DNA Sequence Signatures for Rapid Detection of Six Target Bacterial Pathogens Using PCR Assays

Kenjiro Nagamine*,†, Guo-Chiuan Hung*, Bingjie Li and Shyh-Ching Lo

Tissue Microbiology Laboratory, Division of Cellular and Gene Therapies and Division of Human Tissues, Office of Cellular, Tissue and Gene Therapies, Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD, USA. *Both are first authors and both equally contributed to this study. †Current address: Department of Oral and Maxillofacial Surgery, Nagoya University Graduate School of Medicine, Showa-ku, Nagoya Aichi, Japan.

ABSTRACT: Using *Streptococcus pyogenes* as a model, we previously established a stepwise computational workflow to effectively identify species-specific DNA signatures that could be used as PCR primer sets to detect target bacteria with high specificity and sensitivity. In this study, we extended the workflow for the rapid development of PCR assays targeting *Enterococcus faecalis*, *Enterococcus faecium*, *Clostridium perfringens*, *Clostridium difficile*, *Clostridium tetani*, and *Staphylococcus aureus*, which are of safety concern for human tissue intended for transplantation. Twenty-one primer sets that had sensitivity of detecting 5–50 fg DNA from target bacteria with high specificity were selected. These selected primer sets can be used in a PCR array for detecting target bacteria with high sensitivity and specificity. The workflow could be widely applicable for the rapid development of PCR-based assays for a wide range of target bacteria, including those of biothreat agents.

KEYWORDS: DNA signature, species-specific PCR, rapid detection

Introduction

The rapid and economic detection, identification, and quantification of infectious agents in various clinical settings are essential. Currently, the "gold standards" for the detection of bacterial and fungal agents have been the traditional microbiological culturing methods that are generally laborious and time-consuming.[1,2] Therefore, it is highly desirable to develop alternative molecular methods, such as nucleic acid-based tests (NATs), which have exquisite sensitivity and a relatively short turnaround time.

Recent advances in detection technology have produced several new systems that allow the detection of a wide range of pathogens. For example, TessArae uses high-density and high-resolution microarrays for microbial identification based on resequencing of selected target genes by hybridization.[3] Taking advantage of the high accuracy of Matrix-assisted laser desorption/ionization (MALDI) to measure the mass of DNA fragments with high precision, PLEX-ID utilizes the base composition of specific DNA fragments measured by MALDI for microbial identification.[4–6] The capacity for these two systems to detect a wide range of pathogens, from viruses and bacteria to protozoa, is very high. However, additional processing time is required for hybridization and MALDI for TessArae and PLEX-ID, respectively, and the results are in general not quantitative. In comparison, real-time PCR assays are more desirable for rapid detection of the targeted or selected pathogens. Multiple primer sets can be arranged in an array format on a 96- or 384-well plate for multiplex detection.

In developing any NATs, it is most critical to effectively select "DNA signatures" (ie, species-specific sequences) from the target microbe of interest for highly sensitive and specific assays.[7] DNA signatures should be highly conserved within the target species and absent in any other species of microorganisms, including those of closely related microbial species. In spite of the importance of identifying DNA signatures in assay development, the processes for selecting DNA signatures have generally not been well described in the literature.[8,9]

In our previous study, we established a computational workflow using Insignia, dCAS, and NCBI-BlastN sequentially for the identification of DNA signatures from whole genome sequence data that can be used as the basis for designing real-time PCR assays using *Streptococcus pyogenes* as a model organism.[10] In this study, we extended the use of this approach for rapid development of PCR tests against six additional bacterial pathogens. *Enterococcus faecalis*, *Enterococcus faecium*, *Staphylococcus aureus*, *Clostridium perfringens*, *Clostridium difficile*, and *Clostridium tetani* have been defined as "high-risk" bacteria American Association of Tissue Banks (AATB)

in tissue grafts.[11] It is our final objective that a PCR array can be developed for multiplex detection of all the high-risk bacteria pathogens in testing during and after tissue processing.

Materials and Methods

Microbial strains. Microbial strain *E. faecalis* (ATCC 29212), *E. faecium* (ATCC 35667), *S. aureus* (ATCC 25923), *C. perfringens* (ATCC 3628, ATCC 13124), and *C. difficile* (ATCC 17857, ATCC 43255) were obtained from American Type Cell Collection (ATCC). *C. tetani* strain Massachusetts C2 was kindly provided by James Keller (FDA). Microorganisms were cultured following protocols recommended by ATCC. The identities of these microbial strains were confirmed by Biolog Microbial Identification System GEN III (Biolog), which utilizes metabolic characteristics such as pH, salt, lactic acid tolerance, reducing power, chemical sensitivity, and the ability of the cell to metabolize all the major classes of biochemicals to identify microbial species.

Isolation of genomic DNA. To isolate genomic DNA from microorganisms, cell pellets were harvested from either broth or solid cultures, and genomic DNAs were isolated using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's protocol. Human genomic DNA was prepared from buffy coat cells of a healthy blood donor from NIH Blood Bank using the DNeasy Blood & Tissue Kit or the conventional phenol/chloroform extraction method. DNA was quantified using Nanodrop (Thermo Scientific) and stored at −20°C prior to PCR amplification.

Identification of final DNA signatures and design of PCR primer sets. The overall computational workflow has been previously described.[10] Briefly, the open-internet-accessible Insignia program[12,13] (http://insignia.cbcb.umd.edu) was selected to produce DNA signatures for *E. faecalis*, *E. faecium*, *S. aureus*, *C. perfringens*, *C. difficile*, and *C. tetani*. Currently, there are a total of 13,928 genome sequences in Insignia's database, with 11,274 from virus/phages and 2,653 from nonvirus organisms. To identify DNA signatures for a target species, the Insignia program first calculated conserved regions within genomes of all available strains of the target species. Subsequently, DNA signatures were generated by excluding shared regions with all nontarget (background) genomes and output in FASTA format using the genomic coordinates of the reference genome as their description. However, we realized that the majority of these initial DNA signatures had high homology with sequences of closely related species. Therefore, it was necessary to use the dCAS program[14] as a second screening to select the top 50 DNA signatures that were more different from the sequences of closely related species. The specificity of the top 50 DNA signatures was then subjected to the third screening against the NCBI nonredundant database to select for final DNA signatures that do not have high homology (generally E-value > 1, bit score < 40) with any other known DNA sequences by NCBI-BlastN.[15] All oligonucleotide primer sets were then designed for these final DNA

signatures using Primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/),[16] and named after the target species (Efl, Efm, Cp, Cd, Ct, and Sa for *E. faecalis*, *E. faecium*, *C. perfringens*, *C. difficile*, *C. tetani*, and *S. aureus*, respectively). Since the PCR assays will be used in the presence of human DNA, primer set specificity was checked to ensure that no likely theoretical amplification from human DNA was predicted.

Conventional and real-time PCR testing and melting curve analysis. To perform conventional PCR assays, AmpliTaq gold DNA polymerase (Applied Biosystems) was used. PCRs (20 μL) were performed in 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 200 μM dNTPs, 50 pmol of each primer with 1 U AmpliTaq gold DNA polymerase (Applied Biosystems) using Eppendorf Mastercycler ep gradient S (Eppendorf) with the following parameters: an initial activation of AmpliTaq gold DNA polymerase at 94°C for ten minutes, followed by 45 cycles of 94°C for 30 seconds (denaturation), 55°C for 30 seconds (annealing), and 72°C for 30 seconds (extension), followed by a final extension at 72°C for seven minutes. Amplicons were separated on 2% agarose gel by electrophoresis, stained with ethidium bromide, and detected upon UV transillumination by an electronic documentation system (GelDoc-It Imaging System, UVP).

To perform real-time PCR assays, Power SYBR® Green PCR Master Mix with AmpliTaq gold DNA polymerase was used. Real-time PCR was performed in triplicates on CFX96 real-time PCR Detection System (Bio-Rad) with the following parameters: an initial activation of AmpliTaq gold DNA polymerase at 94°C for ten minutes, followed by 50 cycles of 94°C for 15 seconds (denaturation) and 60°C for 60 seconds (annealing and extension). Amplicons were subjected to melting curve analysis by increasing temperature from 65°C to 95°C at 0.5°C per second, recording the changes in fluorescence to determine the melting profiles for each detected signal. The threshold of each PCR reaction is determined by the CFX manager software (Bio-Rad) with default setting.

Results

Identification of species-specific DNA signatures for *E. faecalis*, *E. faecium*, *C. perfringens*, *C. difficile*, *C. tetani*, and *S. aureus*, and selection of specific primers. Briefly, for each target species, we selected genomes of all available strains within each target species in the Insignia program ($n = 22$, 9, 8, 10, 1, and 21 for *E. faecalis*, *E. faecium*, *C. perfringens*, *C. difficile*, *C. tetani*, and *S. aureus*, respectively) for DNA signature computation. DNA signatures ranging from 20 to 385 bp in length were produced for each species ($n = 45,963$; 36,645; 26,245; 38,587; 30,644; and 39,206 for *E. faecalis*, *E. faecium*, *C. perfringens*, *C. difficile*, *C. tetani*, and *S. aureus*, respectively). We only selected species-specific DNA signatures that are longer than 102 bp so that a sizable amplicon could be produced in PCR (more in discussion).

After the size adjustment, the number of DNA signatures remaining for *E. faecalis*, *E. faecium*, *C. perfringens*, *C. difficile*,

C. tetani, and *S. aureus*, were 844, 278, 237, 583, 3,322, and 913, respectively. As a second screening, each species' DNA signatures were then compared to a local database containing the genome sequences of their respective closely related species within the same genus by dCAS. Then, the top 50 DNA signatures with highest E-value compared to the closely related species, indicating their sequence distinctness, were selected for a third screening against the NCBI nonredundant database by BlastN. A total of 5, 11, 10, 4, 3, and 10 final DNA signatures were selected for primer design for *E. faecalis*, *E. faecium*, *C. perfringens*, *C. difficile*, *C. tetani*, and *S. aureus*, respectively. Table 1 lists the locations of these DNA signatures, based on the genomic coordinates of the type strain, and their annotated gene information.

The Primer-BLAST program was used to design primer sets based on the final selected DNA signatures. The optimal melting temperature (T_m) for all primers was set at 55°C with a maximum T_m difference between the forward and reverse primers of 3°C, and the length of amplicons was generally limited to 150 bp for maximum PCR efficiency. Table 1 lists the sequences of all primer sets.

Assessment of the sensitivity and specificity of species-specific primer sets in conventional PCR assays. Following the theoretical design of 43 primer sets in silico, we first evaluated the sensitivity and specificity of these primer sets using conventional PCR assays. A typical testing comprised a panel of DNA samples consisting of 5 pg, 500 fg, 50 fg, and 5 fg of DNA from each target pathogen with and without the presence of 5 ng human DNA, and 5 ng of DNA from 39 bacterial species, including closely related species, other common Gram-positive and Gram-negative bacteria, 5 ng of DNA from three common human yeast pathogens *Candida albicans*, *Candida tropicalis*, and *Candida parapsilosis*, and 5 ng of human DNA alone were included in the testing. A representative conventional PCR testing of primer sets Efl6, Efm12, Cp14, Cd3, Ct3, and Sa2 is shown in Figure 1. Specific amplicons of the expected sizes from as low as 5–50 fg of DNA from their respective target species were produced. In most cases, no amplicons of the expected sizes were produced using 5 ng DNA (a million-fold excess) of nontarget species (Fig. 1 and Table 1). Among the 43 selected primer sets, 20 primer sets (Efl3, Cp11–Cp20, Cd3, Sa1–Sa4, Sa6–Sa9) detected 5 fg DNA of the respective target species; 17 primer sets (Elf2, Efl6, Efm3, Efm4, Efm6–Efm8, Efm10–Efm12, Cd1, Cd2, Cd4, Ct1–Ct3, and Sa10) detected 50 fg DNA of the respective target species; and six primer sets (Elf4, Efl5, Efm2, Efm5, Efm9, and Sa5) detected 500 fg DNA of the respective target species (Table 1). In most cases, the limit of detection was not affected by the presence of 5 ng human DNA in the background.

Assessment of the sensitivity and specificity of selected species-specific primer sets in real-time PCR assays. To avoid potential problems with data interpretation, we eliminated 16 primer sets from a total of 43 primer sets (denoted as "–" in Table 1) from the real-time PCR testing: those that had higher limits of detection; those with too many side products amplified from nontarget species; or those that had nontarget amplicons close to those of the target species. A representative real-time PCR testing of primer sets Efl6, Efm12, Cp14, Cd3, Ct3, and Sa2 is shown in Figure 2. Among the remaining 27 primer sets, 22 primer sets had the same sensitivity as the conventional PCR assay, which detected 5–50 fg target DNA (Table 1). We noted with interest that the assay sensitivity for primer sets Efl4, Efm3, Efm9, and Cd4 was improved in real-time PCR testing, compared to the conventional PCR testing. On the contrary, primer set Efl2 was less sensitive in real-time PCR testing (Table 1).

With regard to the assay specificity, primer sets Efl2, Efm3, Cp18, Ct2, Sa7, and Sa8 produced many side amplicons from nontarget DNA with similar melting peaks close to those of the respective target species, and thus were considered "unacceptable" (Table 1). In comparison, four primer sets (Cp19, Cp20, Ct1, and Sa6) showed very high specificity, with no above-threshold amplification observed from all their respective nontarget DNA tested (Table 1 and Fig. 2). The remaining 17 primer sets were also considered "acceptable" (Table 1) as the side peaks produced from their respective nontarget DNA using each of these primer sets are clearly distinguishable from their respective target DNA, based on their distinct melting profiles. These 21 acceptable primer sets will be selected for future study.

Discussion

It is highly advantageous to develop nucleic acid-based molecular methods with both high sensitivity and specificity for rapid detection of pathogenic microbes. However, the true challenge to develop such molecular assays critically depends on the effective identification of truly species-specific DNA signatures as primers for PCR. In order to meet the critical challenge, we previously developed a stepwise computational workflow to effectively identify highly species-specific DNA sequences that could be used to detect the target bacteria in high sensitivity using *S. pyogenes* as a model organism.[10] In this study, we further tested the computational workflow for its effectiveness to identify species-specific DNA sequences intended as primers to be used in real-time PCR assays for the rapid detection of six different bacterial pathogens.

In the initial step of the workflow, tens of thousands of DNA signatures of 20–385 bp in length were identified by Insignia program for each pathogen studied. However, our previous study of *S. pyogenes* revealed that PCR primers designed from many of these signatures would still amplify DNA from many closely related species of bacteria. The Insignia algorithm identified as few as one difference out of every 20 nucleotides against all background genomes as DNA signature. A valid DNA signature could have up to 95% similarity in sequence with its close relatives. We used dCAS program

Table 1. Evaluation of primer sets designed from selected signature sequences of *E. faecalis* (Efl), *E. faecium* (Efm), *C. prefringens* (Cp), *C. difficile* (Cd), *C. tetani* (Ct) and *S. aureus* (Sa) by 2 PCR platforms.

PRIMER NO	GENOMIC COORDINATES	GENE PRODUCT INFORMATION	FORWARD PRIMER	REVERSE PRIMER	CONVENTIONAL PCR		REAL-TIME PCR		ACCEPT-ABILITY
					LOD (fg)	NO. OF SIDE BANDS	LOD (fg)	NO. OF SIDE PEAKS	
Efl2	1344172 to 1344271	Transcriptional regulator	AGGAGTGACACGATGAGTCG	CCATTGATTTTCGGGGATTAAAGGT	50	1	500	Many	No
Efl3	1567016 to 1567155	Hypothetical protein	TCGTTGTTGTACCGCCGTAAT	AGGTTGGGCAAGTCGTCAAC	5	6	–	–	–
Efl4	1375763 to 1375871	Hypothetical protein	AAAGGAGGTACATGCTATGTTTGA	GGAAGACCATACTTGACACGCA	500	4	50	2	Yes
Efl5	1650850 to 1650983	Transcriptional regulator	GAAACCCGTGAGCGTGTCTT	ACCCAAGCCTCAAAAGTTTCAT	500	1	–	–	–
Efl6	361371 to 361477	Diaminopimelate epimerase	TGTTCGCGGGAAACAGGATT	CCATCGCTGTTAATCACTCGC	50	0	50 g	5	Yes
Efm2	1126761 to 1126876	LacI family sugar-binding transcriptional regulator	AAGCAGAAGGAAATCGGGCG	TAGCTTGGTCGGTGTGGAGT	500	9	–	–	–
Efm3	241541 to 241389	Glycosyl Hydrolase Family 88	CTGCGACAGTTGCCACACTA	GCTTTGTTTGACAGAATTGCTCG	50	1	5	Many	No
Efm4	1355375 to 1355272	Hypothetical protein	CGCAAGATTCGGCGTTGAAG	AAATCAGCATTGGTGCGCGG	50	Many	–	–	–
Efm5	547278 to 547377	Nudix hydrolase, YffH family	TGACCAAAGGAAAGTAGAGGGA	AATGACTCGCTTGCTTGTGC	500	11	–	–	–
Efm6	2344582 to 2344483	Hypothetical protein	ACAAGTCCGGATACAAGTGCT	TTTGCTGTGTGCTGACGTGCT	50	14	–	–	–
Efm7	1156608 to 1156714	Serine-type D-Ala-D-Ala carboxypeptidase	GCTCCCAGGCTGATCTCCGTA	TACAGCGAACACCACACAGG	50	Many	–	–	–
Efm8	399854 to 399957	Tetronasin resistance transmembrane protein	CCTAGAAATTTTTGTTGCGAACAGC	TGATTTGTCTAGCTGCTATTTTACCA	50	4	50	5	Yes
Efm9	1355386 to 1355287	Hypothetical protein	GCGCAAAAATACGCAAGATTCG	CGCGGTCCACCATACACTC	500	3	50	1	Yes
Efm10	345637 to 345738	Malonyl CoA-acyl carrier protein transacylase	CCCCCGGCAATACTTTGACA	ACGTGGCGTCATCAATACCA	50	3	50	5	Yes
Efm11	1126774 to 1126876	LacI family sugar-binding transcriptional regulator	TCGGGCGGGTTTATTTGTCA	TAGCTTGGTCGGTGTGGAGT	50	Many	–	–	–
Efm12	2515276 to 2515129	Acyltransferase	GAAAGGACACGATTCTACTTGCT	ATGAACGGAGCCCATCAAACC	50	0	50	2	Yes
Cp11	1964474 to 1964576	Nitrate reductase, NADH oxidase subunit	TATTCCAGCAGCTGATGCCC	ACTCCTCAATGCGTTGCAAAA	5	10	–	–	–

ID	Position	Protein	Forward sequence	Reverse sequence					
Cp12	1616197 to 1616304	Hypothetical protein	CCATCCCATACTTTCGCCCG	GCAGCTGTTGGTATGATGGT	5	Many	–	–	–
Cp13	2563413 to 2563567	Putative maltose/maltodextrin ABC transporter	ACCTCCTCGAATTGTCCACTTA	GCTGGTTATACAGATATTCTTG-CATCT	5	2	5	1	Yes
Cp14	456572 to 456678	Putative arginine/ornithine antiporter	GCAGCAGCAAGGATTCTAAAGG	CAAAACATGGCCCAATCCGC	5	0	5	2	Yes
Cp15	315350 to 315449	Glycosyl transferase, group 1 family protein	TCCAACGTTCTCAACATATGGCA	TCAACAGTCTTTACATGCGGATCA	5	4	–	–	–
Cp16	643102 to 643201	ABC transporter, permease protein	TGCAGTGCTAAGCTTACCTTTCT	AGTACATCCATCTATCATTGCTGCC	5	0	5	2	Yes
Cp17	972508 to 972610	Glycosyl hydrolase, family 38	TGTTTATACTAGAGAGGCAGATTGTGA	AGCACATATTTCACCCTTAGCCA	5	12	–	–	–
Cp18	1574810 to 1574912	Voltage-gated chloride channel family protein	TAAAAGTCCTATGGCAGCACCTTCT	GGAGGGGAAATCTTTGAGTAAT-GAGA	5	2	5	Many	No
Cp19	1891982 to 1892100	Phosphate acetyltransferase	CTCTCTTCATCACCTTCTGGAAGTA	ACACATAAACATCTGCTTTGACTTGA	5	0	5	0	Yes
Cp20	351811 to 351910	Putative membrane protein	CAGTTTGGCATAGTGTTAGTTCCT	TTACTTCTTCATCCTTACTGGCTTT	5	3	5	0	Yes
Cd1	498902 to 499013	Membrane protein	GAAGCATCATGGGTGCAGGA	AGTGCACCTGCAAGAACTGC	50	5	–	–	–
Cd2	1495998 to 1496130	Putative aliphatic sulfonates ABC transporter	CTGTTGATTGCTGCACTTGCT	CCACTGGACCTGCAAGCATT	50	4	–	–	–
Cd3	612589 to 612689	Putative phage-related replicative helicase	GGAGAAAGTGAGGAGGCAGTTA	TGCTTATGCCTAGTCTGTCCA	5	1	5	5	Yes
Cd4	3761159 to 3761304	Putative membrane protein	ACTGAAGCACCCATCCATGA	TTCGCAGGTGCGTATGTAGC	50	1	5	5	Yes
Ct1	804482 to 804590	Membrane-associated protein	AAAGCCCTAAGGAAGATCCAATACT	CAAGATATCCAATCTGCTGACCACT	50	4	50	0	Yes
Ct2	1613628 to 1613759	Fumarate reductase flavoprotein subunit	ATTCCTTGTCCACCGTAACCTCTAA	CACTAGTGACATCAGCAATTGGCT	50	0	50	Many	No
Ct3	2217985 to 2218095	Branched chain amino acid transport system II	AGGTGCTACAAAGGATTCAGGACTA	ATAGCAATGTTAGTGTTAGGACCTC	50	8	50	4	Yes
Sa1	80490 to 80722	Squalene/phytoene synthase family protein	AGCGCCATCATGATTCTACG	GGTTTGGGCAATTTATGCTG	5	0	5	5	Yes
Sa2	2736301 to 2736472	Putative membrane protein	GTAAGCAACCAACGCCAGAT	CCATCTTTATAGGCACCTTTGTG	5	0	5	0	Yes
Sa3	2445849 to 2446014	Putative glycerate kinase	AAATTGTTGCTGTTGATGCG	AATATTTCGGGTACACGACTAGC	5	Many	–	–	–
Sa4	82622 to 82781	Hypothetical protein	GTCGCTCATTAAAGTGGCGT	CAATGCTTCCAATTTGTGGTT	5	0	5	2	Yes

(*continued*)

Table 1. (Continued)

PRIMER NO	GENOMIC COORDINATES	GENE PRODUCT INFORMATION	FORWARD PRIMER	REVERSE PRIMER	CONVENTIONAL PCR		REAL-TIME PCR		ACCEPT-ABILITY
					LOD (fg)	NO. OF SIDE BANDS	LOD (fg)	NO. OF SIDE PEAKS	
Sa5	2388982 to 2389371	Hypothetical protein	TCGCATATTATACAAGAGACACTTCA	TGGTTTAACGTGTCTAGCTCACT	500	0	–	–	–
Sa6	180768 to 180978	Putative transport system permease	TGTGCCACACGTTAAAGATCG	TTGCCATGCGACCACCATTC	5	0	5	0	Yes
Sa7	75822 to 75979	MFS family major facilitator transporter	TTTGGTGGTGTTGTTGATGCATC	CACAAGAGACCCTGCGCTGA	5	5	5	Many	No
Sa8	77197 to 77387	LucA/lucC family siderophore biosynthesis protein	TCCGACTCCTAAGAGTGCAAGT	CGCGTGAGCAGAAACATCTTG	5	0	5	Many	No
Sa9	2377597 to 2377776	Hypothetical protein	GTATAATCGCTGGTTGCAATTCGA	CAATACATTTGCTTGGTGACTAGGA	5	0	5	1	Yes
Sa10	2439014 to 2439200	Dethiobiotin synthase	TTTGATGGCAACACACTGATGAC	TGCTGGGGGAATTGCCGTAC	50	1	50	7	Yes

to effectively differentiate the degree of sequence homology between the DNA signatures produced by Insignia program and the genomic sequences of closely related species by the E-values of their sequence alignments. The top 50 DNA signatures for each of the six pathogens were selected for final verification of sequence uniqueness by BlastN against NCBI nonredundant nucleotide database. Using the workflow, we selected 43 candidates of DNA signatures (>102 bp) that could be used for designing primers (Table 1) and experimentally evaluated them for detection sensitivity against the target bacteria and specificity of not amplifying a panel of nontarget bacteria at a concentration of a million folds higher than those of target bacteria using both conventional PCR testing and real-time PCR testing.

A recent study by Zhang and Sun[17] described successful identification of tens of thousands of uniquely conserved regions (UCRs) of 18 nucleotides from bacteria genome using in-house bioinformatic pipeline. The study similarly showed that the UCRs identified could be used as primer sets for the detection of 15 target bacterial species. Both Insignia program and UCRs pipeline utilize the principle of k-mer. Insignia program relies on the precomputed pairwise comparison database, whereas UCRs pipeline converts nucleotide sequence into numerical value using integer mapping methods for the identification of species-specific sequences. In the study of UCRs pipeline, the limits of detection for target bacteria by the eight primer sets generating amplicons of 576–1110 bp in size were not specifically examined. In our study, the limit of detection of the final 21 selected primer sets for the six pathogens (Table 1) ranged from 5 to 50 fg (equivalent to 1–10 genomic copies) per reaction, which is equivalent to previously reported assays with high sensitivity.[9] We believe the high sensitivity observed in our PCR assays could in part be attributed to the sizes of amplicons produced by the selected DNA signatures that were less than 150 bp. Importantly, the high sensitivity of detection against the target bacteria using these selected primer sets were achieved at the same time with extremely high specificity of the assays. None of the 21 primer sets selected would amplify human DNA and a panel of DNA from 42 nontarget species of bacteria tested at a concentration of a million folds higher than those of the target bacteria.

The target sequences for the DNA signatures we identified and studied for these bacteria were different from those used in various studies previously reported,[8,18,19] most of which selected target genes based on their species-specific biochemical or immunological properties. It is unclear how well the selected target gene sequences are conserved among different strains of the same species of bacteria by this approach, which is evidently different from our computational workflow using whole genome sequences to search for DNA signatures conserved among multiple strains within the target species. The rapid advance of the next generation of sequencing technology has already made many more
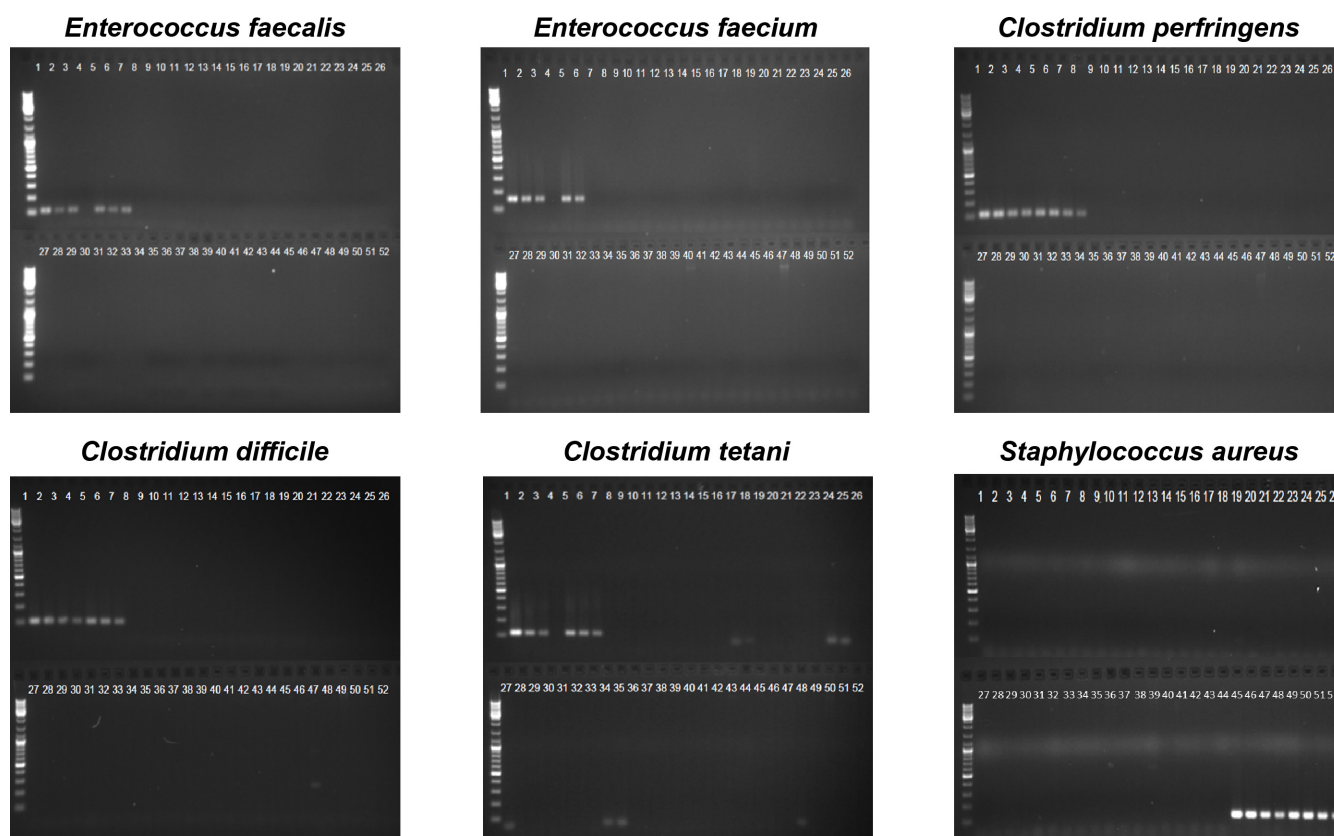
Figure 1. Representative conventional PCR testing of species-specific primer sets (Elf6, Efm12, Cp14, Cd3, Ct3, and Sa2) using 5 pg, 500 fg, 50 fg, and 5 fg of purified target genomic DNA (lanes 1–4 without human DNA and lanes 5–8 with 5 ng of human DNA, respectively, for *E. faecalis*, *E. faecium*, *C. clostridium*, *C. difficile,* and *C. tetani*; lanes 45–48 without human DNA and lanes 49–52 with 5 ng of human DNA for *S. aureus*) and 5 ng of purified DNA from respective nontarget species: *S. pyogenes*, *Streptococcus dysgalactiae* Group G, *Streptococcus agalactiae* Group B, *Streptococcus dysgalactiae* Group C, *Streptococcus* sp. strain H60R Group F, *Streptococcus viridans*, *Streptococcus equi* Group C, *Streptococcus gallolyticus*, *Streptococcus mutans*, *Streptococcus uberis*, *Streptococcus salivarius*, *Streptococcus thermophilus*, *Streptococcus pneumoniae*, *Staphylococcus epidermis*, *Listeria monocytogenes*, *Corynebacterium* sp., *Bacillus subtilis*, *Bacillus cereus*, *Bacillus circulans*, *Bacillus polymyxa*, *Bacillus megaterium*, *Bacillus sphaericus*, *Bacillus thuringiensis*, *Bacillus coagulans*, *Bacillus alvei*, *Escherichia coli*, *Salmonella typhimurium*, *Salmonella* sp. Group D, *Pseudomonas fluorescens*, *Shigella sonnei*, *Enterobacter aerogenes*, *Klebsiella pneumoniae*, *Morganella morganii*, *Proteus mirabilis*, *Alcaligenes odorans*, *Stenotrophomonas maltophilia*, *C. albicans*, *C. tropicalis*, *C. parapsilosis*, human being, and a no template control.

microbial genome sequences available in the public domain. It is important to note that as more new genome sequences become available,[20,21] the currently validated DNA signatures should be reexamined against the new genome sequences to ensure the conservedness. Our approach or protocol of finding species-specific DNA signature should become much more stringent in the future.

In conclusion, we demonstrated that our computational workflow effectively provided a simple genome-wide approach to rapidly identify multiple, highly specific primer sets against six target pathogens. We evaluated individual primer sets under the same condition in real-time PCR assays, these primers and reagents could be easily assembled into a 96- or 384-well plate as a PCR array for multiplex detection of these target bacteria pathogens. Since the capacity of PCR array is high, more validated primer sets for additional pathogens designed in the future could be quickly incorporated for simultaneous detection of an extended list of pathogens of interest. The approach described here could

prove highly valuable in the rapid development of a highly sensitive and specific conventional or real-time PCR assay for the detection of any target pathogen, including those of biothreat agents.

Author Contributions
Conceived and designed the experiments: GCH and SCL. Analyzed the data: KN, GCH, BL and SCL. Wrote the first draft of the manuscript: GCH. Contributed to the writing of the manuscript: KN, GCH, BL and SCL. Agree with
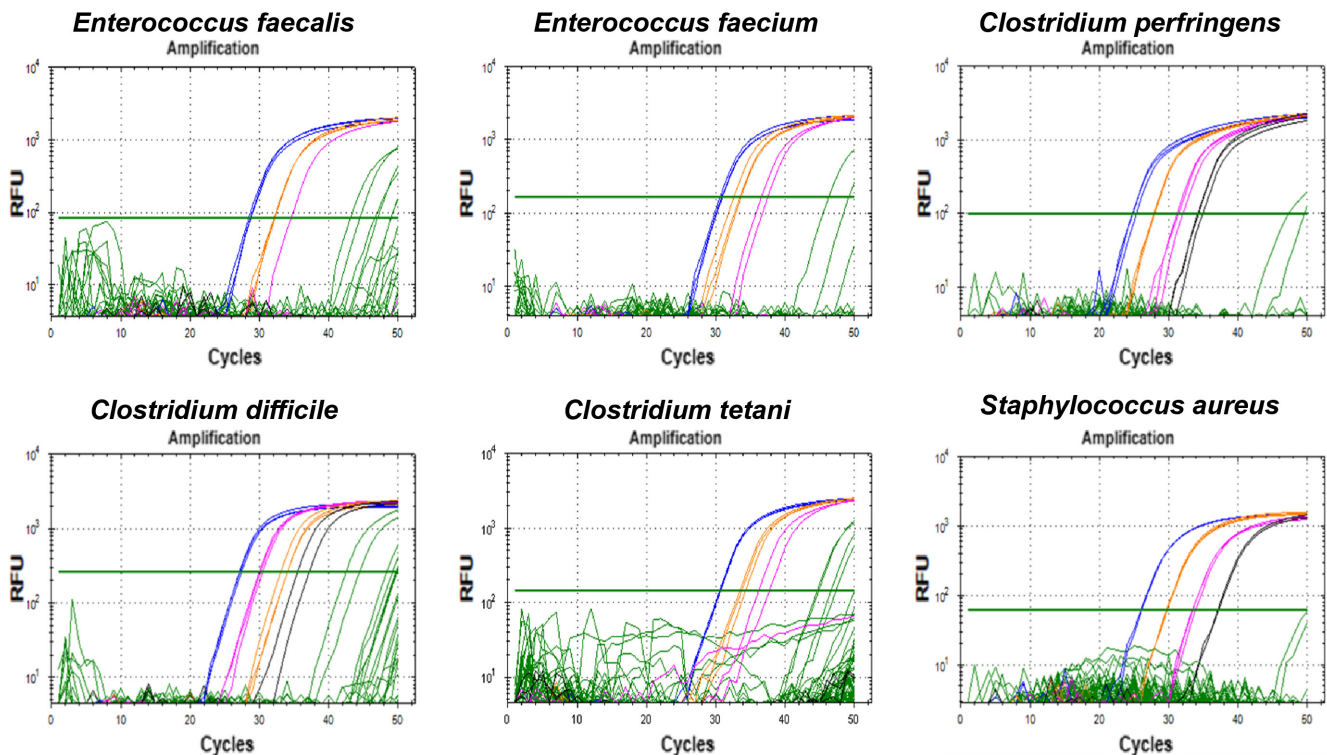
Figure 2. Representative real-time PCR testing conducted with species-specific primer sets (Elf6, Efm12, Cp14, Cd3, Ct3, and Sa2) using 5 pg (blue), 500 fg (orange), 50 fg (pink), and 5 fg (black) of target genomic DNA, and 5 ng of purified DNA from respective nontarget species: *S. pyogenes, Streptococcus dysgalactiae* Group G, *Streptococcus agalactiae* Group B, *S. dysgalactiae* Group C, *Streptococcus* sp. strain H60R Group F, *Streptococcus viridans, Streptococcus equi* Group C, *Streptococcus gallolyticus, Streptococcus mutans, Streptococcus uberis, Streptococcus salivarius, Streptococcus thermophilus, Streptococcus pneumoniae, Staphylococcus epidermis, Listeria monocytogenes, Corynebacterium* sp., *Bacillus subtilis, Bacillus cereus, Bacillus circulans, Bacillus polymyxa, Bacillus megaterium, Bacillus sphaericus, Bacillus thuringiensis, Bacillus coagulans, Bacillus alvei, Escherichia coli, Salmonella typhimurium, Salmonella* sp. Group D, *Pseudomonas fluorescens, Shigella sonnei, Enterobacter aerogenes, Klebsiella pneumoniae, Morganella morganii, Proteus mirabilis, Alcaligenes odorans, Stenotrophomonas maltophilia, C. albicans, C. tropicalis, C. parapsilosis,* human being, and a no template control (green).

manuscript results and conclusions: KN, GCH, BL and SCL. Made critical revisions and approved final version: GCH and SCL. All authors reviewed and approved of the final manuscript.

REFERENCES

1. Volokhov DV, Graham LJ, Brorson KA, Chizhikov VE. *Mycoplasma* testing of cell substrates and biologics: review of alternative non-microbiological techniques. *Mol Cell Probes*. 2011;25:69–77.
2. Wengenack NL, Binnicker MJ. Fungal molecular diagnostics. *Clin Chest Med*. 2009;30:391–408, viii.
3. Wang Z, Malanoski AP, Lin B, et al. Resequencing microarray probe design for typing genetically diverse viruses: human rhinoviruses and enteroviruses. *BMC Genomics*. 2008;9:577.
4. Baldwin CD, Howe GB, Sampath R, et al. Usefulness of multilocus polymerase chain reaction followed by electrospray ionization mass spectrometry to identify a diverse panel of bacterial isolates. *Diagn Microbiol Infect Dis*. 2009;63:403–408.
5. Eshoo MW, Crowder CD, Li H, et al. Detection and identification of *Ehrlichia* species in blood by use of PCR and electrospray ionization mass spectrometry. *J Clin Microbiol*. 2010;48:472–478.
6. Sampath R, Blyn LB, Ecker DJ. Rapid molecular assays for microbial contaminant monitoring in the bioprocess industry. *PDA J Pharm Sci Technol*. 2010;64: 458–464.
7. Albuquerque P, Mendes MV, Santos CL, Moradas-Ferreira P, Tavares F. DNA signature-based approaches for bacterial detection and identification. *Sci Total Environ*. 2009;407:3641–3651.
8. Fukushima H, Tsunomori Y, Seki R. Duplex real-time SYBR green PCR assays for detection of 17 species of food- or waterborne pathogens in stools. *J Clin Microbiol*. 2003;41:5134–5146.
9. Maurin M. Real-time PCR as a diagnostic tool for bacterial diseases. *Expert Rev Mol Diagn*. 2012;12:731–754.
10. Hung GC, Nagamine K, Li B, Lo SC. Identification of DNA signatures suitable for use in development of real-time PCR assays by whole-genome sequence approaches: use of *Streptococcus pyogenes* in a pilot study. *J Clin Microbiol*. 2012;50: 2770–2773.
11. AATB. *Standards for Tissue Banking*. American Association of Tissue Banks, McLean, Virginia; 2012.
12. Phillippy AM, Mason JA, Ayanbule K, et al. Comprehensive DNA signature discovery and validation. *PLoS Comput Biol*. 2007;3:e98.
13. Phillippy AM, Ayanbule K, Edwards NJ, Salzberg SL. Insignia: a DNA signature search web server for diagnostic assay development. *Nucleic Acids Res*. 2009;37:W229–W234.
14. Guo Y, Ribeiro JM, Anderson JM, Bour S. dCAS: a desktop application for cDNA sequence annotation. *Bioinformatics*. 2009;25:1195–1196.
15. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–410.
16. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012;13:134.
17. Zhang Y, Sun Y. A method for de novo nucleic acid diagnostic target discovery. *Bioinformatics*. 2014;30:3174–3180.
18. Brakstad OG, Aasbakk K, Maeland JA. Detection of *Staphylococcus aureus* by polymerase chain reaction amplification of the nuc gene. *J Clin Microbiol*. 1992;30:1654–1660.
19. Ke D, Picard FJ, Martineau F, et al. Development of a PCR assay for rapid detection of enterococci. *J Clin Microbiol*. 1999;37:3497–3503.
20. Lecuit M, Eloit M. The diagnosis of infectious diseases by whole genome next generation sequencing: a new era is open. *Front Cell Infect Microbiol*. 2014;4:25.
21. Sloots TP, Nissen MD, Ginn AN, Iredell JR. Rapid identification of pathogens using molecular techniques. *Pathology*. 2015;47:191–198.