



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.



CoVrimer: A tool for aligning SARS-CoV-2 primer sequences and selection of conserved/degenerate primers

Merve Vural-Ozdeniz^a, Aslinur Akturk^{b,1}, Mert Demirdizen^{b,1}, Ronaldo Leka^b, Rana Acar^b, Ozlen Konu^{a,b,*}

^a Interdisciplinary Neuroscience Program, Bilkent University, Ankara, Turkey

^b Department of Molecular Biology and Genetics, Bilkent University, Ankara, Turkey

ARTICLE INFO

Keywords:

SARS-CoV-2
COVID-19
Primer design
Conserved primers
Mutation
Genome alignment

ABSTRACT

As mutations in SARS-CoV-2 virus accumulate rapidly, novel primers that amplify this virus sensitively and specifically are in demand. We have developed a webserver named CoVrimer by which users can search for and align existing or newly designed conserved/degenerate primer pair sequences against the viral genome and assess the mutation load of both primers and amplicons. CoVrimer uses mutation data obtained from an online platform established by NGDC-CNCB (12 May 2021) to identify genomic regions, either conserved or with low levels of mutations, from which potential primer pairs are designed and provided to the user for filtering based on generalized and SARS-CoV-2 specific parameters. Alignments of primers and probes can be visualized with respect to the reference genome, indicating variant details and the level of conservation. Consequently, CoVrimer is likely to help researchers with the challenges posed by viral evolution and is freely available at <http://konulabapps.bilkent.edu.tr:3838/CoVrimer/>.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged in Wuhan, China, in late 2019 and gave rise to a global health challenge by becoming the causative agent of the novel coronavirus disease 2019 (COVID-19) [1]. Following the emergence of this outbreak, many diagnostic and treatment methods for COVID-19 have been made available by governments and/or companies and organizations. Most of these use quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR) as the standard methodology, employing specific primers designed to amplify parts of SARS-CoV-2 viral genome [2]. Although a few web-based tools have recently become available to identify these primers [3–6], an online tool that provides means to select novel conserved or degenerate primers derived from the alignment of available SARS-CoV-2 sequences is yet to be developed.

One commonly used approach for specific and sensitive detection of SARS-CoV-2 by qRT-PCR is through TaqMan-based assays [7]. Relatively less costly detection methods are also available, including SYBR Green and conventional RT-PCR [8]. Other detection strategies include reverse transcription loop-mediated isothermal amplification (RT-

LAMP) [9–11], reverse transcription recombinase-aided amplification (RT-RAA) [12,13], and recombinase polymerase amplification-lateral flow reading (RPA-LF) [14] assays as well as those based on sequencing [15], droplet digital PCR (ddPCR) [16], and RT-LAMP-CRISPR platform [17]. Despite the abundance of diagnostic methods, the quantitative nature of qRT-PCR may be necessary when especially the information on viral load is required [11]. Accordingly, the availability of conserved and/or degenerate qRT-PCR primers specific to the rapidly evolving SARS-CoV-2 has become a necessity.

The single-stranded RNA genome of SARS-CoV-2 contains 14 open-reading frames (ORFs) encoding 27 proteins [18,19]. With respect to RT-PCR diagnostic assays, the most commonly targeted genes of the virus include ORF 1a or 1b, RNA-dependent RNA polymerase (RdRp)/helicase (Hel), spike (S), envelope (E), and nucleocapsid (N) [20]. The key step in the development of functional RT-PCR assays is the design of specific and efficient primer pairs [21]. Important aspects for increasing primer efficiency include the availability of a full-length reference sequence to choose targets from, avoiding regions subject to positive selection or variation, integration of degenerate nucleotides into the primers to include possible variants of the viral genome, and keeping the

* Corresponding author at: Interdisciplinary Neuroscience Program, Bilkent University, Ankara, Turkey.

E-mail address: konu@fen.bilkent.edu.tr (O. Konu).

¹ Equal contribution.

<https://doi.org/10.1016/j.ygeno.2021.07.020>

Received 20 February 2021; Received in revised form 30 May 2021; Accepted 17 July 2021

Available online 20 July 2021

0888-7543/© 2021 Elsevier Inc. All rights reserved.

primer length between 16 and 28 nucleotides and the melting temperature between 50 and 62 °C [22]. In addition, forward and reverse primers should have similar physiochemical properties; GC content of the 3' end should be optimal; and primer dimers should be avoided by limiting complementarity between primers [21]. To ensure accurate diagnosis of COVID-19, selected primers should ideally be specific only to SARS-CoV-2 and not cross-react with other coronaviruses [23]. In this context, a study using Convolution Neural Network (CNN) based deep learning has obtained 1221 SARS-CoV-2 specific sequences from 533 samples belonging to different strains of the coronavirus family to select a primer set for efficient amplification of viral RNA only from SARS-CoV-2 positive patients [24]. Furthermore, the Centers for Disease Control and Prevention (CDC) has developed the now most widely used amplification test targeting three independent regions of the nucleocapsid, named N1, N2 and N3, and the human ribonuclease P (RNase P) genes for the internal control. After multiple validations, N3 has been excluded due to its potential cross-reactivity with other betacoronaviruses [25]. The importance of a proper assay design strategy and primer specificity is further highlighted in the studies related to assaying COVID-19-RdRp/Hel and COVID-19-nsp2 genes, the primers employed in which are highly specific for SARS-CoV-2 and result in high specificity of the assays [20,26].

One of the challenges faced in COVID-19 research has been the proportion of false-negative test results from RT-PCR assays which may occur in up to 54% of COVID-19 patient samples [27] and may partly result from the presence of mutations in primer/probe binding regions. There are primarily two sources for the rapid emergence of genetic variations in SARS-CoV-2 genome: First, the mutation rate in coronaviruses is higher compared to other RNA-viruses [20], hence SARS-CoV-2 can mutate relatively fast (1.12×10^{-3} mutations per site-year) [28] and lead to surfacing of different strains. Secondly, recombination takes place among coronaviruses, promoting the rate of evolution [20].

As a matter of fact, the characteristically fast evolution rate of SARS-CoV-2 viral genome has already resulted in appearance of a vast number of lineages (i.e., B.1.17, B.1.351 etc.). At the time of this study, combined efforts in epidemiology and phylogenetics have enabled researchers to discover variants of the virus with distinguishing features such as enhanced transmissibility or increased disease severity [29,30]. With regard to these features, B.1.1.7, P.1, B.1.351, and B.1.427/B.1.429 have been classified as variants of concern (VOC) by CDC [31]. It is then critical to consider the rapid emergence of SARS-CoV-2 genetic variants in primer design as VOCs that go undetected, due to shortfalls in diagnostic assays, are highly likely to pose threats on controlling the spread and mortality of the disease. Correspondingly, a few variants in the viral genome have been found to reach 99% frequency in all populations [32–34]. Taking into account that a few mismatches between oligonucleotides and their templates or a single mismatch in the last 5 nucleotides of 3' ends of primers can result in decreased efficiency of qRT-PCR diagnostic tests [35,36], it is crucial to design variant specific/inclusive primers. Inclusion of primer degeneracy in frequently mutated sites also will help track SARS-CoV-2 variations and prevent false negative qRT-PCR results.

On the other hand most viral genomes carry rare mutations, for example, Farkas et al. have reported that more than half of SARS-CoV-2 sequences ($N = 50$) in their study exhibited single founder mutations of which 63% or more were missense [37]. Although rare, when these mutations appear in regions where primers bind, sensitivity of the assay is likely to decrease due to mis-priming. GISAID (Global Initiative on Sharing All Influenza Data) [38] has stated that, in an evaluation of more than 660,000 SARS-CoV-2 genomes, binding regions of eight commonly used RT-PCR assays turned out to coincide with at least one mutation in approximately less than 5% of the genomes [39]. Similarly, an in silico analysis of 27 qRT-PCR primer/probe sets, based on an alignment of more than 17,000 full-length SARS-CoV-2 genomes to the full-length genome of Wuhan-Hu-1 strain (NC_045512.2), has found occurrences of mispriming in seven assays [40]. Although majority of these

mismatches may not be in critical positions, this study shows that it is of necessity to set the strategy right in the process of primer/probe design. Moreover, in a bioinformatics pipeline, which was published in October 2020, 15 sets of primer/probe oligonucleotides collected from published studies were analyzed against 15,001 SARS-CoV-2 genome sequences to find that more than 98% of the SARS-CoV-2 genomes had no mismatches for 12 out of 15 primer/probe pairs and only three sets contained mismatches in the binding regions of at least 99% of genomes [41].

It is thus important to regularly evaluate existing assays in the context of updated data that may disclose new mutations in binding regions. For instance, most mutations are found in N gene primers and probes [42], therefore assays targeting this region should be subject to frequent reassessment so that they can be adjusted with consideration to new mutation knowledge. It should be also noted that primer target regions located in the middle or at the 3' ends of the viral genome have been shown to accumulate mutations and recombination events at the highest rate, hence it can be essential to consider the most suitable sites for primer design as has been done in a study including the nsp1 gene found at the 5' end [20].

Another challenge comes from using unoptimized protocols and/or primers leading to false-negative outcomes. When different primers are used in the same protocol, since each primer pair works efficiently at different annealing temperatures and amplification parameters, the optimization workflow is critical [43]. For example, a two-step PCR protocol recommended by WHO has been found to lead to formation of primer dimers and unspecific binding of primers and probes, causing unspecific signals for the target gene [44]. One solution to achieve optimal sets of primers is to filter out unfeasible primer pairs among a large collection of choices obtained by primer generation programs while at the same time considering multiple parameters specific to SARS-CoV-2 and other coronaviruses.

Studies that design primers use data collected from clinical research accessible in nucleotide and protein databases and generate primers using consensus sequences [20,45,46]. The pipeline involves aligning selected sequences to find conserved consensus regions [26]. A recent study [5] provides an online platform to search for existing primers across available viral genomes in NCBI and displays mutation counts. Similarly, another database (CoV2ID) [3] provides details including a reference sequence, a table of oligonucleotides used in several diagnostic protocols and results of three different alignments with tunable parameters also available. Furthermore, Naeem et al. have developed a web application facilitating inquiry and visualization of mutations (number of mutated strains from around the world or a particular country) in established or user provided primer sequences [6]. Finally, a webtool called CoronaVR which is an integrative resource has been developed mainly for epitope analysis but also reports primers newly designed and from literature [4]. Yet there is no online tool that facilitates generation and selection of novel conserved or degenerate primers based on the given primer selection parameters as well as mutation counts obtained from multiple alignment of available SARS-CoV-2 genomes.

In the present study, we have identified conserved genomic regions by defining a threshold of 0.1% [40,47] on the data downloaded on May 12, 2021 from the 2019 Novel Coronavirus Resource (2019nCoV), an online analysis platform providing open-access information on SARS-CoV-2 from National Genomics Data Center (NGDC) of the China National Center for Bioinformation (CNCB) [32,48,49]. The defining and commonly found mutations in new variants of the virus have been used to generate degenerate primers which can then be utilized in diagnosis of various strains of the virus carrying mutations at corresponding positions. Using the Shiny [50] and openPrimer [51] packages of R programming language, we have made it possible for the user to search and filter existing and novel primers and visualize the alignment to the reference genome (NC_045512.2 [52]) while being presented with the information on mutation load of primers. In CoVrimer, all primers can be filtered based on common primer design parameters such as

minimum Tm value and GC content as well as the degree of homology with other bat coronaviruses. CoVrimer can help COVID-19 researchers to test their own primers as well as select new primers considerate of viral evolution.

2. Materials and methods

CoVrimer is a web-based software application which has been created using the open-source programming language R [53] and its Shiny [50] package. The interface layout has been designed using the shinydashboard package [54]. CoVrimer contains two main modules, namely, *Align Published Primer Sets* and *Select New Primer Pair*, with two submodules.

2.1. Mutation analysis and filtering

On the day data were downloaded, the number of sequences used in mutation analysis was reported to be 954,725, which we have based our mutation frequency calculations on. In the downloaded data, 28,499 nucleotide positions of the whole genome sequence were shown to contain at least one mutation. However, to exclude those with low prevalence and potential sequencing errors in primer generation, we have defined a stringent threshold of 0.1% where positions with <0.1% alteration incidence were considered as conserved [40,47].

2.2. Identification of conserved primer pairs

Conserved primers have been designed within the conserved regions

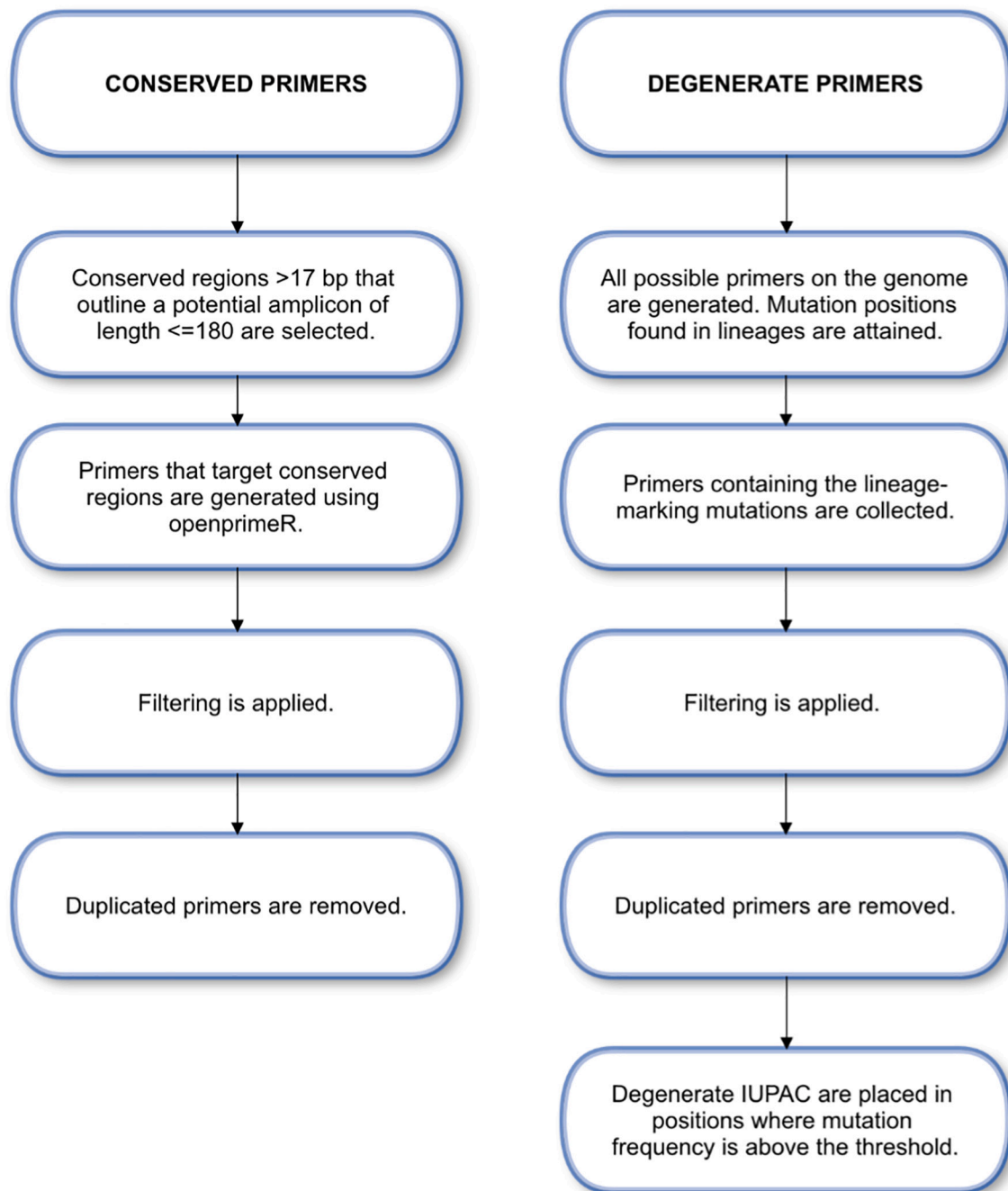


Fig. 1. Pipeline of conserved and degenerate primer design.

which we have defined by the selected threshold of 0.1%. Our algorithm involves first the calculation of lengths of all conserved regions and then selection of segments longer than 17 and less than 181 because of thresholds set for minimum primer and maximum amplicon length as 18 bp and 180 bp, respectively.

For each region, all possible primer pairs that could bind to those regions have been generated using `get_initial_primers` function of `openPrimer` package [51]. We have created six additional functions: `forw_degenerate`, `reverse_degenerate`, `calculate_Tm`, `calculate_GC`, `find_Tm` and `find_Gc`. `find_Tm` and `find_Gc` have been used to identify the minimum and maximum possible Tm and GC values for degenerate primers. After obtaining all conserved primer pairs, filters based on Tm and GC values have been applied in order to reduce number of primer pairs. Additional filtering has been done based on the SARS-CoV-2 specificity values. In the final stage of filtering, primer pairs having SARS-CoV-2 specificity values greater than 0 were selected. To remove repetitive pairs that cover same nucleotide positions with only a few base pair differences, the unique forward primers based on their sequences and start/end positions have been grouped before retaining the reverse primers having maximum SARS-Specificity from each group. This algorithm has also been applied for grouping of the reverse primers and retaining the forward primers with the maximum SARS-Specificity. The workflow of designing primers and the filtering parameters are shown in Fig. 1 and Table 1, respectively.

2.3. Identification of degenerate primer pairs

10 of the current SARS-CoV-2 lineages have been selected as subjects for getting degenerate primer pairs. The mutations in these variants have been identified and used to create degenerate primers inclusive of the mutated positions. For example, among the selected variants in CoVrimer, D614G and P314L variations found in all lineages and T265I, which is found in 5 of 10 lineages, have mutation frequencies %98.25, %98.19 and %14.85, respectively.

Initially, we first identified all possible primer pairs that cover the coding regions of the viral genome based on the reference sequence (NC_045512.2). Next, the pairs that include the mutated positions of different lineages in either the forward or the reverse primer have been selected. The same filtering procedure as that used for conserved primers has been applied to these degenerate primer pairs (Table 1); and those pairs showing conservation in the last 5 nucleotides at 3' end have further been chosen. In the primer pair pool, some degenerate primers were paired with conserved primers sharing the same primer start or end positions resulting in overlapping primer selection regions in the viral genome. Accordingly, the forward primers with identical sequences but pairing with different conserved reverse primers were grouped together, and then primer pairs from each group whose reverse primer with the maximum SARS specificity have been kept. The same procedure has also been applied for the identical reverse primers that have conserved forward primer pairs. A degenerate primer with no conserved pair was matched with the primer having maximum SARS specificity. The mutated positions were converted to degenerate nucleotides to be used in virus detection from samples of different lineages. All obtained primer sets have been integrated into CoVrimer upon removal of duplicated

Table 1
Parameters used in filtering the primers.

Length	18–25 nucleotides
GC Content	40–60%
Tm	55–65 °C
Runs	no more than 4 bp
GC Clamp	no more than 3 in last 5 bases at 3' end
End	no T residue at the 3' end
Tm differences	< 5
Amplicon length	60–180 bp
SARS-Specificity	no 0 values in both primers

pairs.

2.4. Graphics for alignments

In CoVrimer, the amplicon targeted by the selected primer pair can be visualized with respect to nucleotide positions and percentages of possible mutations. In the first graph, the reference nucleotides are placed at the bottom (on x-axis) and frequency of variation is shown on the y-axis. Primer binding sites are displayed in blue and the region in between in green. Dots representing mutations below and above the threshold are orange and red, respectively. This visualization has been made possible using `tidyverse` [55] and `ggplot` [56] packages where nucleotides from the reference sequence are positioned below those of the primer sequence query. The user can zoom in on a region of selection to better visualize the nucleotides' identities. Above this graph, the detailed information about each mutated position within the amplicon is provided in table format. All variation details regarding SNP, indel, insertion and deletion counts, and alterations can be obtained from this table. Another panel generated via use of `msaR` package [57] displays the alignment of the selected primer pair to the reference genome.

2.5. Visualization of conservation of primers among 44 bat viruses

UCSC SARS-CoV-2 Genome Browser provides 50 tracks with different molecular details about the virus [58]. PhyloP data, which are derived from 44 bat viruses, and in which positive scores indicate that a given site is predicted to be conserved while negative scores are assigned to the sites predicted to be fast evolving, have been obtained from UCSC genome browser (<http://genome.ucsc.edu/>). In CoVrimer, PhyloP score of each nucleotide of the selected primer pair is displayed on the graphs which also show the positions of the amplicon exhibiting mutations, if present.

We have also aligned sequences of all 44 bat viruses used in PhyloP data present on UCSC SARS-CoV-2 Genome Browser and generated a variation scoring table indicating the frequency of observed variation at each position (v_score). Based on this table, we calculated the SARS-CoV-2 specificity of each primer using the following formula where L is the length of the primer in bp:

$$\frac{\text{sum}(v_scores \text{ of nucleotides in } L-5) / (L-5) + \text{sum}(v_scores \text{ of last } 5 \text{ nucleotides} * (1 + v_scores \text{ of last } 5 \text{ nucleotides}))}{5}$$

We increased the weights of the last 5 nucleotides at the 3' end where mismatches exert an enhanced negative effect on priming by disrupting the nearby polymerase active site [59]. The importance of the last 5 nucleotides at 3' end had been highlighted in a study which ranked oligonucleotides by taking mean value of three different measures: percentage of identical sites in the last 5 nucleotides at the 3' end, percentage of identical sites in the primer and percentage of pairwise identity [3,60]. In our analysis, we have used the frequency of substitutions on each position and reached the final score by increasing the contribution of last 5 nucleotides to the score. The proposed scoring scheme hence assigns higher scores to primers with greater v_scores towards the 3' end of the sequence.

2.6. Table of published primer sets

The database for published primer sets can be found under the module *Align Published Primer Sets*. We have manually curated a database from literature containing published RT-PCR primer sets for which sequence information is available and crosschecked with the recently published list in COV2ID website [3]. In one column of our primer table, the degenerate versions of primers representing possible mutations in their target regions are displayed. In other columns we report the number of possible mutations and the mutation frequencies in the target regions of primers. When a row is selected upon clicking, alignment graphs are displayed at the bottom of the table.

2.7. Conserved and degenerate primer selection

The second main module of CoVrimer, named *Select New Primer Pairs*, has been designed to allow the user to select conserved as well as degenerate primers. Gene regions suitable for designing conserved primers could be identified are reported along with an option to visualize “All”. Additionally, in the second tab of this section, degenerate primers exhibiting mutations found in different virus lineages are presented along with the name of the lineages having the mutations in primers. Moreover, CoVrimer allows the user to select the desired features and limits by using sliders in the primer table such that a reduced number of primer sets can then be displayed.

With the help of the *Get Degenerate Primer Pairs* module, the user can provide the nucleotide sequence of a primer pair of choice in the respective search boxes and inquire available features on that primer pair. Each forward and reverse primer are analyzed separately to visualize their alignment with the reference genome. Degenerate IUPAC codes are placed in the primer/primer pair if the primers’ target regions contain mutations above the previously determined frequency threshold; and a table presenting information on the position and length of primer binding regions, position specific number and frequency of observed mutations in these regions, Tm and GC content values, and amplicon length is displayed. In all modules, alignments of primers and probes, if present, to the reference genome and the sequence of amplicon can be visualized.

3. Results

3.1. Visualization of alignments can be used to identify mismatches on published primer sets

Primer sets designed in numerous studies (the first-time appearance) have been included in CoVrimer with Pubmed IDs and links to the articles (Fig. 2). In total, 90 pairs/sets of published primer sequences are currently housed in CoVrimer (literature search: 18/04/2021). Degenerate versions of published primers have been generated on positions where mutation frequencies are greater than 0.1%, whereas the mutation frequency is considered 0 when below 0.1%. CoVrimer makes possible the visualization of mutation load in the primers, probes and amplicons by presenting a table including all variation counts and alterations, and showing the alignment of each set to the reference sequence when the user highlights a row on the primer table (Fig. 3, top table and first plot). The reference nucleotides are displayed on the x-axis while the y-axis shows the percentage of mutations along with the altered version of the nucleotide at each position. The second plot of

Fig. 3 shows PhyloP (44 bat viruses) value of each nucleotide in the primer pair and the amplicon followed by visualization of the alignment of primer pair and probes to the reference genome (Fig. 3, third plot).

3.2. Select New Primers Pairs module can provide SARS-CoV-2 specific optimized conserved and degenerate primers

All observed mutations having frequency greater than 0.1% have been saved into a table and used within the CoVrimer pipeline leading to generation of 551 primer pairs corresponding to the segments that are conserved. These primers are housed in the *Select New Primers* module, allowing the user further filtering through general primer optimization parameters, such as the length of amplicon and primer, Tm and GC content as well as SARS-CoV-2 specific parameters (Fig. 4). Conserved primers identified for all viral genes except nsp7, 2'-O-ribose methyltransferase, Envelope, ORF7, ORF7b, ORF8 and ORF10 are displayed.

Before filtering, we have had more than 100,000 primer pairs having lengths in between 60 and 180 bp. After the first filtration process, restrictions of which are described in Table 1, this number has been reduced to 13,000 primer pairs. Later, removal of pairs with duplicate sequences and start/end sites and selection of the primers bearing maximum SARS-Specificity on their pairs have resulted in only 551 pairs, which are placed under *Select Conserved/Degenerate Primers* tab. We have provided all possible pairs of forward and reverse primers that meet all criteria in the table under this tab. In CoVrimer, additional columns showing positions and frequencies of mutations make annotation of primers possible. In the *Degenerate primer pairs* sub-tab, we have provided information about 610 degenerate filtered primer pairs designed based on common virus strains, such as position of mutations, degenerate version of primers, mutation frequencies, and lineage names annotated with mutation names (Fig. 5). SARS-CoV-2 Specificity, calculated as mentioned in methods, increases confidence in primer specificity and resulting binding efficiency to SARS-CoV-2 genome. All four plots described in Fig. 3 can be generated for the conserved primer and degenerate primer table entries when a row in the table corresponding to a primer set is selected.

We have also analyzed the correlation between SARS-CoV-2 specificity indices and mean of PhyloP scores of each primer in the conserved, degenerated and published primer datasets (Supplementary Fig. 1). Results indicate that all primers we have generated share similar profiles with the published primers. However, the degenerate primers exhibit higher SARS-Specificity indices in general. The mean values (+/- Std) for conserved, degenerate as well as published primers in terms of Tm, GC and SARS-CoV-2 specificity indices (Supplementary Table 1) indicate that published primers have lower specificity indices in comparison

Table of published primer sets											
Please select a primer pair to display mutation and conservation plots for the amplicon and alignment to the reference genome											
COUNTRY	PANEL	TARGET GENE	LABEL	PRIMER SETS	DEGENERATE VERSIONS	POSITIONS	AMPLICON SIZE	MUTATION NUMBER IN FW	MUTATION FREQUENCY IN FW (if <0.1%)	MUTATION NUMBER IN RV	MUTATION FREQUENCY IN RV (if <0.1%)
China	NA	orf1ab	Forward Primer, Reverse Primer	CAGACCTCGTCATGCTCTTAAGGC CCCTGGTCAAGGTAATATAGGCA	cgagcctcgctcgtcctttaaaggc ccctggtaaatataggca	13814-13837 14165-14188	375	0	000000000000000000000000	2	0000001440000000001230000000000000
China	NA	S	Forward Primer, Reverse Primer	CTTCCTCGATGTCAGCACCTC AKCCAGTGTGTGGCATTGSA	cttccctcgtgtagcactc aacccatggtggcattgsa	24719-24734 24851-24879	156	0	000000000000000000000000	1	0000000117000000000000000000
Germany	WHO panel	RdRP	RdRp_SARS-F, RdRp_SARS-R, RdRp_SARS-P1, RdRp_SARS-P2	GTGAAATGGTCAATGTGTGGGGG GARATGTAAGAGACATWGCATA FAM: CCAGGTGGWACHTCAFCMGGTGTATG- BHQ FAM: CAGGTGGAACCTTCATCAGGATATG- BHQ	gtgaaatggtcaatgtgtggggg garatgtaagagacatwgcata fam: ccaggtggwachtcafcmmgggtgatg- bhq fam: caggtggaaccttcatacaggatgatg- bhq	15431-15432 15505-15530 15469-15494 15470-15494	100	2	00000000002000000000000000001400	0	00000000000000000000000000000000

Fig. 2. Table of published primer sets. The table shows each primer set that has appeared for the first time in its respective study, country of origin and the panel name if given, the target region of the primers, and the sequences of primer sets. CoVrimer provides degenerated versions of the published primers, where mutations have frequencies above 0.1%, as well as the amplicon length between forward and reverse primers. Other listed details include computer programs used for the designing of primers, the assays for which the primers have been designed and further relevant information.

Detailed Variation Information

Show entries

Search:

	position	reference nt	SNP	SNP count	SNP freq	insertion	insertion count	insertion freq	deletion	deletion count
1	28881	G	G>C:8,G>A:561215,G>K:3,G>M:1,G>R:364,G>W:3,G>T:2260	563854	1				GGGGAACCTC>G:1,GGGGAACCTCTCCTGC>G:1	2
2	28882	G	G>C:8,G>A:559679,G>K:3,G>M:1,G>R:366,G>S:1,G>T:91	560149	1					
3	28883	G	G>C:559802,G>A:181,G>M:1,G>R:3,G>S:370,G>T:3	560360	1					
4	28884	G	G>C:27788,G>A:69,G>K:6,G>M:1,G>R:3,G>S:17,G>T:1443	29327	1					
5	28885	A	A>T:3,A>W:1,A>R:4,A>G:24,A>C:4,A>M:1	37	0.974				AACTTCTCCTGCTAGAATG>A:1	1

Showing 1 to 5 of 98 entries

Previous 2 3 4 5 ... 20 Next

Red dots (above the threshold (0.1%)) and orange dots (below the threshold (0.1%)) represent mutations in nucleotides.
 Blue and green letters refer to the nucleotides in the reference genome (NC_045512) and indicate primer binding sites and the region between primers, respectively.
 Next to the points are the degenerate IUPAC codes for each alteration.
 PhyloP values (44 Bat virus strains Basewise Conservation) were obtained from UCSC.
 *Select the area and double-click for zoomed version.
 *Before selecting the new primer pair, please double-click on the empty place to reset the zoom.

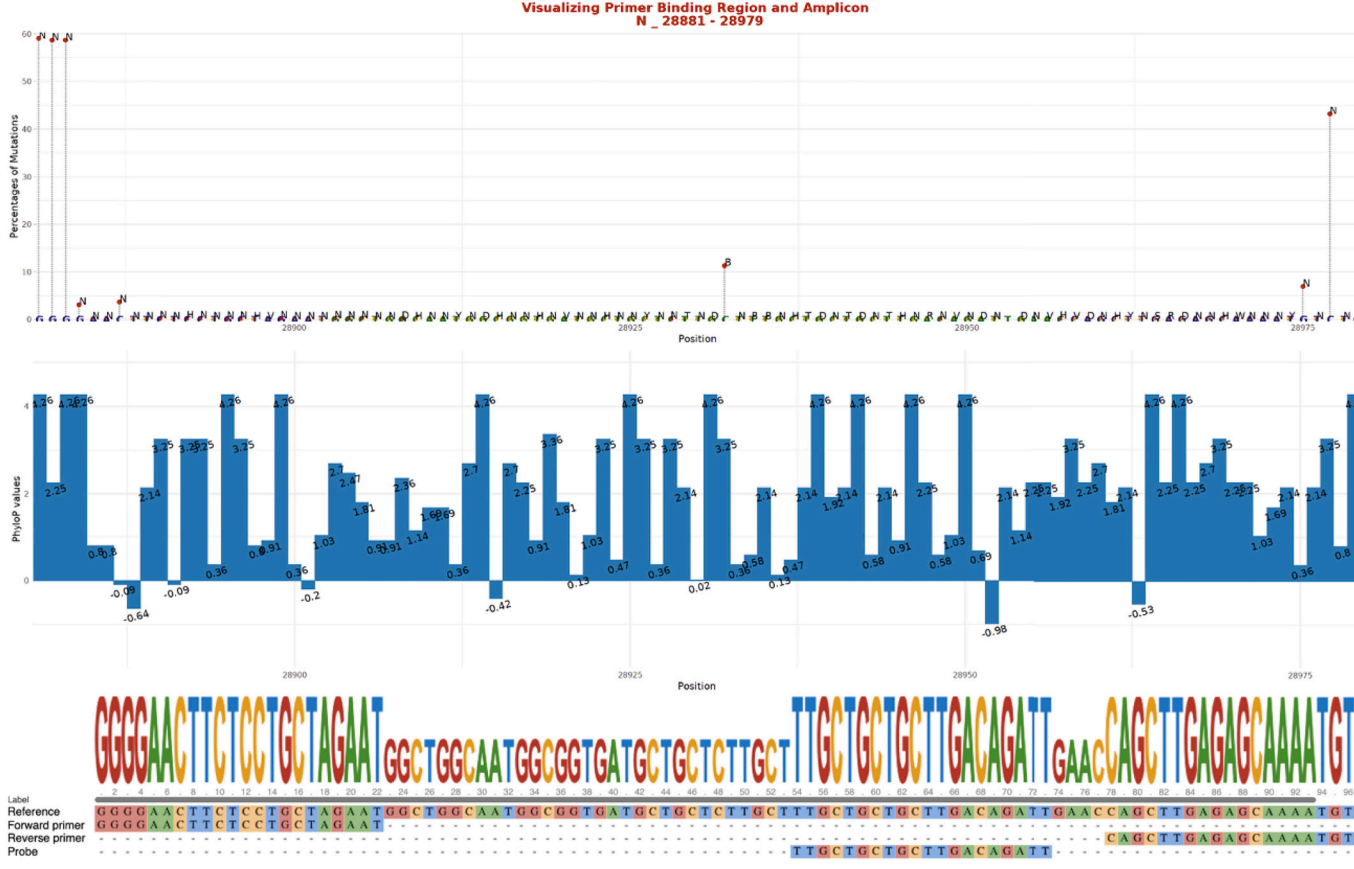


Fig. 3. Mutation details, percentage, PhyloP scores and pairwise alignment of a selected primer pair. “Detailed Variation Information” section displays the mutation details such as variation types and counts within amplicon. The first plot after the table shows the mutation percentages along with the selected primer binding regions and the amplicon where mutations have occurred based on multiple sequence alignment of viral genomes. Next plot shows the PhyloP value of each nucleotide within the window visualized. The bottom plot presents the msr alignment showing the pairwise alignment of the selected primer pair to the reference sequence (NC_045512.2).

to both the degenerate and conserved primers contained in CoVrimer.

3.3. Users can identify degenerate primers using Get Degenerate Primer Pairs module

Users can search their own primer pairs at hand in this module that displays available information in CoVrimer related to the primer pair being analyzed. If the binding regions of primers have mutation frequencies above our threshold at any position, CoVrimer constructs degenerated version(s) of the primer(s) and shows the presence of

mutations at each position in a numbered format. For example, in “000100020100”, each number indicates the number of nucleotides that can replace the original nucleotide found in the reference genome on that particular position-in other words, in how many different ways the position can be altered (Fig. 6). “1” indicates that only one of the possible nucleotide substitutions has been reported for the reference nucleotide while 2 indicates changes have been reported with two different nucleotides. CoVrimer also shows the mutation frequency observed at each position. In the last part of this module, the alignment of a searched primer set to the reference genome and the corresponding

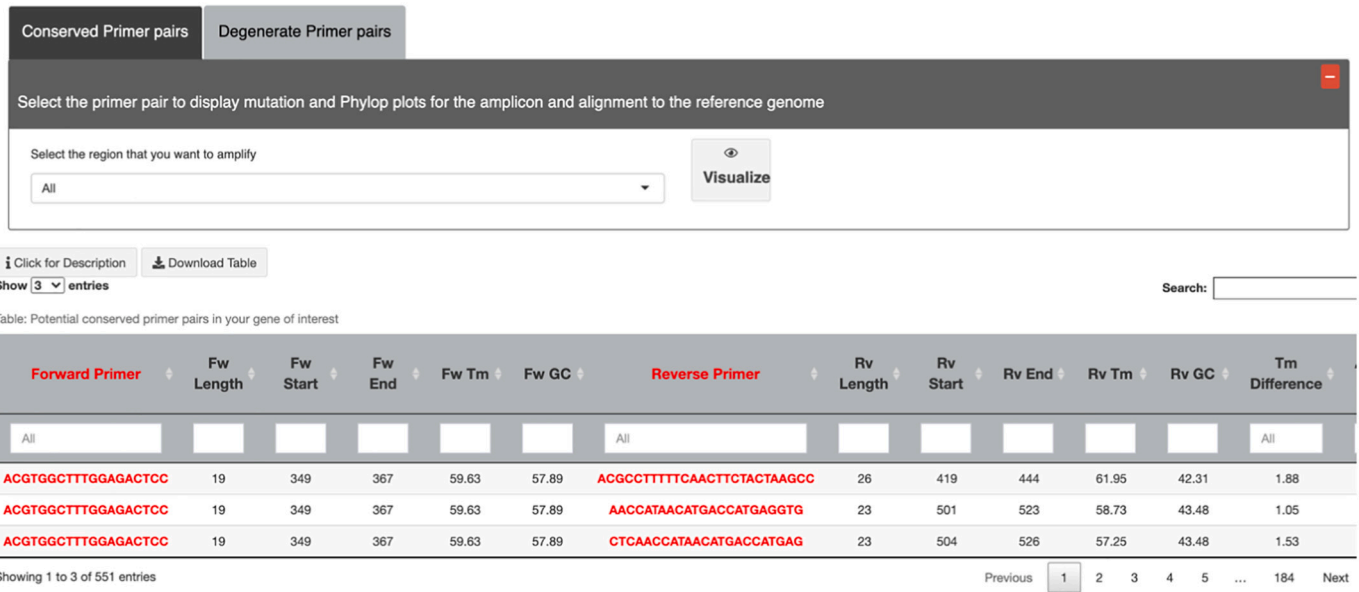


Fig. 4. Tab showing the table of conserved primer pairs. All possible conserved primer pairs indicating the number of mutations in the amplicon region are presented along with SARS-CoV-2 specificity indices calculated (see Methods).

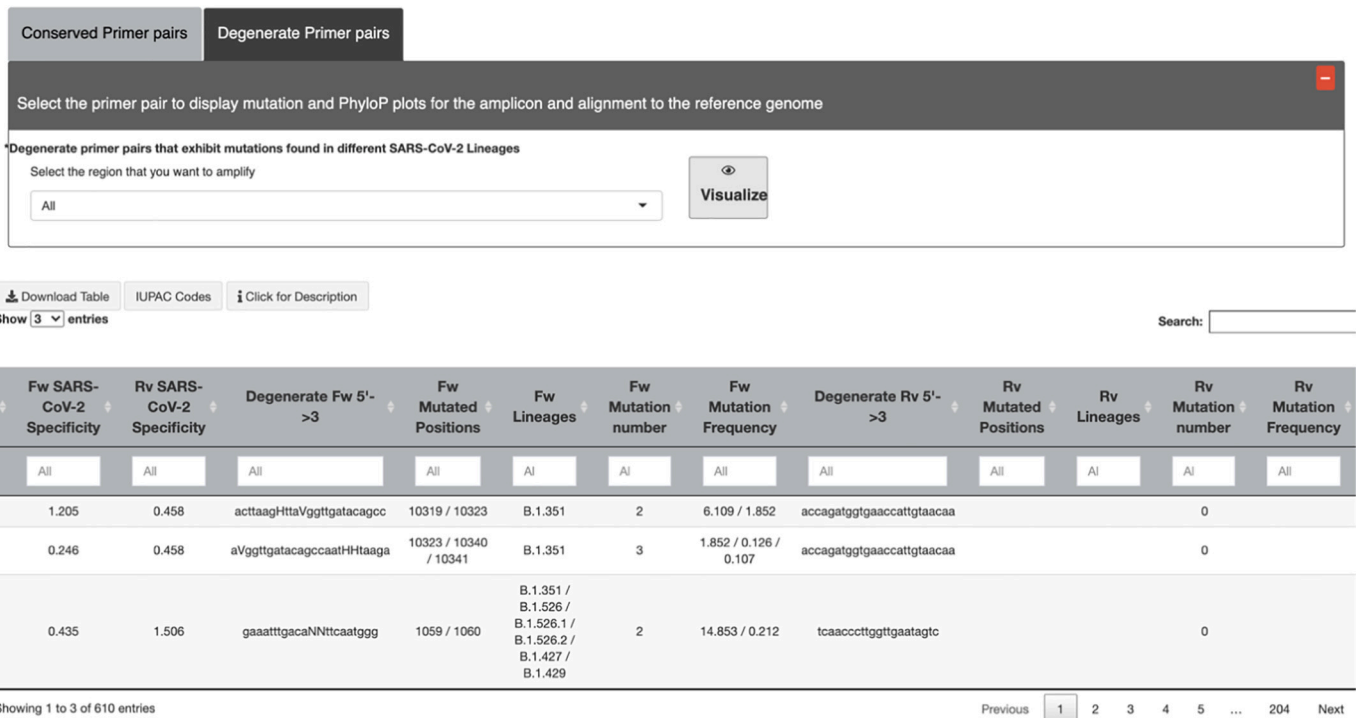


Fig. 5. Tab showing primer pairs with degeneracy. Last columns of the table contain degenerated versions of primers, exact positions of mutations, lineages which mutations belong to, number of mutations in each primer and mutation frequencies.

amplicon region with detailed information on mutated nucleotides can be visualized by clicking on the *Align* button.

4. Discussion

Numerous information portals and dashboards for COVID-19 have been launched by healthcare professionals, researchers, and organizations. Some of them provide general information on the epidemiology of the disease, signs, diagnosis, preventive steps, treatment and the ratios

of morbidity and mortality [3,4,31,61,62]. Others provide information about the genome of the virus with respect to sequence variations, alignments with other viruses and geographical mapping [49,58,63]. Along with these portals, there has been a tremendous need for online primer design/screening tools for helping researchers develop efficient diagnostic assays and therapeutic drugs. With the intention to aid in responding to this need, we have developed an application named CoVrimer, which contains multiple modules for selection of novel conserved primer pairs as well as modification of existing primers

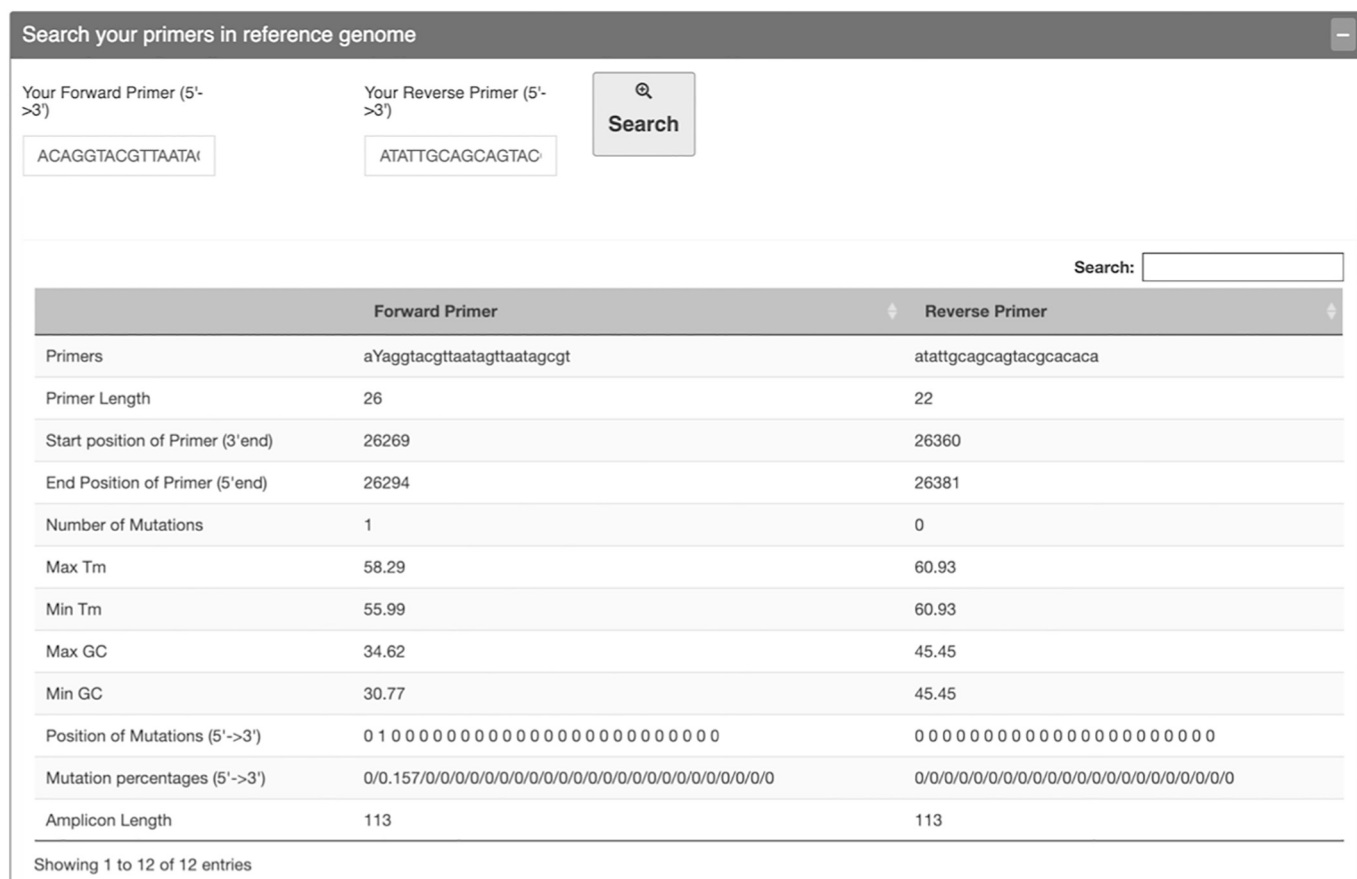


Fig. 6. *Get Degenerate Primer Pairs* tab allows users to search their primer pairs by pasting the sequence from 5' end to 3' end into the boxes. The table shows degenerated versions if the binding region has mutations present. The binding locations and positions of mutations are presented in the table along with the information on predicted Tm and GC values.

through insertion of degenerate nucleotides to accommodate the evolution and sequence heterogeneity of SARS-CoV-2 observed across the world.

In a recent study, Jain et al. have obtained complete SARS-CoV-2 genome sequences from GISAID, filtered them by certain criteria and aligned the remaining 45,830 sequences to the Wuhan-Hu-1 reference genome [64]. 132 primer/probe sequences were mapped onto the reference genome and overlapped with the variant positions. They have identified 5862 unique variants present on primer/probe binding sites, and a total of 27 primers showing a cumulative variant frequency of more than 1%, including those targeting S, M, ORF1ab, and ORF3a. Strikingly, the 2019-nCoV-NFP binding site was found to have a cumulative variant frequency of 93.5%. Moreover, they have identified 268 regions of at least 20 bp that were conserved. In CoVrimer, we have identified 553 conserved regions (having less than 0.1% variation frequency) with lengths varying between 18 bp and 140 bp based on downloaded mutation data. However, the availability of conserved regions is likely to decrease as the virus genome continues to mutate and when new mutation data are incorporated in the future updates of CoVrimer, the selected threshold value may need to be changed.

We also have performed a statistical analysis of the mutation data on published primers and found that the highest number of mutations on a primer binding site is observed for N. The median mutation occurrence for N-targeting primers was 5, the highest number among the primers targeting other genes (2.19, 3.1 and 2 for E, Orf1ab, S and M, respectively). Proportions of primers that have at least 5 potential mismatches in their binding sites were 60.86%, 23.80% and 11.11% for N, S and Orf1ab targeting primers, respectively. Therefore, we have concluded that N primer target sites present in literature are more frequently

mutated in comparison to target sites present on other genes, which may pose a risk on the efficiency of N targeting primers.

Our analysis has shown that majority of the published primer pairs might exhibit up to two or more possible mutations with frequencies greater than 0.1% (Supplementary Fig. 2). Similar results have been found in a study performing mutation analysis of 41 primers/probes across more than 7000 complete genomes collected between Jan 5 and May 1, 2020 [38]. In all primers they analyzed, we have found smaller number of mutations, most likely due to our use of a conservation level threshold in CoVrimer pipeline. Since we have defined this threshold as 0.1%, it is inevitable that some mutations do not appear in the sequences. Otherwise, we would have expected to observe mutations at almost every position since 28,499 positions of the virus genome exhibit at least one mutation.

Our findings confirm that as mutations in SARS-CoV-2 genome accumulate, the previously established primer pairs may need to be modified. CoVrimer can help with this task by providing degenerate versions of existing primers with specific lineage information. In addition, we demonstrate that designing conserved primers may be feasible when the regions where variation frequencies are below a certain threshold are targeted.

There are other SARS-CoV-2 oriented primer and mutation databases [3,5,6] reporting on a smaller number of primers while another study displays limited number of newly designed potential primers [4]. In this respect, CoVrimer is a conserved and degenerate primer prioritization tool across lineages and specific for SARS-CoV-2, hence is extensive, modular and updateable in structure and in essence. In addition, it has the capability to perform alignments of these novel and published primers online while at the same time displaying mutation frequencies,

alterations in nucleotides and PhyloP values. Genomic mutations that survive the natural selection are constantly acquired by all replicating viruses, including coronaviruses. CoVrimer has the unique feature of allowing for selection of conserved primers that are more specific for SARS-CoV-2 in comparison to other similar coronaviruses. Currently, CoVrimer incorporates conservation data across 44 bat viruses from UCSC Genome browser in PhyloP format; this can be observed for the alignment of primers and amplicons, thereby helping users select the regions most conserved within SARS-CoV-2 and the least among other related viruses. Future studies are planned to add tracks for human coronaviruses and to incorporate additional alignments against viral genome (e.g., blast) for increasing stringency in filtering primers.

In conclusion, CoVrimer is the first conserved and/or degenerate primer selection and alignment webserver for SARS-CoV-2 based on genomic information from close to a million isolates and considers viral evolution and incorporates methods to overcome associated challenges. Frequent updates are planned to follow the viral evolution closely so that the accuracy and functionality of primer analysis and design processes can be maintained. As revealed by large scale genome database content, 28,499 positions of SARS-CoV-2 genome are found to contain at least one variation and 1788 positions have mutation frequency above the threshold (0.1%), confirming the need for new approaches to primer design and optimization of SARS-CoV-2 diagnostics. CoVrimer houses existing primer sequences and makes it possible for the user to analyze primer pairs, probes, and the amplicon in the context of number of mutations present as well as other SARS-CoV-2 specific parameters. Incorporation of variation information on the amplicons is important for accurate estimation and reduction of multiple amplicon peaks while facilitating visualization of the alignment of primers/amplicons against the reference genome. Thus, it may be possible to use CoVrimer to identify amplicons that contain the most variable regions to amplify since there can be more than one variant present in an amplicon. The current version of CoVrimer allows for static viewing of the mutation positions, frequencies, and the IUPAC codes on plots while detailed information on mutations are given in a table. The future updates on plots will include the incorporation of an interactive visualization option where mutation types/frequencies are accessible on the plot itself. Moreover, we plan to add more flexibility by allowing user-input for determining a threshold value in the selection of mutation frequency above which a position in the primer will be degenerated. Our results suggest that degeneracy of nucleotides which are mutated in different virus lineages within a primer is needed for reliable detection as the virus evolves. Currently, CoVrimer houses degenerate primers that can detect the presence of virus regardless of the selected variants. However, CoVrimer can also be used to detect different SARS-CoV-2 lineages based the “Detailed Variation Information” table showing each genomic alteration regardless of a threshold. As new primer pairs emerge in the literature, they will be housed in the existing primer data table and an updated version of CoVrimer will be available to trace the changes in primer/probe/amplicon sequences as the virus evolves incorporating the viral heterogeneity.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ygeno.2021.07.020>.

Declaration of Competing Interest

The authors report no conflict of interest.

Acknowledgments

We would like to thank Ayse G. Keskus for helpful comments and suggestions, and Can Ugur Ayfer and Bilkent University for the support provided to host CoVrimer.

References

- [1] World Health Organization (WHO), Novel Coronavirus - 2019-nCoV SITUATION REPORT - 1 21 JANUARY 2020, *Mater. Methods*. 10 (2020).
- [2] S.A. Bustin, T. Nolan, RT-qPCR testing of SARS-CoV-2: a primer, *Int. J. Mol. Sci.* 21 (2020), E3004, <https://doi.org/10.3390/ijms21083004>.
- [3] J. Carneiro, C. Gomes, C. Couto, F. Pereira, CoV2ID: Detection and Therapeutics Oligo Database for SARS-CoV-2, *BioRxiv*. (2020), <https://doi.org/10.1101/2020.04.19.048991>, 2020.04.19.048991.
- [4] A.K. Gupta, M.S. Khan, S. Choudhury, A. Mukhopadhyay, A. Rastogi Sakshi, A. Thakur, P. Kumari, M. Kaur, C. Saini Shalu, V. Sapahia, P.K. Patel Barkha, K. T. Bhamare, M. Kumar, CoronaVR: a computational resource and analysis of epitopes and therapeutics for severe acute respiratory syndrome coronavirus-2, *Front. Microbiol.* 11 (2020), <https://doi.org/10.3389/fmicb.2020.01858>.
- [5] A. Rakha, H. Rasheed, Z. Batool, J. Akram, A. Adnan, J. Du, COVID-19 variants database: a repository for human SARS-CoV-2 polymorphism data, *BioRxiv*. (2020), <https://doi.org/10.1101/2020.06.10.145292>, 2020.06.10.145292.
- [6] R. Naeem, A. Pain, assayM: A Web Application to Monitor Mutations in COVID-19 Diagnostic Assays, 2020, <https://doi.org/10.1101/2020.12.18.423467>.
- [7] V.M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D.K.W. Chu, T. Bleicker, S. Brünink, J. Schneider, M.L. Schmidt, D.G.J.C. Mulders, B. L. Haagmans, B. Van Der Veer, S. Van Den Brink, L. Wijsman, G. Goderski, J. L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M.P. G. Koopmans, C. Drosten, Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, *Eurosurveillance*. 25 (2020), <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>.
- [8] E.G. Dorlase, C.O. Monteiro, A.O. Viana, C.P. Soares, R.R.G. Machado, L. M. Thomazelli, D.B. Araujo, F.B. Leal, E.D. Candido, B.L. Telezynski, C.A. Valério, V.N. Chalup, R. Mello, F.J. Almeida, A.S. Aguiar, A.C.M. Barrientos, C. Succupira, M. De Paulis, M.A.P. Sáfiadi, D.G.B.P. Silva, J.J.M. Sodré, M.P. Soledade, S. F. Matos, S.R. Ferreira, C.M.N. Pinez, C.P. Buonafina, L.N.F. Pieroni, F.M. Malta, R. A.F. Santana, E.C. Souza, R.A. Fock, J.R.R. Pinho, L.C.S. Ferreira, V.F. Botosso, E. L. Durigon, D.B.L. Oliveira, Lower cost alternatives for molecular diagnosis of COVID-19: conventional RT-PCR and SYBR Green-based RT-qPCR, *Braz. J. Microbiol.* 51 (2020), <https://doi.org/10.1007/s42770-020-00347-5>.
- [9] S. Bhadra, T.E. Riedel, S. Lakhotia, N.D. Tran, A.D. Ellington, High-surety isothermal amplification and detection of SARS-CoV-2, including with crude enzymes, *BioRxiv*. (2020), <https://doi.org/10.1101/2020.04.13.039941>, 2020.04.13.039941.
- [10] L.E. Lamb, S.N. Bartolone, E. Ward, M.B. Chancellor, Rapid detection of novel coronavirus/Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) by reverse transcription-loop-mediated isothermal amplification, *PLoS One* 15 (2020), <https://doi.org/10.1371/journal.pone.0234682>.
- [11] M. Jiang, W. Pan, A. Arasthfer, W. Fang, L. Ling, H. Fang, F. Daneshnia, J. Yu, W. Liao, H. Pei, X. Li, C. Lass-Flörl, Development and validation of a rapid, single-step reverse transcriptase loop-mediated isothermal amplification (RT-LAMP) system potentially to be used for reliable and high-throughput screening of COVID-19, *Front. Cell. Infect. Microbiol.* 10 (2020), <https://doi.org/10.3389/fcimb.2020.00331>.
- [12] G. Xue, S. Li, W. Zhang, B. Du, J. Cui, C. Yan, L. Huang, L. Chen, L. Zhao, Y. Sun, N. Li, H. Zhao, Y. Feng, Z. Wang, S. Liu, Q. Zhang, X. Xie, D. Liu, H. Yao, J. Yuan, Reverse-transcription recombinase-aided amplification assay for rapid detection of the 2019 novel coronavirus (SARS-CoV-2), *Anal. Chem.* 92 (2020), <https://doi.org/10.1021/acs.analchem.0c01032>.
- [13] T. Wu, Y. Ge, K. Zhao, X. Zhu, Y. Chen, B. Wu, F. Zhu, B. Zhu, L. Cui, A reverse-transcription recombinase-aided amplification assay for the rapid detection of N gene of severe acute respiratory syndrome coronavirus 2(SARS-CoV-2), *Virology*. 549 (2020), <https://doi.org/10.1016/j.virol.2020.07.006>.
- [14] T. Shelite, A. Uscanga-Palomeque, A. Castellanos, P. Melby, B. Travi, Isothermal Recombinase Polymerase Amplification-lateral Flow Detection of SARS-CoV-2, the Etiological Agent of COVID-19, 2020, <https://doi.org/10.21203/rs.3.rs-78408/v1>.
- [15] J. Gómez, S. Melón, J.A. Boga, M.E. Alvarez-Argüelles, S. Rojo-Alba, A. Leal-Negredo, C. Castello-Abietar, V. Alvarez, E. Cuesta-Llavona, E. Coto, Capillary electrophoresis of PCR fragments with 5'-labelled primers for testing the SARS-CoV-2, *J. Virol. Methods* 284 (2020), <https://doi.org/10.1016/j.jviromet.2020.113937>.
- [16] T. Suo, X. Liu, J. Feng, M. Guo, W. Hu, D. Guo, H. Ullah, Y. Yang, Q. Zhang, X. Wang, M. Sajid, Z. Huang, L. Deng, T. Chen, F. Liu, K. Xu, Y. Liu, Q. Zhang, Y. Liu, Y. Xiong, G. Chen, K. Lan, Y. Chen, ddPCR: a more accurate tool for SARS-CoV-2 detection in low viral load specimens, *Emerg. Microbes Infect.* 9 (2020), <https://doi.org/10.1080/22221751.2020.1772678>.
- [17] R. Wang, C. Qian, Y. Pang, M. Li, Y. Yang, H. Ma, M. Zhao, F. Qian, H. Yu, Z. Liu, T. Ni, Y. Zheng, Y. Wang, opvCRISPR: One-pot visual RT-LAMP-CRISPR platform for SARS-cov-2 detection, *Biosens. Bioelectron.* 172 (2021), <https://doi.org/10.1016/j.bios.2020.112766>.
- [18] D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K.M. White, M. J. O'Meara, V.V. Rezelj, J.Z. Guo, D.L. Swaney, T.A. Tummino, R. Hüttenhain, R. M. Kaake, A.L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B.J. Polacco, H. Braberg, J.M. Fabius, M. Eckhardt, M. Soucheray, M.J. Bennett, M. Cakir, M.J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. Mac Kain, L. Miorin, E. Moreno, Z.Z.C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I.T. Kirby, J.E. Melnyk, J.S. Chiorba, K. Lou, S.A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C.J.P. Mathy, T. Perica, K.B. Pilla, S. J. Ganesan, D.J. Saltzberg, R. Rakesh, X. Liu, S.B. Rosenthal, L. Calviello, S. Venkataramanan, J. Liboy-Lugo, Y. Lin, X.P. Huang, Y.F. Liu, S.A. Wankowicz, M. Bohn, M. Safari, F.S. Ugur, C. Koh, N.S. Savar, Q.D. Tran, D. Shengjuler, S.

- J. Fletcher, M.C. O'Neal, Y. Cai, J.C.J. Chang, D.J. Broadhurst, S. Klippsten, P. P. Sharp, N.A. Wenzell, D. Kuzuoglu-Ozturk, H.Y. Wang, R. Trenker, J.M. Young, D. A. Caverio, J. Hiatt, T.L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R. M. Stroud, A.D. Frankel, O.S. Rosenberg, K.A. Verba, D.A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d'Enfert, S. Mukherjee, M. Jacobson, H.S. Malik, D.G. Fujimori, T. Ideker, C. S. Craik, S.N. Floor, J.S. Fraser, J.D. Gross, A. Sali, B.L. Roth, D. Ruggiero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. García-Sastre, K.M. Shokat, B. K. Shoichet, N.J. Krogan, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, *Nature*. 583 (2020), <https://doi.org/10.1038/s41586-020-2286-9>.
- [19] P. Ellis, F. Somogyvári, D.P. Virok, M. Nosedá, G.R. McLean, Decoding Covid-19 with the SARS-CoV-2 genome, *Curr. Genet. Med. Rep.* 9 (2021) 1–12, <https://doi.org/10.1007/s40142-020-00197-5>.
- [20] W.M. Chan, J.D. Ip, A.W.H. Chu, C.C.Y. Yip, L.S. Lo, K.H. Chan, A.C.K. Ng, R.W. S. Poon, W.K. To, O.T.Y. Tsang, W.S. Leung, M.Y.W. Kwan, G.T. Chua, T.W. H. Chung, I.F.N. Hung, K.H. Kok, V.C.C. Cheng, J.F.W. Chan, K.Y. Yuen, K.K.W. To, Identification of nsp1 gene as the target of SARS-CoV-2 real-time RT-PCR using nanopore whole-genome sequencing, *J. Med. Virol.* 92 (2020), <https://doi.org/10.1002/jmv.26140>.
- [21] D. Li, J. Zhang, J. Li, Primer design for quantitative real-time PCR for the emerging Coronavirus SARS-CoV-2, *Theranostics*. 10 (2020), <https://doi.org/10.7150/thno.47649>.
- [22] C.W. Dieffenbach, T.M.J. Lowe, G.S. Dvorkesler, General concepts for PCR primer design. General Concepts for PCR Primer Design PARAMETERS USED IN BASIC PCR PRIMER DESIGN, Cold Spring Harb. Lab. Press, 2008. May 8.
- [23] Z. Ma, P. Li, Y. Ji, A. Ikram, Q. Pan, Cross-reactivity towards SARS-CoV-2: the potential role of low-pathogenic human coronaviruses, *Lancet Microbe*. 1 (2020), e151, [https://doi.org/10.1016/S2666-5247\(20\)30098-7](https://doi.org/10.1016/S2666-5247(20)30098-7).
- [24] A. Lopez-Rincon, A. Tonda, L. Mendoza-Maldonado, D.G.J.C. Mulders, R. Molenkamp, C.A. Perez-Romero, E. Claassen, J. Garssen, A.D. Kraneveld, Classification and specific primer design for accurate detection of SARS-CoV-2 using deep learning, *Sci. Rep.* 11 (2021), <https://doi.org/10.1038/s41598-020-80363-5>.
- [25] A. Babiker, C.W. Myers, C.E. Hill, J. Guarnier, SARS-CoV-2 testing: trials and tribulations, *Am. J. Clin. Pathol.* 153 (2020) 706–708, <https://doi.org/10.1093/ajcp/aqaa052>.
- [26] C.C.Y. Yip, C.C. Ho, J.F.W. Chan, K.K.W. To, H.S.Y. Chan, S.C.Y. Wong, K.H. Leung, A.Y.F. Fung, A.C.K. Ng, Z. Zou, A.R. Tam, T.W.H. Chung, K.H. Chan, I.F.N. Hung, S.C.C. Cheng, O.T.Y. Tsang, S.K.W. Tsui, K.Y. Yuen, Development of a novel, genome subtraction-derived, sars-cov-2-specific covid-19-nsp2 real-time rt-pcr assay and its evaluation using clinical specimen, *Int. J. Mol. Sci.* 21 (2020) 2574, <https://doi.org/10.3390/ijms21072574>.
- [27] I. Arevalo-Rodriguez, D. Buitrago-Garcia, D. Simancas-Racines, P. Zambrano-Achig, R. Del Campo, A. Ciapponi, O. Sued, L. Martinez-Garcia, A.W. Rutjes, N. Low, P.M. Bossuyt, J.A. Perez-Molina, J. Zamora, False-negative results of initial RT-PCR assays for COVID-19: a systematic review, *PLoS One* 15 (2020), <https://doi.org/10.1371/journal.pone.0242958>.
- [28] T. Koyama, D. Platt, L. Parida, Variant analysis of SARS-cov-2 genomes, *Bull. World Health Organ.* 98 (2020), <https://doi.org/10.2471/BLT.20.253591>.
- [29] A. Fontanet, B. Autran, B. Lina, M.P. Kieny, S.S.A. Karim, D. Sridhar, SARS-CoV-2 variants and ending the COVID-19 pandemic, *Lancet*. 397 (2021) 952–954, [https://doi.org/10.1016/S0140-6736\(21\)00370-6](https://doi.org/10.1016/S0140-6736(21)00370-6).
- [30] S.S. Abdool Karim, T. de Oliveira, New SARS-CoV-2 variants — clinical, public health, and vaccine implications, *N. Engl. J. Med.* 384 (2021) 1866–1868, <https://doi.org/10.1056/nejmc2100362>.
- [31] Centers for Disease Control and Prevention (CDC), COVID-19, 2020.
- [32] S. Song, L. Ma, D. Zou, D. Tian, C. Li, J. Zhu, M. Chen, A. Wang, Y. Ma, M. Li, X. Teng, Y. Cui, G. Duan, M. Zhang, T. Jin, C. Shi, Z. Du, Y. Zhang, C. Liu, R. Li, J. Zeng, L. Hao, S. Jiang, H. Chen, D. Han, J. Xiao, Z. Zhang, W. Zhao, Y. Xue, Y. Bao, The global landscape of SARS-CoV-2 genomes, variants, and haplotypes in 2019nCoV, *Genom. Proteom. Bioinform.* (2020), <https://doi.org/10.1016/j.gpb.2020.09.001>.
- [33] Z. Gong, J.-W. Zhu, C.-P. Li, S. Jiang, L.-N. Ma, B.-X. Tang, D. Zou, M.-L. Chen, Y.-B. Sun, S.-H. Song, Z. Zhang, J.-F. Xiao, Y.-B. Xue, Y.-M. Bao, Z.-L. Du, W.-M. Zhao, An online coronavirus analysis platform from the National Genomics Data Center, *Zool. Res.* 41 (2020) 705–708, <https://doi.org/10.24272/j.issn.2095-8137.2020.065>.
- [34] W.M. Zhao, S.H. Song, M.L. Chen, D. Zou, L.N. Ma, Y.K. Ma, R.J. Li, L.L. Hao, C. P. Li, D.M. Tian, B.X. Tang, Y.Q. Wang, J.W. Zhu, H.X. Chen, Z. Zhang, Y.B. Xue, Y. M. Bao, The 2019 novel coronavirus resource, *Yi Chuan = Hered.* 42 (2020) 212–221, <https://doi.org/10.16288/j.yczx.20-030>.
- [35] S. Kwok, D.E. Kellogg, N. McKinney, D. Spasic, L. Goda, C. Levenson, J.J. Sninsky, Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies, *Nucleic Acids Res.* 18 (1990) 999–1005, <https://doi.org/10.1093/nar/18.4.999>.
- [36] S. Lefever, F. Pattyn, J. Hellemans, J. Vandessepele, Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays, *Clin. Chem.* 59 (2013) 1470–1480, <https://doi.org/10.1373/clinchem.2013.203653>.
- [37] C. Farkas, F. Fuentes-Villalobos, J.L. Garrido, J. Haigh, M.I. Barría, Insights on early mutational events in SARS-CoV-2 virus reveal founder effects across geographical regions, *PeerJ*. 2020 (2020), <https://doi.org/10.7717/peerj.9255>.
- [38] Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data – from vision to reality, *Eurosurveillance*. 22 (2017), <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>.
- [39] GISAID, Common Primer Check for High Quality Genomes. <https://www.epicov.org/epi3/frontend#lightbox-1432699498>, 2021.
- [40] K.A. Khan, P. Cheung, Presence of mismatches between diagnostic PCR assays and coronavirus SARS-CoV-2 genome, *R. Soc. Open Sci.* 7 (2020), <https://doi.org/10.1098/rsos.200636>.
- [41] K.S. Kuchinski, A.N. Jassem, N.A. Prystajek, Assessing oligonucleotide designs from early lab developed PCR diagnostic tests for SARS-CoV-2 using the PCR strainer pipeline, *J. Clin. Virol.* 131 (2020), <https://doi.org/10.1016/j.jcv.2020.104581>.
- [42] R. Wang, Y. Hozumi, C. Yin, G.W. Wei, Mutations on COVID-19 diagnostic targets, *Genomics*. 112 (2020), <https://doi.org/10.1016/j.ygeno.2020.09.028>.
- [43] M. Park, J. Won, B.Y. Choi, C.J. Lee, Optimization of primer sets and detection protocols for SARS-CoV-2 of coronavirus disease 2019 (COVID-19) using PCR and real-time PCR, *Exp. Mol. Med.* 52 (2020) 963–977, <https://doi.org/10.1038/s12276-020-0452-7>.
- [44] T. Toptan, S. Hoehl, S. Westhaus, D. Bojkova, A. Berger, B. Rotter, K. Hoffmeier, J. Cinatl, S. Ciesek, M. Widera, Optimized qRT-PCR approach for the detection of intra-and extra-cellular SARS-CoV-2 RNAs, *Int. J. Mol. Sci.* 21 (2020), <https://doi.org/10.3390/ijms21124396>.
- [45] J. Won, S. Lee, M. Park, T.Y. Kim, M.G. Park, B.Y. Choi, D. Kim, H. Chang, V. N. Kim, C.J. Lee, Development of a laboratory-safe and low-cost detection protocol for SARS-CoV-2 of the Coronavirus Disease 2019 (COVID-19), *Exp. Neurobiol.* 29 (2020), <https://doi.org/10.5607/en20009>.
- [46] R. Lu, X. Wu, Z. Wan, Y. Li, X. Jin, C. Zhang, A novel reverse transcription loop-mediated isothermal amplification method for rapid detection of sars-cov-2, *Int. J. Mol. Sci.* 21 (2020), E2826, <https://doi.org/10.3390/ijms21082826>.
- [47] K. Khan, P. Cheung, Evaluation of the sequence variability within the PCR Primer/Probe target regions of the SARS-CoV-2 genome, *Bio-Protocol*. 10 (2020), <https://doi.org/10.21769/bioprotoc.3871>.
- [48] Z. Gong, J.W. Zhu, C.P. Li, S. Jiang, L.N. Ma, B.X. Tang, D. Zou, M.L. Chen, Y. Bin Sun, S.H. Song, Z. Zhang, J.F. Xiao, Y.B. Xue, Y.M. Bao, Z.L. Du, W.M. Zhao, An online coronavirus analysis platform from the National Genomics Data Center, *Zool. Res.* 51 (2020) 705–708, <https://doi.org/10.24272/J.ISSN.2095-8137.2020.065>.
- [49] W.M. Zhao, S.H. Song, M.L. Chen, D. Zou, L.N. Ma, Y.K. Ma, R.J. Li, L.L. Hao, C. P. Li, D.M. Tian, B.X. Tang, Y.Q. Wang, J.W. Zhu, H.X. Chen, Z. Zhang, Y.B. Xue, Y. M. Bao, The 2019 novel coronavirus resource, *Yi Chuan = Hered.* 42 (2020), <https://doi.org/10.16288/j.yczx.20-030>.
- [50] W. Chang, J. Cheng, J. Allaire, Y. Xie, J. McPherson, shiny: Web Application Framework for R., *R Packag. Version 1.5.0*, 2020.
- [51] C. Kreer, M. Döring, N. Lehnen, M.S. Ercanoglu, L. Giesemann, D. Luca, K. Jain, P. Schommers, N. Pfeifer, F. Klein, openPrimer for multiplex amplification of highly diverse templates, *J. Immunol. Methods* 480 (2020), <https://doi.org/10.1016/j.jim.2020.112752>.
- [52] F. Wu, S. Zhao, B. Yu, Y.M. Chen, W. Wang, Z.G. Song, Y. Hu, Z.W. Tao, J.H. Tian, Y.Y. Pei, M.L. Yuan, Y.L. Zhang, F.H. Dai, Y. Liu, Q.M. Wang, J.J. Zheng, L. Xu, E. C. Holmes, Y.Z. Zhang, A new coronavirus associated with human respiratory disease in China, *Nature*. 579 (2020), <https://doi.org/10.1038/s41586-020-2008-3>.
- [53] R Core Team, R: A Language and Environment for Statistical Computing, 2013.
- [54] W. Chang, R.B. Borges, shinydashboard: Create Dashboards with 'Shiny', *R Packag. Version 0.7.1*. <https://cran.r-project.org/package=shinydashboard>, 2018.
- [55] H. Wickham, M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the tidyverse, *J. Open Source Softw.* 4 (2019), <https://doi.org/10.21105/joss.01686>.
- [56] R.A.M. Villanueva, Z.J. Chen, ggplot2: Elegant Graphics for Data Analysis 2nd ed, 17, 2019, <https://doi.org/10.1080/15366367.2019.1565254>. Meas. Interdiscip. Res. Perspect.
- [57] R. Benedikt, Z. Charlop-Powers, msAR: Multiple Sequence Alignment for R Shiny, 2017.
- [58] J.D. Fernandes, A.S. Hinrichs, H. Clawson, J.N. Gonzalez, B.T. Lee, L.R. Nassar, B. J. Raney, K.R. Rosenbloom, S. Nerli, A.A. Rao, D. Schmelzer, A. Fyfe, N. Maulding, A.S. Zweig, T.M. Lowe, M. Ares, R. Corbet-Detig, W.J. Kent, D. Haussler, M. Haeussler, The UCSC SARS-CoV-2 genome browser, *Nat. Genet.* 52 (2020), <https://doi.org/10.1038/s41588-020-0700-8>.
- [59] R. Stadhouders, S.D. Pas, J. Anber, J. Voermans, T.H.M. Mes, M. Schutten, The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay, *J. Mol. Diagn.* 12 (2010) 109–117, <https://doi.org/10.2353/jmoldx.2010.090035>.
- [60] J. Carneiro, F. Pereira, EbolaID: an online database of informative genomic regions for ebola identification and treatment, *PLoS Negl. Trop. Dis.* 10 (2016), e0004757, <https://doi.org/10.1371/journal.pntd.0004757>.
- [61] N. Mamoon, G. Rasskin, COVID-19 Visualizer, Retrieved from, www.Covidvisualizer.Com/, 2020.
- [62] World Health Organization (WHO), WHO Coronavirus Disease (COVID-19) Dashboard, Retrieved from, covid19.who.int, 2020.
- [63] J. Hadfield, C. Megill, S.M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R.A. Neher, NextStrain: Real-time tracking of pathogen evolution, *Bioinformatics*. 34 (2018), <https://doi.org/10.1093/bioinformatics/bty407>.
- [64] A. Jain, M. Rophina, S. Mahajan, B.B. Krishnan, M. Sharma, S. Mandal, T. Fernandez, S. Sultanji, B. Jolly, S. Mathew, S. Sivasubbu, V. Scaria, Analysis of the potential impact of genomic variants in global SARS-CoV-2 genomes on

molecular diagnostic assays, *Int. J. Infect. Dis.* 102 (2021), <https://doi.org/10.1016/j.ijid.2020.10.086>.