Comparison of multiplex PCR hybridization-based and singleplex real-time PCR-based assays for detection of low prevalence pathogens in spiked samples☆

Donna Hockman [a], Ming Dong [b], Hong Zheng [b], Sanjai Kumar [b], Matthew D. Huff [a], Elena Grigorenko [a], Maureen Beanan [c], Robert Duncan [b],*

[a] Diatherix Laboratories, LLC, Huntsville, AL, USA
[b] Laboratory of Emerging Pathogens, Division of Emerging and Transfusion Transmitted Diseases, Center for Biologics Evaluation and Research (CBER), Food and Drug Administration (FDA), Silver Spring, MD, USA
[c] Office of Biodefense, Research Resources, and Translational Research, Division of Microbiology and Infectious Diseases (DMID), National Institute of Allergy and Infectious Diseases (NIAID), Bethesda, MD, USA

ARTICLE INFO

ABSTRACT

Molecular diagnostic devices are increasingly finding utility in clinical laboratories. Demonstration of the effectiveness of these devices is dependent upon comparing results from clinical samples tested with the new device to an alternative testing method. The preparation of mock clinical specimens will be necessary for the validation of molecular diagnostic devices when a sufficient number of clinical specimens is unobtainable. Examples include rare pathogens, some of which are pathogens posing a biological weapon threat. Here we describe standardized steps for developers to follow for the culture and quantification of three organisms used to spike human whole blood to create mock specimens. The three organisms chosen for this study were the Live Vaccine Strain (LVS) of *Francisella tularensis*, surrogate for a potential biothreat pathogen, *Escherichia coli*, a representative Gram-negative bacterium and *Babesia microti* (Franca) Reichenow Peabody strain, representing a protozoan parasite. Mock specimens were prepared with blood from both healthy donors and donors with nonspecific symptoms including fever, malaise, and flu-like symptoms. There was no significant difference in detection results between the two groups for any pathogen. Testing of the mock samples was compared on two platforms, Target Enriched Multiplex-PCR (TEM-PCR™) and singleplex real-time PCR (RT-PCR). Results were reproducible on both platforms. The reproducibility demonstrated by obtaining the same results between two testing methods and between healthy and symptomatic mock specimens, indicates the standardized methods described for creating the mock specimens are valid and effective for evaluating diagnostic devices.

Published by Elsevier B.V.

1. Introduction

It has been more than two decades since the introduction of PCR (Mullis et al., 1986; Shampo and Kyle, 2002). Although this technology was implemented almost immediately in research laboratories, it has only been within the past decade that new diagnostic PCR-based technologies are increasingly being integrated into clinical microbiology laboratory practice (Sobel et al., 2008). These technologies enable the detection and quantification of pathogens with increased sensitivity and speed. The implementation of PCR-based nucleic acid tests in blood donor screening and for diagnosing infectious diseases has substantially improved the ability to obtain rapid, actionable information (Brittain-Long et al., 2011; Roth et al., 2012). There are numerous competitive nucleic acid based technologies that have been utilized for the detection and characterization of microorganisms. These technologies can be separated into several broad categories such as direct hybridization, nucleic acid amplification, and a variety of methods for post-amplification analysis such as sequencing, melt-curve analysis and others. Several methods combine the sensitivity of end-point PCR amplification with hybridization-based methodology, allowing the detection of multiple pathogens in one reaction (multiplex) which increases the throughput of the testing platform without compromising assay performance. This work is aimed at comparing the performance of multiplex end-point TEM-PCR coupled with hybridization-based detection to

singleplex RT-PCR, for detection of low prevalence pathogens spiked into human matrix such as whole blood.

The development of novel molecular devices for detection of emerging low prevalence pathogens such as *Francisella tularensis* and *Babesia microti* is hindered by the lack of clinical samples necessary to conduct clinical sensitivity studies required for U.S. Food and Drug Administration (FDA) clearance or approval (FDA/CDRH, 2014). The use of appropriate specimen matrix for assessment of analytical sensitivity is part of the regulatory requirements for laboratory developed tests or in vitro diagnostics (FDA/CDRH, 2014). Although sample transport medium or another simulated matrix can be used for demonstration of analytical sensitivity, analytical performance of molecular assays should be tested using pathogen-free clinical matrices (CLSI, 2008, FDA/CDRH, 2014). In addition to comparison of analytical sensitivity for multiplex and singleplex PCR tests, we compared human blood matrix collected from patients with fever or other signs of disease to blood collected from healthy donors because the composition of symptomatic blood can have an impact on analytical performance of both molecular methods that is different from the performance when testing healthy blood. The utilization of both testing platforms was recently demonstrated in work to standardize procedures for spiking low prevalence pathogens in human matrices (Dong et al., 2016). As in the first study, the crucial importance of this work is not so much the novelty of the methods described, but the standardization and validation of the methods for use by other investigators. By utilizing these standardized methods evaluation of new technology can be effectively compared to previously established platforms. This study expands upon previously described standardized procedures by developing methods for producing *E. coli*, *F. tularensis*, and *B. microti* mock specimens.

2. Materials and methods

The production of mock specimens for all three organisms shared a common work flow (Fig. 1). The organisms were cultured, aliquoted and cryopreserved. DNA extracted from the aliquots was serially diluted and tested with RT-PCR to determine the limiting dilution at which the sample could be detected, which was the basis for quantification of the cryopreserved aliquots used for spiking into blood.

2.1. Francisella tularensis (Live Vaccine Strain)

2.1.1. Source
The Live Vaccine Strain (ATCC # 29684) of *F. tularensis*, originally obtained from the American Type Culture Collection, was expanded in 5 mL culture and frozen as a master stock in the laboratory of Karen Elkins, PhD at the FDA. An aliquot of the master stock was spread on a plate and a single colony used to inoculate our culture.

2.1.2. Determination of the growth curve
The colony was inoculated into Mueller-Hinton Broth as previously described (Fortier et al., 1991); otherwise the procedure was the same as described for *E. coli*. In our culture, *F. tularensis* reached mid-log at OD_{600} of 1.0 in 14 h of incubation shaking (180–200 rpm) at 37 °C.

2.1.3. Preparation of cryopreserved aliquots
After harvesting the culture when it reached an OD_{600} of 1.0, the culture suspension was mixed with equal parts of a sterile solution of 2.6% (w/v) gelatin, and 500 μL aliquots were immediately frozen in dry ice and stored at −80 °C.

2.1.4. Determination of molecular pathogen concentration of frozen aliquots
The chosen method of quantification based on RT-PCR achieved high reproducibility and less variability among laboratories compared to culture based methods (Dong et al., 2016). The pathogen concentration was calculated by PCR detection of a limiting dilution of the nucleic acid extracted from the frozen aliquot and the concentration was expressed as PCR Detectible Units (PDU) per mL. PDU/mL is an emerging unit of measurement to replace genome copies/mL for evaluating diagnostic and blood screening devices (Anez et al., 2015).

We tested the frozen aliquots using an appropriate PCR primer and probe set to detect *F. tularensis*. Our assay was a Taqman style reaction
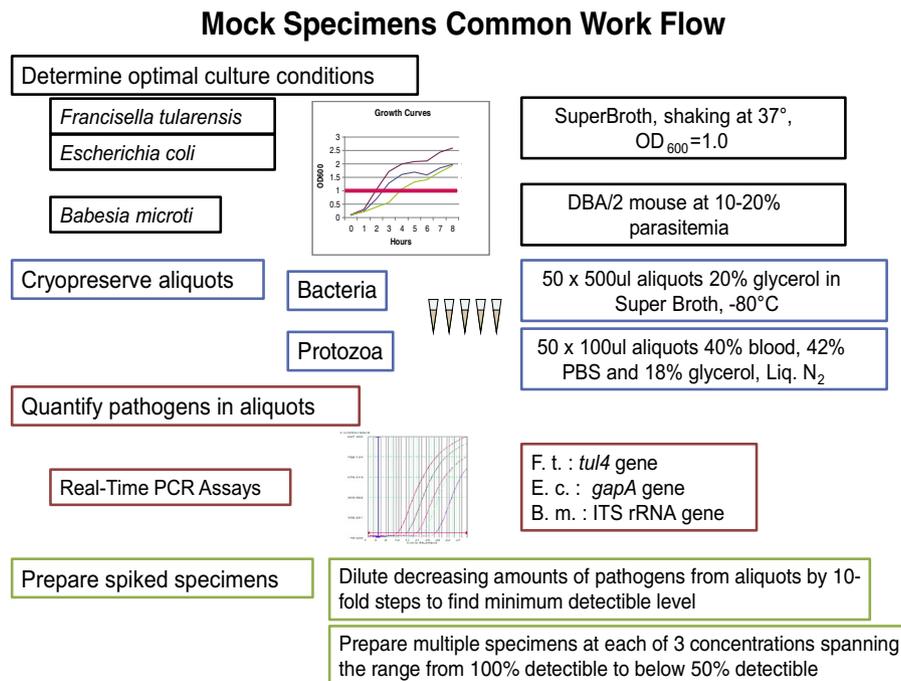
Mock Specimens Common Work Flow

Fig. 1. Graphic summary of the flow of tasks to prepare, qualify and use mock clinical (spiked) specimens for evaluation of diagnostic and blood screening devices for infectious agents. The steps proceed from top to bottom with examples of critical parameters in boxes to the right.

targeting the *tul4* gene (Table 1) which would successfully detect both the LVS and the fully virulent strains. In our study, the mean aliquot PDU/mL was 2.55×10^9. The variability among the aliquots was so little that the 95% confidence interval falls below the significant decimal places.

2.1.5. Preparation of spiked specimens

An initial study was performed to determine the detectible range in spiked whole blood specimens for a given extraction method and assay. An initial estimate was made from the range of concentrations detectible when the pure frozen aliquot was the starting material. Therefore, an appropriate quantity of a thawed aliquot was diluted in fresh human blood to a range of concentrations in PDU/mL that, based on previous analytical studies, was detectible by the diagnostic test and included some dilutions in the series that should be below the limit-of-detection for the diagnostic test. The process of dilution to the desired concentration involved a number of dilution steps from the highly concentrated frozen stock including a few dilution steps in an appropriate buffer such as Phosphate Buffered Saline (PBS). After the final step, a dilution used as a specimen had 99.9% of its volume comprised of blood to ensure sufficient correspondence to a clinical specimen. This was accomplished by making 2 ten-fold dilutions into blood before the final ten-fold dilution into the specimen. Three samples of each specimen in this initial dilution panel were then extracted using the Qiagen DNA Blood Mini Kit (catalog #51104, Qiagen, Valencia, CA) or its equivalent and an undiluted sample assayed by RT-PCR. These results indicated the detectible range near the limit of detection of the assay and platform. Further evaluation of the platform and assay was done with multiple specimens in the detectible range prepared from the additional frozen aliquots in the same manner.

The concentrations of spiked blood samples tested were 1×10^4 ($n = 10$), 1×10^3 ($n = 10$), and 1×10^2 PDU/mL ($n = 10$). Unspiked blood specimens were extracted in the same manner as negative controls ($n = 10$).

2.2. Escherichia coli (TOP10 strain)

2.2.1. Source

E. coli were obtained as One Shot® Competent cells (# C4040-10, Life Technologies, Carlsbad, CA).

2.2.2. Determination of the growth curve

All steps were performed using standard benchtop microbiology aseptic technique (Collins et al., 2004) following the methods described previously for other Gram-negative bacteria (Dong et al., 2016). We found the region of the growth curve where the slope of a line drawn through the plot of OD vs. time in incubation was constant, indicating a uniform growth rate. We determined the time in culture (2 h) and the optical density (O.D.$_{600}$ = 1.0) that was in the center of the linear portion of the growth curve (exponential phase).

2.2.3. Preparation of cryopreserved aliquots

Liquid culture harvested at the chosen mid-exponential growth stage was mixed with sterile 100% glycerol to make the suspension 20% (v/v) glycerol. With frequent mixing to maintain a uniform suspension, 0.5 mL per tube was aliquoted into sterile labeled tubes making at least 50 aliquots. Tubes were placed immediately in crushed dry ice and were stored tightly capped at −80 °C.

2.2.4. Determination of molecular pathogen concentration of frozen aliquots

The chosen method of quantification based on RT-PCR achieved high reproducibility and less variability among laboratories compared to culture based methods (Dong et al., 2016). The pathogen concentration was calculated by PCR detection of a limiting dilution of the nucleic acid extracted from the frozen aliquot and the concentration was expressed as PCR Detectible Units (PDU) per mL.

Steps included thawing a frozen aliquot, extracting total DNA from three separate 100 μL samples of the aliquot suspension by a chosen method (DNeasy kit, catalog #69504, Qiagen, Valencia, CA, following the manufacturer's protocol was used for this step with all bacteria and eukaryotes in this study). The extraction was repeated with at least three other frozen aliquots of *E. coli* from the same batch. A series of ten-fold serial dilutions was made for each extracted DNA sample. Quantitative RT-PCR was performed on all the extracts of all the aliquots in the range of dilutions that span from 100% detectible down to below detectability. Our assay was a Taqman style reaction targeting the *gapA* gene of all Gram-negative bacteria (Table 1). The lowest dilution detectible for each DNA extraction was used to express the PDU/mL in the frozen aliquot. The PDU/mL was calculated for four aliquots and the overall mean and standard deviation assigned as the characteristic of that set of frozen aliquots as described previously (Dong et al., 2016). For this batch of *E. coli* aliquots there was an average of 6.29×10^7 PDU/mL with 95% confidence interval $\pm 3.47 \times 10^3$.

2.2.5. Preparation of spiked specimens

Procedures for preparing spiked specimens were the same as described above for *F. tularensis*. Using the same process, a dilution series of *E. coli* in blood samples obtained from healthy donors was prepared and extracted. The dilution series consisted of *E. coli* at 1×10^3 PDU/mL ($n = 10$), 1×10^2 PDU/mL ($n = 10$) and 1×10^1 PDU/mL ($n = 10$) concentrations. Unspiked blood specimens were extracted in the same manner to be used as negative controls ($n = 10$). The specimens were tested by RT-PCR. To compare the methods on an alternative platform, the same samples, with coded labels, were prepared in the FDA laboratory and sent to the Diatherix laboratory to be tested by Target Enriched Multiplex-PCR (TEM-PCR). The RT-PCR and TEM-PCR platforms are described below.

To assess whether the diagnostic test had similar performance in symptomatic patients, an analysis was performed by spiking blood samples obtained from individuals with symptoms similar to the very early stage of an infection (fever, malaise or flu-like symptoms). The study was performed as described above using the same frozen aliquots of pathogen spiked in the same concentrations tested for

Table 1
RT-PCR primer and probe sequences.

Pathogen	Gene target		Sequence (5′-3′)
Babesia microti	ITS ribosomal RNA (Ohmori et al., 2011)	Forward	CCTGGGCCTCTGTGTTAGATCA
		Reverse	ACGTTGCAGCAGCCAAGAA
		Probe	CTAACGGGCATTGGGGGATTATACTGT
Francisella tularensis	*tul4* gene (Sjostedt et al., 1997; Tomioka et al., 2005)	Forward	GCTGTATCATCATTTAATAAACTGCTG
		Reverse	TTGGGAAGCTTGTATCATGGCACT
		Probe	GATTACAATGGCAGGCTCCAG
Gram-negative bacteria	gapA gene	Forward	TGATGGCCCGTCTCACAAAG
		Reverse	GCTTTAGCAGCACCGGTAGA
		Probe	ACTGGCGCGGCGGC

healthy donor specimens, the only difference being the source of donor matrix. Collection of both donor matrices was previously described (Dong et al., 2016).

2.3. Babesia microti

2.3.1. Source and determination of the growth curve

The *Babesia microti* (Franca) Reichenow Peabody strain can be cultured only in a living host. Infected mouse blood was obtained from the Kumar laboratory (FDA/CBER/DETTD, Silver Spring, MD) and detailed procedures were described elsewhere (Moitra et al., 2015). Briefly, DBA/2 mice are injected with a blood sample containing intraerythrocytic *Babesia microti* (ATCC PRA-99), allowed to infect until microscopic observation of a blood sample reveals 10–20% parasitemia.

2.3.2. Preparation of cryopreserved aliquots

A sample of Babesia-infected mouse blood was determined to have a concentration of 1000 parasites per microliter by microscopy. Fifty microliters of infected red blood cells (RBCs) were combined with 1950 μL of human whole blood and 3000 μL of freezing medium (30% v/v glycerol in PBS), mixed gently but thoroughly, aliquoted with 100 μL per vial, and immediately frozen in dry ice and stored in a liquid nitrogen freezer. The frozen aliquot contained a final estimated concentration of 1×10^4 parasites/mL.

2.3.3. Determination of molecular pathogen concentration of frozen aliquots

Steps included thawing a frozen aliquot and extracting total DNA with a Qiagen DNA Blood Mini Kit (catalog #51104, Qiagen, Valencia, CA) following the manufacturer's protocol. The average concentration was determined as above with a Taqman style reaction targeting the ITS in the ribosomal RNA gene (Table 1) to be $4.64 \times 10^4 \pm 1.94 \times 10^3$ PDU/mL.

2.3.4. Preparation of spiked specimens

Procedures for preparing spiked specimens were the same as described above for *E. coli* except that because the frozen aliquot was composed of whole blood at a lower pathogen concentration than other aliquots, fewer dilution steps were made to create the mock specimens and to span the 3-log range that is our standard. Consequently, sample preparation was slightly altered to achieve higher sensitivity. The alteration was to extract the sample with Qiagen DNA Blood Mini Kit by the same procedures as above except the column bound DNA was eluted in a smaller volume resulting in a more concentrated DNA sample. The concentrations of spiked blood samples were 2.5×10^2 ($n = 10$), 2.5×10^1 ($n = 10$), and 2.5 PDU/mL ($n = 10$). Negative controls were prepared from unspiked blood samples and the DNA eluted in the same volume as the Babesia spiked specimens to make comparable controls ($n = 10$).

2.3.5. RT-PCR detection of pathogens

The singleplex RT-PCR assays (Table 1) were evaluated with the DNA extracted from the human blood spiked with each pathogen at the 3-log range of concentrations. Specimens of pathogen-negative human blood were extracted as negative control samples. The cutoff of each PCR assay was established as 2 cycles less than the mean of the negative control Cq values. All reactions were performed with a Bio-Rad CFX96 RT-PCR System. The *E. coli*, *F. tularensis* and *B. microti* RT-PCR assays had PRIMER concentrations of 200 nmol/L in the final reaction volume and the reagent mix was Premix Ex Taq, catalog #RR039A, Takara Bio, Kusatsu, Shiga, Japan. Two replicate PCR reactions were run with the DNA extracted from each specimen and both were required to be over the threshold for the pathogen to be classified as "detected."

2.3.6. TEM-PCR™ panel (Diatherix Laboratories, LLC) for detection of low prevalence pathogens

The multiplex panel detects up to three pathogens in a single tube. There are two gene targets per pathogen and an internal control. The samples were tested in triplicate. Only one positive replicate from either gene target was required for the pathogen to be classified as "detected." The internal control must be positive for the negative pathogen results to be valid. Assay sequence information is shown in Tables 2 and 3.

PCR reactions for spiked blood samples contained 4 μL of template DNA at each concentration of pathogen, 2.4 μL of primer mix, and reagents from a Multiplex PCR Kit (Qiagen, Valencia, CA) according to the manufacturer's recommendations with total volume of 20 μL. TEM-PCR reactions took place in 96-well plates in a GeneAmp® PCR System 9700 thermocycler (ThermoFisher, Waltham, MA). Low concentration nested gene-specific primers were designed to enrich the targets during an enrichment PCR step. Each gene specific primer had a final reaction concentration of 20 nmol/L. Later, a pair of universal primers containing proprietary sequences was used to amplify all targets in a single PCR reaction. The reverse universal primer is labeled with biotin to facilitate the subsequent detection of target-specific amplicons hybridized with detection (probe) oligonucleotides coupled to Carboxyl Barcoded Magnetic Beads (BMBs) from Applied BioCode, Inc. (Santa Fe Springs, CA). The hybridization products are detected with streptavidin-phycoerythrin and results are reported as median fluorescence intensities for each detected target. BMB-probe coupling, BMB-probe mix preparation, hybridization of BMB-probes to amplicons, washes and detection were performed according to manufacturer's instructions (Applied BioCode Technical note: Protocols for Capture Probe Coupling, Hybridization with Biotinylated Target, and SA-PE Reaction).

A self-priming internal control oligonucleotide (proprietary component of the Diatherix TEM-PCR assay) was added to each reaction at 0.001 fmol/reaction. Failure to detect the internal control indicates a poor PCR or hybridization reaction. Threshold values were set for each target at ten times the Limit of Blank (LoB). The Limit of Blank was determined using results from spiked samples from healthy donors for which the analyte of interest was not present. The Limit of Blank was calculated to be: LoB = mean + 1.64 ∗ standard deviation (Burd, 2010).

2.3.7. Data organization & statistical analysis

Each pathogen was tested at specific high, medium, and low concentrations, described above for each pathogen, along with negative controls. Detection results from each pathogen at each concentration were combined into one "high concentration", one "medium concentration", one "low concentration", and one "negative control" group. The combined results were further aggregated into 4 additional groups; (1) all specimens that were tested with TEM-PCR, (2) all specimens that were tested with RT-PCR, (3) all healthy donor samples tested and (4) all symptomatic samples tested.

McNemar's test for paired samples with a Yates continuity correction of 1.0 was used to compute *p*-values for the null hypothesis that there are no differences in detection results between groups 1 and 2 and a two-tailed Fisher's exact test was used between groups 3 and 4 at high, medium, low, and negative control concentrations (Fisher, 1922; McDonald, 2014; McNemar, 1947). The reasoning behind the difference in statistical tests was because the samples that were tested between groups 1 and 2 were actually the same samples, but the specimens' volume were aliquoted and tested separately with each assay. Therefore, those samples are considered paired and McNemar's test applies (McNemar, 1947). The comparison between healthy and symptomatic donors included samples taken from different patients and are considered unpaired. Therefore, Fisher's exact test was more appropriate. Significance was defined as a *p*-value < 0.05 for both tests. Statistical testing was performed in the R language version 3.2.3 (Vienna, Austria).

Table 2
Target-specific primer sequences used in TEM-PCR primer mixes.

Pathogen/gene target	Forward out/in primer sequence (5′-3′)	Reverse out/in primer sequence (5′-3′)
Babesia microti	GGGCGTTCTCACTCCCAAGG	GCAGCCAAGAAAGCCACAGG
ITS ribosomal RNA	ftag-TCCCTGGGCCTCTGTGTTAG	rtag-CAGCGGAAACAGTATAATCC
Babesia microti	GGTGTTATGAGGAATCCATTC	GGTTGAGAATGCTGATGAAG
beta-tubulin gene	ftag-TTAAGCATTTGCTCATCAAC	rtag-TTCTTTATGATTGGATTTGC
Francisella tularensis	GCTACTACTGAGCAAGCTGC	GGAAGCTTGTATCATGGCAC
tul4 gene (Sjostedt et al., 1997; Tomioka et al., 2005)	ftag-GCTGTATCTAAGCCAACTGC	rtag-CCTTCTGGAGCCTGCCATTG
Francisella tularensis	ATTGGTTGTACTGTACAGCG	TAGCGCTTTGACTAACAAGG
fopA gene	ftag-TGATGTTATAACCGATTGTG	rtag-AACGGTCTAGCTGGAACTCC
Escherichia coli	TTGCCATTCAGTTCTGGCAG	ATGACCACCGTTCACGCTAC
gapA gene	ftag-TAGCAGCACCGGTAGAGGAC	rtag-TTCACGCTACTACCGCTACT
Escherichia coli	TACACCACGCCGAATACCTG	AGTTCAACGCTGACATCACC
uidA gene	ftag-TGGACGATATCACCGTGGTG	rtag-GCCAGTCAACAGATGCGTGG

3. Results

3.1. Testing negative control blood specimens

The performance of two amplification methods, singleplex real time PCR (RT-PCR) and Target Enriched Multiplex PCR (TEM-PCR) was evaluated based on percentage of positive samples detected (a measure of sensitivity) and percentage of true negative samples classified correctly in these assays. Coded samples were tested and results were compared after completion of testing. The quantitative output of each assay that indicates positive detection of the pathogen (the cut off) was set in both amplification methods to be greater than values calculated from known negative specimens. Furthermore, the percentage of negative results when testing true negative specimens represents a measure of assay specificity. In order to set the cutoffs, to compare two amplification methods (multiplex-PCR and singleplex PCR) and the effect of blood matrix type (healthy donor blood or patient blood) on classification of true negatives, DNA extracted from unspiked-whole blood specimens was tested. Among the 65 negative specimens of the healthy donor matrix and symptomatic donor matrix combined (Table 4), the RT-PCR method classified all 65 as negative (100%) and the TEM-PCR method classified 60 as negative (92.3%). The five false positives with the TEM-PCR method were split between three healthy donor specimens and two symptomatic donor specimens (Table S4—supplemental). Although not significant, the percentage of negative control samples correctly identified with TEM-PCR was less than with singleplex PCR ($p = 0.134$).

3.2. TEM-PCR Platform versus RT-PCR comparison with spiked specimens

To compare the performance of the two detection platforms, mock specimens were created for *B. microti*, *E. coli* and *F. tularensis* pathogens at a high concentration that should be consistently positive, a medium concentration that approached the minimum detectible and a low concentration that was below the level of consistent detection. The exact concentrations of the three levels is pathogen dependent (see Materials and methods section) and are labeled as

"High", "Medium" and "Low" to simplify presentation of results. The results from 10 healthy donor specimens for each of the three pathogens at the highest concentration were combined with the results from 10 symptomatic donor specimens for each pathogen, 60 total, in a 2×2 table for comparison of TEM-PCR to RT-PCR (Table 4). Of the 60 specimens in this group, all 60 (100%) had positive detections with both platforms. Thus, at this concentration they both exhibit 100% sensitivity and there is no difference between the methods ($p = 1.000$). At both medium and low concentrations of pathogen, the data suggested a trend toward greater sensitivity with TEM-PCR than RT-PCR based on the higher number of discrepant results with TEM-PCR positive and RT-PCR negative among these known positive specimens (Table 4), though the only significant difference was at the lowest concentration ($p = 0.044$). This was a subtle overall trend because there were no significant differences when the comparisons were made for each pathogen, concentration and matrix type separately (Table S4).

3.3. Blood matrix from healthy donors versus symptomatic donors

For the blood matrix comparisons, the data from both platforms were combined. Thirty mock specimens (10/pathogen) at high, medium and low concentrations were prepared using blood from healthy as well as symptomatic donors. Each of these sets of 30 specimens was tested with TEM-PCR and RT-PCR for a total of 60 data points per concentration. Of the 60 data points from healthy donors and the 60 data points from symptomatic donors at the high concentration, all 120 (100%, $p = 1.000$) were positively detected (Table 5). At the medium concentration, the symptomatic donors had the higher correct call percentage, but at the lower concentration, the healthy donors had

Table 3
Probe sequences for the detection of TEM-PCR amplicons.

Pathogen	Gene target	TEM-PCR Detection probe sequence (5′-3′)
Babesia microti	ITS ribosomal RNA (Ohmori et al., 2011)	TCATTCTAACGGGCATTGGG
	Beta-tubulin gene	GCACACATCATATTCTTGGC
Francisella tularensis	tul4 gene (Sjostedt et al., 1997; Tomioka et al., 2005)	AACCCACAAGGAAGTGTAAG
	fopA gene	CCAGCACCTGATGGAGAGTT
Escherichia coli	gapA gene	GGACGGGATGATGTTCTGGGAA
	uidA gene	TGTCGCGCAAGACTGTAACC

Table 4
2×2 contingency tables: TEM-PCR vs. RT-PCR.
Data from all pathogens, *B. microti*, *F. tularensis* and *E. coli*, are combined and grouped by high, medium, and low concentrations. Data from both healthy and symptomatic mock specimens are combined to show the impact of test platform. See Table S4 for uncombined data. The reported sensitivity at each pathogen concentration was calculated based on the knowledge that the spiked specimens all contained the targeted pathogen and the specificity is calculated on negative specimens, which were known to be pathogen-free.

Known concentration	Healthy & symptomatic donors				
High	$p = 1.000$	RT+	RT−	TEM sensitivity	RT sensitivity
	TEM +[a]	60	0	100%	100%
	TEM −	0	0		
Medium	$p = 0.421$	RT+	RT−		
	TEM +	44	11	91.7%	81.7%
	TEM −	5	0		
Low	$p = 0.842$	RT+	RT−		
	TEM +	8	15	38.3%	21.7%
	TEM −	5	32		
Negative	$p = 1.000$	RT+	RT−	TEM ~~sensitivity~~	RT ~~sensitivity~~
	TEM +	0	5	92.3%	100%
	TEM −	0	60		

[a] TEM-PCR positive; TEM −, TEM-PCR negative; RT+, RT-PCR positive; RT−, RT-PCR negative.

Table 5

2 × 2 contingency tables: healthy donor specimens vs. symptomatic donor specimens. All pathogens, *B. microti*, *F. tularensis* and *E. coli* are combined and grouped by high, medium, and low concentrations. Data from both platforms are combined to reveal the impact of the two sample types. See Table S5 for uncombined data. The observed sensitivity at each pathogen concentration was calculated based on the knowledge that the spiked specimens contained the targeted pathogen and the specificity is calculated on negative specimens, which were known to be pathogen-free.

Concentration	TEM-PCR & RT-PCR platforms				
High	$p = 1.000$	Symptomatic	Healthy	Correct, symptomatic	Correct, healthy
	Positive	60	60	100%	100%
	Negative	0	0		
Medium	$p = 0.421$	Symptomatic	Healthy		
	Positive	54	50	90.0%	83.3%
	Negative	6	10		
Low	$p = 0.842$	Symptomatic	Healthy		
	Positive	17	19	28.3%	31.7%
	Negative	43	41		
Negative	$p = 1.000$	Symptomatic	Healthy	Correct, symptomatic	Correct, healthy
	Positive	2	3		
	Negative	48	77	96.0%	96.3%

the highest correct call frequency, although the differences were not significant (Medium: $p = 0.421$, healthy donors 50/60 [83.3%] vs. symptomatic donors 54/60 [90.0%], Low: $p = 0.842$, healthy donors 19/60 [31.7%] vs. symptomatic donors 17/60 [28.3%]). The correct identification of negative specimens was not different between symptomatic and healthy ($p = 1.000$). Evaluating the 2 × 2 comparisons of each condition (pathogen, concentration or test platform) separately also did not reveal any significant differences (Table S5—supplemental).

4. Discussion

RT-PCR and multiplex PCR are commonly used methods for detecting low abundance pathogens. Singleplex RT-PCR pathogen detection is considered to be a gold standard methodology for high sensitivity, reproducibility and ease of use. To reduce the burden of clinical validation for low abundance pathogens, mock specimens can be created from human matrices and cultured pathogens. However, spiking should be done by standardized methods to obtain results that can be compared appropriately (Dong et al., 2016). In this study, standardized methods for preparing mock whole blood specimens were employed with *B. microti*, *E. coli* and *F. tularensis*, encompassing an emerging protozoan parasite, a common Gram-negative bacterium and a potential biothreat bacterium. These mock specimens were utilized to perform comparison between TEM-PCR and singleplex RT-PCR and between healthy donor blood matrix and symptomatic donor blood matrix.

Target Enriched Multiplex PCR offers many advantages in a diagnostic or blood donor screening setting. The potential for increased sensitivity and simultaneous screening for multiple pathogens of concern suggests improved clinical outcomes. Furthermore, a well-designed multiplex platform could facilitate addition of new pathogen detection assays to the platform with reduced development time and regulatory burden. The initial development of a multiplex platform will involve the optimization of primer mixes to achieve maximum amplification while avoiding nonspecific products due to the presence of many primers in the same PCR reaction. Typically, empirical testing and a trial-and-error approach are being used for testing several primer pairs, because there are limited means to predict the performance characteristics of selected primer mixes.

New technology such as TEM-PCR will be evaluated by comparison to a reference test such as singleplex RT-PCR. In our comparison, the mock specimens with three distinct types of pathogens at multiple concentrations across the detectible range and two matrix types showed that the new technology TEM-PCR performed as well as the

singleplex assays. The observation that fewer false negatives and more false positives occurred with the TEM-PCR are consistent with common expectations of a multiplex assay. However, a much larger number of samples would have to be tested to show significance of these slight differences between the performance of the TEM-PCR and the singleplex RT-PCR.

The FDA has recommended that diagnostic tests be evaluated not only with healthy donor samples, but also samples from individuals with signs and symptoms of disease such as fever, aches and weakness (FDA/CDRH, 2014). The symptomatic specimens are expected to have a composition that presents challenges to the performance of a diagnostic device and may be more similar to the clinical specimens that the device is intended to test.

In our comparison, the same cultured pathogens were spiked either into healthy donor blood or symptomatic blood at three concentrations spanning a 3-log range. DNA samples from these mock specimens were extracted by a uniform method. These samples were tested for the presence of the pathogens by two platforms, the TEM-PCR and the RT-PCR. In this study, there were no differences in the performance between the two matrix types on either platform. The uniformity of results may be explained by the extraction step in a molecular assay removing enough of the components that make symptomatic specimens challenging in, for example, serological assays where no specimen processing is involved. In our results there was no clear trend toward more sensitive detection or higher specificity among the healthy donor specimens.

We realize that the sample size for this study is rather small and given infinite time and resources, large samples are always preferred to increase the statistical power of the results. However, researchers are often confronted with scenarios where small sample size analysis is unavoidable, especially in studies where extremely low prevalence pathogens are analyzed and the number of proper donor specimens must be ethically justified. In statistical literature, there is much debate on whether or not p-values should even be calculated on small sample sizes to avoid drawing incorrect conclusions. However, a few statistical tests have been reported to function correctly even when the sample size is small (de Winter, 2013; McNemar, 1947). We have chosen the McNemar's test, and Fisher's exact test, which perform well when comparing small sample size groups allowing us to calculate p-values between various groupings (Fisher, 1922; McDonald, 2014; McNemar, 1947). The construction of the 2 × 2 tables and these statistical methods penetrated the complexity of the multiple variables to reveal the correct comparison of assay platform and sample matrix.

5. Conclusion

The standardized methods of preparing mock (spiked) clinical specimens for evaluating diagnostic devices intended to test for low prevalence pathogens have effectively demonstrated the equivalence of a proprietary multiplex platform to individual real-time PCR. The equivalence between TEM-PCR, optimized to screen for the presence of *Babesia microti*, *Escherichia coli* and *Francisella tularensis*, and individual real-time PCR assays for each pathogen over a three-log range of pathogen concentrations was demonstrated in healthy donor and symptomatic donor specimens.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.mimet.2016.11.005.

References

Anez, G., Jiang, Z., Heisey, D.A., Kerby, S., Rios, M., Chikungunya Virus Collaborative Study Group, 2015. Collaborative study for the characterization of a chikungunya virus RNA reference reagent for use in nucleic acid testing. Vox Sang. 109, 312–318.

Brittain-Long, R., Westin, J., Olofsson, S., Lindh, M., Andersson, L.M., 2011. Access to a polymerase chain reaction assay method targeting 13 respiratory viruses can reduce antibiotics: a randomised, controlled trial. BMC Med. 9, 44.

Burd, E.M., 2010. Validation of laboratory-developed molecular assays for infectious diseases. Clin. Microbiol. Rev. 23, 550–576.

CLSI, 2008. Verification and Validation of Multiplex Nucleic Acid Assays; Approved Guideline. Clinical and Laboratory Standards Institute, Wayne, PA .

Collins, C.H., Lyne, P.M., Grange, J.M., Falkinham III, J.O., 2004. Microbiological Methods. Arnold, A member of the Hodder Headline Group, London .

de Winter, J., 2013. Using the Student's t-test with extremely small sample sizes. Pract. Assess. Res. Eval. 2016.

Dong, M., Fisher, C., Anez, G., Rios, M., Hobson, J.P., Beanan, M., Nakhasi, H.L., Hockman, D., Grigorenko, E., Duncan, R., 2016. Methods to generate mock (spiked) clinical specimens by spiking blood or plasma with cultured pathogens. J. Appl. Microbiol. http://dx.doi.org/10.1111/jam.13082 .

FDA/CDRH, 2014. Highly Multiplexed Microbiological/Medical Countermeasure In Vitro Nucleic Acid Based Diagnostic Devices .

Fisher, R.A., 1922. On the interpretation of X^2 from contingency tables, and the calculation of P. J. R. Stat. Soc. 85, 87–94.

Fortier, A.H., Slayter, M.V., Ziemba, R., Meltzer, M.S., Nacy, C.A., 1991. Live vaccine strain of Francisella tularensis: infection and immunity in mice. Infect. Immun. 59, 2922–2928.

McDonald, J., 2014. Handbook of Biological Statistics. 2015. Sparky House Publishing, Baltimore, MD.

McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 12, 153–157.

Moitra, P., Zheng, H., Anantharaman, V., Banerjee, R., Takeda, K., Kozakai, Y., Lepore, T., Krause, P.J., Aravind, L., Kumar, S., 2015. Expression, purification, and biological characterization of Babesia microti apical membrane antigen 1. Infect. Immun. 83, 3890–3901.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. 51 (Pt 1), 263–273.

Ohmori, S., Kawai, A., Takada, N., Saito-Ito, A., 2011. Development of real-time PCR assay for differential detection and quantification for multiple Babesia microti-genotypes. Parasitol. Int. 60, 403–409.

Roth, W.K., Busch, M.P., Schuller, A., Ismay, S., Cheng, A., Seed, C.R., Jungbauer, C., Minsk, P.M., Sondag-Thull, D., Wendel, S., Levi, J.E., Fearon, M., Delage, G., Xie, Y., Jukic, I., Turek, P., Ullum, H., Tefanova, V., Tilk, M., Reimal, R., Castren, J., Naukkarinen, M., Assal, A., Jork, C., Hourfar, M.K., Michel, P., Offergeld, R., Pichl, L., Schmidt, M., Schottstedt, V., Seifried, E., Wagner, F., Weber-Schehl, M., Politis, C., Lin, C.K., Tsoi, W.C., O'Riordan, J., Gottreich, A., Shinar, E., Yahalom, V., Velati, C., Satake, M., Sanad, N., Sisene, I., Bon, A.H., Koppelmann, M., Flanagan, P., Flesland, O., Brojer, E., Letowska, M., Nascimento, F., Zhiburt, E., Chua, S.S., Teo, D., Stezinar, S.L., Vermeulen, M., Reddy, R., Park, Q., Castro, E., Eiras, A., Gonzales Fraile, I., Torres, P., Ekermo, B., Niederhauser, C., Chen, H., Oota, S., Brant, L.J., Eglin, R., Jarvis, L., Mohabir, L., Brodsky, J., Foster, G., Jennings, C., Notari, E., Stramer, S., Kessler, D., Hillyer, C., Kamel, H., Katz, L., Taylor, C., Panzer, S., Reesink, H.W., 2012. International survey on NAT testing of blood donations: expanding implementation and yield from 1999 to 2009. Vox Sang. 102, 82–90.

Shampo, M.A., Kyle, R.A., 2002. Kary B. Mullis—Nobel Laureate for procedure to replicate DNA. Mayo Clinic Proceedings. 77, p. 606.

Sjostedt, A., Eriksson, U., Berglund, L., Tarnvik, A., 1997. Detection of Francisella tularensis in ulcers of patients with tularemia by PCR. J. Clin. Microbiol. 35, 1045–1048.

Sobel, M.E., Bagg, A., Caliendo, A.M., Ladanyi, M., Zehnbauer, B., 2008. The evolution of molecular genetic pathology: advancing 20th-century diagnostic methods into potent tools for the new millennium. J Mol Diagn. 10, 480–483.

Tomioka, K., Peredelchuk, M., Zhu, X., Arena, R., Volokhov, D., Selvapandiyan, A., Stabler, K., Mellquist-Riemenschneider, J., Chizhikov, V., Kaplan, G., Nakhasi, H., Duncan, R., 2005. A multiplex polymerase chain reaction microarray assay to detect bioterror pathogens in blood. J. Mol. Diagn. 7, 486–494.