1

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

Running title: Identifying SARS-CoV-2 clades by multiplex ARMS-PCR

Md. Tanvir Islam[1#], A. S. M. Rubayet Ul Alam[1#], Najmuj Sakib[1#], Md. Shazid Hasan[1], Tanay Chakrovarty[1], Md. Tawyabur[1], Ovinu Kibria Islam[1], Hassan M. Al-Emran[2], Iqbal Kabir Jahid[1*], M. Anwar Hossain[3,4]*

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

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

[3]Vice-Chancellor, Jashore University of Science and Technology, Jashore, Bangladesh

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

*Corresponding authors:

M. Anwar Hossain, Vice-Chancellor, Jashore University of Science and Technology, Jashore, Bangladesh, E-mail: hossaina@du.ac.bd, Phone: +8801709818101

and

Iqbal Kabir Jahid, Department of Microbiology, Faculty of Biological Sciences and Technology, Jashore University of Science and Technology, Jashore-7408, Bangladesh, E-mail: ikjahid_mb@just.edu.bd, Phone: +8801709818178

[#]**Authors contributed equally**

Summary

32

33   Tracing the globally circulating SARS-CoV-2 mutants is essential for the outbreak alerts and

34   far-reaching epidemiological surveillance. The available technique to identify the phylogenetic

35   clades through high-throughput sequencing is costly, time-consuming, and labor-intensive that

36   hinders the viral genotyping in low-income countries. Here, we propose a rapid, simple and

37   cost-effective amplification-refractory mutation system (ARMS)-based multiplex reverse-

38   transcriptase PCR assay to identify six distinct phylogenetic clades: S, L, V, G, GH, and GR.

39   This approach is applied on 24 COVID-19 positive samples as confirmed by CDC approved

40   real-time PCR assay for SARS-CoV-2. Our multiplex PCR is designed in a mutually exclusive

41   way to identify V-S and G-GH-GR clade variants separately. The pentaplex assay included all

42   five variants and the quadruplex comprised of the triplex variants alongside either V or S clade

43   mutations that created two separate subsets. The procedure was optimized in the primer

44   concentration (0.2-0.6 µM) and annealing temperature (56-60°C) of PCR using 3-5 ng/µl

45   cDNA template synthesized upon random- and oligo(dT)-primer based reverse transcription.

46   The different primer concentration for the triplex and quadruplex adjusted to different strengths

47   ensured an even amplification with a maximum resolution of all targeted amplicons. The

48   targeted Sanger sequencing further confirmed the presence of the clade-featured mutations with

49   another set of our designed primers. This multiplex ARMS-PCR assay is sample, cost-

50   effective, and convenient that can successfully discriminate the circulating phylogenetic clades

51   of SARS-CoV-2.

52   **Keywords:** ARMS, multiplex PCR, SARS-CoV-2, clade, mutations, COVID-19

53

54

55

56

57

58

59

60

61

62

63

64

Introduction

SARS-CoV-2 has spread across 188 countries/regions within the first six months of COVID-19 pandemic infecting more than 354 million people (Dong, Du, & Gardner, 2020). This highly infectious virus poses a single-stranded positive-sense RNA genome of nearly 30 kbp (Mousavizadeh & Ghasemi, 2020). Both synonymous and non-synonymous mutations were identified in the genomic region that code for non-structural 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) (M. R. Islam et al., 2020; Kamitani, 2020; Liu, Fung, Chong, Shukla, & Hilgenfeld, 2014; Ou et al., 2020). Researchers have demonstrated that the predominant mutations may attribute to virulence (Alam, Islam, Hasan, et al., 2020; Rahman et al., 2020; Q. Wang et al., 2020). 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) (Shu & McCauley, 2017) 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 as G clade. Additional mutation to the G clade at N protein:p.RG203-204KR (GGG28881-28883AAC) and ORF3a:p.Q57H (G25563T) refers 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 (Mercatelli & Giorgi, 2020).

Previous studies showed the prevalence of phylogenetic clades were different by regions and times and were closely related to variable death-case ratio (Alam et al., 2020; Toyoshima et al 2020, nature). G clade variant was dominant in Europe (Korber et al., 2020) and USA (Brufsky, 2020) on the eve of pandemic which caused high mortality in USA. This mutation variant has gradually been circulated in Southeast Asia (Alam et al., 2020; Islam et al., 2020) and Oceania (Mercatelli & Giorgi, 2020). On the contrary GR and GH clades were emerged at the end of February 2020 and GR mutant are now the leading type that cause more than one-third of infection globally (Mercatelli & Giorgi, 2020). 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 belongs to separate clades (Li et al., 2020; To et al., 2020). The dominance of a particular viral clade over others might determine the virulence, disease severity, and infection dynamics (Alam et al., 2020). However,

3

98   the implications of different clades on effective drug and vaccine development is yet to be

99   clearly elucidated (Chellapandi & Saranya, 2020).

100 The identification of phylogenetic clades requires the identification of specific mutations into

101 viral genome. This identification is performed by the whole genome sequence through NGS

102 technique that has now scaled up the deposited sequences number in GISAID to 139,000 as of

103 October 6, 2020. Another high-throughput NGS alternative is based on clade-based genetic

104 barcoding that targets PCR amplicons encompassing the featured mutation as described by

105 Guan et al. (2020). However, this state of art technique has limited access to most laboratories

106 in low-income countries. A short-throughput and small-scale genotyping would be the Sanger

107 based targeted sequencing approach (Alam, Islam, Rahman, Islam, & Hossain, 2020), but this

108 is labor intensive, time-consuming, inconvenient, and difficult to perform at low cost.

109 Therefore, we have hardly observed the worldwide distribution of circulating clades in many

110 countries, like Afghanistan, Maldives, Iraq, Syria, Yemen, Ethiopia, Sudan, Zimbabwe,

111 Bolivia, Paraguay, and Chile, most probably due to the lack of sequencing facilities and

112 appropriate technical personnel to perform this state-of-the-art technique. PCR-based point

113 mutation discriminating technique, which is also known as the amplified refractory mutation

114 system (ARMS), has been proven to be useful in identifying subtypes or clades of other

115 respiratory viruses previously (Brister, Barnum, Reedy, Chambers, & Pusterla, 2019; Lee,

116 Kim, Shin, & Song, 2016; W. Wang et al., 2009). In this study, we aimed to develop and

117 validate an ARMS-based novel multiplex-PCR to identify the clade-specific point mutations

118 of the circulating SARS-CoV-2 clades.

119 **Methods and materials**

120 *Sample collection and cDNA preparation*

121 Nasal and oral samples were collected in the health care facilities in the south-west part of

122 Bangladesh and sent to the genome center, Jashore University of Science and Technology,

123 Bangladesh. RNA was extracted from those samples using nucleic acid extraction kit,

124 Invitrogen inc. The extracted RNA was then tested for SARS CoV-2 using a commercial kit

125 from Sansure Biotceh Co., Ltd (China). The left-over RNA were preserved at -40°C in the

126 genome center lab. A total of 24 randomly selected SARS CoV-2 positive samples were tested

127 for the analysis (**supplementary Table s1**). A representative SARS CoV-2 negative sample

128 (of 5 of the negative samples randomly selected) was used as a negative control in this study

129 (**supplementary Table s2**). cDNA was prepared for each selected sample using the GoScript™

4

130   Reverse Transcription System (Promega, USA) following the manufacturer's protocol. In

131   brief, primer/ RNA mix was prepared by mixing 10µl of extracted RNA with 1µl of Random

132   primer and 1µl of Oligo(dT)$_{15}$ primer (total volume 12µl). Then the mixture was heated at 70°C

133   for 5 minutes, followed by immediate chilling on ice for 5 minutes and a quick spin. The

134   mixture for reverse transcription reaction was prepared by making a cocktail of the components

135   from GoScript™ Reverse Transcription System in a sterile 1.5ml micro centrifuge tube keeping

136   on ice. The final reaction mix was 40µl for each cDNA synthesis reaction to be performed.

137   *Design and in silico validation of variant-specific (3′-SNP) multiplex primers*

138   A set of 15 primers (**Table 1**) was designed based on the ARMS for differentiating six major

139   clades of SARS-CoV-2: S, L, V, G, GH, and GR. We designated here the L clade strains as the

140   wild type and others as mutants. For each clade apart from L, we selected a single representative

141   SNP variant, including 23403 A>G (p.D614G), 25563 G>T (p.Q57H), 28882 G>A (p.R203K),

142   26144 G>T (p.G251V) and 28144 T>C (p.L84S) from the multiple co-evolving mutations of

143   the clades (G, GH, GR, V and S, respectively). For example, the S clade is deviated from the

144   L clade by two mutations: C8782T, T28144C. The 'T' or 'C' at 28144 positions was rendered

145   as wild (L clade) or mutant type variant (indicating S clade), respectively. Details of other

146   clade-specific mutations can be derived from the GISAID site and this literature. As established

147   for ARMS technique, this specificity was directed towards the 3′-end of the annealed primer-

148   template (Fig. 1). The forward or reverse type-specific primers were paired with counterpart

149   reverse or forward primer. The amplicons were simultaneously distinguished by their

150   molecular weight (bp) in multiplex PCR in different combinations. The positive amplification

151   of wild type targeting primers was determined as the L type. The other types were determined

152   based on the co-evolving mutation at respective sites.

153   The primer sets were designed using Primer3Plus (Untergasser et al., 2012) and Primer-

154   BLAST (Ye et al., 2012) with the following stringent parameters and standard PCR conditions:

155   avoiding hypothetical primer dimer (self or hetero) formation with less than -9 Kcal/mol, sized

156   18-22 nucleotide in length, Tm of (58-60)°C, (40-60)% GC content, G/C within the last five

157   bases, no repeat of four or more of any base, amplicon size ranging from 200 to 600 bp, and

158   avoiding hairpin loop structure. The primer specificity against SARS-CoV-2 and other

159   organisms was checked by the Primer-BLAST. We performed the *in silico* PCR with the

160   primers in the UCSC genome browser (https://genome.ucsc.edu/). Finally, the primer set was

161   synthesized from the IDT company (https://www.idtdna.com/).

162   *Standardization of annealing temperature for Single-variant specific PCRs*

163   A gradient PCR (SimpliAmp Thermal Cycler, Applied Biosystems, USA) was performed for

164   each of the variant separately with freshly prepared cDNA template to standardize the

165   annealing temperature. Couple of distinct tubes was prepared for each of the variant using the

166   respective primer pairs to differentiate between the wild type and the mutant. The PCR was

167   carried out in 10µl reaction volume containing 3-5 ng/µl DNA, 5µl master mixture (GoTaq®

168   G2 Green Master; Promega, USA), 0.2µM of each forward and reverse primer and 2.8 µl

169   nuclease free water. The thermocycling conditions were as follows: initial denaturation at 95°C

170   for 1 min followed by 30 cycles at 95°C for 30s, annealing at a range of 55-65°C for 30s and

171   72°C for 30s followed by a final extension at 72°C for 5min. The PCR products were

172   electrophoresed on a 1% (w/v) agarose gel stained with ethidium bromide (UltraPure™

173   Ethidium Bromide, 10 mg/mL; Thermo Fisher, USA) and visualized using a gel documentation

174   system (Bio-Rad, USA).

175   *Multiplex PCR assays for simultaneous identification of the variants*

176   Four sets (duplex, triplex, quadruplex and pentaplex) of multiple variant-specific reactions

177   were arranged for simultaneous detection of a clade. A duplex PCR was performed by using a

178   mix of 26144 G>T(p.G251V) and 28144 T>C (p.L84S) variant specific primer pairs viz.

179   NS3_26144_F-NS3_26144_wR (wild type)/NS3_26144_F-NS3_26144_mR (mutant) and

180   NS8_28144_F-NS8_28144_wR (wild type)/ NS8_28144_F-NS8_28144_mR (mutant),

181   respectively while mixing for wild types and the mutants in separate PCR tubes. A triplex PCR

182   assay was performed by using a blend of primer pairs viz. S_23403_wF-S_23403_R (wild type

183   primers)/ S_23403_mF- S_23403_R (mutant primers), NS3_25563_w1F-NS3_25563_1R

184   (wild type primers)/ NS3_25563_m1F-NS3_25563_1R (mutant), and N_28882_F-

185   N_28882_wR (wild type)/ N_28882_F-N_28882_mR specific to 23403 A>G (p.D614G),

186   25563 G>T (p.Q57H) and 28882 G>A (p.R203K) SNP variants respectively. Quadruplex and

187   pentaplex PCR assays were further performed in similar manner. The pentaplex consisted the

188   mix for all five SNP variants, On the other hand, the quadruplex contained the variants of

189   triplex plus either 26144 G>T (p.G251V) or 28144 T>C (p.L84S), thus making two different

190   subsets. Among the 24 samples, one was used as a representative of all sets of multiplexes and

191   reproducibility (described in more detail below) was checked over the rest of the samples.

192   *Validation of multiplex PCR assays*

6

193   The multiplex PCR assays were performed over all the 24 positive samples. To validate the

194   reliability of the assays, another five pairs of primer set (**Table 2**) were designed for the clades

195   keeping the probable mutation points within the middle of the amplicons by using Primer3Plus

196   (Untergasser et al., 2012) and Primer-BLAST (Ye et al., 2012). Above mentioned parameter

197   settings were followed to design those (except for the amplicon size ranging from 200-400bp

198   and Tm of 52-54°C). Amplicons were subjected to Sanger sequencing using BigDye™

199   Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific) to confirm the specific clades

200   (wild type/mutant). The commercial kit protocol was optimized to reduce the cost because of

201   the high cost of this BigDye Terminator reagent (Platt, Woodhall, & George, 2007). 1.0μL (per

202   10 10μL reaction) undiluted BigDye Terminator v3.1 Ready Reaction mix was used instead of

203   4μL mentioned in commercial kit protocol. Along with the 1.0 μL BigDye Terminator v3.1

204   Ready Reaction mix, 1.75 μL 5x sequencing buffer, 1μL primer, 2μL template DNA and

205   4.25μL nuclease free water was added (to make the final reaction volume 10μL). The cycle

206   sequencing PCR condition was setup accordingly to the kit protocol. The sequences were

207   aligned with and verified by the reference sequence (NC_045512.2_SARS-CoV-2_Wuhan-

208   Hu-1 complete genome) using Molecular Evolutionary Genetics Analysis (MEGA X) software

209   (Kumar, Stecher, Li, Knyaz, & Tamura, 2018). For the most part, the cost of the processes were

210   optimized and compared to our in-house NGS system (Ion torrent; ThermoFisher Scientific,

211   USA).

212   **Prediction of Primer dimer formation and RNA secondary structure**

213   We carried out RNA secondary structure prediction of the ORF3a or NS3 RNA using the Mfold

214   web server (Zuker, 2003). The full NS3 sequence was extracted from the SARS-CoV-2

215   reference sequence of NCBI GenBank. The default parameters were used in generating

216   structure. Besides, we used oligoanalyzer v3.1 of integrated DNA technologies (IDT) to

217   examine possible primer duplexes and calculate the primer-dimer (both self and hetero)

218   formation energy in case of the primers, NS3_26144_F, NS3_26144_wR, and

219   NS3_26144_mR, targeting ORF3a multiplexed amplicons. The possibility of dimers and

220   energy values for the targeted primers were checked against other twelve primer sets.

221   **Results**

222   *Identification of the clades by single-variant specific PCRs*

223   ARMS-based PCRs for single SNP variants (singleplex PCRs) were optimized at an annealing

224   temperature of 57°C. The primer pairs for the variants 23403 A>G (p.D614G) and 28882 G>A

225   (p.R203K) amplified the related mutant bands (208bp and 387bp, respectively). On the other

226   hand, the primer pairs designed to denote the variants 25563 G>T (p.Q57H), 26144

227   G>T(p.G251V) and 28144 T>C (p.L84S) had wild type amplicons (279bp, 568bp and 496bp,

228   respectively) (**Fig.2 IA-IE**). Hence, the single-variant specific PCRs were able to identify the

229   SARS-CoV-2 positive sample containing GR-clade of the virus.

230   *Optimization of multiplex PCR assays*

231   The singleplex PCRs showed successful annealing at 57°C, however, temperatures for the

232   duplex, triplex and quadruplex assays were needed to be further optimized to 60°C, 57°C and

233   56°C, respectively (**Fig.2 IIA-IIE**). The primer concentrations for duplex were similar (0.2µM

234   of each primer pair for both the SNP variants) to the single-variant specific PCRs, they were

235   adjusted to different strengths for the triplex (for both forward and reverse primers: 0.2µM for

236   23403 A>G (p.D614G), 0.3µM for 25563 G>T (p.Q57H), and 0.4µM for 28882 G>A

237   (p.R203K)) and the quadruplex (for both forward and reverse primers: 0.4µM for 23403 A>G

238   (p.D614G), 0.3µM for 25563 G>T (p.Q57H), 0.6µM for 28882 G>A (p.R203K) and 0.2µM

239   for 28144 T>C (p.L84S)) to get the maximum possible resolution. The duplex PCR assay

240   simultaneously amplified the desired wild type bands of 568bp and 496bp; the band for 28144

241   T>C (p.L84S) was faint in the duplex setting comparing to the single run of the variant

242   described before. Besides, the triplex PCR also amplified the products as expected (i.e., 208bp,

243   279bp and 387bp). One of the subsets of quadruplex that contained the variants of triplex plus

244   28144C T>C (p.L84S) was able to distinguish the desired bands individually. However,

245   quadruplex (that had 26144 G>T (p.G251V)) and pentaplex arrangement could not

246   discriminate the bands between wild types and mutants (**supplementary Fig.s1**).

247   *Validation of multiplex PCR assays*

248   All the 24 positive samples confirmed the test reproducibility of the assays; four of them

249   excluding the one used before for multiplex assays were taken as representative to display the

250   reproducibility in this article (**Fig.2 IIIA-IIIC**). In this study, only GR clade was found in all

251   positive samples tested. The homology of the nucleotide sequences for the PCR products

252   showed more than 99% identity with the respective positions of the clades that validated the

253   assays (**supplementary Fig.s2**). Accession IDs to the submitted sequences for one positive

8

254   sample as an archetype are available in GISAID EpiFlu™ database (EPI_ISL_548260,

255   EPI_ISL_561630, EPI_ISL_561375, EPI_ISL_561376 and EPI_ISL_561377).

256   **Discussion**

257   This study proposes a simple and exclusive ARMS-based SNP-discriminating method using

258   conventional PCR to establish multiplex-assays in detecting SARS-CoV-2 mutation clades.

259   This concept was adopted from the other studies applied to identify the genetic profile of

260   respiratory or gastrointestinal coronaviruses of pigs, human cancer risk related SNPs, virus that

261   causes systemic infection in canines, the resistance profile of a bacteria etc. (Chulakasian et al.,

262   2010; Lai, Welter, & Welter, 1995; Shi et al., 2013; C. Zhang et al., 2013). This study designed

263   point-mutation specific primers to detect the six different SARS-CoV-2 clades as described by

264   GISAID. The clade-based discrimination during COVID-19 pandemic was exceedingly

265   important because the prevalence of SARS-CoV-2 clades were varied by regions and times,

266   and were closely related to variable death-case ratio (Alam et al., 2020; Toyoshima et al 2020,

267   nature). G clade variant was dominant in Europe (Korber et al., 2020) and USA (Brufsky, 2020)

268   on the beginning of the pandemic that caused high mortality in USA. This mutation variant

269   was gradually circulating in Southeast Asia (Alam, Islam, Hasan, et al., 2020; O. K. Islam et

270   al., 2020) and Oceania (Mercatelli & Giorgi, 2020). Among the two derivative of G clade (GR

271   and GH) that emerged at the end of February 2020, GR mutant are now the leading type that

272   cause more than one-third of infection globally (Mercatelli & Giorgi, 2020).

273   In this study, we attempted to validate two sets of multiplex PCR covering G, GH and GR in

274   the first set, and V and S in the second set. Based on the available data of clade prevalence we

275   propose to run the first set of multiplex PCR in the beginning that can confirm the most three

276   prevalent clades (Alam, Islam, Hasan, et al., 2020). Our attempt for pentaplex and/or the

277   quadruplex (that included the SNP variant 26144 G>T (p.G251V)) was unsuccessful, where

278   the template regarding the variant 26144 G>T (p.G251V) did not amplify, possibly due to

279   primer-dimer formation with higher thermodynamic stability than other variant-specific primer

280   sets. The forward primer could bind to the N_28882_mR primer with a G value of <-7

281   Kcal/mol, but can make longer products ~40 bp. In case of reverse primers that target mutation,

282   only NS3_26144_wR would form a self-dimer with high free energy (-12.9 Kcal/mol). These

283   homo- and hetero-dimer formation would make more primer duplex and might reduce the

284   chance of effective pentaplex PCR. Another possibility could be that the NS3 binding region

285   of the primers (205-222 and 752-772) has resided in a stable stem site hindered the effective

286   annealing. The RNA structure showed the complex stem-loop region and open sites as well.

287   Our targeted primer binding sites for the variant 26144 G>T (p.G251V) were within the stem-

9

288   loop region whereas the primer annealing sites for 25563 G>T (p.Q57H) variant reside within

289   the open region of the template. Here we assume, the complex structural configuration of NS3

290   may block the PCR reaction during a competitive multiplexing (**supplementary Fig.s3**). A

291   longer RNA denaturation step during cDNA synthesis and more stringent cycle denaturation

292   of cDNA template might solve this issue. However, it could damage low-concentrated, sample-

293   extracted viral RNA and inhibit amplification of other clade-specific templates by affecting

294   overall optimized multiplex condition.

295   The advantage of our ARMS-based multiplex assays is rapid. The turnaround time for our

296   designed assay would range from approximately 3-4 hours for 96 samples. The NGS and

297   Sanger methods, on the other hand, had a turnaround time of more than 24 hours and 10-12

298   hours, respectively (Tsuchihashi & Dracopoli, 2002; J. Zhang et al., 2020). An ARMS based

299   multiplex PCR assay similar to the current study would render a more convenient way to detect

300   clade specific mutation (SNPs) due to the process being faster and cost effective (Ahlawat,

301   Sharma, Maitra, Roy, & Tantia, 2014). The cost of the assay for a single reaction was $7 per

302   run (the cost includes import Tax and VAT etc. for Bangladesh) that is much less than targeted

303   and whole-genome based NGS methods in identifying the clades. The cost will be further

304   reduced if an optimized one-step PCR system is used and we are currently working on it to cut

305   the overall cost down to less than $2. Thus, our method can overcome a serious limitation to

306   effectively identify viral clades with a prospective broader application. The requirement of

307   technical skill would also be low for this assay wherein the training of personnel is a minimal

308   requirement and interpretation of results is generic (Syrmis et al., 2004; L. Wang et al., 2011).

309   Besides, the presence of the template as well as their quantity and quality are determined at the

310   same time. The false-negative result for the absence of a template can also be determined in a

311   facile manner (Edwards & Gibbs, 1994). In general, mutating the primer at its 3'prime end

312   makes it refractory to the 'wild type template' whereas the absence of mutation in the primer

313   is retractable to the 'mutant template' amending a reliable technique over sequencing

314   (Chulakasian et al., 2010). On the other hand, next generation sequencing technology such as

315   whole genome sequencing (WGS) or metagenomics approach can generate millions of high-

316   throughput data that enabled researchers to unroll new dimensions in the field of genome

317   sequencing applications (El-Metwally, Hamza, Zakaria, & Helmy, 2013). The lack of technical

318   personnel to analyze NGS data is also a reason to prefer alternative approach other than NGS

319   technology in low-income countries. Therefore, the ARMS technology with the conventional

320   multiplex PCR methods in identifying the clades would be more applicable in low and

321   minimum resource settings.

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 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.

Declaration of competing interest

The authors declare no completing interest to this work.

Authors' contribution

OKI, ARA, and IKJ conceived and designed the study outline. MTI and MSH setup initial experiments. MTI established the methodology with the assistance of NS, TC and MT. MTI, NS processed the experimental data, interpreted the results and generated the figures. ARA designed the primers for multiplex ARMS-PCR and MSH designed the primers for targeted SNP positions. ARA performed bioinformatics and *in silico* primer validation. MSH analyzed sequence data. NS, ARA, and MSH wrote the manuscript in consultation with OKI, HMA, IKJ and MTI. HMA and IKJ further reviewed and edited the manuscript extensively. IKJ and MAH supervised the project and edited the manuscript. Final approval was provided by all authors.

Availability of data and materials

Data can be checked from the GISAID EpiFlu™ database and the accession numbers are EPI_ISL_548260, EPI_ISL_561630, EPI_ISL_561375, EPI_ISL_561376 and EPI_ISL_561377.

Ethics approval and consent to participate

353 This study was approved by the ethical review committee of Jashore University of Science and

354 Technology (Ref: ERC/FBS/JUST/2020-45, Date: 06/10/2020). The samples have been used

355 from the routine diagnosis.

356 **Consent for publication**

357 Not applicable.

358 **References**

359 Ahlawat, S., Sharma, R., Maitra, A., Roy, M., & Tantia, M. S. (2014). Designing, optimization
360 and validation of tetra-primer ARMS PCR protocol for genotyping mutations in caprine
361 Fec genes. *Meta Gene*, *2*, 439–449. https://doi.org/10.1016/j.mgene.2014.05.004

362 Alam, A. S. M. R. U., Islam, M. R., Rahman, M. S., Islam, O. K., & Hossain, M. A. (2020).
363 Understanding the possible origin and genotyping of the first Bangladeshi SARS-CoV-2
364 strain. *Journal of Medical Virology*, Epub ahead of print.
365 https://doi.org/10.1002/jmv.26115

366 Alam, A. S. M. R. U., Islam, O. K., Hasan, M. S., Al-Emran, H. M., Jahid, M. I. K., & Hossain,
367 M. A. (2020). *Evolving Infection Paradox of SARS-CoV-2: Fitness Costs Virulence?*
368 https://doi.org/10.31219/OSF.IO/Y36VE (preprint)

369 Brister, H., Barnum, S. M., Reedy, S., Chambers, T. M., & Pusterla, N. (2019). Validation of
370 two multiplex real-time PCR assays based on single nucleotide polymorphisms of the
371 HA1 gene of equine influenza A virus in order to differentiate between clade 1 and clade
372 2 Florida sublineage isolates. *Journal of Veterinary Diagnostic Investigation*, *31*(1), 137–
373 141. https://doi.org/10.1177/1040638718822693

374 Brufsky, A. (2020). Distinct viral clades of SARS-CoV-2: Implications for modeling of viral
375 spread. *Journal of Medical Virology*, Vol. 92, pp. 1386–1390.
376 https://doi.org/10.1002/jmv.25902

377 Chellapandi, P., & Saranya, S. (2020). Genomics insights of SARS-CoV-2 (COVID-19) into
378 target-based drug discovery. *Medicinal Chemistry Research*, Vol. 29, pp. 1777–1791.
379 https://doi.org/10.1007/s00044-020-02610-8

380 Chulakasian, S., Lee, M. S., Wang, C. Y., Chiou, S. S., Lin, K. H., Lin, F. Y., … Hsu, W. L.
381 (2010). Multiplex Amplification Refractory Mutation System Polymerase Chain Reaction
382 (ARMS-PCR) for diagnosis of natural infection with canine distemper virus. *Virology*
383 *Journal*, *7*(1), 122. https://doi.org/10.1186/1743-422X-7-122

384 Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-
385 19 in real time. *The Lancet Infectious Diseases*, Vol. 20, pp. 533–534.
386 https://doi.org/10.1016/S1473-3099(20)30120-1

387 Edwards, M. C., & Gibbs, R. A. (1994). Multiplex PCR: Advantages, development, and
388 applications. *Genome Research*, *3*(4). https://doi.org/10.1101/gr.3.4.S65

389 El-Metwally, S., Hamza, T., Zakaria, M., & Helmy, M. (2013). Next-Generation Sequence
390 Assembly: Four Stages of Data Processing and Computational Challenges. *PLoS*
391 *Computational Biology*, Vol. 9. https://doi.org/10.1371/journal.pcbi.1003345

392 Islam, M. R., Hoque, M. N., Rahman, M. S., Alam, A. S. M. R. U., Akther, M., Puspo, J. A.,

393 … Hossain, M. A. (2020). Genome-wide analysis of SARS-CoV-2 virus strains
394 circulating worldwide implicates heterogeneity. *Scientific Reports*, *10*(1).
395 https://doi.org/10.1038/s41598-020-70812-6

396 Islam, O. K., Al-Emran, H. M., Hasan, M. S., Anwar, A., Jahid, M. I. K., & Hossain, M. A.
397 (2020). Emergence of European and North American mutant variants of SARS-CoV-2 in
398 South-East Asia. *Transboundary and Emerging Diseases*, Epub ahead of print.
399 https://doi.org/10.1111/tbed.13748

400 Kamitani, W. (2020). Nonstructural proteins of Novel Coronavirus (SARS-CoV-2)新型コロ
401 ナウイルス等のウイルス複製に必要な蛋白質. *Proceedings for Annual Meeting of*
402 *The Japanese Pharmacological Society*, *93*(0), 2-ES-2.
403 https://doi.org/10.1254/jpssuppl.93.0_2-es-2

404 Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., …
405 Montefiori, D. C. (2020). Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G
406 Increases Infectivity of the COVID-19 Virus. *Cell*, *182*(4), 812-827.e19.
407 https://doi.org/10.1016/j.cell.2020.06.043

408 Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular
409 evolutionary genetics analysis across computing platforms. *Molecular Biology and*
410 *Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

411 Lai, C. H., Welter, M. W., & Welter, L. M. (1995). The use of ARMS PCR and RFLP analysis
412 in identifying genetic profiles of virulent, attenuated or vaccine strains of TGEV and
413 PRCV. *Advances in Experimental Medicine and Biology*, *380*, 243–250.
414 https://doi.org/10.1007/978-1-4615-1899-0_39

415 Lee, E. J., Kim, E. J., Shin, Y. K., & Song, J. Y. (2016). Design and testing of multiplex RT-
416 PCR primers for the rapid detection of influenza A virus genomic segments: Application
417 to equine influenza virus. *Journal of Virological Methods*, *228*, 114–122.
418 https://doi.org/10.1016/j.jviromet.2015.11.012

419 Li, Y., Ji, D., Cai, W., Hu, Y., Bai, Y., Wu, J., & Xu, J. (2020). Clinical characteristics, cause
420 analysis and infectivity of COVID-19 nucleic acid re-positive patients: A literature
421 review. *Journal of Medical Virology*, *n/a*(n/a). https://doi.org/10.1002/jmv.26491

422 Liu, D. X., Fung, T. S., Chong, K. K. L., Shukla, A., & Hilgenfeld, R. (2014). Accessory
423 proteins of SARS-CoV and other coronaviruses. *Antiviral Research*, Vol. 109, pp. 97–
424 109. https://doi.org/10.1016/j.antiviral.2014.06.013

425 Mercatelli, D., & Giorgi, F. M. (2020). Geographic and Genomic Distribution of SARS-CoV-
426 2 Mutations. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.01800

427 Mousavizadeh, L., & Ghasemi, S. (2020). Genotype and phenotype of COVID-19: Their roles
428 in pathogenesis. *Journal of Microbiology, Immunology and Infection*, Epub ahead of print.
429 https://doi.org/10.1016/j.jmii.2020.03.022

430 Ou, X., Liu, Y., Lei, X., Li, P., Mi, D., Ren, L., … Qian, Z. (2020). Characterization of spike
431 glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-
432 CoV. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15562-9

433 Platt, A. R., Woodhall, R. W., & George, A. L. (2007). Improved DNA sequencing quality and
434 efficiency using an optimized fast cycle sequencing protocol. *BioTechniques*, *43*(1), 58–

62. https://doi.org/10.2144/000112499

Rahman, M. S., Islam, M. R., Hoque, M. N., Alam, A. S. M. R. U., Akther, M., Puspo, J. A., … Hossain, M. A. (2020). Comprehensive annotations of the mutational spectra of SARS-CoV-2 spike protein: a fast and accurate pipeline. *Transboundary and Emerging Diseases*, Epub ahead of print. https://doi.org/10.1111/tbed.13834

Shi, X., Zhang, C., Shi, M., Yang, M., Zhang, Y., Wang, J., … Ma, X. (2013). Development of a single multiplex amplification refractory mutation system PCR for the detection of rifampin-resistant Mycobacterium tuberculosis. *Gene*, *530*(1), 95–99. https://doi.org/10.1016/j.gene.2013.07.060

Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, Vol. 22. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

Syrmis, M. W., Whiley, D. M., Thomas, M., Mackay, I. M., Williamson, J., Siebert, D. J., … Sloots, T. P. (2004). A sensitive, specific, and cost-effective multiplex reverse transcriptase-PCR assay for the detection of seven common respiratory viruses in respiratory samples. *Journal of Molecular Diagnostics*, *6*(2), 125–131. https://doi.org/10.1016/S1525-1578(10)60500-4

To, K. K.-W., Hung, I. F.-N., Ip, J. D., Chu, A. W.-H., Chan, W.-M., Tam, A. R., … Yuen, K.-Y. (2020). COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. *Clinical Infectious Diseases,* Epub ahead of print. https://doi.org/10.1093/cid/ciaa1275

Tsuchihashi, Z., & Dracopoli, N. C. (2002). Progress in high throughput SNP genotyping methods. *Pharmacogenomics Journal*, Vol. 2, pp. 103–110. https://doi.org/10.1038/sj.tpj.6500094

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15). https://doi.org/10.1093/nar/gks596

Wang, L., Si, Y., Dedow, L. K., Shao, Y., Liu, P., & Brutnell, T. P. (2011). A low-cost library construction protocol and data analysis pipeline for illumina-based strand-specific multiplex RNA-seq. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0026426

Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., … Qi, J. (2020). Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell*, *181*(4), 894-904.e9. https://doi.org/10.1016/j.cell.2020.03.045

Wang, W., Ren, P., Mardi, S., Hou, L., Tsai, C., Chan, K. H., … Deubel, V. (2009). Design of multiplexed detection assays for identification of avian influenza a virus subtypes pathogenic to humans by SmartCycler real-time reverse transcription-PCR. *Journal of Clinical Microbiology*, *47*(1), 86–92. https://doi.org/10.1128/JCM.01090-08

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, *13*, 134. https://doi.org/10.1186/1471-2105-13-134

Zhang, C., Liu, Y., Ring, B. Z., Nie, K., Yang, M., Wang, M., … Ma, X. (2013). A Novel Multiplex Tetra-Primer ARMS-PCR for the Simultaneous Genotyping of Six Single Nucleotide Polymorphisms Associated with Female Cancers. *PLoS ONE*, *8*(4).

14

478   https://doi.org/10.1371/journal.pone.0062126

479   Zhang, J., Yang, J., Zhang, L., Luo, J., Zhao, H., Zhang, J., & Wen, C. (2020). A new SNP
480   genotyping technology Target SNP-seq and its application in genetic analysis of
481   cucumber varieties. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-62518-
482   6

483   Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction.
484   *Nucleic Acids Research*, *31*(13), 3406–3415. https://doi.org/10.1093/nar/gkg595

485

486   **Table 1. Primer sets for targeted SNP-based single- and/or multi-plex PCR**

Clade	SNP position	Primer*	direction	Sequence (5'-3')	Position in CDS#	Amplicon size (bp)	Tm (°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	GATTAACAACTCCGGATGAAC	ORF3a_752-772		58.7
		NS3_26144_mR		GATTAACAACTCCGGATGAAA	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

15

Clade	Name	Primer Name		Sequence	Name#	Amplicon Size (bp)	Tm (°C)
		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	TGGCATCATAAAGTAATGGG	ORF3a_411-430		57.8
GR	28881-28883	N_28882_F	Forward	CCAGATGACCAAATTGGC	N protein_238-255	387	58
		N_28882_wR	Reverse	TAGCAGGAGAAGTTCCCC	N protein_608-625		60
		N_28882_mR		TAGCAGGAGAAGTTCGTT	N protein_608-625		58

487 ** The 'w' and 'm' in the primer names denotes, respectively, the wild and mutant type allele*

488 *corresponding to 'no' and 'single' base change for the wild and mutant type.*

489 *# The nucleotide position of the coding sequence for each protein where the primers bind to.*

490

491 **Table 2. List of primers for targeted Sanger sequencing to validate the multiplex assays**

Clade	Name	Primer Name	Sequence	bp	Amplicon Size (bp)	Tm (°C)
S		NS8_1_F	GTGGATGAGGCTGGTTCTAA	20	217	54.6

|	NS8_ 28144_ T to C	NS8_1_R	TGGGGTCCATTATCAGACAT	20		53.6
V	NS3_ 26144_G to T	NS3_3_F	CTGGTGTTGAACATGTTACCTT	20	209	53.2
		NS3__3_R	CTCTTCCGAAACGAATGAGTA	20		52
G	Spike_ 23403_ A to G	Spike_1_F	CGTGATCCACAGACACTTGA	20	228	54.6
		Spike_1_R	CCCTATTAAACAGCCTGCAC	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

492

493

494

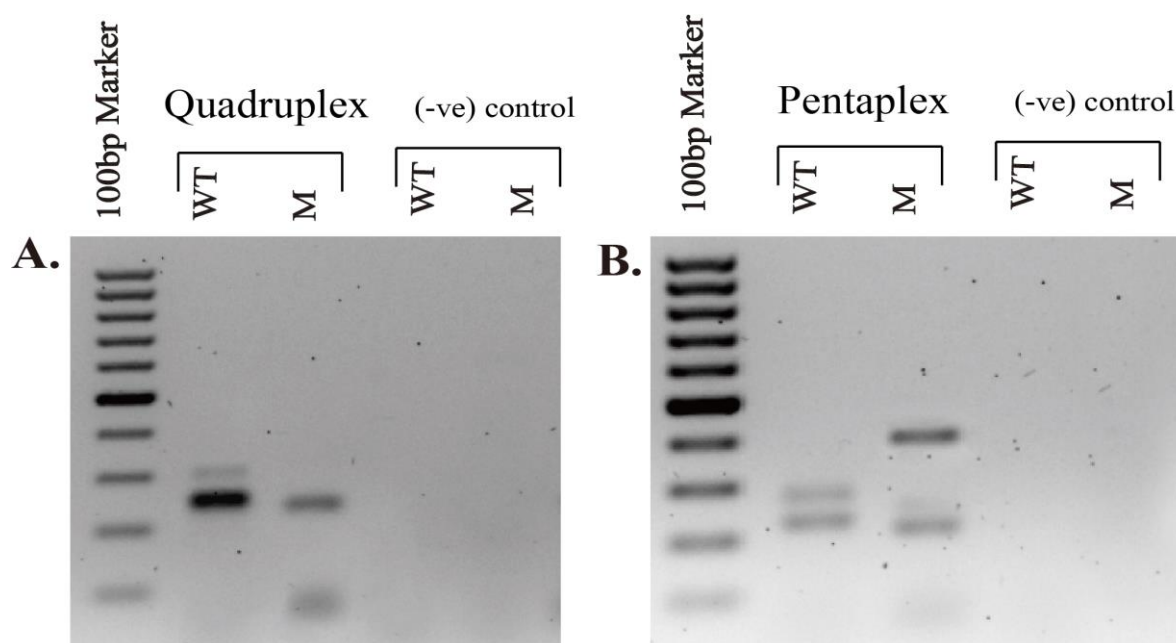495

496

497

498

499

500

501

502

503
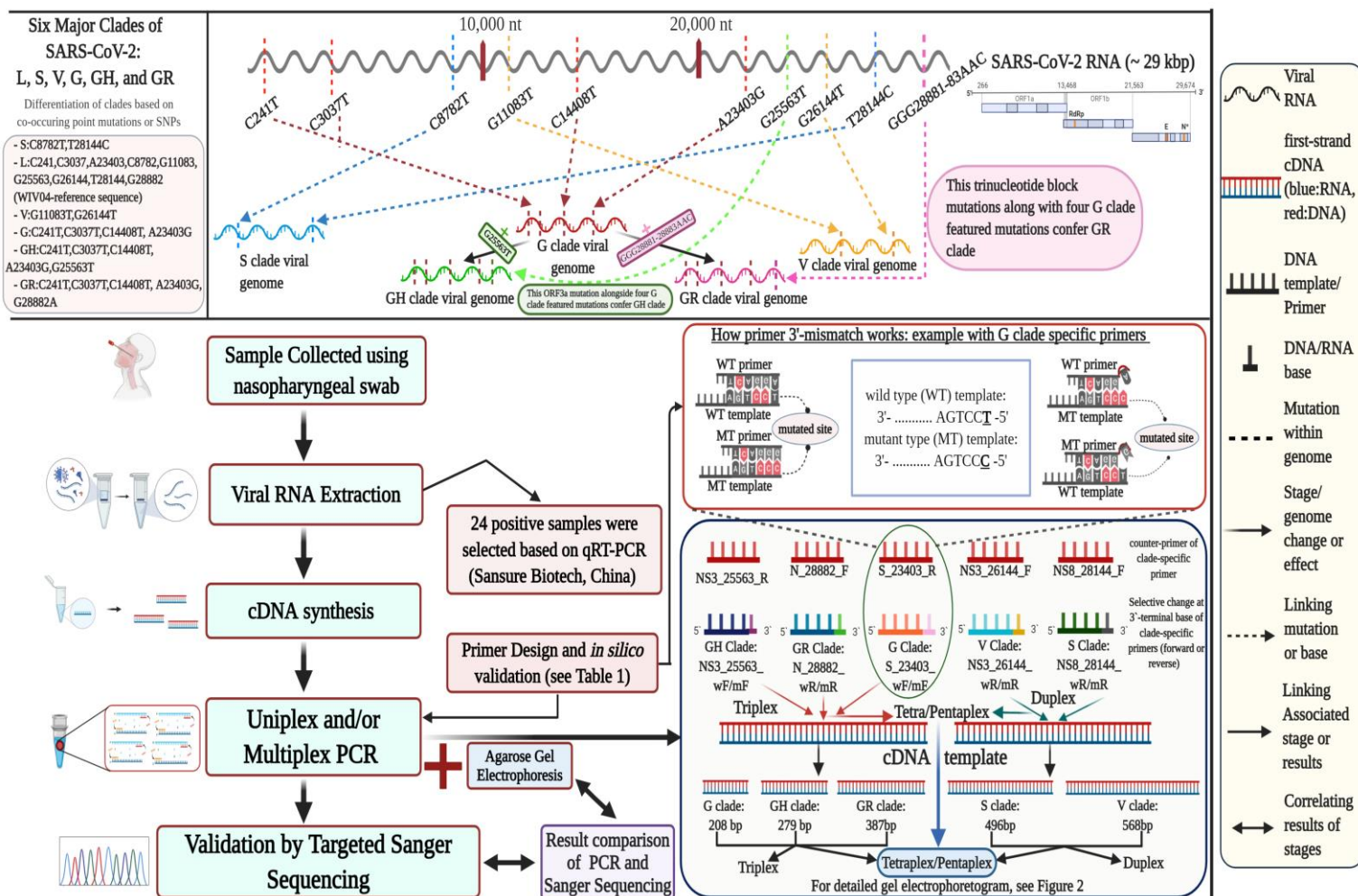
Figure 1. Schematic workflow of ARMS-based multiplex PCR assays for the identification of SARS-CoV-2 clades. The upper portion of the figure showed the concept of clade as described in the GISAID with a comprehensive genomic visualization. The lower segment is dedicated to the overall workflow and the primer design.

504

505

506

507

508

509

510

511

512

513

514
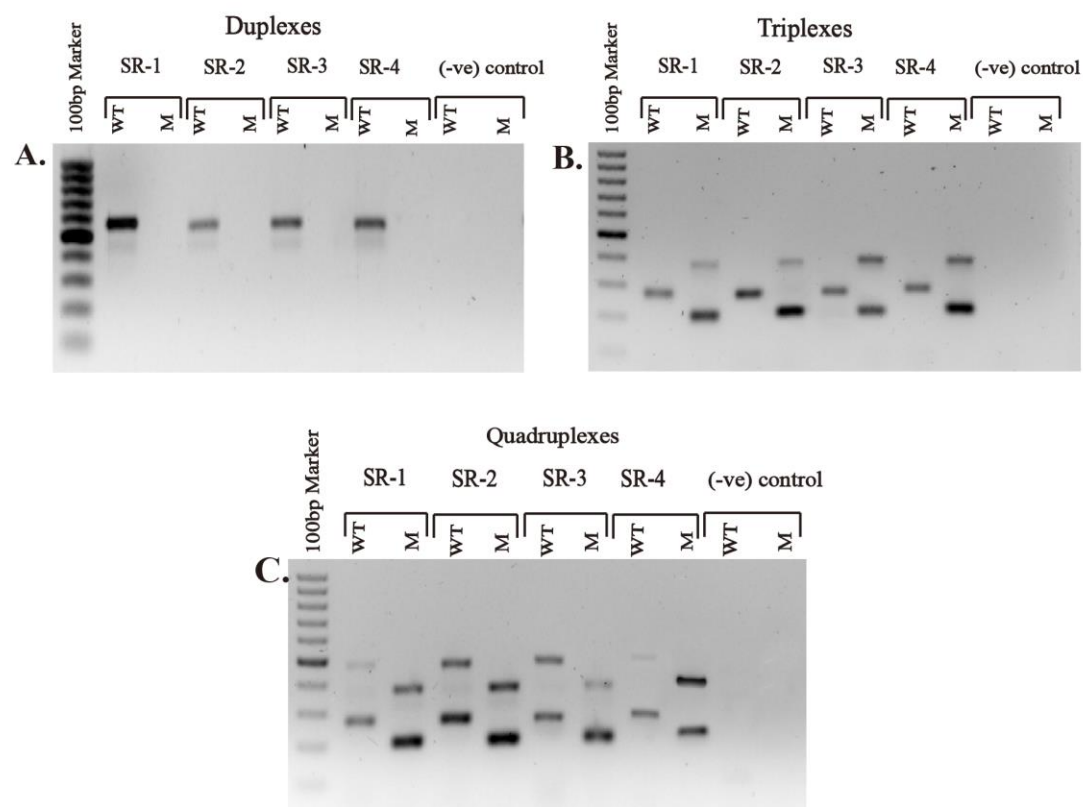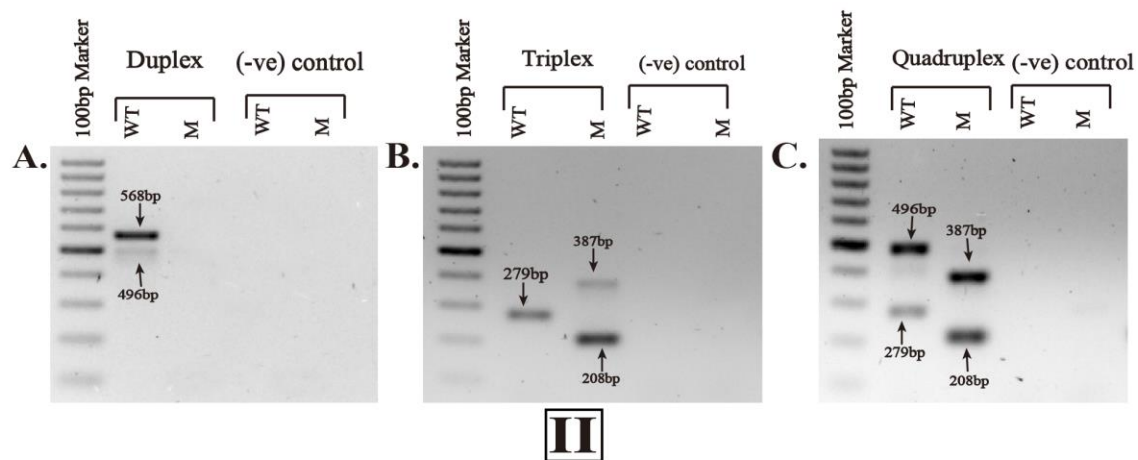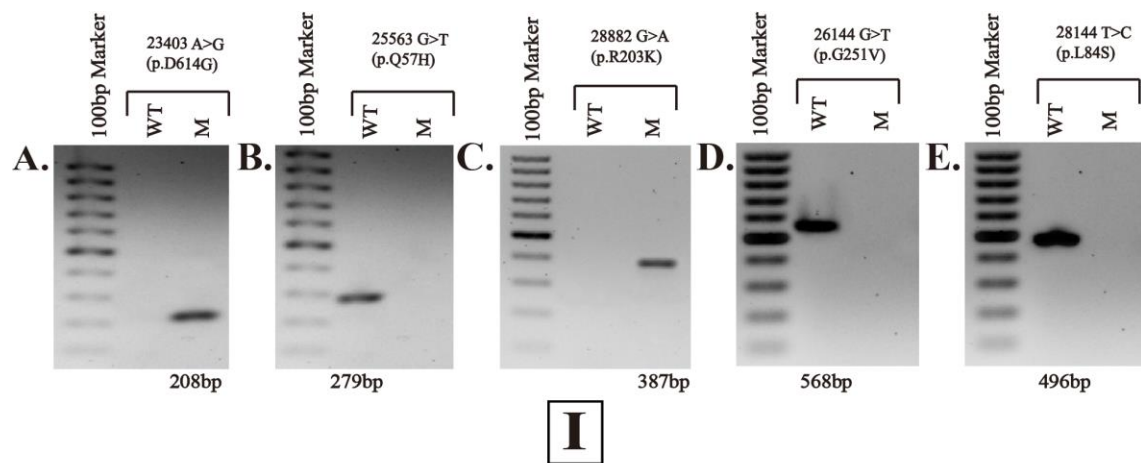
516 **Figure 2. Strategy and validation of ARMS-based multiplex PCR assays**. **(I)** Single-
517 variant specific PCRs with the respective amplicons in bp at the bottom of each gel image,
518 showing the variants. **(II)** Multiplex PCR assays, containing PCR products for different clade
519 combinations: duplex for 26144 G>T (p.G251V) and, 28144 T>C (p.L84S) variants at an
520 annealing temperature of 60°C; triplex for the variants 23403 A>G (p.D614G), 25563 G>T
521 (p.Q57H) and, 28882 G>A (p.R203K) at 57°C; quadruplex for the variants 23403 A>G
522 (p.D614G), 25563 G>T (p.Q57H), 28882 G>A (p.R203K) and, 28144C T>C (p.L84S) at 56°C.
523 SARS-CoV-2 positive sample, GC 46.003 was used as a representative to perform the PCRs
524 for (I) and (II). **(III)** Validation of the multiplex PCR assays, containing the identical settings
525 for duplex, triplex and quadruplex PCRs with GC 44.201 (denoted as SR-1), GC 88.025
526 (denoted SR-2), GC 90.175 (denoted SR-3) and GC 92.172 (denoted as SR-4) as
527 representatives to display the reproducibility of the assays. SARS-CoV-2 negative sample, GC
528 116.09 was used as (-ve) control for the comparison. 'GC' indicates the Genome Center
529 identification number generated at the Genome Center of Jashore University of Science and
530 Technology for COVID-19 suspected patients. WT, wild type band; M, Mutant band. 100-
531 1000bp marker was used in the first lane of the 1% agarose gels. The primers are listed in Table
532 1.

533

534 **Figure s1. Failed attempts in generating optimized quadruplex and pentaplex assays.** No
535 optimization or adjustment could discriminate the desired location of the bands in quadruplex
536 (A) and in the pentaplex (B). Quadruplex comprises the SNP variants 23403 A>G (p.D614G),

20

537   25563 G>T (p.Q57H), 28882 G>A (p.R203K) and, 26144 G>T (p.G251V). Pentaplex includes

538   the SNP variants 23403 A>G (p.D614G), 25563 G>T (p.Q57H), 28882 G>A (p.R203K),

539   28144C T>C (p.L84S) and 26144 G>T (p.G251V). SARS-CoV-2 positive sample, GC 46.003

540   was used as a representative to perform the PCRs. SARS-CoV-2 negative sample, GC 116.09

541   was used as (-ve) control for the comparison. 'GC' indicates the Genome Center identification

542   number generated at the Genome Center of Jashore University of Science and Technology for

543   COVID-19 suspected patients. WT, wild type band; M, Mutant band. 100-1000bp marker was

544   used in the first lane of the 1% agarose gels. The primers are listed in Table 1.
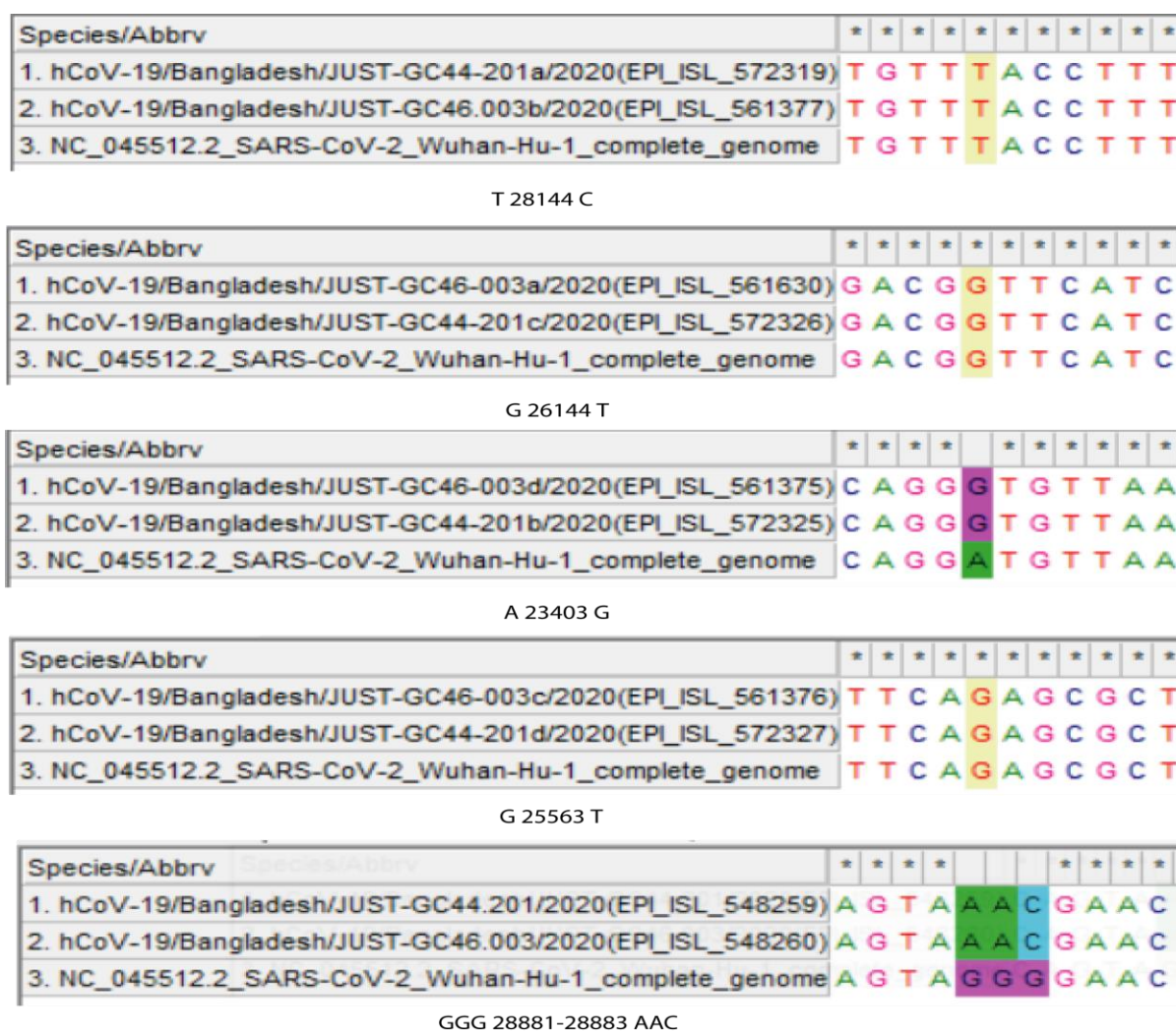
546   **Figure s2. Sanger sequencing of variant-specific mutation (SNPs) positions to detect the**

547   **clade of SARS-CoV-2.** Alignment of the partial sequences with the reference sequence

548   (NC_045512.2_SARS-CoV-2_Wuhan-Hu-1_complete_genome) showed that both the

549   GC44.201and GC46.003 samples are GR Clade viruses. 'GC' indicates the Genome Center

550   identification number generated at the Genome Center of Jashore University of Science and

551   Technology for COVID-19 suspected patients.

21

dG = -242.57 [Initially -257.80] ORF3a

552

553
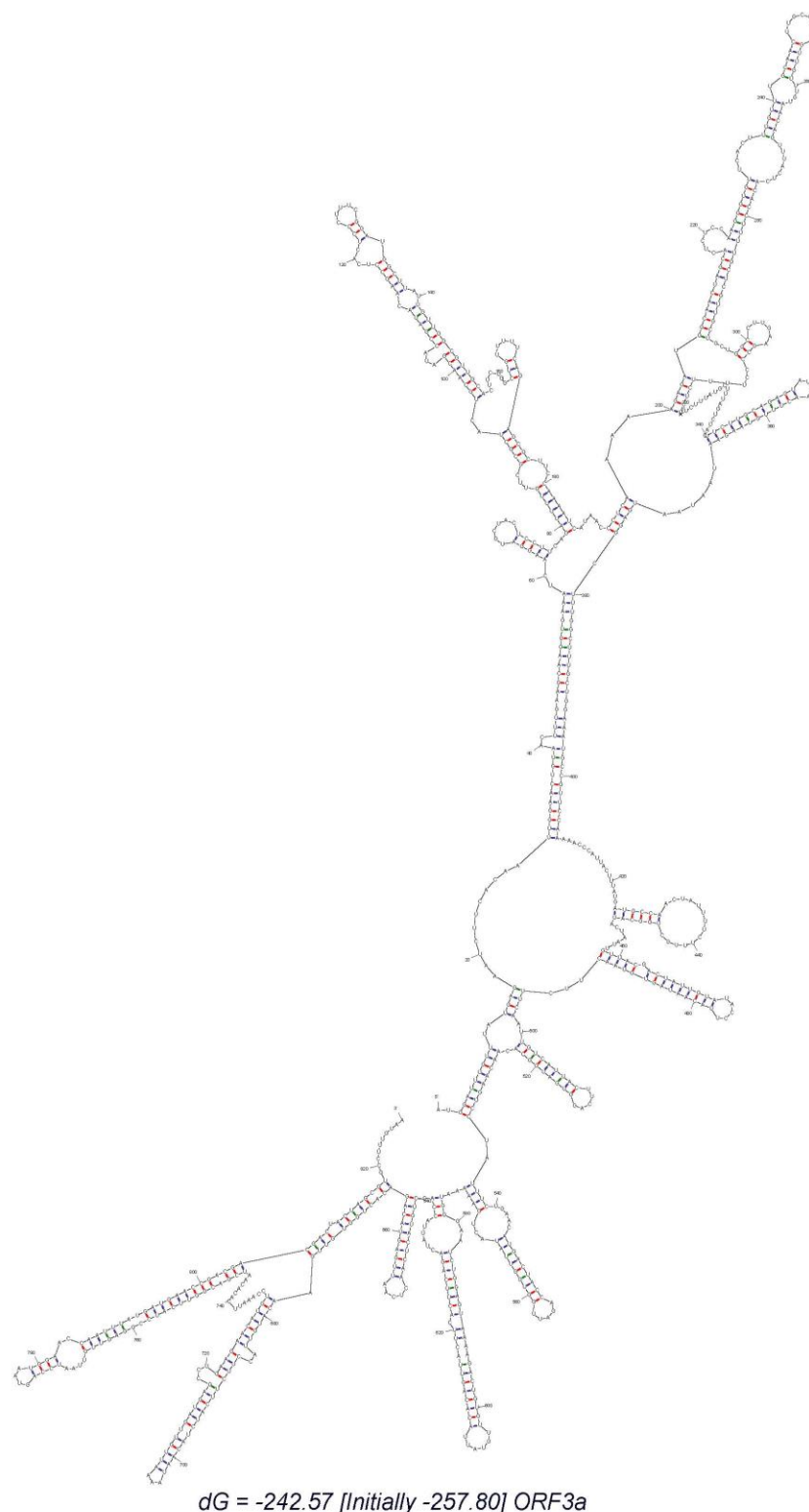
554 **Figure s3. The secondary RNA structure of NS3 of ORF3a raw template**. The interval is

555 20 nucleotides. The free energy is -242.57 Kcal/mol.

556

557 **Table s1. List of SARS-CoV-2 positive samples examined at the Genome Center of**

558 **Jashore University of Science and Technology**

	Sample/ Specimen ID*	District	Date of Sample Collection (dd-mm-yyyy)	Date of Test (dd-mm-yyyy)	Result
1	GC36.122	Jashore	7/6/2020	8/6/2020	Positive
2	GC37.01	Jashore	8/6/2020	8/6/2020	Positive
3	GC42.44	Jashore	13/06/2020	14/06/2020	Positive
4	GC43.374	Jashore	14/6/2020	16/06/2020	Positive
5	GC44.201	Sathkhira	15/06/2020	18/06/2020	Positive
6	GC46.003	Jashore	17/06/2020	19/06/2020	Positive
7	GC46.026	Jashore	17/06/2020	19/06/2020	Positive
8	GC48.152	Bagerhat	19/06/2020	20/06/2020	Positive
9	GC48.159	Bagerhat	19/06/2020	20/06/2020	Positive
10	GC51.183	Jashore	22/06/2020	23/06/2020	Positive
11	GC53.091	Satkhira	24/06/2020	25.06.2020	Positive
12	GC86.011	Jashore	03/08/2020	04/08/2020	Positive
13	GC86.020	Jashore	03/08/2020	04/08/2020	Positive
14	GC87.025	Jashore	4/08/2020	5/08/2020	Positive
15	GC88.005	Magura	5/08/2020	6/08/2020	Positive
16	GC88.131	Jashore	5/08/2020	6/08/2020	Positive
17	GC88.222	Jashore	5/08/2020	6/08/2020	Positive
18	GC88.025	Magura	5/08/2020	6/08/2020	Positive
19	GC88.056	Narail	5/08/2020	6/08/2020	Positive

20	GC91.124	Jashore	8/08/2020	9/08/2020	Positive
21	GC90.175	Narail	7/08/2020	8/08/2020	Positive
22	GC92.172	Jashore	9/08/2020	10/08/2020	Positive
23	GC52.200	Bagerhat	23/06/2020	24/06/2020	Positive
24	GC53.110	Magura	24/06/2020	25/06/2020	Positive

559 *'GC' indicates the Genome Center identification number generated at the Genome Center

560 of Jashore University of Science and Technology for COVID-19 suspected patients.

561

562

563 **Table s2: List of SARS-CoV-2 negative samples examined at the Genome Center of**

564 **Jashore University of Science and Technology**

	Sample/ Specimen ID*	District	Date of Sample Collection (dd-mm-yyyy)	Date of Test (dd-mm-yyyy)	Result
1	GC116.08	Jashore	8/09/2020	9/09/2020	Negative
2	GC116.09	Jashore	8/09/2020	9/09/2020	Negative
3	GC116.11	Jashore	8/09/2020	9/09/2020	Negative
4	GC116.12	Jashore	8/09/2020	9/09/2020	Negative
5	GC116.13	Jashore	8/09/2020	9/09/2020	Negative

565 *'GC' indicates the Genome Center identification number generated at the Genome Center of

566 Jashore University of Science and Technology for COVID-19 suspected patients.