




RESEARCH ARTICLE

A rapid and cost-effective multiplex ARMS-PCR method for the simultaneous genotyping of the circulating SARS-CoV-2 phylogenetic clades

Mohammad Tanvir Islam¹ | ASM Rubayet UI Alam¹  | Najmuj Sakib¹ |
Mohammad Shazid Hasan¹  | Tanay Chakrovarty¹ | Mohammad Tawyabur¹ |
Ovinu Kibria Islam¹ | Hassan M. Al-Emran² | Mohammad Iqbal Kabir Jahid¹ |
Mohammad Anwar Hossain^{3,4} 

¹Department of Microbiology, Jashore University of Science and Technology, Jashore, Bangladesh

²Department of Biomedical Engineering, Jashore University of Science and Technology, Jashore, Bangladesh

³Department of Microbiology, Jashore University of Science and Technology, Jashore, Bangladesh

⁴Department of Microbiology, University of Dhaka, Dhaka, Bangladesh

Correspondence

Mohammad Iqbal Kabir Jahid, Department of Microbiology, Faculty of Biological Sciences and Technology, Jashore University of Science and Technology, Jashore 7408, Bangladesh. Email: ikjahid_mb@just.edu.bd

Mohammad Anwar Hossain, Department of Microbiology, Jashore University of Science and Technology, Jashore 7408, Bangladesh. Email: hossaina@du.ac.bd

Funding information

Jashore University of Science and Technology Research Grant, Grant/Award Number: FoBST-06

Abstract

Tracing the globally circulating severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) phylogenetic clades by high-throughput sequencing is costly, time-consuming, and labor-intensive. We here propose a rapid, simple, and cost-effective amplification refractory mutation system (ARMS)-based multiplex reverse-transcription polymerase chain reaction (PCR) assay to identify six distinct phylogenetic clades: S, L, V, G, GH, and GR. Our multiplex PCR is designed in a mutually exclusive way to identify V-S and G-GH-GR clade variants separately. The pentaplex assay included all five variants and the quadruplex comprised of the triplex variants alongside either V or S clade mutations that created two separate subsets. The procedure was optimized with 0.2–0.6 μM primer concentration, 56–60°C annealing temperature, and 3–5 ng/μl complementary DNA to validate on 24 COVID-19-positive samples. Targeted Sanger sequencing further confirmed the presence of the clade-featured mutations with another set of primers. This multiplex ARMS-PCR assay is a fast, low-cost alternative and convenient to discriminate the circulating phylogenetic clades of SARS-CoV-2.

KEYWORDS

ARMS, clade, COVID-19, multiplex PCR, mutations, SARS-CoV-2

1 | INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread across 188 countries/regions within the first 6 months of the coronavirus disease 2019 (COVID-19) pandemic, infecting more than 354 million people.¹ This highly infectious virus poses a single-stranded-positive sense RNA genome of nearly 30 kbp.² Both synonymous and nonsynonymous mutations

were identified in the genomic region that code for nonstructural proteins (NSP1–16), structural proteins (spike, membrane, envelope, and nucleocapsid proteins), and/or seven other accessory proteins (ORF3a, ORF6, ORF7a, ORF7b, ORF8a, ORF8b, ORF8, and ORF10).^{3–7} Researchers have demonstrated that the predominant mutations may attribute to virulence.^{8–10} The virus has been classified into six clades, namely GH, GR, G, V, L, and S by the global initiative on sharing all influenza data

(GISAID)¹¹ by the clustered, co-evolving, and clade-featured point mutations.

The mutations at position C241T along with C3037T, C14408T (RdRp:p.P323L), and A23403G (S:p.D614G) was referred to as G clade. Additional mutation to the G clade at N protein:p.RG203-204KR (GGG28881-28883AAC) and ORF3a:p.Q57H (G25563T) refer to GR and GH clade, respectively. The V clade was classified by co-evolving mutations at G11083T (NSP6:p.L37F) and G26144T (ORF3a:p.G251V) where S clade strains contain C8782T and T28144C (NS8:p.L84S) variations, respectively. The L clade strains are the original or wild version for the featured mutations of five clades.¹²

Previous studies showed that the prevalence of phylogenetic clades was different by regions and times and was closely related to variable death–case ratio.^{9,13} G clade variant was dominant in Europe¹⁴ and United States¹⁵ on the eve of the pandemic, which caused high mortality in the United States. This mutation variant has gradually been circulated in Southeast Asia^{9,16} and Oceania.¹² On the contrary, GR and GH clades emerged at the end of February 2020, and GR mutants are now the leading type that causes more than one-third of infection globally.¹² Therefore, it is indispensable to identify the circulating clades in a specific region. Besides, several reports speculated the occurrence of SARS-CoV-2 reinfection by phylogenetically different strains that belong to separate clades.^{17,18} The dominance of a particular viral clade over others might determine the virulence, disease severity, and infection dynamics.⁹ However, the implications of different clades on effective drug and vaccine development are yet to be clearly elucidated.¹⁹

The identification of phylogenetic clades requires the identification of specific mutations into the viral genome. This identification is performed by the whole genome sequence through the next-generation sequencing (NGS) technique that has now scaled up the deposited sequences number in GISAID to 139,000 as of October 6, 2020. Another high-throughput NGS alternative is based on clade-based genetic barcoding that targets polymerase chain reaction (PCR) amplicons encompassing the featured mutation as described by Guan et al.²⁰ However, this state-of-the-art technique has limited access to most laboratories in low-income countries. A short-throughput and small-scale genotyping would be the Sanger-based targeted sequencing approach,⁷ but this is labor-intensive, time-consuming, inconvenient, and difficult to perform at low cost. Therefore, we have hardly observed the worldwide distribution of circulating clades in many countries, like Afghanistan, Maldives, Iraq, Syria, Yemen, Ethiopia, Sudan, Zimbabwe, Bolivia, Paraguay, and Chile, most probably due to the lack of sequencing facilities and appropriate technical personnel to perform this state-of-the-art technique. PCR-based point mutation discriminating technique, which is also known as the amplification refractory mutation system (ARMS), has been proven to be useful in identifying subtypes or clades of other respiratory viruses previously.^{21–23} In this study, we aimed to develop and validate an ARMS-based novel multiplex-PCR to identify the clade-specific point mutations of the circulating SARS-CoV-2 clades.

2 | METHODS AND MATERIALS

2.1 | Sample collection and complementary DNA (cDNA) preparation

Nasal and oral samples were collected in the health care facilities in the south-west part of Bangladesh and sent to the genome center, Jashore University of Science and Technology, Bangladesh. RNA was extracted from those samples using a nucleic acid extraction kit, Invitrogen Inc. The extracted RNA was then tested for SARS-CoV-2 using a commercial kit from Sansure Biotceh Co. Ltd (China). The left-over RNA was preserved at -40°C in the genome center lab.

From June 6, 2020 to June 30, 2020 and August 3, 2020 to August 8, 2020, we tested 6334 samples from five different districts of Bangladesh, of which 1849 (29%) were SARS-CoV-2 positive. Among the positive cases, only 503 possessed C_t value <30 , from which we selected 25 samples using a random number generator in Microsoft Excel Inc. (Table S1). One positive sample was excluded because it was a duplicate follow-up sample. Five recent samples from SARS-CoV-2 negative cases were also included in this study. Details of the selected SARS-CoV-2-positive samples can be found in Table S2.

cDNA was prepared for each selected sample using the Go-Script™ Reverse Transcription System (Promega) following the manufacturer's protocol. In brief, primer/RNA mix was prepared by mixing 10 μl of extracted RNA with 1 μl of Random primer and 1 μl of Oligo(dT)₁₅ primer (total volume 12 μl). Then the mixture was heated at 70°C for 5 min, followed by immediate chilling on ice for 5 min and a quick spin. The mixture for reverse transcription reaction was prepared by making a cocktail of the components from the Go-Script™ Reverse Transcription System in a sterile 1.5 ml microcentrifuge tube kept on ice. The final reaction mix was 40 μl for each cDNA synthesis reaction to be performed.

2.2 | Design and in silico validation of variant-specific (3'-SNP) multiplex primers

A set of 15 primers (Table 1) was designed based on the ARMS for differentiating six major clades of SARS-CoV-2: S, L, V, G, GH, and GR. We designated here the L clade strains as the wild type and others as mutants. For each clade apart from L, we selected a single representative single-nucleotide polymorphism (SNP) variant, including 23403A>G (p.D614G), 25563G>T (p.Q57H), 28882G>A (p.R203K), 26144G>T (p.G251V), and 28144T>C (p.L84S) from the multiple co-evolving mutations of the clades (G, GH, GR, V, and S, respectively). For example, the S clade is deviated from the L clade by two mutations: C8782T and T28144C. The "T" or "C" at 28144 positions was rendered as wild (L clade) or mutant type variant (indicating S clade), respectively. Details of other clade-specific mutations can be derived from the GISAID site and this literature. As established for the ARMS technique, this specificity was directed toward the 3'-end of the annealed primer template (Figure 1).

TABLE 1 Primer sets for targeted SNP-based single and/or multiplex PCR

Clade	SNP position	Primer ^a	Direction	Sequence (5'-3')	Position in CDS ^b	Amplicon size (bp)	T _m (°C)
S	28144	NS8_28144_F	Forward	AAGTTCAAGAACTTTACTCTCC	ORF7a_275-296	496	58.1
		NS8_28144_wR	Reverse	TGGCAATTAATTGTAAAAGGTA	ORF8_251-272		58
		NS8_28144_mR		TGGCAATTAATTGTAAAAGGTG	ORF8_251-272		57.8
V	26144	NS3_26144_F	Forward	TGGCAACTAGCACTCTCC	ORF3a_205-222	568	60.9
		NS3_26144_wR	Reverse	GATTAACAACCTCCGGATGAAC	ORF3a_752-772		58.7
		NS3_26144_mR		GATTAACAACCTCCGGATGAAA	ORF3a_752-772		58.3
G	23403	S_23403_wF	Forward	GTTGCTGTTCTTTATCAGGA	Spike_1822-1841	208	58
		S_23403_mF		GTTGCTGTTCTTTATCAGGG	Spike_1822-1841		58
		S_23404_R	Reverse	TGAGTCTGATAACTAGCGC	Spike_2012-2030		58
GH	25563	NS3_25563_wF	Forward	CACTTCTTGCTGTTTTTCAG	ORF3a_152-171	279	58
		NS3_25563_mF		CACTTCTTGCTGTTTTTCAT	ORF3a_152-171		57
		NS3_25563_R	Reverse	TGGCATCATAAAGTAAATGGG	ORF3a_411-430		57.8
GR	28881-28883	N_28882_F	Forward	CCAGATGACCAATGGC	N protein_238-255	387	58
		N_28882_wR	Reverse	TAGCAGGAGAAGTCCCC	N protein_608-625		60
		N_28882_mR		TAGCAGGAGAAGTTCGTT	N protein_608-625		58

Abbreviations: PCR, polymerase chain reaction; SNP, single-nucleotide polymorphism.

^aThe “w” and “m” in the primer names denote, respectively, the wild- and mutant-type allele corresponding to “no” and “single” base change for the wild and mutant type.

^bThe nucleotide position of the coding sequence for each primer where the primers bind to.

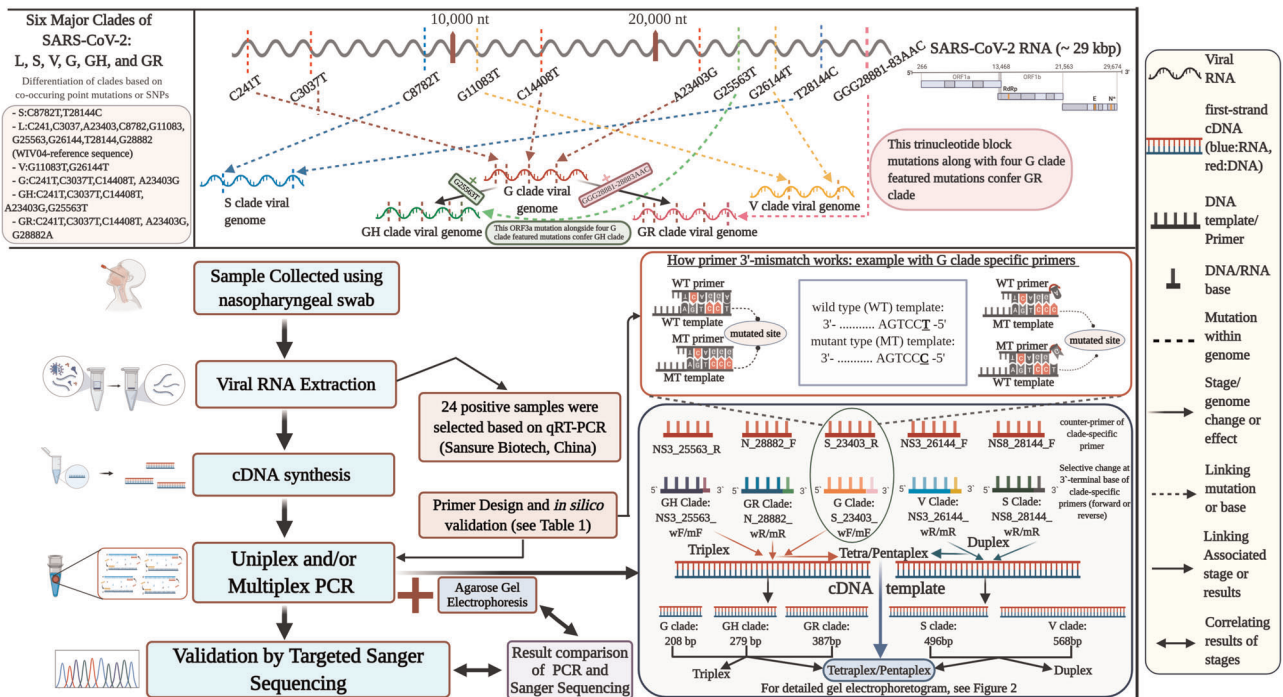


FIGURE 1 Schematic workflow of ARMS-based multiplex PCR assays for the identification of SARS-CoV-2 clades. The upper portion of the figure showed the concept of the clade as described in the GISAID with a comprehensive genomic visualization. The lower segment is dedicated to the overall workflow and the primer design. ARMS, amplification refractory mutation system; cDNA, complementary DNA; GISAID, global initiative on sharing all influenza data; qRT-PCR, quantitative real-time reverse-transcription polymerase chain reaction; PCR, polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

The forward- or reverse-type-specific primers were paired with counterpart reverse or forward primer. The amplicons were simultaneously distinguished by their molecular weight (bp) in multiplex PCR in different combinations. The positive amplification of wild-type-targeting primers was determined as the L type. The other types were determined based on the co-evolving mutation at respective sites.

The primer sets were designed using Primer3Plus²⁴ and Primer-BLAST²⁵ with the following stringent parameters and standard PCR conditions: avoiding hypothetical primer dimer (self or hetero) formation with less than -9 kcal/mol, sized 18–22 nucleotide in length, T_m of 58–60°C, 40%–60% GC content, G/C within the last five bases, no repeat of four or more of any base, amplicon size ranging from 200 to 600 bp, and avoiding hairpin loop structure. The primer specificity against SARS-CoV-2 and other organisms was checked by the Primer-BLAST. We performed the *in silico* PCR with the primers in the UCSC genome browser (<https://genome.ucsc.edu/>). Finally, the primer set was synthesized from the IDT company (<https://www.idtdna.com/>).

2.3 | Standardization of annealing temperature for single-variant-specific PCRs

A gradient PCR (SimpliAmp Thermal Cycler; Applied Biosystems) was performed for each of the variants separately with a freshly prepared cDNA template to standardize the annealing temperature. A couple of distinct tubes was prepared for each of the variant using the respective primer pairs to differentiate between the wild type and the mutant. The PCR was carried out in 10 μ l reaction volume containing 3–5 ng/ μ l DNA, 5 μ l master mixture (GoTaq® G2 Green Master; Promega), 0.2 μ M of each forward and reverse primer, and 2.8 μ l nuclease-free water. The thermocycling conditions were as follows: initial denaturation at 95°C for 1 min followed by 30 cycles at 95°C for 30 s, annealing at a range of 55–65°C for 30 s and 72°C for 30 s followed by a final extension at 72°C for 5 min. The PCR products were electrophoresed on a 1% (w/v) agarose gel stained with ethidium bromide (UltraPure™ Ethidium Bromide, 10 mg/ml; Thermo Fisher Scientific) and visualized using a gel documentation system (Bio-Rad).

2.4 | Multiplex PCR assays for simultaneous identification of the variants

Four sets (duplex, triplex, quadruplex, and pentaplex) of multiple-variant-specific reactions were arranged for simultaneous detection of a clade. A duplex PCR was performed using a mix of 26144G>T (p.G251V) and 28144T>C (p.L84S) variant-specific primer pairs namely NS3_26144_F-NS3_26144_wR (wild type)/NS3_26144_F-NS3_26144_mR (mutant) and NS8_28144_F-NS8_28144_wR (wild type)/NS8_28144_F-NS8_28144_mR (mutant), respectively, while mixing for wild types and the mutants in separate PCR tubes.

A triplex PCR assay was performed by using a blend of primer pairs namely S_23403_wF-S_23403_R (wild-type primers)/S_23403_mF-S_23403_R (mutant primers), NS3_25563_w1F-NS3_25563_1R (wild-type primers)/NS3_25563_m1F-NS3_25563_1R (mutant), and N_28882_F-N_28882_wR (wild type)/N_28882_F-N_28882_mR specific to 23403A>G (p.D614G), 25563G>T (p.Q57H) and 28882G>A (p.R203K) SNP variants, respectively. Quadruplex and pentaplex PCR assays were further performed in a similar manner. The pentaplex consisted of the primer mixture targeted for all five SNP variants whereas the quadruplex contained the variants of triplex in addition to either 26144G>T (p.G251V) or 28144T>C (p.L84S), thus making two different subsets. Among the 24 samples, one was used as a representative of all sets of multiplexes, and the reproducibility (described in more detail below) was checked over the rest of the samples.

2.5 | Validation of multiplex PCR assays

The multiplex PCR assays were performed over all the 24 positive samples. To validate the reliability of the assays, another five pairs of primer set (Table 2) were designed for the clades keeping the probable mutation points within the middle of the amplicons by using Primer3Plus²⁴ and Primer-BLAST.²⁵ The above-mentioned parameter settings were followed to design those (except for the amplicon size ranging from 200 to 400 bp and T_m of 52–54°C). Amplicons were subjected to Sanger sequencing using BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific) to confirm the specific clades (wild type/mutant). The commercial kit protocol was optimized to reduce the cost because of the high cost of this BigDye Terminator reagent.²⁶ In all, 1.0 μ l (per 10 μ l reaction) undiluted BigDye Terminator v3.1 Ready Reaction mix was used instead of 4 μ l mentioned in commercial kit protocol. Along with the 1.0 μ l BigDye Terminator v3.1 Ready Reaction mix, 1.75 μ l 5 \times sequencing buffer, 1 μ l primer, 2 μ l template DNA, and 4.25 μ l nuclease-free water was added (to make the final reaction volume 10 μ l). The cycle sequencing PCR condition was set up accordingly to the kit protocol. The sequences were aligned with and verified by the reference sequence (NC_045512.2_SARS-CoV-2_Wuhan-Hu-1 complete genome) using Molecular Evolutionary Genetics Analysis (MEGA X) software.²⁷ For the most part, the cost of the processes was optimized and compared to our in-house NGS system (Ion torrent; Thermo Fisher Scientific).

2.6 | Prediction of primer dimer formation and RNA secondary structure

We carried out RNA secondary structure prediction of the ORF3a or NS3 RNA using the Mfold web server.²⁸ The full NS3 sequence was extracted from the SARS-CoV-2 reference sequence of NCBI GenBank. The default parameters were used in generating the structure.

TABLE 2 List of primers for targeted Sanger sequencing to validate the multiplex assays

Clade	Name	Primer name	Sequence	bp	Amplicon size (bp)	T _m (°C)
S	NS8_28144_T to C	NS8_1_F	GTGGATGAGGCTGGTTCTAA	20	217	54.6
		NS8_1_R	TGGGGTCCATTATCAGACAT	20		53.6
V	NS3_26144_G to T	NS3_3_F	CTGGTGTGAACATGTTACCTT	20	209	53.2
		NS3_3_R	CTCTCCGAAACGAATGAGTA	20		52
G	Spike_23403_A to G	Spike_1_F	CGTGATCCACAGACACTTGA	20	228	54.6
		Spike_1_R	CCCTATTAACAGCCTGCAC	20		53.6
GH	NS3_25563_G to T	NS3_2_F	CAAGGTGAAATCAAGGATGC	20	207	51.8
		NS3_2_R	CAACAGCAAGTTGCAAACAA	20		52.6
GR	N protein 28882_G to A	N protein_1_F	AGGAACAACATTGCCAAAAG	20	231	51.7
		N protein_1_R	TGTTGGCCTTTACCAGACAT	20		54.2

Besides, we used oligoanalyzer v3.1 of integrated DNA technologies (IDT) to examine possible primer duplexes and calculate the primer-dimer (both self and hetero) formation energy in the case of the primers, NS3_26144_F, NS3_26144_wR, and NS3_26144_mR, targeting ORF3a multiplexed amplicons. The possibility of dimers and energy values for the targeted primers were checked against the other 12 primer sets.

3 | RESULTS

3.1 | Identification of the clades by single-variant-specific PCRs

ARMS-based PCRs for SNP variants (singleplex PCRs) were optimized at an annealing temperature of 57°C. The primer pairs for the variants 23403A>G (p.D614G) and 28882G>A (p.R203K) amplified the related mutant bands (208 and 387 bp, respectively). On the other hand, the primer pairs designed to denote the variants 25563G>T (p.Q57H), 26144G>T (p.G251V), and 28144T>C (p.L84S) had wild-type amplicons (279, 568, and 496 bp, respectively; Figure 2A(a)–(e)). Hence, the single-variant-specific PCRs were able to identify the SARS-CoV-2-positive sample containing GR clade of the virus.

3.2 | Optimization of multiplex PCR assays

The singleplex PCRs showed successful annealing at 57°C; however, temperatures for the duplex, triplex, and quadruplex assays were needed to be further optimized to 60°C, 57°C, and 56°C, respectively (Figure 2B(a)–(c)). The primer concentrations for a duplex were similar (0.2 μM of each primer pair for both the SNP variants) to the single-variant-specific PCRs, they were adjusted to different strengths for the triplex (for both forward and reverse primers: 0.2 μM for 23403A>G (p.D614G), 0.3 μM for 25563G>T (p.Q57H), and 0.4 μM for 28882G>A (p.R203K)) and the quadruplex (for both forward and reverse primers: 0.4 μM for 23403A>G (p.D614G),

0.3 μM for 25563G>T (p.Q57H), 0.6 μM for 28882G>A (p.R203K) and 0.2 μM for 4T>C (p.L84S)) to get the maximum possible resolution. The duplex PCR assay simultaneously amplified the desired wild type bands of 568 and 496 bp; the band for 28144T>C (p.L84S) was faint in the duplex setting compared to the single run of the variant described before. Besides, the triplex PCR also amplified the products as expected (i.e., 208, 279, and 387 bp). One of the subsets of quadruplex that contained the variants of triplex plus 28144CT>C (p.L84S) was able to distinguish the desired bands individually. However, the quadruplex (that had 26144G>T (p.G251V)) and the pentaplex arrangement could not discriminate the bands between wild types and mutants (Figure S1).

3.3 | Validation of multiplex PCR assays

All the 24 positive samples confirmed the test reproducibility of the assays; four of them, excluding the one used before for multiplex assays, were taken as a representative to display the reproducibility in this article (Figure 2C(a)–(c)). In this study, only GR clade was found in all positive samples tested. The homology of the nucleotide sequences for the PCR products showed more than 99% identity with the respective positions of the clades that validated the assays (Figure S2). Accession IDs to the submitted sequences for one positive sample as an archetype are available in the GISAID EpiFlu™ database (EPI_ISL_548260, EPI_ISL_561630, EPI_ISL_561375, EPI_ISL_561376, and EPI_ISL_561377).

4 | DISCUSSION

This study proposes a simple and exclusive ARMS-based SNP-discriminating method using conventional PCR to establish multiplex assays in detecting SARS-CoV-2 mutation clades. This concept was adopted from the other studies applied to identify the genetic profile of respiratory or gastrointestinal coronaviruses of pigs, human cancer risk-related SNPs, the virus that causes systemic infection in canines, the resistance profile of a bacteria, and so forth.^{29–32}

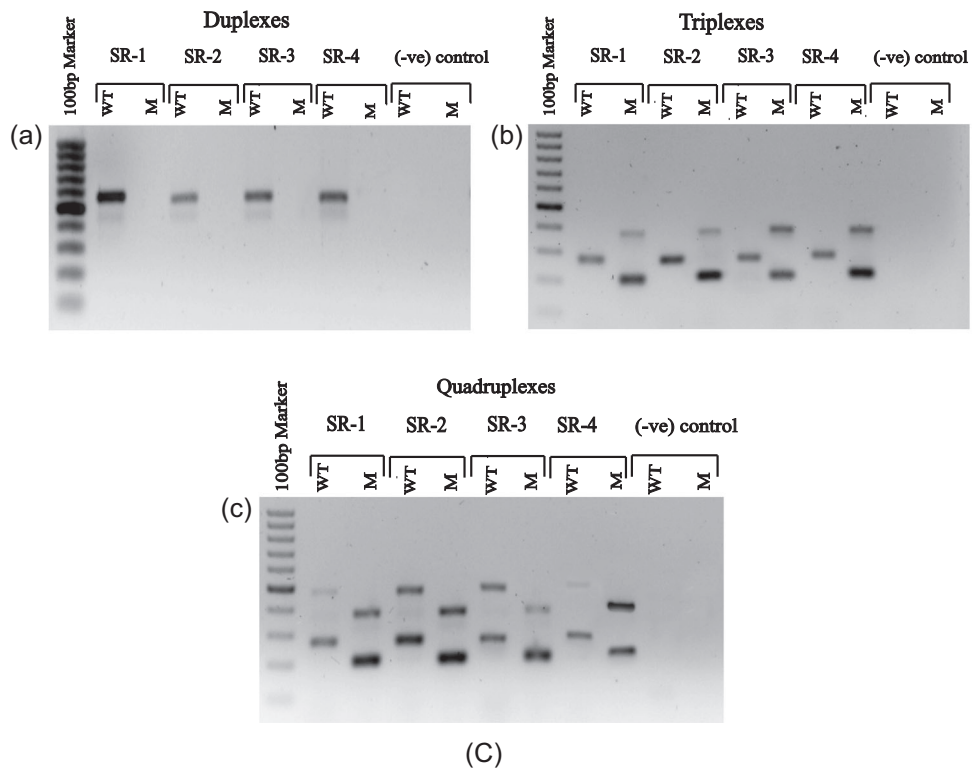
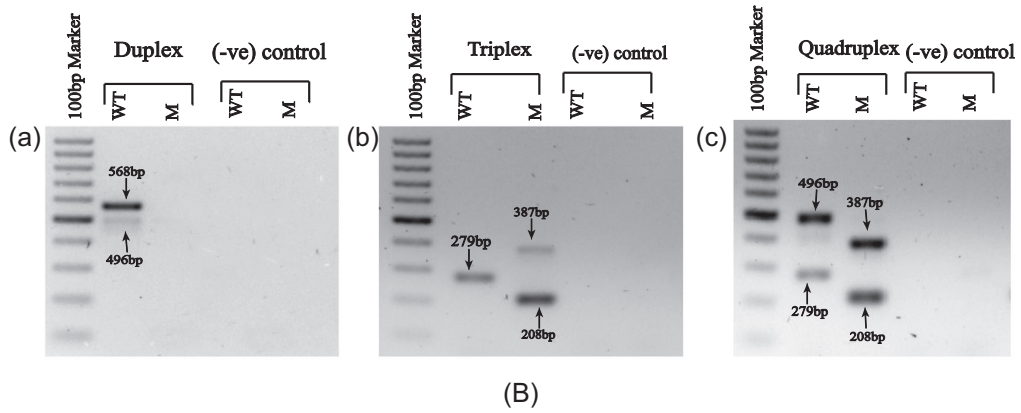
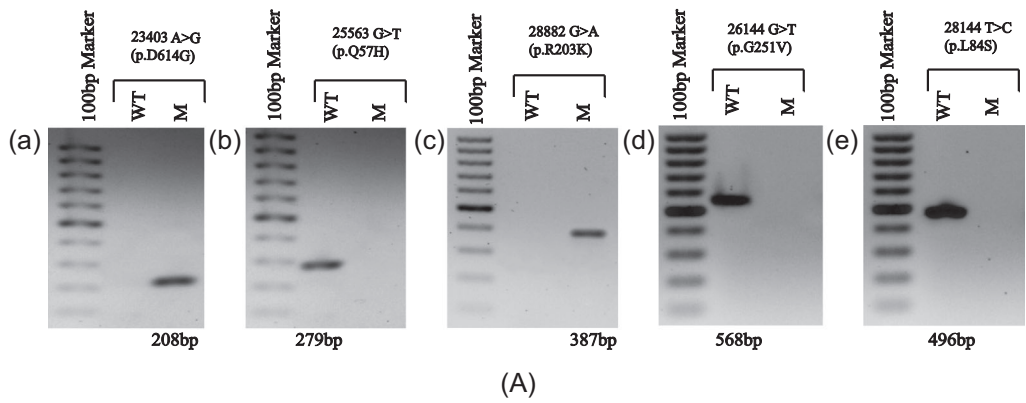


FIGURE 2 (See caption on next page)

This study designed point-mutation-specific primers to detect the six different SARS-CoV-2 clades as described by GISAID. The clade-based discrimination during the COVID-19 pandemic was exceedingly important because the prevalence of SARS-CoV-2 clades was varied by regions and times, and was closely related to variable case-fatality rate.^{9,13}

In this study, we attempted to validate two sets of multiplex PCR covering G, GH, and GR in the first set and V and S in the second set. Based on the available data of clade prevalence, we propose to run the first set of multiplex PCR at the beginning that can confirm the most three prevalent clades.⁹ Our attempt for pentaplex and/or the quadruplex (that included the SNP variant 26144G>T (p.G251V)) was unsuccessful, where the template regarding the variant 26144G>T (p.G251V) did not amplify, possibly due to the primer-dimer formation with higher thermodynamic stability than other variant-specific primer sets. The forward primer could bind to the N_28882_mR primer with a G value of less than -7 kcal/mol but can make longer products ~ 40 bp. In the case of reverse primers that target mutation, only NS3_26144_wR would form a self-dimer with high free energy (-12.9 kcal/mol). These homo- and heterodimer formation would make more primer duplex and might reduce the chance of effective pentaplex PCR.

Another possibility could be that the NS3 binding region of the primers (205–222 and 752–772) has resided within a stable stem site hindering the effective annealing. The RNA structure showed the complex stem-loop region and open sites as well. Our targeted primer-binding sites for the variant 26144G>T (p.G251V) were within the stem-loop region, whereas the primer annealing sites for 25563G>T (p.Q57H) variant reside within the open region of the template. Here, we assume that the complex structural configuration of NS3 may block the PCR reaction during competitive multiplexing (Figure S3). A longer RNA denaturation step during cDNA synthesis and a more stringent cycle denaturation of cDNA template might solve this issue. However, it could damage low-concentrated, sample-extracted viral RNA and inhibit amplification of other clade-specific templates by affecting overall optimized multiplex conditions.

Besides, the low concentration of our template cDNA or initial RNA due to low loads in samples may ultimately reduce the PCR amplification of the largest product from the “V-specific” clades,

which is 568 bp product targeting 26144G>T (p.G251V). We could not use a long extension time in PCR since it will increase the primer-dimer formation and will inhibit the amplification of other targets, that is, G, GR, GH, and S clades. We also were not able to detect viral load in the sample or cDNA quantity precisely since it would be dependent on the sampling, cell culture, and availability of the control RNA. Since we did not have positive and/or negative control RNA (synthesized), we could not detect how sensitive or specific our method was as well. However, we employed Sanger sequencing from each representative group to check the reproducibility of the results, and as Bangladeshi strains are mostly GR clade that is included within the G clade as well, we identified both of these clades in terms of our in-house multiplex ARMS PCR and Sanger sequencing.

The advantage of our ARMS-based multiplex assays is rapid. The turnaround time for our designed assay would range from approximately 3–4 h for 96 samples. The NGS and Sanger methods, on the other hand, had a turnaround time of more than 24 h and 10–12 h, respectively.^{33,34} An ARMS-based multiplex PCR assay similar to the current study would render a more convenient way to detect clade-specific mutation (SNPs) due to the process being faster and cost effective.³⁵ The cost of the assay for a single reaction was \$7 per run (the cost includes import Tax and VAT, etc. for Bangladesh), which is much less than targeted and whole-genome-based NGS methods in identifying the clades. The cost will be further reduced if an optimized one-step PCR system is used and we are currently working on it to cut the overall cost down to less than \$2. Thus, our method can overcome a serious limitation to effectively identify viral clades with a prospective broader application. The requirement of technical skill would also be less for this assay wherein the training of personnel is a minimal requirement and interpretation of results is generic.^{36,37}

Besides, the presence of the template as well as their quantity and quality are determined at the same time. The false-negative result for the absence of a template can also be determined in a facile manner.³⁸ In general, mutating the primer at its 3'-end makes it refractory to the “wild-type template,” whereas the absence of mutation in the primer is retractable to the “mutant template” amending a reliable technique over sequencing.³² On the other hand, NGS,

FIGURE 2 Strategy and validation of ARMS-based multiplex PCR assays. (I) Single-variant-specific PCRs with the respective amplicons in bp at the bottom of each gel image, showing the variants. (II) Multiplex PCR assays, containing PCR products for different clade combinations: duplex for 26144G>T (p.G251V) and 28144T>C (p.L84S) variants at an annealing temperature of 60°C; triplex for the variants 23403A>G (p.D614G), 25563G>T (p.Q57H), and 28882G>A (p.R203K) at 57°C; quadruplex for the variants 23403A>G (p.D614G), 25563G>T (p.Q57H), 28882G>A (p.R203K), and 28144CT>C (p.L84S) at 56°C. SARS-CoV-2-positive sample, GC 46.003, was used as a representative to perform the PCRs for (I) and (II). (III) Validation of the multiplex PCR assays, containing the identical settings for duplex, triplex, and quadruplex PCRs with GC 44.201 (denoted as SR-1), GC 88.025 (denoted SR-2), GC 90.175 (denoted SR-3), and GC 92.172 (denoted as SR-4) as representatives to display the reproducibility of the assays. SARS-CoV-2 negative sample, GC 116.09, was used as (–ve) control for the comparison. “GC” indicates the Genome Center identification number generated at the Genome Center of Jashore University of Science and Technology for COVID-19-suspected patients. 100–1000 bp marker was used in the first lane of the 1% agarose gels. The primers are listed in Table 1. ARMS, amplification refractory mutation system; COVID-19, coronavirus disease 2019; M, mutant band; PCR, polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; WT, wild-type band

such as whole-genome sequencing (WGS) or metagenomics approach, can generate millions of high-throughput data that enabled researchers to unroll new dimensions in the field of genome-sequencing applications.³⁹ The lack of technical personnel to analyze NGS data is also a reason to prefer an alternative approach other than NGS technology in low-income countries. Therefore, the ARMS technology with the conventional multiplex PCR methods in identifying the clades would be more applicable in low and minimum resource settings.

5 | CONCLUSION

Our assay can enhance the identification of genotypic variants of SARS-CoV-2 worldwide, especially in low-resource settings where NGS and Sanger-sequencing techniques are difficult to reach out. This rapid barcoding method may assist to reveal disease epidemiology, patient management, and protein-based drug designing and also contribute to modify the future national policy and vaccine development. A more cost-effective one-step procedure based on modified tetra ARMS assay (T-ARMS) is under development by our group that will considerably reduce the labor and cost further.

ACKNOWLEDGMENTS

We acknowledge GISAID for sharing the sequence data and IDT for giving the opportunity to use the tools for validating primers in silico. We also acknowledge the Ministry of Health and Family Welfare, Bangladesh, for giving us permission for the SARS-CoV-2 diagnosis. The present study was funded by the Jashore University of Science and Technology Research Grant (#FoBST-06) supported by the University Grant Commission, Bangladesh.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Ovinu Kibria Islam, A. S. M. Rubayet UI Alam, and Mohammad Iqbal Kabir Jahid conceived and designed the study outline. Mohammad Tanvir Islam and Mohammad Shazid Hasan setup initial experiments. Mohammad Tanvir Islam established the methodology with the assistance of Najmuj Sakib, Tanay Chakrovarty, and Mohammad Tawyabur. Mohammad Tanvir Islam and Najmuj Sakib processed the experimental data, interpreted the results, and generated the figures. A. S. M. Rubayet UI Alam designed the primers for multiplex ARMS-PCR and Mohammad Shazid Hasan designed the primers for targeted SNP positions. A. S. M. Rubayet UI Alam performed bioinformatics and in silico primer validation. Mohammad Shazid Hasan analyzed sequence data. Najmuj Sakib, A. S. M. Rubayet UI Alam, and Mohammad Shazid Hasan wrote the manuscript in consultation with Ovinu Kibria Islam, Hassan M. Al-Emran, Mohammad Iqbal Kabir Jahid, and Mohammad Tanvir Islam. Hassan M. Al-Emran and

Mohammad Iqbal Kabir Jahid further reviewed and edited the manuscript extensively. Mohammad Iqbal Kabir Jahid and Mohammad Anwar Hossain supervised the project and edited the manuscript. Final approval was provided by all authors.

ETHICS STATEMENT

This study was approved by the ethical review committee of Jashore University of Science and Technology (Ref: ERC/FBS/JUST/2020-45, Date: 06/10/2020). The samples have been used from the routine diagnosis.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1002/jmv.26818>.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in the GISAID EpiFlu™ database at <https://www.gisaid.org/>, with reference number (the accession numbers are EPI_ISL_548260, EPI_ISL_561630, EPI_ISL_561375, EPI_ISL_561376, and EPI_ISL_561377). These data were derived from the following resources available in the public domain GISAID (<https://www.gisaid.org/>).

ORCID

ASM Rubayet UI Alam  <https://orcid.org/0000-0001-9295-9865>

Mohammad Shazid Hasan  <https://orcid.org/0000-0002-2021-4693>

Mohammad Anwar Hossain  <https://orcid.org/0000-0001-9777-0332>

REFERENCES

- Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020;20(5):533-534.
- Mousavizadeh L, Ghasemi S. Genotype and phenotype of COVID-19: their roles in pathogenesis. *J Microbiol Immunol Infect*. 2020. Epub ahead of print. <https://doi.org/10.1016/j.jmii.2020.03.022>
- Liu DX, Fung TS, Chong KKL, Shukla A, Hilgenfeld R. Accessory proteins of SARS-CoV and other coronaviruses. *Antiviral Res*. 2014;109(1):97-109.
- Ou X, Liu Y, Lei X, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nat Commun*. 2020;11(1):1620.
- Kamitani W. Nonstructural proteins of novel coronavirus (SARS-CoV-2) 新型コロナウイルスにルス等のウにルス複製に必要な蛋白質. *Proc Annu Meet Japanese Pharmacol Soc*. 2020;93(0):2-ES-2.
- Islam MR, Hoque MN, Rahman MS, et al. Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Sci Rep*. 2020;10(1):14004.
- ASMRU Alam, Islam MR, Rahman MS, Islam OK, Hossain MA. Understanding the possible origin and genotyping of the first Bangladeshi SARS-CoV-2 strain. *J Med Virol*. 2020. Epub ahead of print. <https://doi.org/10.1002/jmv.26115>
- Wang Q, Zhang Y, Wu L, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*. 2020;181(4):894-904.
- Alam ASMRU, Islam MR, Rahman MS, Islam OK, Hossain MA. Evolving infection paradox of SARS-CoV-2: fitness costs virulence? Preprint; 2020. <https://doi.org/10.31219/OSF.IO/Y36VE>

10. Rahman MS, Islam MR, Hoque MN, et al. Comprehensive annotations of the mutational spectra of SARS-CoV-2 spike protein: a fast and accurate pipeline. *Transbound Emerg Dis*. 2020; Epub ahead of print. <https://doi.org/10.1111/tbed.13834>
11. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data—from vision to reality. *Euro Surveill*. 2017;22:13.
12. Mercatelli D, Giorgi FM. Geographic and genomic distribution of SARS-CoV-2 mutations. *Front Microbiol*. 2020;11:1800.
13. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J Hum Genet*. 2020;65:1075-1082. Epub ahead of print. <https://doi.org/10.1038/s10038-020-0808-9>
14. Korber B, Fischer WM, Gnanakaran S, et al. Tracking changes in SARS-CoV-2 Spike: Evidence that D614G increases infectivity of the COVID-19 virus. *Cell*. 2020;182(4):812-827.
15. Brufsky A. Distinct viral clades of SARS-CoV-2: implications for modeling of viral spread. *J Med Virol*. 2020;92(9):1386-1390.
16. Islam OK, Al-Emran HM, Hasan MS, Anwar A, Jahid MIK, Hossain MA. Emergence of European and North American mutant variants of SARS-CoV-2 in South-East Asia. *Transbound Emerg Dis*. 2020. Epub ahead of print. <https://doi.org/10.1111/tbed.13748>
17. KK-W To, IF-N Hung, Ip JD, et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. *Clin Infect Dis*. 2020. Epub ahead of print. <https://doi.org/10.1093/cid/ciaa1275>
18. Li Y, Ji D, Cai W, et al. Clinical characteristics, cause analysis and infectivity of COVID-19 nucleic acid re-positive patients: a literature review. *J Med Virol*. 2020. Epub ahead of print. <https://doi.org/10.1002/jmv.26491>
19. Chellapandi P, Saranya S. Genomics insights of SARS-CoV-2 (COVID-19) into target-based drug discovery. *Med Chem Res*. 2020;29(10):1777-1791.
20. Guan Q, Sadykov M, Mfarrej S, et al. A genetic barcode of SARS-CoV-2 for monitoring global distribution of different clades during the COVID-19 pandemic. *Int J Infect Dis*. 2020;100:216-223.
21. Lee EJ, Kim EJ, Shin YK, Song JY. Design and testing of multiplex RT-PCR primers for the rapid detection of influenza A virus genomic segments: application to equine influenza virus. *J Virol Methods*. 2016;228:114-122.
22. Wang W, Ren P, Mardi S, et al. Design of multiplexed detection assays for identification of avian influenza A virus subtypes pathogenic to humans by SmartCycler real-time reverse transcription-PCR. *J Clin Microbiol*. 2009;47(1):86-92.
23. Brister H, Barnum SM, Reedy S, Chambers TM, Pusterla N. Validation of two multiplex real-time PCR assays based on single nucleotide polymorphisms of the HA1 gene of equine influenza A virus in order to differentiate between clade 1 and clade 2 Florida sub-lineage isolates. *J Vet Diagn Invest*. 2019;31(1):137-141.
24. Untergasser A, Cutcutache I, Koressaar T, et al. Primer3-new capabilities and interfaces. *Nucleic Acids Res*. 2012;40(15):e115.
25. 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.
26. Platt AR, Woodhall RW, George AL. Improved DNA sequencing quality and efficiency using an optimized fast cycle sequencing protocol. *Biotechniques*. 2007;43(1):58-62.
27. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35(6):1547-1549.
28. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*. 2003;31(13):3406-3415.
29. Shi X, Zhang C, Shi M, et al. Development of a single multiplex amplification refractory mutation system PCR for the detection of rifampin-resistant *Mycobacterium tuberculosis*. *Gene*. 2013;530(1):95-99.
30. Lai CH, Welter MW, Welter LM. The use of ARMS PCR and RFLP analysis in identifying genetic profiles of virulent, attenuated or vaccine strains of TGEV and PRCV. *Adv Exp Med Biol*. 1995;380:243-250.
31. Zhang C, Liu Y, Ring BZ, et al. A novel multiplex tetra-primer ARMS-PCR for the simultaneous genotyping of six single nucleotide polymorphisms associated with female cancers. *PLOS One*. 2013;8(4):e62126.
32. Chulakasian S, Lee MS, Wang CY, et al. Multiplex amplification refractory mutation system polymerase chain reaction (ARMS-PCR) for diagnosis of natural infection with canine distemper virus. *Virol J*. 2010;7(1):122.
33. Tsuchihashi Z, Dracopoli NC. Progress in high throughput SNP genotyping methods. *Pharmacogenomics J*. 2002;2(2):103-110.
34. Zhang J, Yang J, Zhang L, et al. A new SNP genotyping technology Target SNP-seq and its application in genetic analysis of cucumber varieties. *Sci Rep*. 2020;10. Article number: 5623.
35. Ahlawat S, Sharma R, Maitra A, Roy M, Tantia MS. Designing, optimization and validation of tetra-primer ARMS PCR protocol for genotyping mutations in caprine Fec genes. *Meta Gene*. 2014;2:439-449.
36. Syrmis MW, Whitley DM, Thomas M, et al. A sensitive, specific, and cost-effective multiplex reverse transcriptase-PCR assay for the detection of seven common respiratory viruses in respiratory samples. *J Mol Diagnostics*. 2004;6(2):125-131.
37. Wang L, Si Y, Dedow LK, Shao Y, Liu P, Brutnell TP. A low-cost library construction protocol and data analysis pipeline for illumina-based strand-specific multiplex RNA-seq. *PLOS One*. 2011;6:10.
38. Edwards MC, Gibbs RA. Multiplex PCR: advantages, development, and applications. *Genome Res*. 1994;3(4):S65-S75.
39. El-Metwally S, Hamza T, Zakaria M, Helmy M. Next-generation sequence assembly: four stages of data processing and computational challenges. *PLOS Comput Biol*. 2013;9:12.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Islam MT, Alam ASMRU, Sakib N, et al. A rapid and cost-effective multiplex ARMS-PCR method for the simultaneous genotyping of the circulating SARS-CoV-2 phylogenetic clades. *J Med Virol*. 2021;93:2962-2970. <https://doi.org/10.1002/jmv.26818>