

RESEARCH ARTICLE

Developing a multiplex PCR-based assay kit for bloodstream infection by analyzing genomic big data

Dijun Zhang^{1,2}  | Yong Luo² | Xianping Zeng² | Yunsong Yu^{1,3,4} | Yong Wu²

¹Department of Infectious Diseases, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China

²Ningbo Health Gene Technologies Co., Ltd., Ningbo, China

³Key Laboratory of Microbial Technology and Bioinformatics of Zhejiang Province, Hangzhou, China

⁴Regional Medical Center for National Institute of Respiratory Diseases, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China

Correspondence

Yunsong Yu, Department of Infectious Diseases, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China.

Email: yvys119@zju.edu.cn

Yong Wu, Ningbo Health Gene Technologies Co., Ltd., Ningbo, China.
Email: yong.wu@hgt.cn

Funding information

Ningbo Science and Technology Innovation 2025 Major Special Project, Grant/Award Number: 2019B10056

Abstract

Background: In recent years, the incidence of bloodstream infections (BSI) has increased, the composition of pathogenic bacteria has changed, and drug resistance among bacteria has gradually increased due to the widespread use of interventional techniques, broad-spectrum antibacterial drugs, hormones, and immunosuppressive agents. Here, we have developed a multiplex PCR assay kit for the detection of pathogens (14 Gram-negative bacteria, 15 Gram-positive bacteria, and 4 fungi) in whole blood from patients with BSI using five-color fluorescent multiplex PCR followed by capillary electrophoresis. Our assay exhibits a diagnosis of higher quality and an improved detection rate for common pathogens.

Methods: A local genome DNA database of 33 pathogenic bacteria was constructed. Next, “Exhaustive” primer search of the full coding sequence of the reference genomes of these bacteria was performed. Panels with minimal interactions between primers and amplicons were selected by random sampling and testing by a recursive algorithm. Primers and Mg²⁺ concentrations and PCR reaction procedures were optimized to maximize the detection efficacy.

Results: The LOD of the kit was determined as 100 copies/μl. Using clinical samples, results generated by this kit and regular blood culture method were found to be 95.08% consistent. Additionally, six pathogens which were unidentifiable by blood culture were successfully detected by this kit.

Conclusion: Our study provided a bioinformatics approach to the challenge of primer design in multiplex PCR, and combined with optimized wet lab practice, a multiplex PCR-based assay kit for BSI with higher sensitivity and accuracy than blood culture was produced.

KEYWORDS

alignment, bioinformatics, bloodstream infection, multiplex PCR

1 | INTRODUCTION

Bloodstream infection (BSI) refers to the invasion of various pathogenic microorganisms into the circulation. They then multiply and release metabolites or toxins which induce cytokine release in the host, causing systemic infection and inflammation, and leading to changes in coagulation ability and the fibrinolytic system. In turn, this leads to clinical manifestations of sepsis such as sudden chills, high fever, tachycardia, rash, hepatosplenomegaly, mental and psychiatric changes, as well as shock, diffuse intravascular coagulation, and multi-organ failure in severe cases, even leading to death.¹ BSI has a global incidence of approximately 113–204 cases per 100,000 persons. Regularly, they include *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella*, *Pseudomonas aeruginosa*, *Enterococcus*, *Salmonella*, *Streptococcus*, coagulase-negative *Staphylococcus* (CoNS), and *Candida*, with *E. coli* having the highest detection rate, followed by *S. aureus* and *P. aeruginosa*.^{1–5} Pathogenic composition of hospital-acquired infections is noticeably different from community-acquired ones, with *P. aeruginosa* and CoNS often associated with hospital-acquired infections and *E. coli* and *Streptococcus* commonly associated with community-acquired infections.¹ In most underdeveloped areas, *Salmonella* is the principal pathogen in BSI, with a detection rate of 50% or more, which is significantly higher than developed countries.^{6,7} CoNS is the most common isolate in blood culture. However, as it is a common skin colonizer, inadvertent contamination often occurs, and conventional detection methods are unable to effectively distinguish the source of infection. In general, positive blood cultures from two or more sites are required to make a relatively valid determination.⁸

Currently, blood culture is still the gold standard for diagnosis of BSI. It is generally combined with biochemical identification and 16S rRNA/18S rRNA gene sequencing. However, low detection rates, long turn-around times, and prone to contamination render it insufficient to meet the demand for early and/or rapid clinical diagnosis.⁹ With continued development in molecular biotechnologies, numerous nucleic acid detection technologies for the diagnosis of BSI have emerged, such as fluorescence in situ hybridization (FISH), microarray gene chips, nested multiplex PCR, and quantitative real-time PCR (qRT-PCR).^{10–14}

Multiplex PCR has been widely used in a variety of fields of genetic testing and showed higher detection efficiency and lower cost compared to conventional PCR.^{15–17} However, at least two shortcomings remain. The first is poor consistency of amplification efficiency between different targets,¹⁸ and the second is the false-positive/false-negative caused by primer-primer interactions and primer-amplicon interactions.¹⁹ The rate of these two issues grows exponentially as the number of targets or primers increases, because primers originated from different genomes are pooled together in a multiple PCR reaction. Only one or a few pairs of primers may function as designed to guide extension during pathogen multiplex PCR assays, whereas the rest primers serve as “bystanders”

or even as “troublemakers.” Although these primers can be blocked from disrupting the entire multiplex PCR process by adjusting reaction conditions or cycling parameters, any optimization will not be beneficial if the primers are not proactively designed to prevent interactions.

In this study, we developed a multi-color fluorescence-based multiplex assay kit for BSI detection through bacterial genome analysis and an “Exhaustive” primer search strategy in bacterial entire coding sequences in conjunction with advanced fragment analysis (AFA). The kit can detect selected 14 Gram-negative bacteria including *Acinetobacter baumannii*, *Burkholderia cepacia*, *Enterobacter cloacae*, *E. coli*, *Haemophilus influenzae*, *Klebsiella aerogenes*, *Klebsiella oxytoca*, *K. pneumoniae*, *Moraxella catarrhalis*, *Proteus mirabilis*, *P. aeruginosa*, *Salmonella enterica*, *Serratia marcescens*, and *Stenotrophomonas maltophilia*, selected 15 Gram-positive bacteria including *Enterococcus faecalis*, *Enterococcus faecium*, *S. aureus*, *Staphylococcus capitis*, *Staphylococcus epidermidis*, *Staphylococcus haemolyticus*, *Staphylococcus hominis*, *Streptococcus agalactiae*, *Streptococcus milleri*, *Streptococcus mitis*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus salivarius*, and *Streptococcus sanguinis*, and selected 4 fungi including *Candida albicans*, *Candida glabrata*, *Candida parapsilosis*, and *Candida tropicalis*. Pathogenic detection of BSI by this kit could serve in precision medicine practice and reduce the morbidity and mortality of BSI.

2 | MATERIALS AND METHODS

2.1 | Sequences data download, local database construction, and primer design

Genome sequences and reference genome coding sequences of selected 14 Gram-negative bacteria, 15 Gram-positive bacteria, and 4 fungi were downloaded from NCBI Genome Refseq database by ncbi-genome-download (<https://github.com/kbclin/ncbi-genome-download>) up to date of December 31, 2020. Local genome databases were constructed on a per-species basis (Table 1).

Reference genome coding sequences were entered into Primer3 program as templates for primer design (Table S1).²⁰ Yielded primers with 3'-Dim $\Delta G \geq -2.5$ kcal/mol, P.E. (Priming Efficiency) ≥ 450 , and $59^\circ\text{C} \leq T_m \leq 61^\circ\text{C}$ were qualified by commercial software Oligo7 screening.

2.2 | Primer and amplicon sequence analysis

Each primer for its target bacteria was conducted with a short sequence pattern alignment to its own genome database with 100% coverage and 100% identity. Then, primer pairs which qualified $genome\ number * 95\% \leq number\ of\ hits \leq genome\ number * 105\%$ were extracted and entered the specificity analysis. These primers were conducted with a short sequence pattern alignment against other

TABLE 1 Summary of NCBI data downloaded and local database construction for each target

Target	NCBI Taxonomy	Stains number	Database Volumes	Sequences number	Reference genome assembly accession	CDS number
<i>A. baumannii</i>	403	5466	6	757,281	GCF_008632635.1	3637
<i>B. cepacia</i>	10,703	181	1	22,575	GCF_009586235.1	7524
<i>C. albicans</i>	21	46	1	63,209	GCF_000182965.3	6030
<i>C. glabrata</i>	192	18	1	3214	GCF_000002545.3	5213
<i>C. parapsilosis</i>	930	11	1	1250	GCF_000182765.1	5856
<i>C. tropicalis</i>	212	2	1	879	GCF_000006335.3	6254
<i>E. cloacae</i>	1219	217	1	26,016	GCF_000770155.1	4474
<i>E. faecalis</i>	808	1631	2	105,865	GCF_018986755.2	3412
<i>E. faecium</i>	871	1910	2	343,970	GCF_009734005.1	2780
<i>E. coli</i>	167	21,544	28	3,922,086	GCF_000005845.2	4302
<i>H. influenzae</i>	165	752	1	40,396	GCF_004802225.1	1711
<i>K. aerogenes</i>	3417	269	1	37,021	GCF_007632255.1	4871
<i>K. oxytoca</i>	1165	148	1	16,228	GCF_002984395.1	5530
<i>K. pneumoniae</i>	815	9911	14	1,281,258	GCF_000240185.1	5779
<i>M. catarrhalis</i>	1232	205	1	11,473	GCF_002080125.1	1823
<i>P. mirabilis</i>	1162	389	1	51,550	GCF_000069965.1	3690
<i>P. aeruginosa</i>	187	5710	10	996,548	GCF_000763245.3	6348
<i>S. enterica</i>	152	11,231	14	1,017,576	GCF_000783815.2	4546
<i>S. marcescens</i>	1112	778	2	84,779	GCF_003516165.1	4897
<i>S. aureus</i>	154	10,842	8	648,769	GCF_000013425.1	2767
<i>S. capitis</i>	2054	127	1	11,585	GCF_001028645.1	2329
<i>S. epidermidis</i>	155	2329	2	217,399	GCF_006094375.1	2267
<i>S. haemolyticus</i>	1141	316	1	32,630	GCF_001611955.1	2424
<i>S. hominis</i>	2014	139	1	10,782	GCF_003812505.1	2160
<i>S. maltophilia</i>	880	646	1	110,512	GCF_900475405.1	4012
<i>S. agalactiae</i>	186	1311	1	72,775	GCF_001552035.1	1953
<i>S. milleri</i>	71,339	3	1	4	GCF_900636715.1	1799
<i>S. mitis</i>	530	123	1	5592	GCF_000960005.1	1820
<i>S. mutans</i>	856	244	1	18,015	GCF_009738105.1	1912
<i>S. pneumoniae</i>	176	9076	5	663,203	GCF_002076835.1	2136
<i>S. pyogenes</i>	175	1979	1	40,688	GCF_001267845.1	1686
<i>S. salivarius</i>	507	128	1	7069	GCF_000785515.1	1965
<i>S. sanguinis</i>	1345	44	1	1668	GCF_000191105.1	2192
Total	-	87,726	115	10,623,865	-	120,099

bacteria genome databases with 75% coverage and 75% identity. Then, primer pairs that have no hit in these databases were kept for subsequent analysis.

The amplicons corresponding to the selected primer pairs were BLAST against its own genome database with 100% coverage and 98% identity. Amplicons qualified $genome\ number * 95\% \leq number\ of\ hits \leq genome\ number * 105\%$ and have no base insertions or deletions were selected for subsequent analysis. Meanwhile, amplicons having five or more consecutive mononucleotide repeat will be disqualified together with their corresponding primer pairs.

2.3 | Internal control and human DNA reference primers design

A 100,000 random artificial sequences between 400–600bp in length were generated. Those sequences with even base distribution (24%–26% of each base) and no more than five consecutive mononucleotide repeats were selected as primer design templates for reaction internal control (IC). Human DNA internal control (including GAPDH, β -actin, 18S RNA, B2M, HPRT, and TBP) were used as design templates for human internal references. Primer3 was

used for primer design, and oligo7 was used for screening as previously described. Primer pairs and amplicons met designated standard as previously described and were saved.

2.4 | Panel calculation and screening

Fluorescent color channels were assigned to each bacterium (Table S2). All sets of primer pairs/amplicon for each bacterium, IC, and human DNA internal control were randomly sampled from its set to make a combination, and the compatibility and quality were tested by a recursive algorithm with the following parameters: (1) the minimum spacing between targets in the same fluorescent channel should be greater than or equal to 5 bp, (2) the minimum spacing between targets in different fluorescent color channels should be greater than or equal to 4 bp, (3) the eight bases at the 3' end of each primer within the panel are not allowed to be an exact sequence match with amplicons except itself, (4) the T_m of the interaction between any two primers should be less than 15°C (calculated using calcHeterodimer() in the primer3.py package).²⁰ If the preset compatibility conditions were met, the result will be saved.

Random sampling and recursive algorithm execution produced us 200,000 sets of candidate panels. Panels with the lowest interactions were selected for another round of checking on conservativeness, specificity, and amplicon length using primer-blast.²¹ After all, the optimal panel was formed.

2.5 | Primer synthesis

Primers were synthesized as specified in Table 2. All primers were synthesized by Sangon Biotech (Shanghai, China) in a precise final concentration of 100 μM.

2.6 | Amplified fragment plasmid synthesis and digital PCR quantitation

Amplified fragment plasmids containing the target bacterium, IC, and human DNA internal control were synthesized by TsingKe Biotechnology Co., Ltd. (Beijing, China). The pUC57 vector and the Stab-Top10 expression strain were selected. The synthesized plasmids were precisely quantitated using Droplet Digital PCR (Sniper DQ24). The details were described in Data S1. Based on the quantitated results, all positive control plasmids were diluted to ~400 copies/μl and stored at -20°C for later use.

2.7 | Standard strain culture and genome extraction

Standard target strains from ATCC were resuscitated and cultured in media recommended by ATCC (Table S4). Bacterial genome DNA

extraction kits (Tiangen Biotech, Beijing, China) were used for Gram-negative/Gram-positive bacteria genome extraction, and fungal DNAout2.0 column kits (Tiandz, Beijing, China) were used for fungi genome extraction (Data S2).

2.8 | Primer conservativeness and specificity testing

The reaction system was prepared using the EasyTaq® DNA Polymerase kit (TransGen Biotech, Beijing, China): 1 μl EasyTaq® DNA Polymerase, 5 μl template, 2 μl 10× EasyTaq® Buffer, 2 μl 2.5 mM dNTPs, 2 μl primer mix, 8 μl ddH₂O. Each 100 μl of primer mix contains 1 μl of primers of all the 33 targets bacterium, IC, and human DNA reference control and 30 μl of TE. The PCR amplification procedure was as follows: pre-denaturation at 95°C for 10 min, denaturation at 95°C for 15 s, annealing at 60°C for 30 s, elongation at 72°C for 45 s, 40 cycles, elongation at 72°C for 10 min. After PCR was completed, 1 μl of the amplification product was mixed with 9 μl of loading buffer (containing 1:12 Size-500: deionized formamide) and loaded in an ABI 3500 genetic analyzer with the following parameters: injection voltage: 1.2 kV; injection time: 15 s; electrophoresis time: 1500 s. The PCR amplification template for conservativeness analysis was high concentration positive control plasmids (~400 copies/μl) and target bacterium genome that was extracted and diluted by a factor of 10⁵. The PCR amplification template for specificity analysis was as described above with the addition of other microbial genomes that may be present in BSI (Table S5).

2.9 | Reaction system and procedure optimization

All positive control plasmids (~400 copies/μl) were combined as the positive control plasmid mixture for reaction system and procedure optimization. Effect of reaction parameters was evaluated by equation (1) below, and the reaction parameter set with the highest total panel score was selected for subsequent testing.

$$\text{Total panel score} = \frac{\sum(RFU \div 32,000) \times 100}{35} \quad (1)$$

where RFU is the fluorescence signal of each target in the panel. The maximum value of fluorescence signal on ABI 3500 genetic analyzer is 32,000.

The chosen reaction system, PCR program, and electrophoresis method which applied in primer conservativeness and specificity testing were applied to the positive control plasmid mixture. Primer concentrations were adjusted in the second round of reactions based on the performance of each target in the first-round test and the primer interaction analysis. Finally, the primer concentration combination with the highest score was selected for further optimization.

Mg²⁺ concentration (0–6 mM), PCR annealing temperature (55–65°C), and amplification cycles (30–40 cycles) were optimized by

TABLE 2 Primer information for each target

Target	Forward primer (5'-3')	Reverse primer (5'-3')	Amplicon length (bp)
<i>A. baumannii</i>	VIC-GTTGGCCTAGGACATGAAATC	GTTGTGACTAATTGTTGTAGTACGG	206
<i>B. cepacia</i>	VIC-ATTCTGATCAACAAGGACACGAA	TAGATCGGAATGCCTTCGAAATC	164
<i>C. albicans</i>	FAM-TTGATTCTAGAAGTCGCAGTGAAG	GGTTTGACTGTTTGTAACCTCTTAAT	249
<i>C. glabrata</i>	FAM-ATTCCTCTATTACTCAAGGGTTACA	ACTCTCTGAACTGCTACTCACT	180
<i>C. parapsilosis</i>	FAM-GCCAGCATATTAGGTATAAATCAGG	ATGGTATAATGATATATGGTTGGGATG	198
<i>C. tropicalis</i>	FAM-GGTATGTCGATGAAGATTGCTAATAT	ATTGCCAATACATTCTTGACTCTTG	275
<i>E. cloacae</i>	VIC-CCGCCATCTATGTATAAAGTTATGTAA	AATGAACGGTAAGCATCAGTTG	130
<i>E. faecalis</i>	AF591-GCTCGTCATCTAAATTAATCGGTAA	ATAGCCATGCTTCAAATTGAACA	185
<i>E. faecium</i>	AF591-GACGAAACGAGGAGAAACAATC	GTGCTACTAAATAATTTGGATGGCTT	218
<i>E. coli</i>	VIC-AAATAACTATCGTCCGGAGTTGC	TAAATGTTTCGGGTTATTCAACGG	100
<i>H. influenzae</i>	VIC-GGTTGCTGAAATACAAGATCG	AGCGTGTCTAAAGAGAGTTCC	295
<i>K. aerogenes</i>	VIC-GGCGTATTGGTTAAACTTATCTCA	CCTAATAAATCATGGTGGTAATCGC	260
<i>K. oxytoca</i>	VIC-AGAATCGGCTTAACTCATCTCT	AATTGCGCCCTCGGTTAATA	141
<i>K. pneumoniae</i>	VIC-TGCTCAAATCTGAAAGTAAAGAGAC	GGTTTAGCTGATACTGAGTACCAA	240
<i>M. catarrhalis</i>	FAM-GAGATGTACGATCCGATCACC	TTGCACCAGGATTGATACAATTAAC	268
<i>P. mirabilis</i>	FAM-CTTTGATTACAACAGGCGTACTC	TGGGATCTGTCTTAACCTGTGC	170
<i>P. aeruginosa</i>	VIC-CTACGATGTGTACCAGAAAGTCTA	TTTGTCTTCCAGTATTCTCCAG	289
<i>S. enterica</i>	VIC-CGGTATGGGCTGACGAAATA	AACGTATCATCATAAGCGCTTT	119
<i>S. marcescens</i>	VIC-TGATTTCTCTATTTTCGCTCCCTAC	CGTTTCATGCTGCTGTTCAAA	154
<i>S. aureus</i>	NED-TTATTTGCAAGGCTATGATGAACC	TCTCCACTTCTAATTCAAACATCAT	135
<i>S. capitis</i>	NED-AGACGATGGGTAAGTTAGCAATG	GCTCTCCGGTTAAATAATAATAGACAC	264
<i>S. epidermidis</i>	NED-GTTTATTTACTGGTATCATTGGTCCTT	TCATAAACATACATATCACGCCTACT	150
<i>S. haemolyticus</i>	NED-ATGTTAAGCAGTTAATTAGATGTTGTCG	TAAATGTTTCCATACTCGCACCT	124
<i>S. hominis</i>	NED-TAGGTATAGAAAGAGAACAAGTAGGAGA	CTAATCTTAAAGCACTGACTGTACTATG	210
<i>S. maltophilia</i>	FAM-CATGCAGCCAAAGAAATACGC	TTGTCCAATCCCTTGATGTCC	111
<i>S. agalactiae</i>	AF591-GCGGCTTACATTAATGGTTTCTTAA	TAATAAGCGGTGAACAATACGTG	300
<i>S. milleri</i>	ROX-AAAGTTTGATTATGAGATGGTCTT	TTTCTCTGTCATCGTACCATTTAATTG	146
<i>S. mitis</i>	ROX-AGTATTATCTGTTGTGATTGTGATTGC	TCTACTTGATAATACCGTCAATGCTA	202
<i>S. mutans</i>	ROX-GGCTTTCCTACTATGTCAAATTCA	CTACTCCTAACGTTGAGGCTTT	190
<i>S. pneumoniae</i>	AF591-AAGTTTATCGACGAGTGATTATTCCTAA	CGATAATCAGCTCCACCTAGAAT	160
<i>S. pyogenes</i>	AF591-GCATCAACCATGTTATAAACCTGTG	CTTCATACCAATAGATGCATTACTATCA	107
<i>S. salivarius</i>	ROX-TTATCAGCACTCTGAATTCATTGT	GGTTTCAATTGTATCAAGAGCTCTAAC	228
<i>S. sanguinis</i>	ROX-ACTGATTTCCGGTGTCTTGTC	GATGAGAATTCCTAACGCGAA	115
Human_DNA	AF591-TGAAGATGCCGCATTTGGAT	TCTTAACAAGCTTTGAGTGCAAG	310
IC	NED-AGGGCGATCATTTCATAACCTC	ATCAGCGGATTACCCAGATAATG	220

gradient experiments. The condition which produced the highest total panel score and non-specific amplification observed was considered the best.

2.10 | Determination of LOD

Using a positive control plasmid concentration series, the LOD of the kit was determined based on the optimized primer concentration, Mg²⁺ concentration, and reaction procedure. The test was repeated 20 times for each concentration. The LOD was defined as the

lowest concentration when 19 of 20 repeated tests were successful (RFU ≥ 2000).

2.11 | Testing of clinical specimens

A total of 122 clinical blood samples, including 20 positives and 102 negatives by culture, were continuously collected starting from 1 February 2022 to 15 February 2022 in Yinzhou People's Hospital, Ningbo, China. Positive cultures included three cases of *E. coli*, two cases each of *S. capitis*, *B. cepacia*, *S. aureus*, *S. hominis*, *E. faecium*,

C. albicans, and *S. enterica*, and one each of *P. mirabilis*, *E. cloacae*, and *S. marcescens*. All clinical samples were extracted twice using the bacteria genome DNA extraction kits and fungal DNAout2.0 column kits, and the optimized kit was used to test the samples to see if the result was consistent with that of the culture. In case of poor consistency, sequencing using bacterial 16S rRNA universal primer amplification (27F: AGA GTT TGA TCM TGG CTC AG; 1492R: GGT TAC CTT GTT ACG ACT T) or fungal ITS universal primer amplification (ITS1: TCC GTA GGT GAA CCT GCG G; ITS4: TCC GCT TAT TGA TAT GC) was performed for verification. The study was conducted with the approval of the Ethics Committee of the Yinzhou People's Hospital.

3 | RESULTS

3.1 | Primer design and screening

A total of 1,451,686,498 primer pairs were produced, and on average, 12,087 primer pairs were generated for each coding sequence (Figure 1). Due to its large genome size and longer single coding

sequences, *Candida* had highest total number of primers compared to other targets, reaching a total of 10^8 primers. Due to high GC ratio, the total number of primers designed for *S. maltophilia* (66.4%), *P. aeruginosa* (66.2%), and *B. cepacia* (66.6%) was much lower than others. After screening by Oligo7 program (Figure 2A), a total of 276,288,532 primer pairs were qualified, accounting for 19.03% of the total primer pairs generated, with *H. influenzae* (19.91%) having the highest qualification rate and *P. aeruginosa* (16.30%) having the lowest.

Primer conservativeness analysis results are shown in Figure 2B. A 76,363,335 primer pairs met our designated conservativeness standard, which yielded a qualification rate of 5.26% among all primer pairs generated. The conservativeness of *C. parapsilosis* and *C. glabrata* primers was much higher than others, at 17.59% and 14.97%, respectively; they are eukaryotic organisms with low mutation rates, and the database contains only a small number of genomes, which are more homogeneous and less complex comparing to others. Strains with qualification rate below 1% are primarily concentrated among the Viridans Streptococci (*S. mitis*: 0.09%, *S. sanguinis*: 0.14%, *S. salivarius*: 0.40%, *S. milleri*: 0.54%) and Enterobacteria (*E. cloacae*: 0.10%, *E. coli*: 0.24%, *K. oxytoca*: 0.44%, *H. influenzae*: 0.70%), suggesting that mutation rate in these bacteria is higher. A

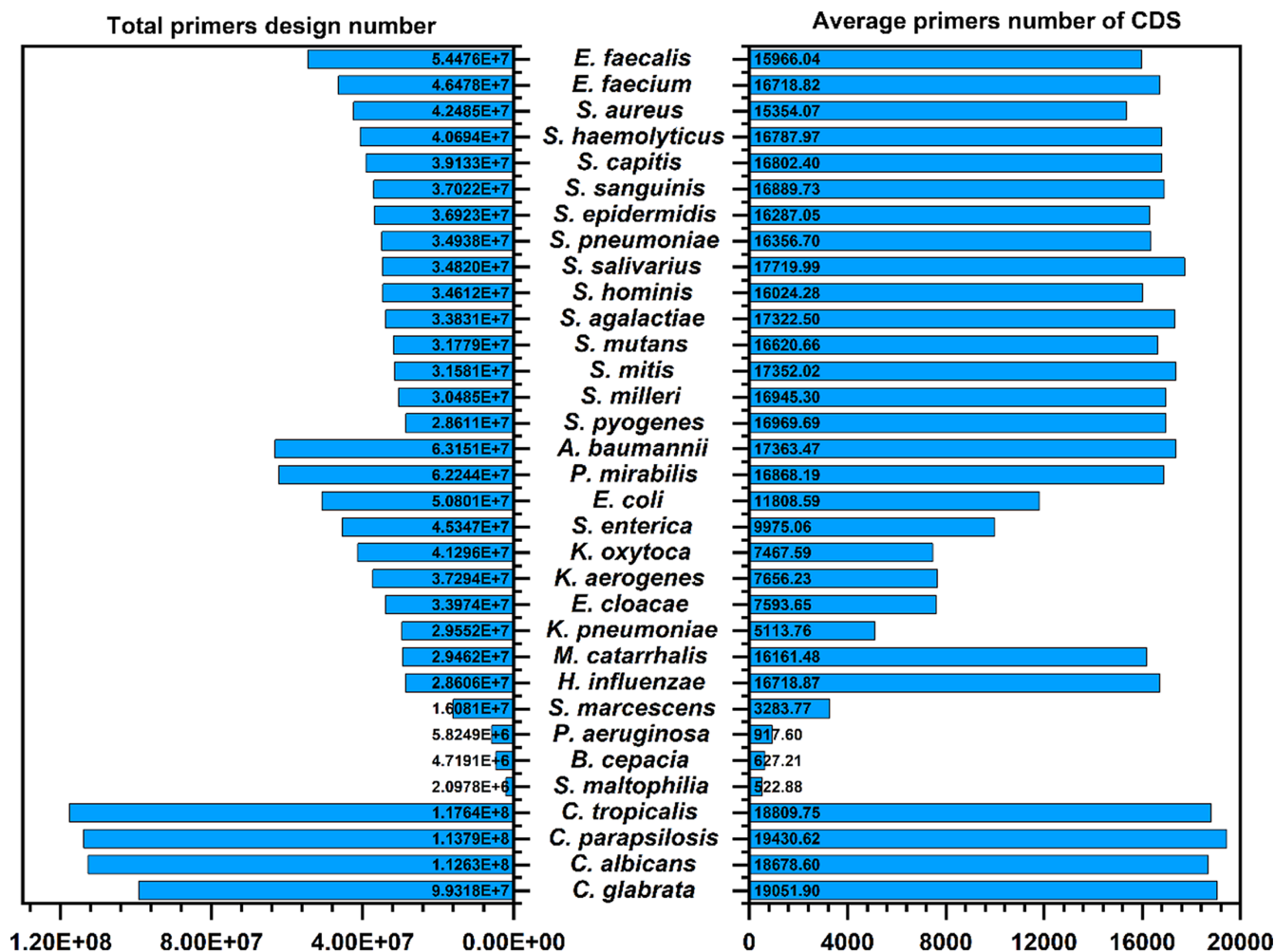


FIGURE 1 Total and average number of primer output for coding sequences of each target.

3,371,930 primer pairs met designated specificity standard, which yielded a qualification rate of 0.23% among all primer pairs generated and 4.41% among the conserved primers (Figure 2C). Highest primer qualification rates were shown in *S. agalactiae* (1.90%), *S. mutans* (1.27%), and *S. pyogenes* (0.985%) and lowest in *E. cloacae* (4.42E-07), *S. mitis* (1.99E-06), and *K. oxytoca* (2.49E-06).

By amplicon sequence conservativeness analysis (Figure 3A), the number of qualified primer pairs was 2,886,769, accounting for 0.20% of all primer pairs generated and 85.61% of the primer pairs qualified by specificity analysis. In addition, amplicon sequences with five or more mononucleotide repeats were disqualified (Figure 3B). The final number of candidate primer pairs was 1,232,128, accounting for 0.08% of all primer pairs generated and 42.68% of the primer pairs qualified by amplicon sequence conservation analysis.

3.2 | IC and human DNA internal reference primer design and screening

A trivial computer program helped generated 100,000 random ATGC character strings with their length ranging between 400–600. Twenty three of them were screened out by criteria of uniformed base distribution and no more than 4 mononucleotide repeats. A total of 334,167 primer pairs for selected random sequences were produced by Primer3. Among them, 58,479 primers pairs were qualified after screening by the same conditions using Oligo7. A total of 2598 primers pairs were qualified after blast against the 33 target genome databases.

A total of 107,370 primer pairs for the human DNA internal reference gene were produced by Primer3. A total of 18,063 primers pairs were qualified after screening by the same conditions using Oligo7. A total of 3539 primers pairs were qualified after blast against the 33 target genome databases.

3.3 | Panel construction and interactions examination

Based on a random sampling and recursive algorithm, 200,000 panels were generated. One panel with the least primer-primer/primer-amplicon interactions was selected (Figure 4). In Figure 5, calculated interaction scores were transferred into blue color gradient with darker end indicating stronger interactions which is undesirable. It suggested the primers concentrations in PCR and annealing temperature need to be concerned with the LOD and the non-specific amplification.

3.4 | Evaluation of primer conservativeness and specificity

First, primers were tested using constructed plasmid as positive control. As shown in Figure 6A, all primers achieved nice PCR result with

little non-specific product signals. They were then tested using positive target genomes. And again, they performed well and presented no off-target or non-specific product signals (Figure 6B). Finally, there is also no non-specific signal observed when negative strain genomes were applied in testing. So, the chosen panel primer set is well off for further kit optimization.

3.5 | Reaction system and procedure optimization

Primer concentrations were adjusted and optimized (Table S6) according to the total panel score after each round (Figure 7A). The total panel score reached a maximum after the fourth round of primer concentration adjustment.

To determine the preferred Mg^{2+} concentration of PCR buffer, we conducted experiment using a series buffer with gradient Mg^{2+} concentrations (Figure 7B). Results showed that the total panel score increased with Mg^{2+} concentration while it is below 3mM and decreased with it while above 4mM. Particularly at 6mM, the total panel score is even lower than buffer completely without Mg^{2+} . This indicated that the effect of Mg^{2+} concentration on PCR amplification efficiency depends on its range. We determined the preferred Mg^{2+} concentration is 3mM.

In determining the preferred annealing temperature, it showed a trend that the total panel score decreased as the temperature going up (Figure 7C). More non-specific signals were encountered while it is below 60°C, and they were shrinking as the temperature increased and completely disappeared when temperature reached 61°C. Thus, 61°C was chosen as the optimized value.

As shown in Figure 7D, with given plasmid concentration, the total panel score went up when the number of PCR cycles increased. However, non-specific amplification started to show up after 37 cycles, which may cause some difficulties in result interpretation. Finally, the PCR cycle number value was optimized at 35.

3.6 | Determination of LOD

Applying a gradient series of positive control plasmid as reference samples, PCR reactions were performed using above optimized conditions. LOD was the reference concentration when fluorescence signal intensity reached its recognition limit (Figure 8). The results showed that targets with LOD of 100copies/ μ l are *S. maltophilia*, *C. parapsilosis*, *C. albicans*, *E. coli*, *S. enterica*, *B. cepacia*, *A. baumannii*, *H. influenzae*, *S. haemolyticus*, *S. capitis*, and *S. salivarius*. Targets with LOD of 50copies/ μ l are *C. glabrata*, *M. catarrhalis*, *C. tropicalis*, *K. oxytoca*, *K. pneumoniae*, *P. aeruginosa*, *S. aureus*, *S. hominis*, *S. mutans*, *S. mitis*, *S. pyogenes*, *S. pneumoniae*, and *E. faecium*. Targets with LOD of 25copies/ μ l are *P. mirabilis*, *E. cloacae*, *S. marcescens*, *K. aerogenes*, *S. epidermidis*, *S. sanguinis*, *S. milleri*, *E. faecalis*, and *S. agalactiae*. Overall, the LOD of the optimized BSI multiplex PCR-based assay kit was 100copies/ μ l.

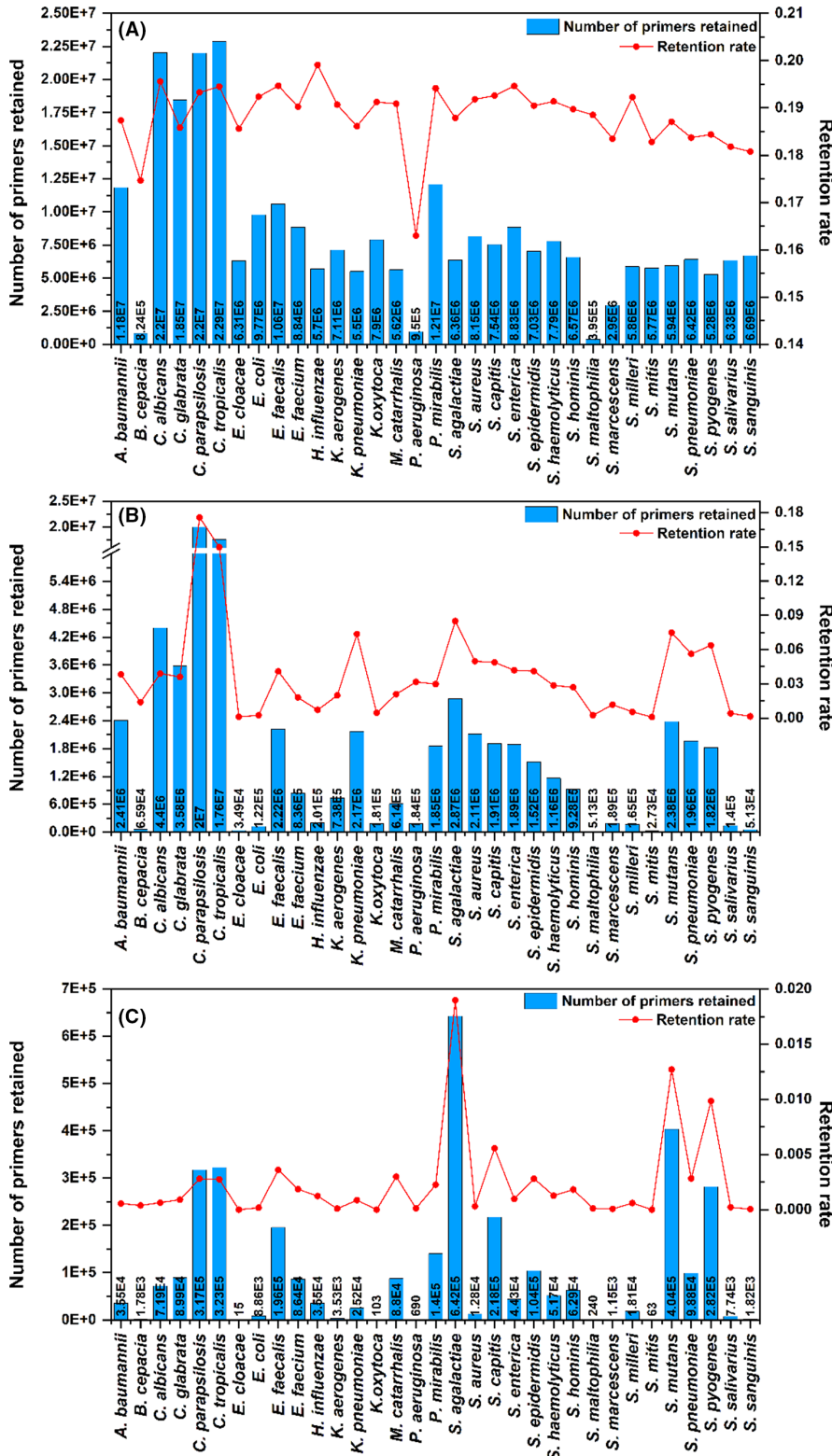


FIGURE 2 Primer screening results. (A) primers qualified after Oligo7 screening; (B) primers qualified after conservation analysis; (C) primers qualified after specificity analysis.

3.7 | Testing of clinical specimens

Total 122 clinical blood samples from Yinzhou People's Hospital were tested using the BSI multiplex PCR-based assay kit. The results were compared to the results of blood culture. In case if they are not consistent, bacterial 16S rRNA or fungal 18S rRNA sequencing

verification was performed (Table 3). Twenty blood culture-positive samples were consistent with results by the kit; 6 of 102 negative samples were tested positive by the kit, namely *E. coli* in sample YZ-10, *S. marcescens* in YZ-46, *C. albicans* in YZ-50, *S. aureus* in YZ-52, *E. faecalis* in YZ-71, and *E. fumigatus* in YZ-101. In sequencing validated samples, sequencing results are in accord with results by the kit

FIGURE 3 Amplicon screening results. (A) Primers qualified after amplicon conservation analysis; (B) primers qualified after mononucleotide repeat analysis.

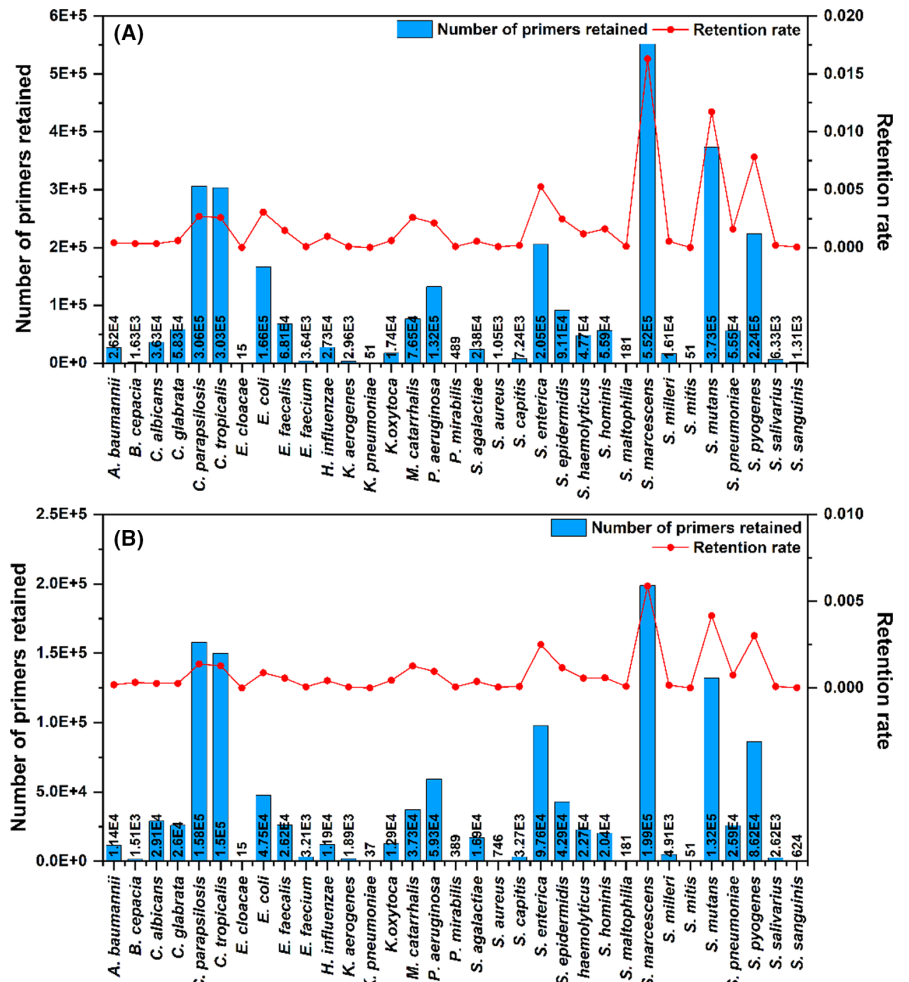
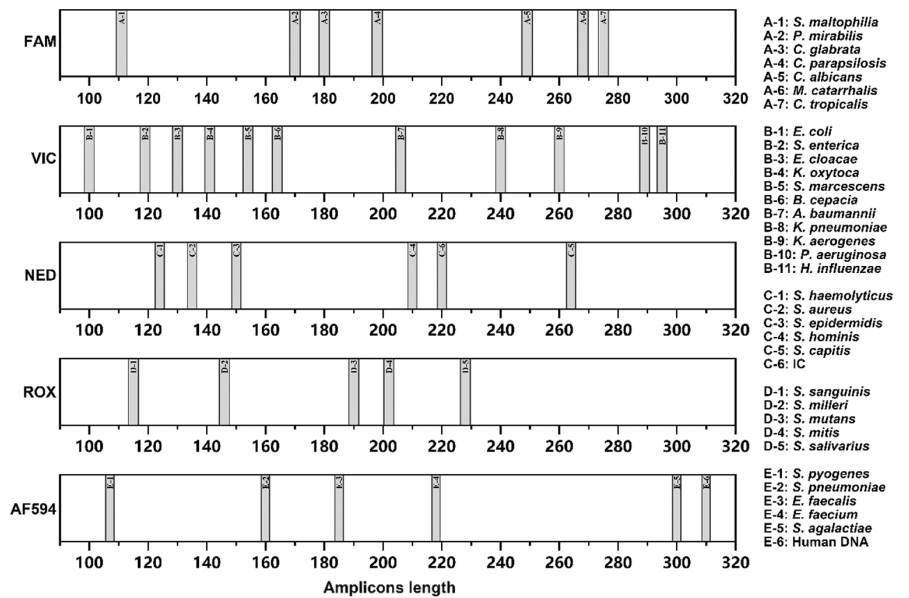


FIGURE 4 Panel distribution. The grey area indicates the expected amplicon length ± 1.5 bp.



except for sample YZ-10, which is likely failed due to an unsuccessful blood culture. The fluorescence signal intensity of YZ-10 was very weak by using the kit. The results were confirmed to be consistent by re-testing. Results produced by this BSI multiplex PCR-based assay

kit developed in this study were 95.08% consistent with clinical blood culture results. Moreover, the assay kit was able to detect pathogenic microorganisms that would be failed to detect by blood culture method owing to the greater detection sensitivity of the assay kit.

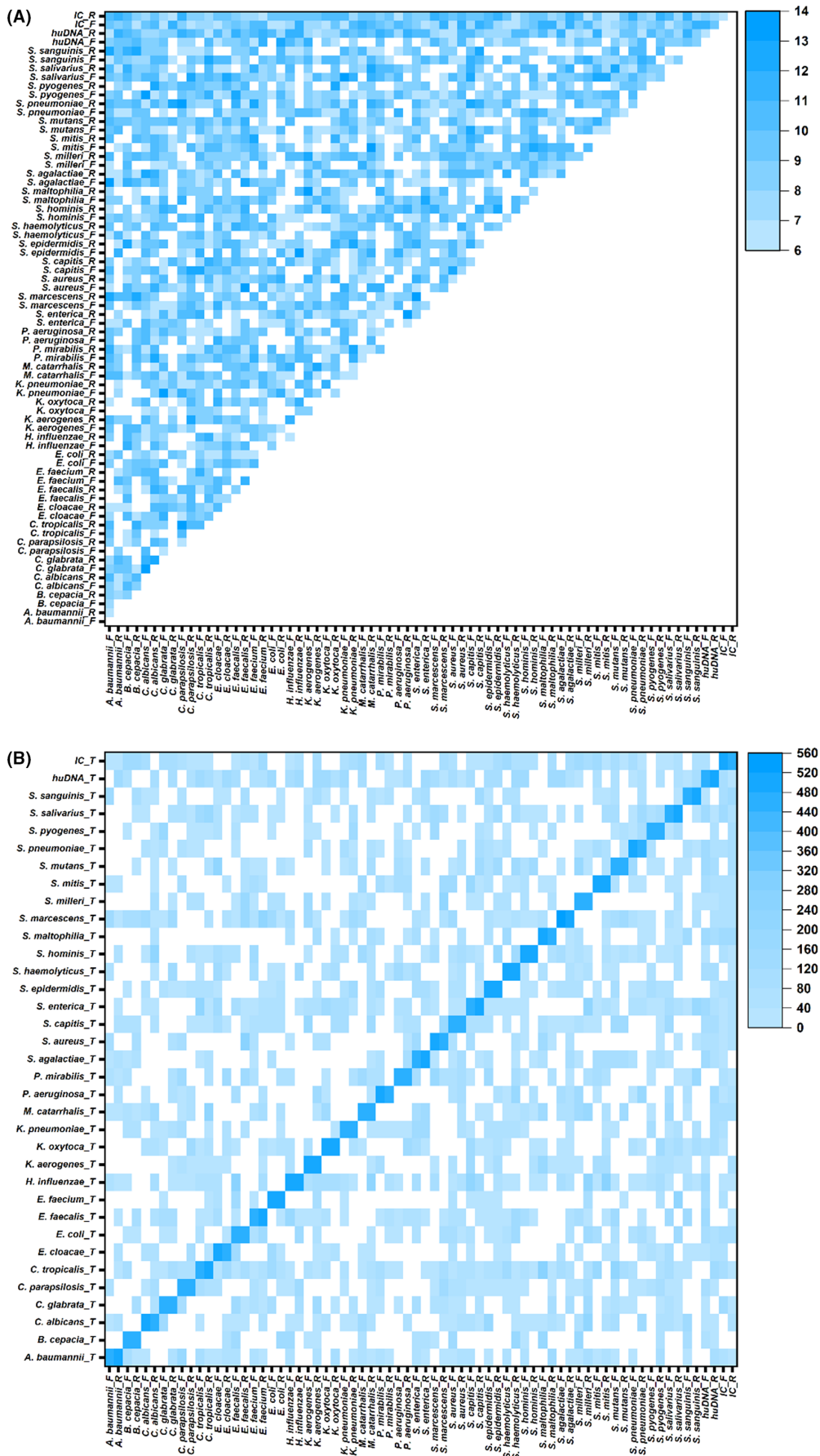
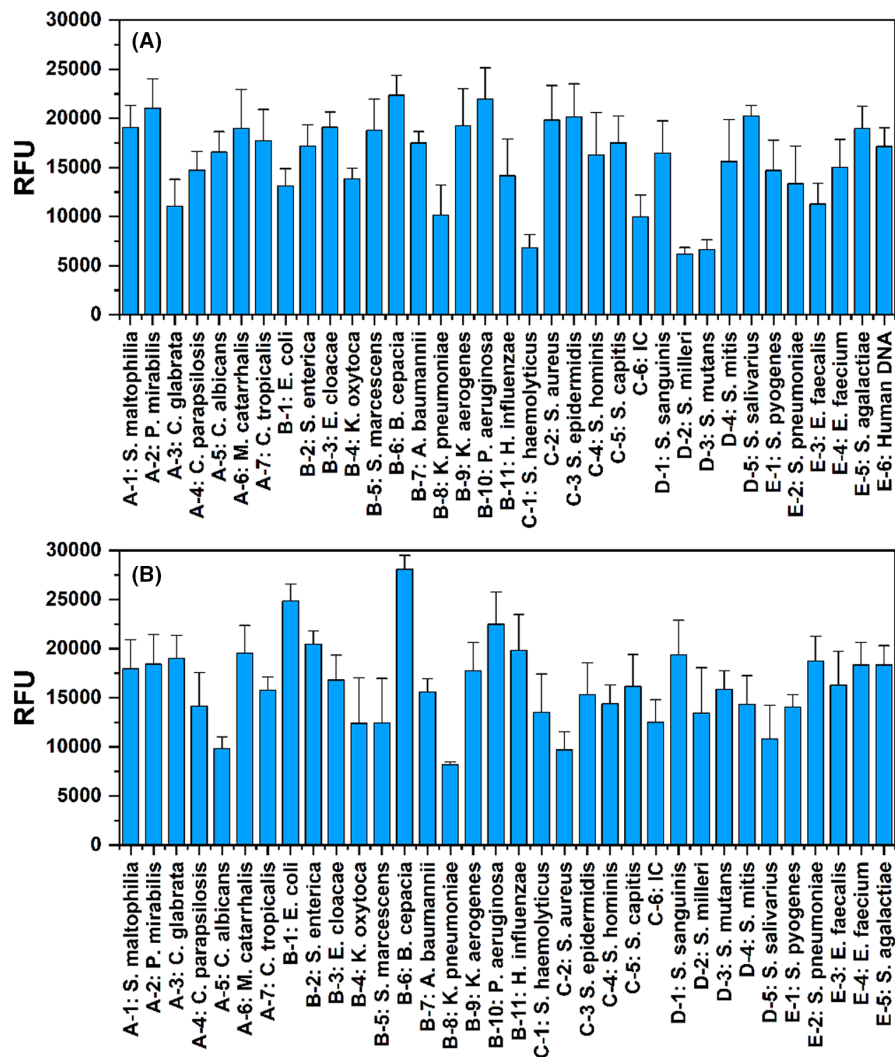


FIGURE 5 Evaluation of panel/bin interactions. (A) Primer-primer interactions analyzed by Autodimer; (B) primer-amplicon interactions analyzed by Oligo 7.

FIGURE 6 Results of primer conservation tests. (A) Synthetic plasmids; (B) standard strain genomes.



4 | DISCUSSION

Multiplex PCR is widely used in many fields of genetic testing because of its higher throughput and lower cost compared to conventional PCR. However, while the complexity of the multiplex PCR is positively correlated to the number of folds of the reaction system, the correlation to detection accuracy and sensitivity are negative. In our study, required characteristics for a primer in multiplex PCR were divided into different aspects or levels. To qualify, they must satisfy the requirement as single primer oligo, as in primer pairs, compatible for unpaired primers, for primers and non-target amplicons, and global compatibility among all primers within the multiplex PCR reactions. Details were analyzed as follows.

For an oligo to be primer, there are known restraint including base composition (GC%), number of mononucleotide repeats, T_m, 3' terminal stability, hairpin structures, and the possibility of homodimer formation. Primer design software performs various checks on the oligo sequence during primer generation, including but not limited to these required restraints. Among the most popular primer design software such as Primer premier 6.0, Oligo7, Primer3, Dnastar, and Vector NTI Suit, only Primer3 provides a Linux version to address the needs of high-throughput primer design. On the other hand, Oligo7

has primer database function that provides screening for 3'Dim ΔG, priming efficiency, and T_m calculation. Primer sequences ought to satisfy all above-mentioned restraint through intensive bioinformatics data processing pipeline of software tools.

The requirements to qualify primer in matching pairs are based primarily on the conservativeness of the pair (i.e., whether a primer pair have the correct matching sequence region in the target genome), the specificity of the primer pair (i.e., whether a primer pair have matching sequence region only on target genome without mistakenly matches to similar ones, especially those closely related bacteria, for example, the distinction between *E. faecium* and *E. faecalis*), and the base composition of the amplicon. For the conservativeness check of primer pairs, we constructed a genome database containing all targets. Each primer pair was checked, and a comprehensive evaluation was made according to sequence coverage, sequence identity, and number of blast hits of the primers in the genome database to ensure that they have covered the variety of target genomes well. For the specificity check of primer pairs, each primer was compared to the genome database which contains all but its target genome separately, to ensure the primer pair has no possible matching region on all other genomes except its target. Moreover, we compared each amplicon with its own genome database to exclude amplicons with

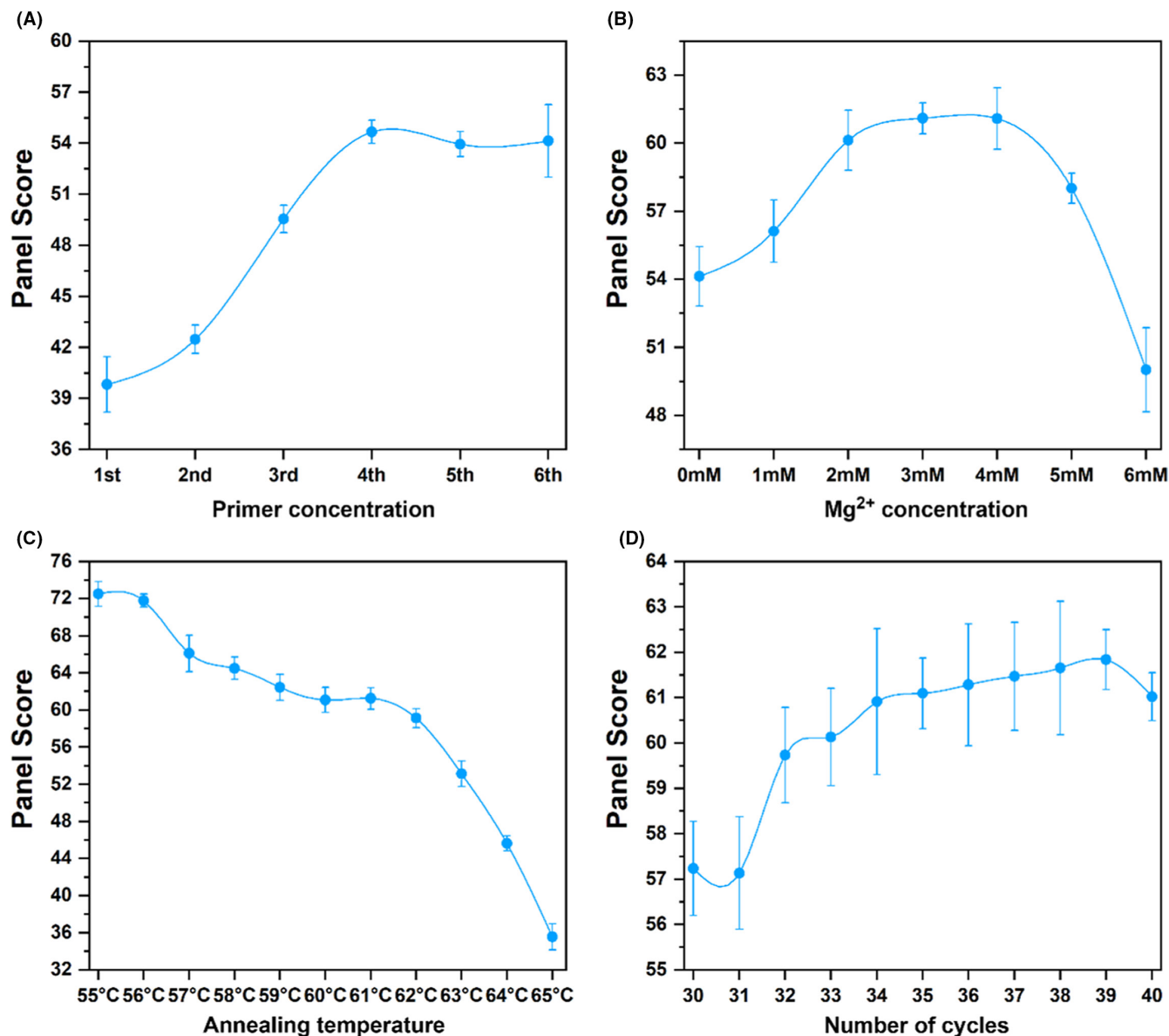


FIGURE 7 Reaction system and PCR program optimization results. (A) Primer concentration optimization; (B) additional Mg²⁺ concentration optimization; (C) annealing temperature optimization; (D) amplification cycle number optimization.

possible length variations. In addition, we decided that amplicons with over four mononucleotide repeats should be disqualified. The reason is that there is certain probability of strand slippage during PCR if a mononucleotide repeat is encountered, and a stutter product with one or more nucleotides fewer than the correct amplicon will be generated.²² Such an amplicon is fatal for the fragment analysis method based on length as it can lead to false-negative results.

The requirements for unpaired primers primarily come from limiting the interactions between primers in the panel. As previously mentioned, primer-primer interactions can have a major impact on the sensitivity and accuracy of multiplex PCR based assay. Software like AutoDimer,²³ the calcHeterodimer() function in primer3-py, and Oligo7 provide multiple primer interaction analysis functions. In this study, we selected calcHeterodimer() to make thermodynamic calculations of primer-primer interactions during panel screening and

applied AutoDimer for base-pairing evaluations in primer-primer interactions during panel re-examination. This joined effort successfully minimized primer-primer interactions and their impact on the final testing sensitivity and accuracy.

For concerns of interaction between a primer and amplicons of other targets, the probability of primer binding to an unrelated amplicon will rise along PCR cycles due to its quantity accumulation. If a primer binds to other target's amplicons and primed the DNA synthesis in one cycle, subsequent cycles will amplify this non-specific amplicon exponentially. This could eventually develop as a signal peak of that amplicon length, and if the length coincides with the expected amplicon length of another target in the panel, a false-positive result would appear. In our study, two approaches were taken to evaluate and further restrict the binding propensity of primers to amplicons of other genome targets. During panel

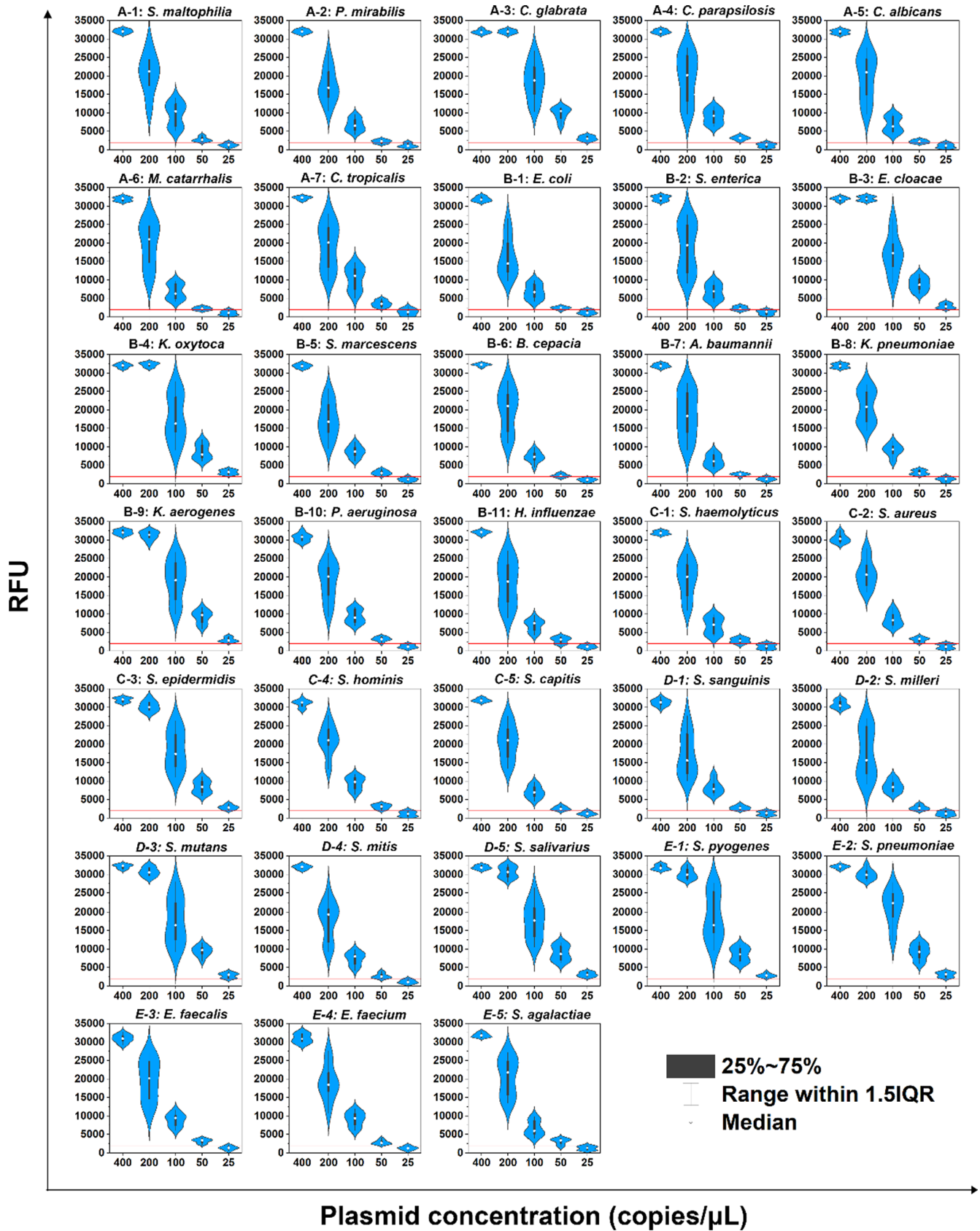


FIGURE 8 LOD of the optimized kit.

Sample number	Blood culture result	Kit result ^a	Sequencing result	Result ^b
YZ-2	Negative	<i>S. enterica</i>	<i>S. enterica</i>	Inconsistent
YZ-5	<i>B. cepacia</i>	<i>B. cepacia</i>	–	Consistent
YZ-10	Negative	<i>E. coli</i>	Sequencing failed	Inconsistent
YZ-11	<i>B. cepacia</i>	<i>B. cepacia</i>	–	Consistent
YZ-45	<i>P. mirabilis</i>	<i>P. mirabilis</i>	–	Consistent
YZ-46	Negative	<i>S. marcescens</i>	<i>S. marcescens</i>	Inconsistent
YZ-49	<i>S. enterica</i>	<i>S. enterica</i>	–	Consistent
YZ-50	Negative	<i>C. albicans</i>	<i>C. albicans</i>	Inconsistent
YZ-51	<i>S. enterica</i>	<i>S. enterica</i>	–	Consistent
YZ-52	Negative	<i>S. aureus</i>	<i>S. aureus</i>	Inconsistent
YZ-58	<i>E. cloacae</i>	<i>E. cloacae</i>	–	Consistent
YZ-60	<i>S. marcescens</i>	<i>S. marcescens</i>	–	Consistent
YZ-64	<i>E. coli</i>	<i>E. coli</i>	–	Consistent
YZ-65	<i>C. albicans</i>	<i>C. albicans</i>	–	Consistent
YZ-66	<i>S. capitis</i>	<i>S. capitis</i>	–	Consistent
YZ-69	<i>E. faecium</i>	<i>E. faecium</i>	–	Consistent
YZ-71	Negative	<i>E. faecalis</i>	<i>E. faecalis</i>	Inconsistent
YZ-72	<i>S. hominis</i>	<i>S. hominis</i>	–	Consistent
YZ-75	<i>C. albicans</i>	<i>C. albicans</i>	–	Consistent
YZ-89	<i>S. aureus</i>	<i>S. aureus</i>	–	Consistent
YZ-91	<i>E. coli</i>	<i>E. coli</i>	–	Consistent
YZ-92	<i>S. capitis</i>	<i>S. capitis</i>	–	Consistent
YZ-93	<i>S. hominis</i>	<i>S. hominis</i>	–	Consistent
YZ-100	<i>E. faecium</i>	<i>E. faecium</i>	–	Consistent
YZ-101	Negative	<i>E. coli</i>	<i>E. coli</i>	Inconsistent
YZ-105	<i>S. aureus</i>	<i>S. aureus</i>	–	Consistent
YZ-112	<i>E. coli</i>	<i>E. coli</i>	–	Consistent

TABLE 3 Inconsistency between positive blood culture specimens and test results

^aKit results are the combined results of two extraction methods due to different targeting between the two methods, so the results for *Candida* are inconsistent.

^bBlood culture results are consistent with kit results.

screening, 8 bases at the 3' end of each primer were restricted to have the same sequence in to be inspected amplicons. Secondly, during panel re-examination, the P.E. and T_m values calculated in the Oligo7 were used to check the priming efficiency and annealing T_m value of primer on amplicon. Higher T_m values indicate higher likelihood of binding; and higher P.E. values indicate the likelihood resulting non-specific PCR product.

Based on these analyses, we obtained a panel that satisfied all requirements and adapted to multiplex PCR reaction conditions in a maximal way. Bioinformatics analysis methods and software tools are in the center played a crucial role in finalizing this panel.

For such a complicated and highly restricted multiplex PCR-primer screening process described here, the number of primers generated by conventional primer design is far from sufficient. Many paired primer candidates are needed to complete the above analysis process if more primer sets are required. In our study, an "Exhaustive" primer design strategy was applied to the entire coding sequences of reference genome, and the number of primer pairs

returned by Primer3 was up to 20,000 to obtain as many primer pairs as possible. With the help of designated hardware computing power, parallel computing system, and NCBI BLAST program, we completed the conservativeness and specificity screening of 1.4 billion primer pairs, roughly equal to the population of China. Primers passed the screening are approximately one in a million. Since the number of final candidate primers is still on the level of millions thereafter, the challenge of panel screening remains. Given the number of targets and average number of primer pairs for each of them, there would be 6×10^{133} possible combinations as the number of panel design, which larger than the number of atoms in the universe, unachievable for "Exhaustive" panel screening or even for name listing. Thus, random sampling and recursive methods were designed for panel construction and screening in our study. Only 200,000 lucky sets of primers that satisfy the predefined conditions were listed, and the best panel was selected for experimental validation.

Result showed that the selected panel designed/optimized through bioinformatics big data analysis, with PCR reaction

system optimized, can detect 33 common pathogens involved in BSI, and its sensitivity is higher than of the conventional blood culture method which is the gold standard and still popularly used everywhere.

Every coin has two sides, no matter how perfect the primer design is and PCR reaction system and procedure is tuned, the interaction between primers and amplicons still exist. As a result, there is still certain possibility of false-positive and false-negative by this kit. Room for improvement is obvious. Moreover, although the sensitivity of this kit is higher than blood culture method, it is still a weakness when compared to qRT-PCR.¹⁴

5 | CONCLUSION

We have constructed a local genome database of 33 pathogens, including 14 Gram-negative bacteria, 15 Gram-positive bacteria, and 4 fungi. We conducted an "Exhaustive" primer search and design on the entire coding sequences of their reference genomes. Their molecule structures, sequences conservativeness and specificity, and base compositions of the primers were all examined, and a panel with minimal interactions between primers and amplicons was selected by random sampling and recursive algorithms for further experimental validation. The conservativeness and specificity of primers were verified by applying positive control plasmids, positive standard strains, and negative standard strains. Primer concentrations, Mg²⁺ concentration, and PCR program were optimized to achieve maximum detection efficiency and an overall lower limit of detection at 100 copies/μl. On clinical specimen testing, results by this kit hold a consistency rate comparing to blood culture method at 95.08%. Furthermore, six cases of detection of organisms by this assay kit were successful while blood culture method failed.

The development of the kit in the present study involved combining intensive computing and wet lab experiments with support from bioinformatics analysis, which helped to prevent serious problems that may be very challenging in doing wet lab experiments. It significantly reduced the probability of unexpected scenario in the experiments, reduced experimental costs, shortened the experimental cycle, and greatly improved the quality of the kit.

AUTHOR CONTRIBUTIONS

Yong Wu and Yunsong Yu designed the experiments and edited the manuscript. Dijun Zhang and Xianping Zeng performed all experimental verification. Dijun Zhang and Yong Luo performed the bioinformatics analysis. Dijun Zhang wrote this article. All the authors have read and approved the final version of this manuscript.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to Yinzhou People's Hospital for providing the clinical samples and Editage (<https://www.editage.cn/>) for the expert linguistic services provided.

FUNDING INFORMATION

This work was supported by Ningbo Science and Technology Innovation 2025 Major Special Project (Grant NO. 2019B10056).

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Dijun Zhang  <https://orcid.org/0000-0003-3417-9849>

REFERENCES

1. Laupland KB, Church DL. Population-based epidemiology and microbiology of community-onset bloodstream infections. *Clin Microbiol Rev.* 2014;27(4):647-664.
2. Thrift AG, Thayabaranathan T, Howard G, et al. Global stroke statistics. *Int J Stroke.* 2017;12(1):13-32.
3. Scheres LJJ, Lijfering WM, Cannegieter SC. Current and future burden of venous thrombosis: not simply predictable. *Res Pract Thromb Haemost.* 2018;2(2):199-208.
4. Conrad N, Judge A, Tran J, et al. Temporal trends and patterns in heart failure incidence: a population-based study of 4 million individuals. *Lancet.* 2018;391(10120):572-580.
5. Anderson DJ, Moehring RW, Sloane R, et al. Bloodstream infections in community hospitals in the 21st century: a multicenter cohort study. *PLoS One.* 2014;9(3):e91713.
6. Marchello CS, Hong CY, Crump JA. Global typhoid fever incidence: a systematic review and meta-analysis. *Clin Infect Dis.* 2019;68(Suppl 2):S105-S116.
7. Reddy EA, Shaw AV, Crump JA. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. *Lancet Infect Dis.* 2010;10(6):417-432.
8. Karakullukçu A, Kuşku MA, Ergin S, Aygün G, Midilli K, Küçükbasımacı Ö. Determination of clinical significance of coagulase-negative staphylococci in blood cultures. *Diagn Microbiol Infect Dis.* 2017;87(3):291-294.
9. Verhoeven PO, Haddar CH, Rigail J, et al. Comparison of the fully automated FilmArray BCID assay to a 4-hour culture test coupled to mass spectrometry for day 0 identification of microorganisms in positive blood cultures. *Biomed Res Int.* 2018;2018:7013470.
10. Fosheim GE, Carey RB, Limbago BM. Evaluation of the AdvanDx VRE EVIGENE assay for detection of vanA in vancomycin-resistant *Staphylococcus aureus*. *J Clin Microbiol.* 2007;45(5):1611-1613.
11. Ward C, Stocker K, Begum J, Wade P, Ebrahimsa U, Goldenberg SD. Performance evaluation of the Verigene(R) (Nanosphere) and FilmArray(R) (BioFire(R)) molecular assays for identification of causative organisms in bacterial bloodstream infections. *Eur J Clin Microbiol Infect Dis.* 2015;34(3):487-496.
12. Rule R, Paruk F, Becker P, Neuhoﬀ M, Chausse J, Said M. Clinical utility of the BioFire FilmArray Blood Culture Identification panel in the adjustment of empiric antimicrobial therapy in the critically ill septic patient. *PLoS One.* 2021;16(7):e0254389.
13. Ben-Zvi H, Drozdinsky G, Kushnir S, et al. Influence of GeneXpert MRSA/SA test implementation on clinical outcomes of *Staphylococcus aureus* bacteremia: a before-after retrospective study. *Diagn Microbiol Infect Dis.* 2019;93(2):120-124.
14. Luo J, Jiang M, Xiong J, et al. Exploring a phage-based real-time PCR assay for diagnosing *Acinetobacter baumannii* bloodstream infections with high sensitivity. *Anal Chim Acta.* 2018;1044:147-153.

15. Sigmund IK, Holinka J, Sevelde F, et al. Performance of automated multiplex polymerase chain reaction (mPCR) using synovial fluid in the diagnosis of native joint septic arthritis in adults. *Bone Joint J.* 2019;101-B(3):288-296.
16. Park GE, Peck KR, Ko JH, et al. Clinical factors influencing the performance of bacterial multiplex polymerase chain reaction in patients with community-onset pneumonia. *Eur J Clin Microbiol Infect Dis.* 2020;39(6):1193-1199.
17. Villamizar-Rodriguez G, Fernandez J, Marin L, Muniz J, Gonzalez I, Lombo F. Multiplex detection of nine food-borne pathogens by mPCR and capillary electrophoresis after using a universal pre-enrichment medium. *Front Microbiol.* 2015;6:1194.
18. Chou Q, Russell M, Birch DE, Raymond J, Bloch W. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Res.* 1992;20(7):1717-1723.
19. Rychlik W. Selection of primers for polymerase chain reaction. *Methods Mol Biol.* 1993;15:31-40.
20. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 2012;40(15):e115.
21. 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.
22. Fazekas A, Steeves R, Newmaster S. Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques.* 2010;48(4):277-285.
23. Vallone PM, Butler JM. AutoDimer: a screening tool for primer-dimer and hairpin structures. *Biotechniques.* 2004;37(2):226-231.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Zhang D, Luo Y, Zeng X, Yu Y, Wu Y. Developing a multiplex PCR-based assay kit for bloodstream infection by analyzing genomic big data. *J Clin Lab Anal.* 2022;36:e24686. doi: [10.1002/jcla.24686](https://doi.org/10.1002/jcla.24686)