Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

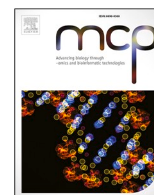
Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Molecular and Cellular Probes 58 (2021) 101748

Contents lists available at ScienceDirect

Molecular and Cellular Probes

journal homepage: www.elsevier.com/locate/ymcpr



In-silico analysis of Covid-19 genome sequences of Indian origin: Impact of mutations in identification of SARS-Co-V2

Kabilan Mani [a,b], Kannan Thirumalmuthu [c], Divya Sri Kathiresan [a,b], Sudha Ramalingam [a,b,d], Ramalingam Sankaran [a,b,e], Sankarganesh Jeyaraj [a,b,*]

[a] *PSG Center for Molecular Medicine & Therapeutics, PSG Institute of Medical Sciences and Research, Coimbatore, Tamil Nadu, 641 004, India*
[b] *PSG Centre for Genetics and Molecular Biology, Off Avinashi Road, Peelamedu, Coimbatore, Tamil Nadu, 641 004, India*
[c] *Lipid and Nutrition Laboratory, Department of Lipid Science, Council of Scientific and Industrial Research-Central Food Technological Research Institute, Mysuru, India*
[d] *Department of Community Medicine, PSG Institute of Medical Sciences and Research, 641004, Coimbatore, Tamil Nadu, India*
[e] *Department of Pharmacology, PSG Institute of Medical Sciences and Research, 641 004, Coimbatore, Tamil Nadu, India*

ARTICLE INFO

Keywords:
Covid-19
Primer and probe mismatches
rRT-PCR
Indian SARS-CoV-2 genome sequences

ABSTRACT

Covid-19 disease caused by SARS-CoV-2 is still being transmitted in developed and developing countries irrespective of healthcare setups. India with 1.3 billion people in the world is severely affected by Covid-19 with 11.3 million cases and 157 000 deaths so far. We have assessed the mismatches in WHO recommended rRT-PCR assays primer and probe binding regions against SARS-CoV-2 Indian genome sequences through in-silico bioinformatics analysis approach. Primers and probe sequences belonging to CN-CDC-ORF1ab from China and HKU-ORF1b from Hong Kong targeting ORF1ab gene while NIH-TH-N from Thailand, HKU-N from Hong Kong and US-CDCN-2 from USA targeting N genes displayed accurate matches (>98.3%) with the 2019 novel corona virus sequences from India. On the other hand, none of the genomic sequences displayed exact match with the primer/probe sequences belonging to Charité-ORF1b from Germany targeting ORF1ab gene. We think it will be worthwhile to release this information to the clinical and medical communities working in Indian Covid-19 frontline taskforce to tackle the recently emerging Covid-19 outbreaks as of March-2021.

1. Introduction

Covid-19 pandemic disease is caused by novel Severe Acute Respiratory Syndrome corona virus 2 (SARS-CoV-2) worldwide. The virus has transmitted to both developed and developing countries irrespective of the healthcare setups [1]. Diagnosis of Covid-19 disease is vital to detect the infected individuals in community regarding quarantine and treatment, where real-time Reverse Transcription Polymerase Chain Reaction (rRT-PCR) is employed to detect the presence of viral pathogen in swab samples collected from Covid-19 suspected patients [2]. The virus has a strong transmission potential and caused Covid-19 outbreak in several countries throughout the world. Prompt and early detection of SARS-COV-2 through rRT-PCR is essential for the disease control strategies planned by national healthcare agencies and Government programmes. To date, there are several rRT-PCR protocols developed by different nations and approved by WHO. SARS-CoV-2 is a RNA virus having the potential to evolve rapidly through high mutation rates [3]. Since its first outbreak in December 2019 in Wuhan province of China it rapidly circulated throughout the world with accumulation of nucleotide polymorphisms in the genomes of SARS-CoV-2 strains [4–6].

Mismatches in the primer and probe regions of SARS-CoV-2 strains may lead to their diminished ability binding to corresponding template DNA molecules during the initial steps of PCR reactions. Mutations in the primer and probe regions have been shown to affect the sensitivity of PCR reactions in case of other viral infections such as human immunodeficiency virus (HIV), Hepatitis B virus (HBV), H1N1 virus and dengue [7–13].

India being the second most populous country in the world severely affected by Covid-19 and as of 10-March-2021, 11 244 786 cases and 157 930 deaths has been reported in India since its first Covid-19 case reported on 30-January-2020 [14]. With 1.3 billion people living in diverse states of different culture and health care inequalities, early detection of SARS-CoV-2 infection with robust diagnostic tests are in great demand to control the disease transmission. Previous studies

* Corresponding author. PSG Center for Molecular Medicine & Therapeutics, PSG Institute of Medical Sciences and Research, Coimbatore, Tamil Nadu, 641 004, India.
E-mail address: sankarganesh@psgimsr.ac.in (S. Jeyaraj).

https://doi.org/10.1016/j.mcp.2021.101748
Received 25 March 2021; Received in revised form 29 May 2021; Accepted 12 June 2021
Available online 16 June 2021
0890-8508/© 2021 Elsevier Ltd. All rights reserved.

reported the mismatches in primer probe regions of Covid-19 PCR primers used the sequences representing SARS-CoV-2 strains from different countries, nevertheless there are no studies focusing on the SARS-CoV-2 genomes from Indian Sub-Continent. In this study, we evaluated mismatches in PCR primers against SARS-CoV-2 genomes from India through in-silico bioinformatics analysis approach [15,16].

2. Materials and methods

Briefly, a total of 3765 whole genome sequences of SARS-CoV-2 strains isolated from India were retrieved in FASTA format from GISAID (Global Initiative on Sharing All Influenza Data) database submitted as on January 30, 2021 (Supplementary Table 1) [17]. Further search criteria employed for collecting the sequences from the database is as follows: (i) "hCoV-19" was selected under virus name (ii) "Human" was selected under host name (iii) location was selected "Asia/India" (iv) complete genomes alone were included (>29000 bp length) and (v) high coverage genomes alone were considered (low coverage entries i.e. >5% Ns were excluded). The collected individual FASTA sequences were then merged as a single FASTA file along with the inclusion of the complete genome of Wuhan-Hu-1 as reference genome in 29903 bp length (NCBI Reference Sequence: NC_045512.2). Primers and probe sequences were collected and shortlisted from the recommended protocols in the literature and from WHO website (Table 1) [18]. A detailed protocol is presented as in supplementary file 1.

A large-scale multiple sequence alignment (LS-MSA) was carried out for the given FASTA files using MAFFT (Multiple Alignment using Fast Fourier Transform) v.7 tool available online (https://mafft.cbrc.jp/alignment/server/) [25]. MSA was carried out with the following options, (i) "Allow unusual symbols", (ii) "same as input" for nucleotide sequence direction, (iii) "Aligned" output order and (iv) "Auto" strategy. The aligned sequences were then downloaded and visualized using AliView program [26]. Sequences representing from forward primer initial nucleotide to reverse primer end nucleotide regions were trimmed from rest of the whole genome alignments and saved separately as FASTA files. Variability within primer-probe sequences were estimated as described elsewhere [16]. Briefly, positional nucleotide numerical summary (PNNS) and entropy was calculated using PNNS (http://entropy.szu.cz:8080/EntropyCalcWeb/pnns) and entropy calculator (http://entropy.szu.cz:8080/EntropyCalcWeb/) available in Alignment Explorer respectively [27]. Further, sequence stratification of primer-probe alignments was carried out using SequenceTracer module in Alignment Explorer with the segregation of sequences into discrete groups of identical sequence variants along with their frequency for each primer/probe. During the execution of the program, sequences with stretches of NNNs and ambiguous sequences in the alignments were excluded from the study. For the removal of extremely low prevalent variants and sequencing errors in the data, a threshold of at least 0.5% incidence of variation was considered as meaningful and those sequences alone were further considered for analysis [16]. Finally, the melting temperature (Tm) of the original primer/probe sequence against the variant sequences was calculated after mapping the whole genome sequences against the same. Since the length of primers included in this analysis is > 13 nucleotides, the following formula was applied for calculating the melting temperature $Tm = 64.9 + 41*(yG + zC - 16.4)/(wA + xT + yG + zC)$ where w, x, y, and z are the number of A, T, G, and C nucleotides, respectively [28].

3. Results

In this study, whole-genome sequences of SARS-CoV-2 strains isolated in India were obtained from GISAID database and their sequence similarity against various primers/probe employed in detection assays recommended by WHO were investigated (Supplementary Table 1). Variations in sequences were identified after performing multiple sequence alignment with MAFFT and visualized using Entropy Calculator and Sequence Tracer. The obtained Results were segregated into distinct groups based on the nucleotide variations. As described earlier [16], to avoid considering mismatches raised due to sequencing errors, a threshold value of ≥0.5% was set, where only sequences showing variation frequency above this limit were considered further. Non-informative sequences were segregated separately as outgroups. Outgroup 1 included sequences having "*n*" in them while incomplete/-short sequences were categorized as 'Outgroup1/2'. Targets with missing or no sequences were excluded from the data processing. The remaining sequences alone called as 'informative sequences' were considered for further analysis. These informative sequences were divided into two groups namely 'perfect matches' and 'mismatches'.

Among the assays targeted against N region of SARS-CoV-2 genome, NIH-TH-N from Thailand displayed highest accuracy with 98.35% of genome sequence matching exactly with corresponding primers and probe (Table 1). This was followed by both US-CDC-N-2 from USA and HKU-N from Hong Kong with 98.3% sequences displaying exact matches. Primers and probe sequences employed in NIID-JP-N assay from Japan displayed exact matches with 3691 (98.03%) SARS-CoV-2 whole genome sequences. The assay US-CDC-N-1 from USA had exact match with 3449 (91.6%) sequences. It had two significant single nucleotide mutations at 28312th position showing mismatch (C28312T) with 194 (5.15%) sequences at probe binding region and 28299th position showing mismatch (C28299A) with 20 (0.53%) sequences in the forward primer binding region. It also showed two insignificant consecutive nucleotide change CA to TT (C28315T and A28316T) of 4 (0.1%) sequences at the probe binding region. Among the assays targeting N gene, CN-CDC-N were found to be least specific. About 2043 (54.26%) sequences displayed exact matches with sequences and probe involved in CN-CDC-N assay while 1471 sequences (39.07%) displayed significant trinucleotide variation (GGG→AAC) in the forward primer region starting at 28881st position and significant single nucleotide mismatch with 45 (1.19%) sequences at forward primer region (C28887T) (Supplementary Fig. 1).

Other assays targeting the ORF1ab and E genes were also analyzed for their exact match accuracies between genomic sequences and primer/probe region. Among the assays targeting ORF1ab genes, HKU-ORF1b assay primers and probe sequences displayed exact matches with 3707 (98.46%) sequences. It is to be noted that the primers and probe used in HKU-ORF1b assay from Hong Kong were found to have degenerative nucleotides (Y=C/T, W = A/T and R = A/G). Further, primers and probe employed in CN-CDC-ORF1abassay had exact matches with 3710 (98.53%) SARS-CoV-2 genome sequences (Table 1). None of the sequences exhibited exact match with the primers and probe involved in Charité-ORF1b assay from Germany. Among 3620 (96.15%) sequences, a significant single nucleotide mutation (C/G15516T) was detected. Both the forward and reverse primers of Charité-ORF1b assay contained degenerative bases (R = A/G in forward primer, S = G/C, Y=C/T in reverse primer) like HKU-ORF1b assay. These bases are included to increase the binding specificity of primers and probes. A single assay targeting the E region, Charité-E, was also evaluated for the alignment accuracy between primer/probe and genomic DNA sequences. About 3667 (97.39%) of sequences exhibited exact matches with SARS-CoV-2 whole genome sequences while 139 (3.691%) of sequences were omitted from analysis as 'outgroups' (Supplementary Fig. 1). It is generally recommended that the range of hybridization temperature of the primer/probe sets should be less than 10 °C. In our study we found that 4 primer/probe sets namely, US-CDC-N1/USA, US-CDC-N2/USA, CN-CDC-ORF1ab/China and Charité-ORF1b/Germany were deviating from the prescribed threshold hybridization temperature. Mutations in the primer/probe binding regions in US-CDC-N1/USA and Charité-ORF1b/Germany assays further increased the hybridization temperature range (Table 1).

Overall, the primers and probe sequences belonging to CN-CDC-ORF1ab from China and HKU-ORF1b from Hong Kong targeting ORF1ab gene while NIH-TH-N from Thailand, HKU-N from Hong Kong

Table 1
Mismatch in the primer and probe sequences of different RT-qPCR assays recommended by WHO used to detect Covid-19.

Gene Target	Assay Name/Country	F/P/R	Sequence (5’ - 3′)	Position	sequences exhibiting exact match		sequences with variance greater than 0.5%		sequences with variance less than 0.5%		Total number of non-informative sequence		Tm Ref/Modified	Ref
					n	%	n	%	n	%	n	%		
N	US-CDC-N-1/USA	F	GACCCCAAAATCAGCGAAAT	28287–>28306	3449	91.6	214	5.68	19	0.5	24	0.63	49.7/47.7	[19]
		P	ACCCCGCATTACGTTTGGTGGACC	28 309–>28332									60.8/59.1	
		R	TCTGGTTACTGCCAGTTGAATCTG	28358<–28335									55.7/NA	
	US-CDC-N-2/USA	F	TTACAAACATTGGCCGCAAA	29164–>29183	3701	98.3	0	0	47	1.24	17	0.45	47.7/NA	[19]
		P	ACAATTTGCCCCCAGCGCTTCAG	29188–>29 210									58.8/NA	
		R	GCGCGACATTCCGAAGAA	29230<–29213									50.3/NA	
	CN-CDC-N/China	F	GGGGAACTTCTCCTGCTAGAAT	28881–>28902	2043	54.26	1516	40.26	92	2.44	114	3.02	54.8/51.1	[20]
		P	TTGCTGCTGCTTGACAGATT	28934–>28953									49.7/47.7	
		R	CAGACATTTTGCTCTCAAGCTG	28979<–28958									53/51.1	
	NIH-TH-N/Thailand	F	CGTTTGGTGGACCCTCAGAT	28320–>28339	3703	98.35	0	0	62	1.64	0	0	53.8/NA	[21]
		P	CAACTGGCAGTAACCA	28341–>28356									43.4/NA	
		R	CCCCACTGCGTTCTCCATT	28376<–28358									53.2/NA	
	NIID-JP-N/Japan	F	AAATTTTGGGGACCAGGAAC	29125–>29144	3691	98.03	0	0	49	1.30	25	0.66	49.7/NA	[22]
		P	ATGTCGCGCATTGGCATGGA	29222–>29241									53.8/NA	
		R	TGGCACCTGTGTAGGTCAAC	29282<–29263									53.8/NA	
	HKU-N/Hong Kong	F	TAATCAGACAAGGAACTGATTA	29145–>29166	3701	98.3	0	0	43	1.14	21	0.55	47.4/NA	[23]
		P	GCAAATTGTGCAATTTGCGG	29177<–29196									49.7/NA	
		R	CGAAGGTGTGACTTCCATG	29254<–29236									51.1/NA	
E	Charité-E/Germany	F	ACAGGTACGTTAATAGTTAATAGCGT	26269–>26294	3667	97.39	0	0	9	0.23	89	2.36	53.2/NA	[24]
		R	ACACTAGCCATCCTTACTGCGCTTCG	26332–>26357									61.1/NA	
		P	ATATTGCAGCAGTACGCACACA	26381<–26360									53/NA	
ORF1ab	Charité-ORF1b/Germany	F	GTGARATGGTCATGTGTGGCGG	15431–>15452	0	0	3620	96.15	12	0.31	139	3.68	56.7–58.6/54.8–56.7*	[24]
		P	CAGGTGGAACCTCATCAGGAGATGC	15470–>15494									61	
		R	CARATGTTAAASACACTATTAGCATA	15530<–15 505									50.1–51.7/48.5–50.1*	
	CN-CDC-ORF1ab/China	F	CCCTGTGGGTTTTACACTTAA	13342–>13362	3710	98.53	0	0	27	0.71	28	0.74	50.5/NA	[20]
		P	CCGTCTGCGGTATGTGGAAAGGTTATGG	13377–>13404									62.9/NA	
		R	ACGATTGTGCATCAGCTGA	13460<–13442									48.9/NA	
	HKU-ORF1b/Hong Kong	F	TGGGGYTTTACRGGTAACCT	18778–>18797	3707	98.45	0	0	16	0.42	42	1.11	49.7–53.8/NA	[23]
		P	TAGTTGTGATGCWATCATGACTAG	18 849–>18 872									52.3/NA	
		R	AACRCGCTTAACAAAGCACTC	18909<–18889									50.5–52.4/NA	

Table represents the Forward primer (F), Reverse primer (R) and Probe (P) sequences, their corresponding position in relation to Wuhan SARS-CoV-2 reference Genome, followed by the number of sequences exhibiting exact match. Sequences with variance of at least 0.5% might decrease the efficiency of primers and probes binding to the corresponding template DNA sequences. About 324 sequences submitted between November 25 2020 to November 30 2020 displayed misalignment with the reverse primer and therefore they have been omitted from the study. * - Range of melting temperature is provided in case of degenerate nucleotide

and US-CDC-N-2 from USA targeting N genes displayed accurate matches (>98.3%) with the 2019 novel corona virus sequences from India. On the other hand, US-CDC-N-1/USA and CN-CDC-N/China primers and probes targeting N gene of SARS-Co-V2 Indian genome sequences displayed only 91.6% and 54.26% respectively. Interestingly, none of the genomic sequences displayed exact match with the primer/probe sequences belonging to Charité-ORF1b from Germany targeting ORF1ab gene.

4. Discussion

Reliability in detection of diagnostically important viruses has been a major concern prevailing for decades in the field of virology. Specific binding of primers and the probes to the gene targets must be achieved to obtain the PCR gene products. Single or multiple nucleotide changes in primer and/or probe binding sites inside the specific gene of interest would result in poor specificity of the primers and/or probes for the test being conducted. This study was conducted to re-evaluate the specificity of primers and probes for reported Indian sequences of 2019 novel Corona virus (SARS-CoV-2). Although more than 25 primer-probe sets are available in the literature, out of them only seven assays showed significant variations at the defined threshold [16]. Hence, specificity was evaluated only for the primers and probes described in different CDC (Center for Disease Control and Prevention) protocols published in the WHO website [18]. Though similar studies have been carried out previously, a data bias has been observed with respect to either geographical locations or fewer samples. In general, most whole genome SARS-CoV-2 sequences that are deposited in the databases are from locations that have an easy access to genome sequencing facilities or from the place of outbreak. Similarly, another study reported the sequencing and depositing the first set of whole genome sequences of SARS-CoV-2 from India, after the isolation and culturing of viral particles from two positive cases with a travel history to Wuhan, China [29]. Since then more than 1000 sequences from India have been deposited to GISAID. Given the fact that corona viruses undergo mutation at moderate rate compared to other RNA viruses, a timely reassessment of the primer/probe sequences employed in the detection of SARS-CoV-2 becomes necessary. The work design and tools employed in this work requires minimal *insilico* infrastructure and bioinformatics skills which except for sequence alignment skills. The same strategy has been successfully applied in the assessment of primer/probes involved in the detection of Influenza A and SARS-CoV-2 viral genomes.

Through stringent data selection parameters and assigning a threshold value of 0.5% as significant nucleotide variation detection parameter, none of the assays displayed exact accuracy with the SARS-CoV-2 sequences obtained from India. Major of the Results obtained were in contrast with the results obtained from studies involving genomic sequences from other countries. A study carried out by Khan and Cheung (2020) demonstrated that primer/probe sequences belonging toUS-CDC-N2, NIH-TH-N, NIID-JP-N, HKU-N, CN-CDC-ORF1ab, HKU-ORF1b and Charité-E displayed 100% accuracy with around 17000 SARS-CoV-2 whole genome sequences collected around the world [16]. Another similar study carried out by Vogels et al., 2020, indicated that the primer/probe sequences belonging to US-CDC-N1, US-CDC-N2 and Charité-E displayed 100% accuracy with around 992 whole genome sequences deposited in GISAID [30]. Previous studies on the genome wide analysis of SARS-CoV-2 strains at a global scale indicated that the single nucleotide transitions act as the major mutational type and on that basis the strains have been classified into three major clades. Further, amino acid changes were also observed in the proteomes of SARS-CoV-2 at a varying rate across six continents. Based on the rates of observed mutations per sample analysis, it has been shown that genomic sequences from India possess relative higher value at 8.40 against Germany (6.09) and Italy (5.92) [31]. Recently, a detailed phylogenetic analysis of SARS-CoV-2 genome sequences Indian origin carried out by Banu et al., 2020 indicated the presence of a distinct clade unique to India, named as clade I/A3i [32]. This cluster is characterised by a core set of 4 genetic variants and at a nucleotide substitution rate of 1.1×10^{-3} variants per site per year. The variations in the clade I/A3i were predominantly observed in the E and N genes compared to S gene (prevalent in the A2a cluster). All these observations validate the result that was obtained in this study. The SARS-CoV-2 sequences obtained from India represent a geographically unique cluster with a differential mutation patterns occurring across the genome compared to other clades across the globe. Therefore, the mutations occurring in the target regions where primer/probe sequences from the assays would bind eventually hamper the successful amplification of the genes during RT-qPCR resulting in possible false-negatives.

Accuracy of some of the assays observed in this study is in accordance with the previously published Results. For instance, CN-CDC-N assay from China targeting N gene displayed an accuracy of 54.26% which was comparable with 81.2% and 87.3% obtained by Khan & Cheung (2020) and Vogels et al. (2020) respectively [16,30]. Similarly, none of the genomic sequences displayed exact match with the primer/probe sequences belonging to Charité-ORF1b from Germany targeting ORF1ab gene. The presence of a degenerative nucleotide (S) was found to be detrimental because of the presence of T in the corresponding position. This could be due to be fact that primers were designed earlier in the outbreak based on the sequences from SARS-CoV and bat-SARS-related CoV genomes. Therefore, removing this degeneracy would help improve the detection limit of this assay. On the contrary, the presence of degenerate bases in both primers and probe sequences involved HKU-ORF1b assay has proved beneficial, turned out to be one of the most accurate assays investigated in the study. We would like to point out that a mere single nucleotide mutation in the genomes would not have serious ramification in the successful amplification of the targets. Mismatches located within 5 bp from 3' end of the primers and in the probe regions are known to have serious implications in the amplification of targets while mismatches at any other position are well tolerated [33]. In our study, we could identify a single assay US-CDC-N1 containing mismatch in the 3' end probe region. In the light we recommend using this primer in RT-qPCR assays with caution.

5. Conclusion

In this study, we have analyzed the binding accuracy of various primers and probes involved in commonly employed RT-qPCR assays for the detection of SARS-CoV-2. We could find that US-CDC-N-1from USA, CN-CDC-N from China, and Charité-ORF1b from Germany contained one or another mismatch, either in primer or probe sequences. The Results obtained in this study indicate that separate strategies, unique to specific geographical locations should be developed in the containment of pandemic. Given the number of available SARS-CoV-2 whole genome sequences in the database, novel primers with a higher accuracy could be developed. Further, this study provides an *in silico* template for re-assessing the diagnosis of COVID-19 through RT-qPCR.

Sankarganesh Jeyaraj: Conceptualization, Methodology, Formal analysis, Writing, Reviewing and Editing; Kabilan Mani.: Methodology, Formal analysis, Interpretation, Writing and Editing; Divya Sri Kathiresan: Formal analysis, Summary report; Kannan Thirumalmuthu: Formal analysis; Sudha Ramalingam: Methodology, Reviewing and Editing; Ramalingam Sankaran: Methodology, Reviewing and Editing; All authors revised and approved the final manuscript

Declarations of competing interest

None.

Acknowledgement

We acknowledge the support by PSG management and Institutional core funding of PSG Institute of Medical Sciences and Research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.mcp.2021.101748.

References

[1] World Health Organization, Corona Virus Disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update, 2021. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. (Accessed 11 March 2021).

[2] N. Sethuraman, S.S. Jeremiah, A. Ryo, Interpreting diagnostic tests for SARS-CoV-2, J. Am. Med. Assoc. 323 (2020) 2249, https://doi.org/10.1001/jama.2020.8259.

[3] C. Wang, Z. Liu, Z. Chen, X. Huang, M. Xu, T. He, Z. Zhang, The establishment of reference sequence for SARS-CoV-2 and variation analysis, J. Med. Virol. 92 (6) (2020) 667–674, https://doi.org/10.1002/jmv.25762.

[4] P. Forster, L. Forster, C. Renfrew, M. Forster, Phylogenetic network analysis of SARS-CoV-2 genomes, Proc. Natl. Acad. Sci. U. S. A 117 (2020) 9241–9243, https://doi.org/10.1073/pnas.2004999117.

[5] T. Phan, Genetic diversity and evolution of SARS-CoV-2, Infect. Genet. Evol. 81 (2020), 104260, https://doi.org/10.1016/j.meegid.2020.104260.

[6] C. Wang, Z. Liu, Z. Chen, X. Huang, M. Xu, T. He, Z. Zhang, The establishment of reference sequence for SARS-CoV-2 and variation analysis, J. Med. Virol. 92 (2020) 667–674, https://doi.org/10.1002/jmv.25762.

[7] C. Christopherson, J. Sninsky, S. Kwok, The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies, Nucleic Acids Res. 25 (1997) 654–658, https://doi.org/10.1093/nar/25.3.654.

[8] C. Liu, L. Chang, T. Jia, F. Guo, L. Zhang, H. Ji, J. Zhao, L. Wang, Real-time PCR assays for hepatitis B virus DNA quantification may require two different targets, Virol. J. 14 (2017) 94, https://doi.org/10.1186/s12985-017-0759-8.

[9] C.K. Chow, K. Qin, L.T. Lau, A. Cheung-Hoi Yu, Significance of a single-nucleotide primer mismatch in hepatitis B virus real-time PCR diagnostic assays, J. Clin. Microbiol. 49 (2011) 4418–4419. https://dio.org/10.1128/JCM.05224-11.

[10] C. Klungthong, P. Chinnawirotpisan, K. Hussem, T. Phonpakobsin, W. Manasatienkij, C. Ajariyakhajorn, K. Rungrojcharoenkit, R.V. Gibbons, R. G. Jarman, The impact of primer and probe-template mismatches on the sensitivity of pandemic influenza A/H1N1/2009 virus detection by real-time RT-PCR, J. Clin. Virol. 48 (2010) 91–95, https://doi.org/10.1016/j.jcv.2010.03.012.

[11] J.R. Yang, C.Y. Kuo, H.Y. Huang, F.T. Wu, Y.L. Huang, C.Y. Cheng, Y.T. Su, F. Y. Chang, H.S. Wu, M.T. Liu, Newly emerging mutations in the matrix genes of the human influenza A (H1N1) pdm09 and A (H3N2) viruses reduce the detection sensitivity of real-time reverse transcription-PCR, J. Clin. Microbiol. 52 (2014) 76–82, https://doi.org/10.1128/JCM.02467-13, 1.

[12] H.K. Lee, C.K. Lee, T.P. Loh, D. Chiang, E.S. Koay, J.W. Tang, Missed diagnosis of influenza B virus due to nucleoprotein sequence mutations, Euro Surveill. 16 (2011), 19943. PMID: 21871229.

[13] C. Koo, S. Kaur, Z.Y. Teh, H. Xu, A. Nasir, Y.L. Lai, E. Khan, L.C. Ng, H. C. Hapuarachchi, Genetic variability in probe binding regions explains false negative Results of a molecular assay for the detection of dengue virus, Vector Borne Zoonotic Dis. 16 (2016) 489–495, https://doi.org/10.1089/vbz.2015.1899.

[14] World Health Organization, WHO health Emergency Dashboard, 2021. https://covid19.who.int/region/searo/country/. (Accessed 10 March 2021).

[15] D.R. Rana, N. Pokhrel, Sequence mismatch in PCR probes may mask the COVID-19 detection in Nepal, Mol. Cell. Probes 53 (2020), 101599, https://doi.org/10.1016/j.mcp.2020.101599.

[16] K.A. Khan, P. Cheung, Presence of mismatches between diagnostic PCR assays and coronavirus SARS-CoV-2 genome, R. Soc. Open. Sci. 7 (2020) 6, https://doi.org/10.1098/rsos.200636, 200636.

[17] Y. Shu, J. McCauley, GISAID: global initiative on sharing all influenza data–from vision to reality, Euro Surveill. 22 (2017) 13, https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494, 30494.

[18] SARS. nCoV-2 PCR. Protocols. https://www.who.int/docs/default-source/coronaviruse/protocol-v2-1.pdf.

[19] X. Lu, L. Wang, S.K. Sakthivel, B. Whitaker, J. Murray, S. Kamili, B. Lynch, L. Malapati, S.A. Burke, J. Harcourt, A. Tamin, US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome coronavirus 2, Emerg. Infect. Dis. 26 (2020) 1654.

[20] National Institute for Viral Disease Control and Prevention China, Specific Primers and Probes for Detection 2019 Novel Coronavirus (China); 21 Jan 2020.Chinese, National Institute for Viral Disease Control and Prevention, Beijing, 2020. Available from: http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html.

[21] P. Okada, R. Buathong, S. Phuygun, T. Thanadachakul, S. Parnmen, W. Wongboot, S. Waicharoen, S. Wacharapluesadee, S. Uttayamakul, A. Vachiraphan, M. Chittaganpitch, Early transmission patterns of coronavirus disease 2019 (COVID-19) in travellers from Wuhan to Thailand, January 2020, Euro Surveill. 25 (2020), 2000097.

[22] K. Shirato, N. Nao, H. Katano, I. Takayama, S. Saito, F. Kato, H. Katoh, M. Sakata, Y. Nakatsu, Y. Mori, T. Kageyama, Development of genetic diagnostic methods for novel coronavirus 2019 (nCoV-2019) in Japan, Jpn. J. Infect. Dis. 73 (4) (2020) 304–307.

[23] D.K. Chu, Y. Pan, S.M. Cheng, K.P. Hui, P. Krishnan, Y. Liu, D.Y. Ng, C.K. Wan, P. Yang, Q. Wang, M. Peiris, Molecular diagnosis of a novel coronavirus (2019-nCoV) causing an outbreak of pneumonia, Clin. Chem. 66 (4) (2020) 549–555.

[24] V.M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D.K. Chu, T. Bleicker, S. Brünink, J. Schneider, M.L. Schmidt, D.G. Mulders, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, Euro Surveill. 25 (3) (2020), 2000045.

[25] K. Katoh, J. Rozewicki, K.D. Yamada, MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization, Briefings Bioinf. 20 (4) (2019) 1160–1166.

[26] A. Larsson, AliView: a fast and lightweight alignment viewer and editor for large datasets, Bioinformatics 30 (2014) 3276–3278, https://doi.org/10.1093/bioinformatics/btu531, 22.

[27] A. Nagy, T. Jiřinec, L. Černíková, H. Jiřincová, M. Havlíčková, Large-scale nucleotide sequence alignment and sequence variability assessment to identify the evolutionarily highly conserved regions for universal screening PCR assay design: an example of influenza A virus, in: C. Basu (Ed.), PCR Primer Design, Humana Press, New York, 2015, pp. 57–72.

[28] R.B. Wallace, J. Shaffer, R.F. Murphy, J. Bonner, T. Hirose, K. Itakura, Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch, Nucleic Acids Res. 6 (1979) 3543–3557, https://doi.org/10.1093/nar/6.11.3543.

[29] P.D. Yadav, V.A. Potdar, M.L. Choudhary, D.A. Nyayanit, M. Agarwal, S.M. Jadhav, T.D. Majumdar, A. Shete-Aich, A. Basu, P. Abraham, S.S. Cherian, Full-genome sequences of the first two SARS-CoV-2 viruses from India, Indian J. Med. Res. 151 (2020) 200–209, https://doi.org/10.4103/ijmr.IJMR_663_20.

[30] C.B.F. Vogels, A.F. Brito, A.L. Wyllie, J.R. Fauver, I.M. Ott, C.C. Kalinich, M. E. Petrone, A. Casanovas-Massana, M. Catherine Muenker, A.J. Moore, J. Klein, P. Lu, A. Lu-Culligan, X. Jiang, D.J. Kim, E. Kudo, T. Mao, M. Moriyama, J.E. Oh, A. Park, J. Silva, E. Song, T. Takahashi, M. Taura, M. Tokuyama, A. Venkataraman, O.E. Weizman, P. Wong, Y. Yang, N.R. Cheemarla, E.B. White, S. Lapidus, R. Earnest, B. Geng, P. Vijayakumar, C. Odio, J. Fournier, S. Bermejo, S. Farhadia, C.S. Dela Cruz, A. Iwasaki, A.I. Ko, M.L. Landry, E.F. Foxman, N.D. Grubaugh, Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT-qPCR primer-probe sets, Nat. Microbiol. 5 (10) (2020) 1299–1305, https://doi.org/10.1038/s41564-020-0761-6. Epub 2020 Jul 10. PMID: 32651556.

[31] D. Mercatelli, F.M. Giorgi, Geographic and genomic distribution of SARS-CoV-2 mutations, Front. Microbiol. 11 (2020) 1800, https://doi.org/10.3389/fmicb.2020.01800.

[32] S. Banu, B. Jolly, P. Mukherjee, P. Singh, S. Khan, L. Zaveri, S. Shambhavi, N. Gaur, R.K. Mishra, V. Scaria, D.T. Sowpati, A distinct phylogenetic cluster of Indian SARS-CoV-2 isolates, Open Forum Infect. Dis. (2020), https://doi.org/10.1093/ofid/ofaa434.

[33] S. Lefever, F. Pattyn, J. Hellemans, J. Vandesompele, Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays, Clin. Chem. (2013), https://doi.org/10.1373/clinchem.2013.203653.