Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1324890

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 6 - Issue 4: 477-485 / October 2023

A PARALLEL ALGORITHM FOR DESIGNING PRIMER AND PROBE FOR ACCURATE DETECTION OF SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS

Hilal ARSLAN[1]*, Rıza DURMAZ[2]

[1]*Ankara Yıldırım Beyazıt University, Faculty of Engineering and Natural Sciences, Department of Software Engineering, 06010, Ankara, Türkiye*
[2]*Ankara Yıldırım Beyazıt University, Faculty of Medicine, Department of Clinical Microbiology, 06010, Ankara, Türkiye*

Abstract: The spread of the SARS-CoV-2 in many countries has led to multiple SARS-CoV-2 variants, and this makes accurate detection of SARS-CoV-2 difficult. The reverse transcription real-time polymerase chain reaction (RT-PCR) is a widely used gold-standard method to detect SARS-CoV-2, and accurate designing of primers and probes is crucial to prevent false negative results, especially with the rise of new dangerous variants. Therefore, it is significant to determine primers and probes targeting conserved regions in the genome sequence to diagnose many variants of SARS-CoV-2. In this paper, we propose a novel and efficient method for identifying PCR primers and probe sequences by evaluating sequences belonging to SARS-CoV-2 variant of concern and variants of interest. We propose 13 primer and probe sets by analyzing 54,524 sequences in Alpha variant, 25,465 sequences in Beta variant, 53,501 sequences in Gamma variant, 46,225 sequences in Delta variant, and 43,682 sequences in Omicron variant from GISAID. Furthermore, we analyzed 1,008 sequences in Lambda variant as well as 5,844 sequences in Mu variant to extract primer and probe sets from GISAID. The proposed primer and probe sets were validated in 406,757 new SARS-CoV-2 unique genomes collected from NCBI. In silico evaluation presented that the proposed set of primers and probes are found inside about 99% of SARS-CoV-2 genome sequences. Designed primers present a higher potential to detect the main SARS-CoV-2 recent variant of concerns and the variants of interests. The superiority of the proposed method is also highlighted by comparing the state-of-the-art PCR primer and probe sets based on the number of mismatches for various types of SARS-CoV-2 genomes.

Keywords: COVID-19, SARS-CoV-2, SARS-CoV-2 variants, Primer and Probe, Real Time PCR

***Corresponding author:** Ankara Yıldırım Beyazıt University, Faculty of Engineering and Natural Sciences, Department of Software Engineering, 06010, Ankara, Türkiye
E mail: hilalarslanceng@gmail.com (H. ARSLAN)
Hilal ARSLAN https://orcid.org/0000-0002-6449-6952
Rıza DURMAZ https://orcid.org/0000-0001-6561-778X

Received: July 09, 2023
Accepted: September 26, 2023
Published: October 15, 2023

Cite as: Arslan H, Durmaz R. 2023. A parallel algorithm for designing primer and probe for accurate detection of severe acute respiratory syndrome coronavirus. BSJ Eng Sci, 6(4): 477-485.

1. Introduction

SARS-CoV-2 detected in 2019 caused a disease called COVID-19 by spreading rapidly around the world. The spread of the SARS-CoV-2 in many countries has led to multiple SARS-CoV-2 variants and accurate detection of SARS-CoV-2 variants is crucial to fight the COVID-19 pandemic. Recent dominant variants of SARS-CoV-2 are B.1.1.7 (Alpha), B.1.351 (Beta), P.1. (Gamma), B.1.617 (Delta), and B.1.1.529 (Omicron). Alpha variant (Volz et al., 2021) was first identified in the United Kingdom in the fall of 2020, and it spread ~ 50% more quickly than the original SARS-CoV-2 (Lauring and Malani, 2021). Although current treatments against the Alpha variant are effective, the Alpha variant may cause more severe COVID-19 disease (Yaniv et al., 2021). The Beta variant (Tegally et al., 2021) first detected in South Africa and the Gamma variant (Sabino et al., 2021) first detected in Brazil at the end of 2020 spread less quickly than the Alpha variant; however, current treatments against the Beta and Gamma variants are less effective. The Delta variant (Mlcochova et al., 2021) first identified in India may cause more severe disease when compared to the other variants. Furthermore, the Delta variant spreads 100% more quickly than the original SARS-CoV-2 (Lauring and Malani, 2021). It is not adequate information on whether it causes more severe COVID-19 disease, or not. The Lambda variant (Baj et al., 2021) first identified in Peru in August 2020, and it was designated as the Lambda variant by the World Health Organization (WHO) in June 2021 (Wink et al., 2021). The Mu variant (Uriu et al., 2021) first identified in Colombia was designated as a variant of interest in August, 2021 by the WHO. It is not known whether the Lambda and Mu variants are more contagious or more pathogenic than other variants. Finally, the Omicron variant (Sahoo and Samal, 2021) first identified in South Africa in November 2021, and it may spread more easily than other variants including the Delta. Xue et al. (2022) investigated factors that affect the recovery of patients and they applied machine learning techniques to estimate the duration of recovery during Omicron pandemic.

Irudayasamy et al. (2022) investigated the effect of Omicron on unvaccinated community. Their results showed that vaccination decreases mortality risk in a significant degree. Arslan (2022b) predicted mortality of patients with COVID-19 in an acceptable accuracy.

The RT-PCR (Cobb et al., 2021; Bustin et al., 2009) is a widely known method to detect SARS-CoV-2 although optimization of the RT-PCR may include a complicated process. Rapid tests for detecting SARS-CoV-2 are based on RT-PCR, and these tests require a forward primer, reverse primer and probe sequences which together are utilized to amplify the signal from the virus within a sample. Although this approach provides specific detection of the virus and does not rely on tissue culture or cell models of animals, new mutations on the primer binding sites may be occurred because of the rapid evolution of the SARS-CoV-2 (Lownick et al., 2021; Osorio and Correia-Neves, 2021). Thus, it is critical to the well design of the primer sets, and these primer sets need to be updated and evaluated regularly (Nayar et al., 2021; Jain et al., 2021).

Various types of studies are published to detect SARS-CoV-2 (Jiang et al., 2020; Zoabi, Deri-Rozov, and Shomron, 2021; Muhammad et al., 2021; Shi et al., 2021; Mohamadou et al., 2020; Arslan and Arslan, 2021; Arslan, 2021a; Arslan and Aygun, 2021; Arslan, 2021b; Togrul and Arslan, 2022). Furthermore, there exist efficient methods for detecting SARS-CoV-2 variants (Ali et al., 2021; Ogiela and Ogiela, 2021; Jamil and Rahman, 2021; Arslan, 2022a; Arslan, 2023). Park et al. (Park et al., 2020) introduced a guideline including three steps to design and optimization of primer sets. After they selected the primer sets for target genes, they performed in silico validation of the primer. Finally, they optimized the PCR conditions for specific hybridization. Alignment-based methods may be used to detect conserved regions that are used for the design of universal primers and probes (Anantharajah et al., 2021). Davi et al. (2021) determined 26 conserved regions in the SARS-CoV-2 genome as a result of the alignments of 2,341 full genome sequences. They selected nine candidate systems including primers and probes. They also analyzed their systems using 211,833 SARS-CoV-2 genome sequences. However, these methods are expensive and require a lot of time. Rincon et al. (2021) used an artificial intelligence technique to identify primers to detect SARS-CoV-2. They identified 12 unique 21-bps sequences to appear only SARS-CoV-2 sequences using Convolutional Neural Network (CNN). They validated their results by using 52,645 SARS-CoV-2 sequences. Langer et al. (2020) investigated the accuracy of artificial intelligence for predicting RT-PCR results for detecting SARS-COV-2 using the main knowledge provided by emergence departments.

In this study, we propose a parallel algorithm to identify the most conserved segments in the SARS-CoV-2 genomes. To determine these segments, we analyze a various number of SARS-CoV-2 genome sequences including the main SARS-CoV-2 variants of concern and interest. After the conserved region is determined, online Primer3Plus (2022) is employed to detect primers and probes using the conserved region. The proposed primer and probe sets are evaluated using specific types of SARS-CoV-2 and also evaluated using 406,757 genome sequences belonging to various types of SARS- CoV-2. The rest of this study is organized as follows. The proposed method is presented in Section 2. Experimental results are evaluated and compared in Section 3. Finally, Section 4 includes the conclusion.

2. Materials and Methods

In this section, we present a rapid and accurate method to determine forward primers, reverse primers and probes used in RT-PCR. The main idea behind the proposed method is to determine the most conserved region of the sequenced SARS-CoV-2 genomes using a sliding window approach in a parallel manner. The algorithm takes complete genome sequences of SARS-CoV-2 and the length of the conserved region and returns forward primers, reverse primers, and probes that are determined for the conserved region. The basic steps of the proposed algorithm are described in Algorithm 1. The length of the conserved region is fixed to 100 base pairs (bps). The first sequence in the dataset is fixed to perform a 100-bp sliding window approach in Step 1. There is a for loop between Steps 4-16 and this loop is executed in parallel by OpenMP threads. In this loop, we iterate over all the possible beginnings of the 100 bp ranges. For each range, we check whether a genome sequence in the dataset includes the 100-bp substring in parallel to find the total number of matches. At the end of for loop, we determine the most repeated 100-bp substring that has the highest match score. After identifying the most repeated 100-bp sequences, we employ online Primer3Plus to pick primers and probes from the most repeated substring in Step 17.

2.1. Primer and Probes Design

We follow the steps defined in Algorithm 1 to determine the most conserved region of the genome sequences of SARS-CoV-2. In the following, we briefly explain the genome sequences of SARS-CoV-2 used in this study. Whole human genome sequences of SARS-CoV-2 are obtained from the Global Initiative on Sharing All Influenza Data (GISAID) (Shu and McCauley, 2017). We download high-quality and complete sequences to minimize sequencing errors. Recently, five variants of concern of SARS-CoV-2, which are B.1.1.7, B.1.351, P.1, B.1.617.2, and B.1.1.529, have been reported. Furthermore, we analyze two variants of interest of SARS-CoV-2, which are C.7 and B.1.621, recently. WHO Label, Scientific Names, date of designation, and the number of sequences used in this study are presented in Table 1. Next, we run online Primer3Plus to design primer pairs and probes adopting the criteria shown in Table 2.

Table 1. Properties of the SARS-CoV-2 sequences used in this study

Variant Name	WHO Label	Scientific Name	Date of Designation	# of sequences
Variant of Concerns	Alpha	B.1.1.7	October, 2020	54,524
	Beta	B.1.351	December, 2020	25,465
	Gamma	P.1	January, 2021	53,501
	Delta	B.1.617.2	October, 202	46,225
	Omicron	B.1.1.529	November, 2021	43,682
Variants of Interest	Lambda	C.7	August, 2020	1,008
	Mu	B.1.621	January, 2021	5,844

Table 2. General conditions for designing primers and probes

Condition	Minimum	Optimum	Maximum
Primer Size	18	20	23
Primer Tm (C)	57	59	62
Primer GC %	30	50	70
Probe Size	18	20	27
Probe Tm (C)	57	60	63
Probe GC %	20	50	80

Algorithm 1. Proposed Parallel Algorithm
Require: dataSeqs: SARS-CoV-2 genome sequences and size: length of the conserved region
Ensure: Determine forward and reverse primers as well as probe
1: seq ← dataSeqs[0]
2: n ← length(seq) - size
3: maxSeq ← 0
4: **parfor** k = 1:n do
5: count ← 0
6: overlapSeq ← seq[k:k+size]
7: **parfor** genSeq in dataSeqs do
8: **if** overlapSeq in genSeq then
9: count = count + 1
10: **end if**
11: **end parfor**
12: **if** count > maxSeq then
13: maxOverlapSeq = overlapSeq
14: maxSeq = count
15: **end if**
16: **end parfor**
17: Use *Primer3Plus* to design primers and probes using maxOverlapSeq

Table 3 presents the primer and probe sequences identified in this study. Furthermore, we present primer and probe sizes, melting temperatures (Tm), percentage of G and C (GC% = G + C). The "SELF" and "ANY" present the possibility which 0the primer will bond to itself composing dimers and hairpins. Alpha-1 and Alpha-2 primers and probes are obtained by running Algorithm 1 on a set of sequences containing B.1.1.7 variant. Similarly, Beta, Gamma, Delta, and Omicron primers and probes are obtained by running Algorithm 1 on a specific set of sequences including B.1.351, P.1, B.1.617.2, and B.1.1.529 variant of concerns. Furthermore, Lambda and Mu primers and probes are obtained by running Algorithm 1 on a set of sequences containing C.7 and B.1.621 variants of interest and shown in Table 3. The suitability of the proposed primers was checked by using PCR Primer Stats program (2022) and we achieved acceptable results. Furthermore, the uniqueness of proposed primers and probes was verified using the tool UCSC In-Slico PCR (Stothard, 2020; Kent et al., 2002).

3. Results and Discussion

In this section, we evaluate primers and probes proposed in this study based on the number of sequences that the primers and probes found. The proposed primers are validated in five datasets containing SARS-CoV-2 variants of concern, which include the Alpha (54,524 sequences), Beta (25,465 sequences), Gamma (53,501 sequences), Delta (46,225 sequences), and Omicron (43,682 sequences). They are also validated in two datasets including SARS-CoV-2 variants of interest, the Lambda (1,008 sequences) and Mu (5,844 sequences). Finally, we validate the proposed primers using 406,757 whole genome sequences including a recent variant of concerns and interests of SARS-CoV-2, which is referred to as the Mixed dataset. We note that we downloaded all possible complete SARS-CoV-2 genome sequences (*.fasta format) from NCBI database (NCBISD, 2021) on March 4th, 2021.
Table 4 summarizes the total number of sequences that proposed primers are found against a specific set of sequences. The percentage is also shown. Considering B.1.1.7 (Alpha) variant of sequences, the appearance of the proposed primer sets ranged from 54,165 sequences (99.34%) to 54,447 sequences (99.86%). The best score is achieved with Mu-1 primers, and only 77 out of 54,524 sequences do not appear. When considering B.1.351 (Beta) variant of sequences, the appearance of the proposed primer sets ranged from 25,069 sequences

(98.44%) to 25,443 sequences (99.91%). The Mu-1 primer achieves the best frequency of appearance and only 22 out of 25,465 sequences do not appear. When considering P.1 (Gamma) variant of sequences, the appearance of the proposed primer sets ranged from 52,467 sequences (98.07%) to 53,456 sequences (99.92%). The Mu-1 primer achieves the best frequency of appearance and only 45 out of 53,501 sequences do not appear. For B.1.617.2 (Delta) variant of sequences, the appearance of the proposed primer sets ranged from 44,904 sequences (97.14%) to 46,141 sequences (99.82%). The Mu-1 primer achieves the best frequency of appearance and only 84 out of 46,225 sequences do not appear. For B.1.1.529 (Omicron) variant of sequences, the appearance of the proposed primer sets ranged from 43,069 sequences (98.6%) to 43,670 sequences (99.97%). The Mu-1 and Beta-4 primers achieve the best frequency of appearance and only 12 out of 43,682 sequences do not appear. When we analyze the frequency results on two SARS-CoV-2 variants of interest, for C.8 (Lambda) variant of sequences, the Beta-2, Beta-4, and Omicron-1 primers achieve 100% frequency of appearance. For B.1.1.529 (Mu) variant of sequences, the Mu-1 primer achieves 100% frequency of appearance.

Table 3. The properties of the primer and probe sequences identified in this study. F forward primer; R reverse primer; P probe

Primer Name	Sequence	Size	Tm	GC(%)	ANY	SELF	Target
Alpha-1-F	GGCCGGCTGTTTTGTAGAT	19	59.3	52	6	0	ORF1ab
Alpha-1-P	TGATTGAACGGTTCGTGTCT	20	59.1	45	7	1	ORF1ab
Alpha-1-R	GGATGTTTAGTAAGTGGGTAAGCA	24	58.7	41.7	5	3	ORF1ab
Alpha-2-F	CAGTTTATGATCCTTTGCAACC	22	58.6	40.9	6	2	S
Alpha-2-P	TGAATTAGACTCATTCAAGGAGGA	24	59.3	37.5	11	3	S
Alpha-2-R	ATGTCACCTAAATCAACATCTGG	23	58	39.1	4	2	S
Beta-1-F	CATGCGAAATGCTGGTATTG	20	60.1	45	4	2	ORF1ab
Beta-1-P	GGTAACTGGTATGATTTCGGTGA	23	60.1	43.5	3	3	ORF1ab
Beta-1-R	CTACCTGGCGTGGTTTGTATG	21	60.4	52.4	4	0	ORF1ab
Beta-2-F	ACAATTCTGTGATGCCATGC	20	59.5	45	5	3	ORF1ab
Beta-2-P	GAAATGCTGGTATTGTTGGTG	21	58	42.9	4	0	ORF1ab
Beta-2-R	TACCTGGCGTGGTTTGTATG	20	59.5	50	4	0	ORF1ab
Beta-3-F	GTAACAGCTTTAAGGGCCAAT	21	57.5	42.9	6	2	ORF1ab
Beta-3-P	CCTGTTGCACTACGACAGATG	21	59.4	52.4	7	2	ORF1ab
Beta-3-R	TTTGTGTAGTACCGGCAGCA	20	60.3	50	5	2	ORF1ab
Beta-4-F	GGCTGTTGCTAATGGTGATT	20	57.7	45	3	2	ORF1ab
Beta-4-P	TCTGAATTTGACCGTGATGC	20	59.7	45	4	2	ORF1ab
Beta-4-R	TTCCAACTTACGTTGCATGG	20	59.6	45	8	2	ORF1ab
Gamma-1-F	GTGCAGGTGCTGCATTACA	19	59.4	52.6	7	3	S
Gamma-1-P	CATTTGCTATGCAAATGGCTT	21	60.1	38.1	12	3	S
Gamma-1-R	GGTTGGCAATCAATTTTTGG	20	60.2	40	5	2	S
Gamma-2-F	CTGGTTGGACCTTTGGTG	18	57.4	55.6	3	1	S
Gamma-2-P	CATTTGCTATGCAAATGGCTT	21	60.1	38.1	12	3	S
Gamma-2-R	TAGAGAACATTCTGTGTAACTCCAAT	26	57.6	34.6	6	3	S
Delta-1-F	AGCTCCAATTTTGGTGCAAT	20	59.6	40	8	2	S
Delta-1-P	AAGTTGAGGCTGAAGTGCAAA	21	60	42.9	5	0	S
Delta-1-R	TGCCTGTGATCAACCTATCAA	21	59.1	42.9	6	1	S
Delta-2-F	AGCTCCAATTTTGGTGCAAT	20	59.6	40	8	2	S
Delta-2-P	TGAGGCTGAAGTGCAAATTG	20	60	45	5	5	S
Delta-2-R	GAAGTCTGCCTGTGATCAACC	20	59.7	52.4	6	2	S
Omicron-1-F	ACAATTCTGTGATGCCATGC	20	59.5	45	5	3	ORF1ab
Omicron-1-P	GAAATGCTGGTATTGTTGGTG	21	58	42.9	4	0	ORF1ab
Omicron-1-R	TGGCGTGGTTTGTATGAAAT	20	58.9	40	3	2	ORF1ab
Lambda-1-F	CCATCATATGCAGCTTTTGC	20	59.3	45	6	3	ORF1ab
Lambda-1-P	AGCAGGCTGTTGCTAATGGT	20	59.9	50	5	0	ORF1ab
Lambda-1-R	GCCACATTCAAAGACTTCTTCA	22	59.4	40.9	6	2	ORF1ab
Mu-1-F	AGGACCTCATGAATTTTGCTC	21	58.3	42.9	6	0	ORF1ab
Mu-1-P	TGTGTACCTTCCTTACCCAGATC	23	59.4	47.8	4	4	ORF1ab
Mu-1-R	GGCCCCTAGGATTCTTGATG	20	60.8	55	6	2	ORF1ab

Table 4. The percentage and total number of sequences that proposed primers are found in the corresponding dataset.

Primer Name	B.1.1.7 (Alpha)	B.1.351 (Beta)	P.1 Gamma)	B.1.617.2 (Delta)	C.8 (Lambda)	B.1.621 (Mu)	B.1.1.529 (Omicron)	Mixed Dataset
Alpha-1	54,257	25,352	52,467	46,026	1,004	5,675	43,634	402,892
	(99.51%)	(99.56%)	(98.07%)	(99.57%)	(99.6%)	(97.11%)	(99.89%)	(99.05%)
Alpha-2	54,380	25,069	53,210	45,440	1,006	5,740	43,275	403,665
	(99.74%)	(98.44%)	(99.46%)	(98.3%)	(99.8%)	(98.22%)	(99.07%)	(99.24%)
Beta-1	54,293	25,369	53,437	45,572	1,006	5,838	43,585	404,793
	(99.58%)	(99.62%)	(99.88%)	(98.59%)	(99.8%)	(99.9%)	(99.78%)	(99.52%)
Beta-2	54,165	25,422	53,359	44,904	1,008	5,841	43,650	404,698
	(99.34%)	(99.83%)	(99.73%)	(97.14%)	(100%)	(99.95%)	(99.93%)	(99.49%)
Beta-3	54,222	25,403	53,320	46,036	979	5,831	43,644	404,808
	(99.45%)	(99.76%)	(99.66%)	(99.59%)	(97.12%)	(99.78%)	(99.91%)	(99.52%)
Beta-4	54,301	25,432	53,383	46,127	1,008	5,841	43,670	405,247
	(99.59%)	(99.87%)	(99.78%)	(99.79%)	(100%)	(99.95%)	(99.97%)	(99.63%)
Gamma-1	54,166	25,280	53,422	45,935	1,007	5,805	43,093	402,564
	(99.34%)	(99.27%)	(99.85%)	(99.37%)	(99.9%)	(99.33%)	(98.65%)	(98.97%)
Gamma-2	54,361	25,362	53,443	45,998	1,006	5,813	43,069	402,615
	(99.7%)	(99.6%)	(99.89%)	(99.51%)	(99.8%)	(99.47%)	(98.6%)	(98.98%)
Delta-1	54,292	25,379	53,434	46,098	1,007	5,821	43,630	403,766
	(99.57%)	(99.66%)	(99.87%)	(99.73%)	(99.9%)	(99.61%)	(99.88%)	(99.26%)
Delta-2	54,292	25,379	53,434	46,098	1,007	5,821	43,630	403,766
	(99.57%)	(99.66%)	(99.87%)	(99.73%)	(99.9%)	(99.61%)	(99.88%)	(99.26%)
Omicron-1	54,165	25,422	53,359	44,904	1,008	5,841	43,650	404,698
	(99.34%)	(99.83%)	(99.73%)	(97.14%)	(100%)	(99.95%)	(99.93%)	(99.49%)
Lambda-1	54,427	25,362	53,428	45,943	1,007	5,833	43,668	404,798
	(99.82%)	(99.6%)	(99.86%)	(99.39%)	(99.9%)	(99.81%)	(99.97%)	(99.52%)
Mu-1	54,447	25,443	53,456	46,141	1,007	5,844	43,670	405,415
	(99.86%)	(99.91%)	(99.92%)	(99.82%)	(99.9%)	(100%)	(99.97%)	(99.67%)

Finally, we analyze the frequency of appearance on the Mixed dataset. The proposed primer sets ranged from 402,564 sequences (98.97%) to 405,415 sequences (99.67%). The Mu-1 primer achieves the best frequency of appearance and only 1342 out of 406,757 sequences do not appear. In silico analysis presents that the sets of primers and probes proposed in this study potentially anneal to a highly conserved region of the SARS-CoV-2.

Designed primers are also analyzed using Oligo 7 software (Rychlik, 2007) based on duplex formation and hairpin formation. Moreover, Oligo 7 program analyzes primers by calculating an efficiency score that is a possibility that a given oligonucleotide is going to prime at a given site on the sequence currently analyzed. It is noted that when efficiency scores of the primers are between 450 and 500, excellent results are obtained for multiplex PCR, and the priming is more likely when this score is over 220 (threshold) (Rychlik, 2007). Figure 1 presents efficiency values of forward primers, reverse primers, and probes designed in this study. When the proposed primers are evaluated based on primer efficiency, as you can see in Figure 1, the efficiencies of the primers and probes are above the threshold and the designed primers either fall in 450-500 primer efficiency range or close to this range. This points out that the proposed primers may be achieved excellent for multiplex PCR.

3.1. Comparison with Existing Primers

In this section, we present results related to the primers reported by Davi et al. (2021). The properties of the primers and probes shown by Davi et. al. (2021) are given in Table 5. Furthermore, the percentages of appearance for each primer designed by Davi et al. (2021) against the different datasets including the main SARS-CoV-2 concerns and two types of SARS-CoV-2 variant of interests are presented in Table 6. The results revealed that the UFRN-5, URFN-6, and URFN-7 primers cannot accurately detect the sequences belonging to the Mu variant. Furthermore, the URFN-8 and URFN-9 primers have a lower percentage of appearance on the Mixed dataset and appear in only 60.12% of the sequences including recent variants of SARS-CoV-2.

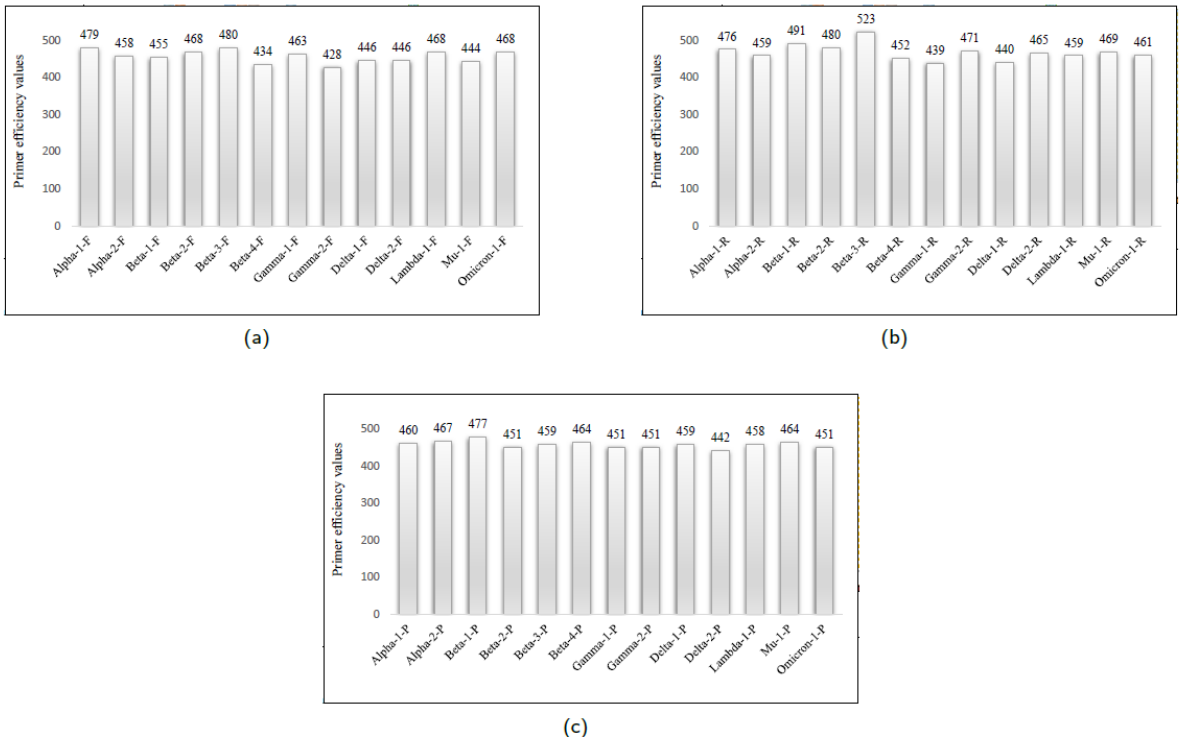

Figure 1. Efficiency scores of (a) forward primers (b) reverse primers (c) probes designed in this study.

Table 5. The information regarding the primers and probes shown by Davi et al. (2021)

Primer Name	Sequence	Length	Tm	GC(%)	Target	Size
UFRN-1-F	GGGCATACACTCGCTATGTC	20	58.22	55	ORF1a	101
UFRN-1-R	GCATGAAGCTTTACCAGCAC	20	57.73	50	ORF1a	101
UFRN-1-P	TCTGTGGCCCTGATGGCTACCCT	23	67.22	60.87	ORF1a	101
UFRN-2-F	GGCTACTAACAATGCCATGC	20	57.22	50	ORF1a	137
UFRN-2-R	TAACATTTGGGCCGACAACA	20	58.02	45	ORF1a	137
UFRN-2-P	GGGTGGTAGTTGTGTTTTAAGCGG	24	62.33	50	ORF1a	137
UFRN-3-F	TTCATGTTGTCGGCCCAAAT	20	58.37	45	ORF1a	98
UFRN-3-R	TGGTGCAAGTAGAACTTCGT	20	57.1	45	ORF1a	98
UFRN-3-P	GAAGACATTCAACTTCTTAAGAGTGC	26	58.71	38.46	ORF1a	98
UFRN-4-F	TGGTGCTAGGAGAGTGTGG	19	58.33	87.89	ORF1a	95
UFRN-4-R	CCCACATGGAAATGGCTTGAT	21	58.89	47.62	ORF1a	95
UFRN-4-P	CTTATGAATGTCTTGACACTCGTTTATA	28	58.01	32.14	ORF1a	95
UFRN-5-F	AGGGCACACTAGAACCAGAA	20	58.27	50	ORF1b	105
UFRN-5-R	CAATTTCAGCAGGACAACGC	20	58.31	50	ORF1b	105
UFRN-5-P	GGTCCAGACATGTTCCTCGGAACT	24	64.18	54.17	ORF1b	105
UFRN-6-F	TCTTCACGACATTGGTAACCC	21	57.95	47.62	ORF1b	90
UFRN-6-R	TCACTACAAGGCTGTGCATC	20	57.9	50	ORF1b	90
UFRN-6-P	TACCTCAAGCTGATGTAGAATGGAAG	26	60.41	42.31	ORF1b	90
UFRN-7-F	CTTCACGACATTGGTAACCCT	21	57.95	47.62	ORF1b	90
UFRN-7-R	GTCACTACAAGGCTGTGCAT	20	58.19	50	ORF1b	90
UFRN-7-P	GTGTACCTCAAGCTGATGTAGAATGG	26	61.4	46.15	ORF1b	90
UFRN-8-F	GGCACAGGTGTTCTTACTGA	20	57.46	50	S	107
UFRN-8-R	TCAAGTGTCTGTGGATCACG	20	57.56	50	S	107
UFRN-8-P	CCAACAATTTGGCAGAGACATTGC	24	61.62	45.83	S	107
UFRN-9-F	AGGCACAGGTGTTCTTACTG	20	57.45	50	S	93
UFRN-9-R	TCACGGACAGCATCAGTAGT	20	58.45	50	S	93
UFRN-9-P	TCCAACAATTTGGCAGAGACATTGC	25	62.75	44	S	93

Table 6. The percentage of appearance for each primer designed by Davi et al. (2021) against the different datasets including the main SARS-CoV-2 concerns and two types of SARS-CoV-2 variant of interests

Primer Name	B.1.1.7 (Alpha)	B.1.351 (Beta)	P.1 (Gamma)	B.1.617.2 (Delta)	C.8 (Lambda)	B.1.621 (Mu)	B.1.1.529 (Omicron)	Mixed Dataset
UFRN-1	98%	98.83%	99.38%	99.05%	99.8%	98.36%	99.86%	98.15%
UFRN-2	98.89%	99.37%	96.95%	99.07%	98.81%	99.16%	99.63%	98.35%
UFRN-3	98.51%	98.7%	88.31%	98.96%	99.7%	99.49%	99.42%	98.28%
UFRN-4	99.31%	98.96%	99.5%	99.1%	100%	99.81%	99.96%	97.07%
UFRN-5	99.34%	99.79%	99.87%	99.37%	99.8%	0.29%	99.95%	98.33%
UFRN-6	99.56%	99.72%	99.71%	99.36%	99.5%	5.18%	99.97%	99.18%
UFRN-7	99.58%	99.71%	99.71%	99.34%	99.40%	5.18%	99.97%	99.18%
UFRN-8	99.78%	99.7%	99.8%	99.48%	99.9%	99.9%	99.89%	60.12%
UFRN-9	99.79%	99.7%	99.8%	99.69%	99.31%	99.9%	99.89%	60.12%

Next, we choose the best-performing primer sets developed by Davi et al. (2021) and WHO to compare against the best primer set obtained by the proposed method. They are chosen based on the highest frequency of appearance on the mixed dataset. Table 7 presents the properties of the primers and probes. Experimental results present that the proposed primer and probe have a higher frequency of appearance on the main SARS-CoV-2 variants of concern and interest. On the other hand, the frequency of appearance of the primer proposed by WHO is lower in the sequences belonging Lambda variant. Furthermore, the primer proposed by Davi et al. (2021) cannot accurately detect the Mu variant.

Table 7. The properties of the best-performing primers and probes developed by Davi et al. (2021), WHO, and the proposed method

Study	Primer Name	Sequence	Length	Tm	GC(%)
Davi et al. (2021)	UFRN-6-F	TCTTCACGACATTGGTAACCC	21	57.95	47.62
	UFRN-6-P	TACCTCAAGCTGATGTAGAATGGAAG	26	60.41	42.31
	UFRN-6-R	TCACTACAAGGCTGTGCATC	20	57.9	50
WHO	N_Sarbeco_F1	CACATTGGCACCCGCAATC	19	60.15	57.89
	N_Sarbeco_P1	ACTTCCTCAAGGAACAACATTGCCA	25	63.15	44
	N_Sarbeco_R1	GAGGAACGAGAAGAGGCTTG	20	58	55
Proposed Study	Mu-1-F	AGGACCTCATGAATTTTGCTC	21	58.3	42.9
	Mu-1-P	TGTGTACCTTCCTTACCCAGATC	23	59.4	47.8
	Mu-1-R	GGCCCCTAGGATTCTTGATG	20	60.8	55

4. Conclusion

In this study, we propose an efficient parallel method to identify primers and probes for real-time PCR. Compared to the alignment-based method, the proposed method is effectively detected to the conserved region. We analyze the performance of the proposed primers and probes based on the number of matches of the PCR primers for genome sequences of SARS-CoV-2. Experimental results present that the proposed primers and probes have about 99% matches for the genome sequences including the recent variants of SARS-CoV-2. In future, other pandemics may occur as a result of the increasing population and growing interaction between people. We believe that the proposed method can be applied to develop more accurate primers and probes for identifying any type of virus and contribute treatment for the rapidly propagating virus as well as help limit the spread of the virus. In future studies, we will also provide laboratory results of the primers and probes. Furthermore, we will extend the proposed technique for other types of virus.

Author Contributions

The percentage of the author(s) contributions is presented below. All authors reviewed and approved the final version of the manuscript.

	H.A.	R.D.
C	50	50
D	100	
S		100
DCP	100	
DAI	20	80
L	50	50
W	80	20
CR	20	80
SR	80	20

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because of there was no study on animals or humans. The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to.

References

Ali S, Tamkanat-E-Ali Khan MA, Khan I, Patterson M. 2021. Effective and scalable clustering of sars-cov-2 sequences. Proceedings of the 5th International Conference on Big Data Research, September 23-25, Qingdao, China, pp: 42-49.

Anantharajah A, Helaers R, Defour J.-P, Olive N, Kabera F, Croonen L, Kabamba-Mukadi B. 2021. How to choose the right real-time RT-PCR primer sets for the SARS-CoV-2 genome detection. J Virol Methods, 295: 114197.

Arslan H. 2021a. COVID-19 prediction based on genome similarity of human SARS-CoV-2 and bat SARS-CoV-like coronavirus. Comput Indust Engin, 161: 107666.

Arslan H. 2021b. Machine learning methods for Covid-19 prediction using human genomic data. Proceed 74 (1): 20.

Arslan H, Arslan H. 2021. A new covid-19 detection method from human genome sequences using cpg island features and knn classifier. Engin Sci Technol Inter J, 24(4): 839-847.

Arslan H, Aygun B. 2021. Performance analysis of machine learning algorithms in detection of covid-19 from common symptoms. Proceeding of 29th Signal Processing and Communications Applications Conference, June 9-11, Istanbul, Türkiye, pp: 1-4.

Arslan H. 2022a. Classification of SARS-CoV-2 Variants in Turkey. J Turkish Operat Manage, 6(1): 1092-1101 .

Arslan H. 2022b. Bagging and boosting for predicting mortality of patients with COVID-19. Dicle Univ J Engin, 13(2): 221-226.

Arslan H. 2023. A k-mer based metaheuristic approach for detecting COVID-19 variants. Dicle Univ J Engin, 14(1): 17-26.

Baj A, Novazzi F, Ferrante FD, Genoni A, Cassani G, Prestia M, Maggi F. 2021. Introduction of SARS-COV-2 c.37 (WHO VOI lambda) from Peru to Italy. J Medical Virol, 93(12): 6460–6461.

Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Wittwer CT. 2009. The MIQE guidelines: Minimum information for publication of quantitative realtime PCR experiments. Clin Chem, 55(4): 611–622.

Cobb BR, Vaks JE, Do T, Vilchez RA. 2011. Evolution in the sensitivity of quantitative HIV-1 viral load tests. J Clin Virol, 52 S77–S82.

Davi MJP, Jeronimo SMB, Lima JPMS, Lanza DCF. 2021. Design and in silico validation of polymerase chain reaction primers to detect severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Sci Rep, 11(1): 12565.

Irudayasamy A, Ganesh D, Natesh M. 2022. Big data analytics on the impact of OMICRON and its influence on unvaccinated community through advanced machine learning concepts. Int J Syst Assur Eng Manag, 2022: 1-10.

Jain A, Rophina M, Mahajan S, Krishnan BB, Sharma M, Mandal S, Scaria V. 2021. Analysis of the potential impact of genomic variants in global sars-cov-2 genomes on molecular diagnostic assays. Inter J Infect Diseas, 102: 460-462.

Jamil S, Rahman M. 2021. A dual-stage vocabulary of features (vof)-based technique for covid-19 variants' classification. Applied Sci, 11 (24): 11902.

Jiang X, Coffee M, Bari A, Wang J, Jiang X, Huang J, Huang Y. 2020. Towards an Artificial Intelligence Framework for Data-Driven Prediction of Coronavirus Clinical Severity. Comput Mater Contin, 62(3): 537–551.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. Genome Res, 12(6): 996–1006.

Langer T, Favarato M, Giudici R, Bassi G, Garberi R, Villa F, Fumagalli R. 2020. Development of machine learning models to predict RT-PCR results for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in patients with influenza-like symptoms using only basic clinical data. Scandinavian J Trauma Resuscit Emerg Med, 28 (1): 113.

Lauring AS, Malani PN. 2021. Variants of SARS-CoV-2. JAMA, 326(9): 880-880.

Lopez-Rincon A, Tonda A, Mendoza-Maldonado L, Mulders DGJC, Molenkamp R, Perez-Romero CA, Kraneveld AD. 2021. Classification and specific primer design for accurate detection of SARS-CoV-2 using deep learning. Sci Reports 11(1): 947.

Lownik JC, Farrar JS, Way GW, McKay A, Roychoudhury P, Greninger AL, Martin RK. 2021. Fast sars-cov-2 variant detection using snapback primer high-resolution melting. Diagnostics, 11(10): 1788.

Mlcochova P, Kemp SA, Dhar MS. 2021. SARS-CoV-2 b.1.617.2 delta variant replication and immune evasion. Nature, 599(7883): 114–119.

Mohamadou Y, Halidou A, Kapen PT. 2020 November. A review of mathematical modeling artificial intelligence and datasets used in the study prediction and management of COVID- 19. Applied Intell, 50(11): 3913–3925.

Muhammad LJ, Algehyne EA, Usman SS, Ahmad A, Chakraborty C, Mohammed IA. 2021 February. Supervised machine learning models for prediction of Covid-19 infection using epidemiology dataset. SN Computer Sci, 2(1): 11.

NCBISD. 2021. National center for biotechnology information search database. URL: https://www.ncbi.nlm.nih.gov/ sars-cov-2/. (accessed date: December 04, 2021).

Nayar G, Seabolt EE, Kunitomi M, Agarwal A, Beck KL, Mukherjee V, Kaufman JH. 2021. Analysis and forecasting of global real time RT-PCR primers and probes for SARS-CoV-2. Sci Reports, 11 (1): 8988.

Ogiela MR, Ogiela U. 2021. Linguistic methods in healthcare application and COVID-19 variants classification. Neural Comput Applicat, 35: 13935–13940

Os´orio NS, Correia-Neves M. 2021. Implication of SARS-CoV-2 evolution in the sensitivity of RT-qPCR diagnostic assays. The Lancet Infect Diseas, 21(2): 166–167.

Park M, Won J, Choi BY, Lee CJ. 2020. Optimiza- 5 tion of primer sets and detection protocols for SARS-CoV-2 of coronavirus disease 2019 (COVID-19) using PCR and real-time PCR. Experiment Molec Med, 52(6): 963–977.

PCR Primer Stats. 2022. URL: https://www.bioinformatics.org/sms2/pcrprimerstats.html. (accessed date October 01, 2022).

Primer3plus. 2022. URL: https://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi. (accessed date October 01, 2022).

Rychlik W. 2007. OLIGO 7 primer analysis software. Humana Pressi In PCR primer design, New York, USA, pp: 35–59.

Sabino EC, Buss LF, Carvalho MPS. 2021. Resurgence of COVID-19 in manaus brazil despite high seroprevalence. The Lancet, 397(10273): 452–455.

Sahoo JP, Samal KC. 2021. World on alert: WHO designated south

african new COVID strain (Omicron/B.1.1.529) as a variant of concern. Biotica Res Today, 3(11): 1086–1088.

Shi F, Wang J, Shi J, Wu Z, Wang Q, Tang Z, Shen D. 2021. Review of artificial intelligence techniques in imaging data acquisition segmentation and diagnosis for covid-19. IEEE Rev Biomedic Eng, 14 4-15.

Shu Y, McCauley J. 2017. GISAID: Global initiative on sharing all influenza data - from vision to reality. Eurosurveillance, 22(13): 30494.

Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques, 28(6): 1102–1104.

Tegally H, Wilkinson E, Giovanetti M. 2021 March. Detection of a SARS-CoV-2 variant of concern in South Africa. Nature, 592(7854): 438–443.

Togrul M, Arslan H. 2022. Detection of SARS-CoV-2 main variants of concerns using deep learning. Proceeding of Innovations in Intelligent Systems and Applications Conference, September 07-09, Antalya, Türkiye, pp: 32.

UCSC In-silico PCR. 2022. URL: https://genome.ucsc.edu/cgi-bin/hgPcr. (accessed date October 01, 2022).

Uriu K, Kimura I, Shirakawa K, Takaori-Kondo A, aki Nakada T, Kaneda A, Sato K. 2021. Neutralization of the SARS-CoV-2 mu variant by convalescent and vaccine serum. New England J Med, 385(25): 2397–2399.

Volz E, Mishra S, Chand M, Barrett J. 2021. Assessing transmissibility of SARS-CoV-2 lineage b.1.1.7 in England. Nature, 593(7858): 266–269.

Wink PL, Volpato FCZ, Monteiro FL, Willig JB, Zavascki AP, Barth AL, Martins AF. 2021. First identification of SARS-CoV-2 lambda (c.37) variant in southern Brazil. Infect Control Hospital Epidemiol, 2021: 1–2.

Xu Y, Ye W, Song Q, Shen L, Liu Y, Guo Y, Liu G, Wu H, Wang X, Sun X, Bai L, Luo C, Liao T, Chen H, Song C, Huang C, Wu Y, Xu Z. 2022. Using machine learning models to predict the duration of the recovery of COVID-19 patients hospitalized in Fangcang shelter hospital during the Omicron BA. 2.2 pandemic. In Frontiers Med, 9: 1001801.

Yaniv K, Ozer E, Shagan M, Lakkakula S, Plotkin N, Bhandarkar NS, Kushmaro A. 2021. Direct RT-QPCR assay for sars-cov-2 variants of concern (alpha b.1.1.7 and beta b.1.351) detection and quantification in wastewater. Environ Res, 201: 111653.

Zoabi Y, Deri-Rozov S, Shomron N. 2021. Machine learning-based prediction of COVID-19 diagnosis based on symptoms. NPJ Digital Medi, 4(1): 3.