Fig. 4).

The E. coli marker gene uidA was present in 97.7% of all strains. The remaining 2.3% all belonged to phylogenetic group F, which is known to be quite divergent from all other E. coli groups (Johnson et al., 2017). An analysis of the occurrence of the c3509, c3686 and chuA genes separately across the 128 samples revealed the following order of prevalence: c3509 (76.6%), c3686 (75.0%) and chuA (72.7%) (Table 3).

The prevalence of each gene as well as the combination of genes within the strain set was then analyzed with respect to the phylogenetic group of the strains (Table 3). All group B1 and B2 strains possessed the c3509 gene (13/13 and 61/61, respectively), while its prevalence was lower in groups D (66.7%; 10/15), F (55.6%; 5/9), A (32.1%; 9/28), and clade V (0%; 0/2). The c3686 gene was found in the clade V samples (100%; 2/2), followed by group B2 (91.8%; 56/61), F (88.9%; 8/9), D (86.7%; 13/15), A (42.9%; 12/28) and B1 (38.5%; 5/13). The chuA gene was present in all group B2, D and F samples (61/61, 15/15, 9/9), with lower occurrence in clade V (50%; 1/2) and groups A (21.4%; 6/28) and B1 (7.7%; 1/13). Group A, B1, B2, clade V and D samples all possessed the uidA gene. In group F, it occurred in 66.7% of the samples (6/9).

We used the presence of any one of the three marker genes in a given sample as the criterion for positive identification of UPEC. This assessment enabled positive UPEC identification in 94.5% (121/128) of the E. coli samples (Table 3). The seven (5.5%) samples that did not produce amplicons with any of the three target genes were all part of phylogenetic group A (which encompasses mainly commensal strains). Inclusion of the uidA marker gene (Table 3) resulted in a positive E. coli detection in 100% (128/128) of the cases.

3.6. Operational improvement of the identification method using a novel (Nextgen) PCR thermal cycler

Using DNA of reference strain CFT073, we compared the detection method on the fast Nextgen PCR thermal cycler versus the previously used regular PCR cycler. A first, side-by-side, test performed with pure DNA extracted from strain CFT073 versus strain MG1655 revealed that the detection was similar between the two machines, given that all three primer sets yielded the expected amplicons at similar target gene levels (Supplementary Fig. S1B).

We then tested the limits of detection of the novel method in the fast Nextgen PCR thermal cycler on DNA from UPEC strains CASC874, CFT073 and XPKO359 in AUM (10^{9} down to 10^{1} CFU/ml, using the modified Deelman DNA extraction method) (Deelman et al., 2012) (Fig. 3B and Supplementary Table 1B). Gene c3509 had a detection limit in the 10^{5} CFU/ml range, while genes c3686, chuA, and uidA had limits of detection ranging from about 1.1 × 10^{3} to 7.7 × 10^{4} CFU/ml. The detection limits of the amplification systems were grossly similar across the machines, except for primer set c3509, in which the regular PCR thermal cycler gave 10-fold higher sensitivity. When tested on all 128 clinical samples, all target gene regions were amplified similarly across the two machines (Fig. 4). Clearly, the main difference between the two machines was in the amplification time, with the Nextgen thermal cycler being considerably faster (16 min.) than the regular thermal cycler (51 min.).

4. Discussion

To develop a fast E. coli identification system, we initially performed an in silico screening of the genomes of a suite of reference UPEC strains. This revealed the presence, in a mosaic fashion, of potential target genes...
genes across these strains. Such mosaic occurrence is a token of the complexity of gene movements that have shaped the emerged UPECs over time, and highlights the necessity of dealing with this variability across the genomes of the target organisms when developing a detection method like the one proposed here. The result of the screening led to the selection of three UPEC-specific genes (c3509, c3686, and chuA) that are potentially associated with virulence. In fact, all three may even be crucial in UTI pathogenesis (Lloyd et al., 2007). In addition, we included gene uidA, to allow for generic E. coli identification (Kiel et al., 2018). Optimization was performed using DNA from the UPEC reference strain CFT073 (phylogenetic group B2), versus that of the commensal K-12 strain MG1655 (group A). Indeed, the selected genes c3509, c3686 and chuA were present in the UPEC strain, but absent from the commensal strain. This constituted a first token of evidence for their specificity for UPECs, which was confirmed in the extended comparative screens.

A fast E. coli identification method requires efficient DNA extraction and purification from urine. Urea is a very critical PCR inhibitor, as it incites degradation of, or damage to, the DNA polymerase used in PCR (Schrader et al., 2012). We tested our DNA analysis method with similar concentrations of urea as present in urine (10 g/L, in urine 9.3–23.3 g/L (Rose et al., 2015). Our modified DNA extraction method enabled a very fast removal of these critical compounds, resulting in a 100-fold improvement of detection of targets, corresponding to 10^3 CFU/ml (based on primer set c3686 on strain CFT073 DNA), compared to 10^5 CFU/ml for the original method (Deelman et al., 2012). Moreover, our optimized method enabled DNA extraction in just 4.5 min and allowed identification of all target genes at 10^3/10^4 CFU/ml.

To examine the potential association of selected genes to virulence in UTIs, we considered the predicted function of the proteins encoded by these genes. The c3509 gene encodes a putative ATP-binding protein of an ABC transport system, involved in the transport of sugars, metals, peptides, amino acids and/or other metabolites, across the membrane. The system is driven by ATP binding and hydrolysis, empowering the translocation of substrates across the membrane (Tanaka et al., 2018). In CFT073, the c3509 gene is located only five genes upstream of a mannitol phosphotransferase system gene, and so it might be involved in mannitol uptake. Mannitol is present in carrots, apples, pineapples and asparagus, and is commonly found in human urine because of its poor absorption by the human gastrointestinal tract (Bouatra et al., 2013). Since gene c3509 was present across all 11 screened UPEC genomes as well as in a majority of the clinical samples (76.6%), it may be important for UPEC metabolism in the UT and thus fitness.

The c3686 (yrbH) gene is predicted to encode a D-arabinose 5-phosphate isomerase (API) (Meredith and Woodard, 2003). API is the first enzyme in the biosynthesis of 3-deoxy-D-manno octulosonate (KDO), a sugar moiety located in the lipopolysaccharide (LPS) layer of Gram-negative bacteria. LPS in UPECs is important in the activation of proinflammatory responses in UTIs (Bien et al., 2012). Invasion of bladder epithelial cells by UPEC stimulates a response via an LPS-modulated mechanism (Schilling et al., 2001). Interestingly, LPS released by UPECs can also subvert host defenses, the invader escaping into the host cell cytoplasm, forming intracellular bacterial communities (Flores-Mireles et al., 2015). It might be that the c3686 gene has a role in the (over) production and eventual release of LPS, thus helping to overcome the host defense in UPEC; c3686 could therefore encode an important virulence factor.

The chuA gene encodes an outer membrane receptor protein, which may be involved in the uptake of compounds like heme. This gene is
part of the genetic locus encoding heme transport (potentially importing iron), which appears to be widely distributed among pathogenic E. coli strains (Torres and Payne, 1997; Wyckoff et al., 1998; Nagy et al., 2001). The chuA gene also appears to be important during intracellular bacterial community formation by UPECs, with many biofilm-like properties. These intracellular biofilms allow establishment of a reservoir of dormant pathogen cells inside bladder epithelial cells, which helps these to outlast a host immune response (Anderson et al., 2004; Reigstad et al., 2007). Use of chuA in a detection system is thus warranted given its presumed role in the bladder during urinary tract infections.

As an internal amplification control, the E. coli specific beta-D-glucuronidase-encoding gene uidA was used. This gene encodes an enzyme specific to E. coli and is therefore widely used in identification kits and as a specific marker for E. coli (Cleuziat and Robert-Baudouy, 1990). It was present in 97.7% of the clinical samples; the three samples that were negative for gene uidA all occurred in phylogenetic group F. Although uidA is globally used as an E. coli marker, consistent with our findings, Johnson et al. (2017) showed that it is typically absent from strains of the E. coli sequence type 648 complex, which belongs to phylogroup F.

The premise behind our multi-gene identification method was that detection of any one of the three marker genes would indicate the presence of a UPEC. Thus, on the basis of only three selected genes, we were able to identify UPECs in 94.5% of the samples within the clinical dataset. The remaining 5.5% were samples that all contained strains belonging to phylogenetic group A. This phylogenetic group is strongly associated with commensal strains (Johnson and Stell, 2000), which generally have smaller genomes than the UPEC strains (Hacker et al., 1997; Ahmed et al., 2008; Touchon et al., 2009). Hence, it does not come as a surprise that these samples were negative for all three marker genes. We hypothesize that these group-A strains escape detection by (1) being inadvertent ‘passengers’ in the UT habitat, or (2) by virtue of possessing different pathogenicity gene sets. We deem it unlikely that such strains possess mutations in all three marker genes that prevent PCR-based detection. Clearly, they do belong to the species E. coli, since they were positive for the E. coli-specific uidA gene. Further research into their potential pathogenicity is warranted. Inclusion of the uidA marker gene in the identification method resulted in a 100% positive E. coli detection.

The lower detection limits of the amplification systems run on the Nextgen PCR thermal cycler were, overall, similar to those obtained with the regular PCR thermal cycler. The Nextgen PCR thermal cycler operates with three different heat blocks, eliminating ramping time and therefore total PCR time. The amplification data obtained with the clinical samples were similar across the two thermal cyclers. In both machines, the detection limit for primer set c3509 was slightly higher than that obtained with the other three target genes. Therefore, further optimization of the system based on primer set c3509 might result in binding that is more specific, consequently raising PCR efficiency and lowering the detection limit. Since this higher detection limit was only observed with primer set c3509, which was always detected in combination with one of the other genes (c3686, chuA, and uidA), the overall detection limit for the method was in reality at 10^4 CFU/ml. We conclude that, for the purpose of detecting UPECs in clinical samples, the novel fast PCR thermal cycler meets the standards of a regular thermal cycler, both in terms of specificity and sensitivity, with a total
time-to-detection of 52 min, which indeed enhances the speed of detection. At a detection limit of 10^4 CFU/ml, we are now able to detect the E. coli pathogen in all of the UTIs according to the ≥10^5 CFU/ml threshold described by Sheffield and Cunningham (2005). However, using the thresholds defined by Orenstein and Wong (1999), the method would not allow identification of UPECs in cases with lower CFU densities (acute cystitis in young men (< 10^4 CFU/ml), recurrent cystitis in young women (10^2 CFU/ml) and in catheter-associated early UTIs (10^2 CFU/ml). In those cases, a short (2-h) culturing step in LB, prior to extraction, might need to be performed.

Our novel method was able to detect E. coli of phylogenetic groups A, B1, B2, D, F, and clade V across diverse sample types, including a vast majority of uropathogenic E. coli. Moreover, the novel (Nextgen) PCR thermal cycler allowed ultrafast PCR amplification, in 16 min (total E. coli identification in 52 min, including DNA extraction, PCR, electrophoresis, staining, destaining and visualization under UV). The method even constitutes a basis for extension to other uropathogens. The Nextgen PCR thermal cycler clearly meets the standards of a regular PCR thermal cycler, greatly reducing the total time-to-detection. Hence, the here developed rapid molecular UTI detection method is a great asset to clinical labs worldwide.

**Declaration of Competing Interest**

There is no conflict of interest to declare.

**Acknowledgments**

We thank Prof. J. Degener (University Medical Center Groningen, the Netherlands) for providing the UPEC strains. Dr. Patricia Stevens is thanked for her preliminary work on E. coli pathogens.

**Appendix A. Supplementary data**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mimet.2019.105799.

**References**


