2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

SCOPE2: A Platform for Sars-COv-2 Primer covErage Evaluation

Fan Tong, Jiangyu Li, Wubin Qu, Wei Song and Dongsheng Zhao

***Abstract*—Currently, there is an increasing number and speed of SARS-CoV-2 mutation taking place around the world, posing a threat to promising public health and challenge to existing diagnostic tools. RT-PCR technology is recognized as the gold standard diagnosing methodology but has shown inaccuracy under some mutated SARS-CoV-2 circumstances. In this study, we developed a platform named SCOPE2 (Sars-COv-2 Primer covErage Evaluation) based on our previous publication. Testing by commonly-used SARS-COV-2 PCR primers, SCOPE2 is proved to effectively and efficiently assess the quality in terms of detection coverage, which may provide a practical tool for primer selection acceleration and primer design improvement.**

***Clinical Relevance*—This assists in single SARS-COV-2 Primer selection and suggestion of different SARS-COV-2 Primer combinations.**

I. Introduction

Multiple severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants are now circulating globally [1]. Researchers have cataloged more than 12,000 mutations in SARS-CoV-2 genomes, and the number increases at a minimum of 2 nucleotides (nt) per month [2]. The observed mutations in SARS-CoV-2 are not predicted to affect the utility of currently deployed vaccines; however, changes in the viral nucleic acid and protein sequences put at risk the utility of certain in vitro diagnostic assays if the mutation occurs in an area critical for a primer in RT-PCR assays [3]. For instance, Katharina Ziegler found two SNPs C28858T/C29200T can escape detection of Xpert Xpress SARS-CoV-2 (GXP) assay (Cepheid Inc., Sunnyvale, United States (US)) [4].

Under such circumstances, each RT-PCR primer needs re-evaluation once a new essential mutation occurs. There has been a push to make the RT-PCR testing process automated, which minimizes user error and improves assay reproducibility, making the test faster and cheaper to run [5]. Victor M Corman presents a validated diagnostic workflow for 2019-nCoV where assays were evaluated based on how well they matched to the 2019-nCoV genome released independently on GISAID [6]. Po-E Li developed a web-based assay validation algorithm that checks existing PCR-based assays against the ever-expanding genome databases for SARS-CoV-2 using both thermodynamic and edit-distance metrics [7].

These above-mentioned tools provide practical methodologies for SARS-CoV-2 primer coverage evaluation. Following these ideas, we developed a platform named SCOPE2 (Sars-COv-2 Primer CovErage Evaluation, http://124.207.243.56:8081/) based on our previous work [8]. Through aligning designated sequences to differential datasets, SCOPE2 calculates the coverage of targeted assay against virus genomes. An optimal one obeys the criteria of high SARS-CoV-2 genome coverage. Since the automatic analysis of the pipeline and continually updating of the genome, SCOPE2 can guide assay selection and assist primer design.

II. Methods and Materials

A. Background Data

GISAID (a global initiative on sharing avian flu data) is recognized as an effective and trusted mechanism for rapid sharing of both published and 'unpublished' influenza data [9]. 17,111 hCoV-19 sequences in this database were collected and retrieved as virus references for coverage calculation of targeted primers. Two widely-recognized and commonly-used primers 1) forward prime "GACCCCAAAATCAGCGAAAT" and reverse primer "TCTGGTTACTGCCAGTTGAATCTG" by US CDC (https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html) and 2) forward prime "CACATTGGCACCCGCAATC" and reverse primer "GAGGAACGAGAAGAGGCTTG" by WHO [10] are introduced as use cases to be evaluated.

B. Evaluation Algorithm

Based on the methodology implemented previously [8,11-12], we defined PCR primer coverage as the ratio of the count of reference sequences which can be mapped by PCR primers against the count of the reference sequence and further developed a pipeline applying to virus sequences dataset. A highly covered virus dataset means this PCR can cope with the different mutations and strains among the SARS-COV-2 genome. In the meantime, we additionally generated a distribution digraph of the existing hCoV-19 sequences in the database covered by the input primer around the world.

C. Integration Framework

SCOPE2 is constructed using Brower/Server structure (Figure. 1). Both the front-end and the back-end of the platform are mainly implemented by Go-lang, ensuring reliability, speed, and user-friendliness across multiple platforms [13]. The background reference sequences are stored locally as fasta-format files, whose indexes are organized in a key-value style and saved in a relational database. The geographic distribution digraph of the mapping hCoV-19 sequences is dynamically drawn by d3js tools [14].

Fan Tong, Jiangyu Li and Dongsheng Zhao is with the Academy of Military Medical Sciences, Beijing, China (corresponding author Prof. Dongsheng Zhao; e-mail: [tongfan, lijiangyu, dszhao]@bmi.ac.cn).

Wubin Qu is with iGeneTech Bioscience Co., Ltd, Beijing, China (corresponding author, e-mail: quwubin@gmail.com).

Wei Song is with the Beijing MedPeer Information Technology Co., Ltd., Beijing, China (e-mail: songwei@medpeer.cn).

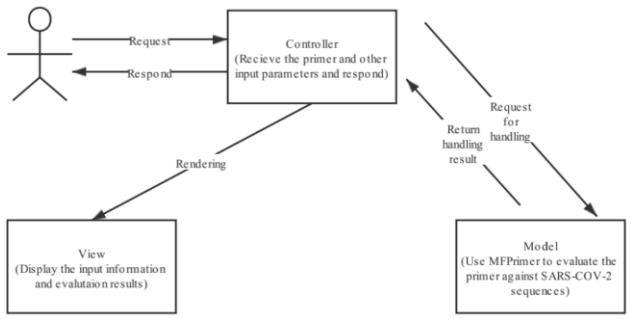

Figure 1. The MVC framework of SCOPE2.

III. Results and Discussion

A. User Interface

As shown in Figure. 2, at the very beginning, users need to submit the job by uploading targeted primer, selecting reference database, and verifying default parameters. Due to the performance limitation of our server, no more than 50 primer sequences in fasta format are recommended to be simultaneously received and respectively analyzed to calculate coverage against the latest virus sequence dataset (SARS-CoV-2 (17,111 sequences, update 2020-09-20)) according to users' intention. Other essential parameters, such as minimum melting temperature and product size range, have been assigned the recommended value as the default input to decrease the professional demand and simplify the complex operation. Once confirming all the above-mentioned information, users can finally press "RUN" and complete submission.

After a few seconds of waiting, users can get a detailed descriptive report of their primer (Figure. 3a-c). The output report consists of the following sections: 1) overall statistics of the input primers including sequence size, GC content, and coverage rate against reference database; 2) detailed description of potential amplicons including mapping sequence, sequence distribution, binding sites at each mapping sequence; 3) specific parameters of the submitted job including reference database, mismatch count and melting temperature as users assigned. Among all these results, the coverage rate against the reference database is the most important indicator showing the potential performance of the primer dealing with a wide variety of SARS-CoV-2 mutants.

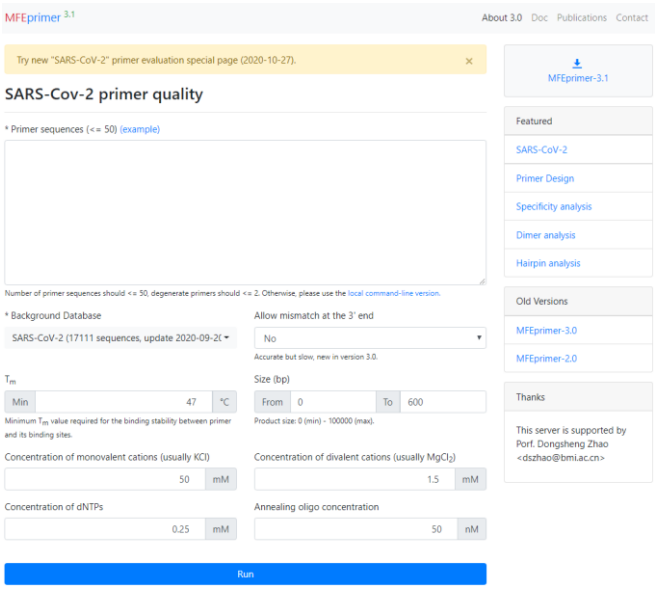

Figure 2. The job submission page of SCOPE2.

There are 17111 records in current SARS-CoV-2 database, this primer set misses 34 records (download failed records), the coverage rate is 99.80% (download all amplified records).

Primer list

*PBN and MBN: primer binding number to the plus and minus strand.

Primer ID	Sequence (5' → 3')	Size (bp)	GC (%)	T_m (°G)	ΔG (kcal/mol)	PBN*	MBN*
2019-nCoV_N1-F	GACCCCAAAATCAGCGAAAT	20	45.00	57.15	-20.50	17098	0
2019-nCoV_N1-R	TCTGGTTACTGCCAGTTGAATCTG	24	45.83	61.34	-24.61	17107	34193

Hairpin list (0)

No hairpins found.

Dimer list (0)

No dimers found.

Histogram of size of first 200 (17077 in total) potential amplicons

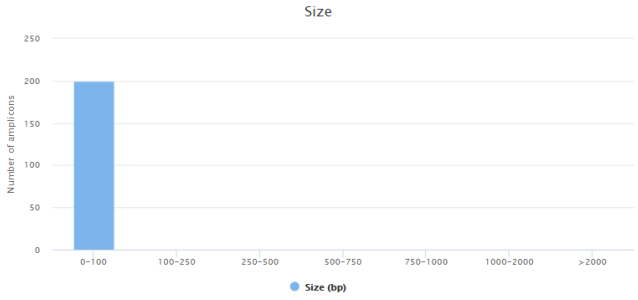

Figure 3a. The result presentation page of SCOPE2 - overall statistics.

Figure 3b. The result presentation page of SCOPE2 - detailed description.

Parameters

Database	SARS-CoV-2-20200927-17111.fasta
Kvalue	9
Mismatch	0
Product Min Size (bp)	0
Product Max Size (bp)	600
T_m (°C)	47.00
Concentration of monovalent cations (mM)	50.00
Concentration of divalent cations (mM)	1.50
Concentration of dNTPs (mM)	0.25
Concentration of annealing oligos (nM)	50.00

Figure 3c. The result presentation page of SCOPE2 - specific parameters.

B. *Using Performance*

We firstly took USCDC primers to evaluate how well it adapted to the majority of existing SARS-CoV-2, this primer set misses 34 records, the coverage rate is 99.80% among 17111 records in the current SARS-CoV-2 database. We subsequently closely examined the matching and missing sequences. As for the matching ones, the primers mainly bound to 28200-28400 loci on the SARS-CoV-2 sequences, suggesting SARS-CoV-2 virus may share some conserved region [15]; while for the missing ones, almost 80% (27/34) originated from Australia and America (Figure. 3), which may be in accordance with the T29095C lineages of mutant SARS-CoV-2 virus mostly discovered in these two countries [16].

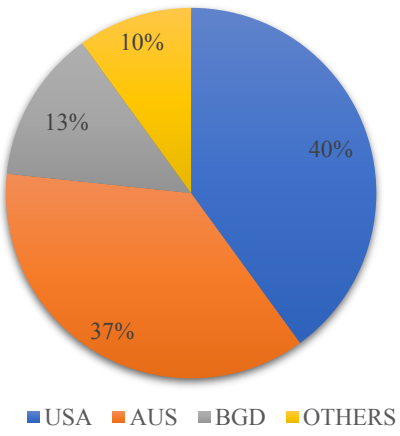

Figure 3. The distribution of uncovered SARS-CoV-2 sequences.

Multiple assays in a real-time RT-PCR SARS-CoV-2 panel can mitigate the risk of loss of sensitivity by new genomic variants during the COVID-19 outbreak [17]. To take a step further, we utilize another SARS-COV-2 primer released by WHO (calculated coverage 98.70% by SCOPE2) to explore whether it can match the missing samples by USCDC primers. 33 extra sequences were successfully mapped by WHO primers while only 1 subject from Iran was failed, proving a combination of primer can indeed improve the coverage from diverse sources, which constitutes a practical solution for SAR-COV-2 testing.

Currently, SCOPE2 is deployed and executed on a high-performance server equipped with 160 cores CPU and 504GB memory, making it possible to load index and map sequence fast and easily. For instance, dealing with the above-mentioned single pair of primers, it takes about 30 seconds to finish the SARS-CoV-2 dataset evaluation job. When the number of primers expands to the maximum limitation, 50, the running time maintain the same as single pair. Each pair of primer was automatically treated as a single submitting job and parallelly calculated at different computing threads in the meantime. As long as the resources are adequate, analysts can get their expected results in an acceptable time.

C. *Downstream Analysis*

After obtaining the primer coverage evaluation result, we can exert downstream analysis to further explore the performance of this primer. For instance, coverage metrics can be extended with the introduction of demographic and virological characteristics. The virus sequence geological distribution is not a random or isolated event that may be associated with the population. [18] The collected data, although retried and updated from all over the world, is not enough to reflect the real world. If the demographic features are included, such as the number of stationary and mobile people, we can exert more weight to the location holding more population and give more attention to these more critical places. Similarly, virological innate features are also important to describe the true COVID-19 epidemic distribution. The phylogeny of SARS-COV-2 can lead us to discover the development of existing SARS-COV-2 and explore the explosion of recent mutants [19]. With the help of this information, we can put more emphasis on the more influential and severe strains.

What's more, specificity metrics can be calculated following the idea of coverage calculation. To be more specific, the combination of the coverage of the primer to SARS-CoV-2 dataset and the irrelevance of this primer to other virus datasets (such as SARS-CoV and MERS-CoV or human influenza viruses including H1N1 and H7N9) can measure the specificity of that primer [20]. Although specificity has been taken into account during the initial primer design process, the virus is evolving at a high speed and rate as we mentioned previously. There is no means for us to be able and affordable to update the overall latest virus sequences dataset. As a result, specificity re-evaluation is highly recommended following the coverage evaluation. High coverage of the SARS-CoV-2 dataset and low coverage of other virus datasets may draw a deduction that the input primer has an acceptable specificity.

D. Transverse Comparison

SCOPE2 may outperform the other existing PCR assessment tools ([6] and [7]) in the following aspects: 1) flexibility, user can complete coverage evaluation of their PCR primers based on their parameters instead of open-access PCR primers based on fixed parameters; 2) scalability, user can execute coverage evaluation to no more than 50 primers in parallel mode instead of once primer at a time; 3) generality, user can take advantage of a well-developed and generally-acknowledged coverage evaluation algorithm enduring long-term utilization and optimization instead of a recently proposed methodology which may still need further validation and has improvement potential.

IV. Conclusion

In this study, we developed and deployed a SARS-CoV-2 primer coverage evaluation platform called SCOPE2 to cope with the question of whether the current PCR primer can maintain validity against the ever-changing SARS-CoV-2 sequences. By testing it using common SARS-CoV-2 primers, we found it has high coverage towards existing SARS-CoV-2 sequences, providing potential evidence that this primer may become a potential candidate for SARS-CoV-2 diagnosis. In addition, be it by single or in parallel, the evaluation can finish in an acceptable time, making SCOPE2 an effective and efficient software for the analyst to accelerate the process and improve the accuracy of SARS-CoV-2 primer evaluation.

Acknowledgment

This work is supported by National Key R&D Program of China 'Disease Research Precision Medical Knowledge Base Construction'.

References

[1] Abdool Karim, S. S., de Oliveira, T., & Loots, G. (2021). Appropriate names for COVID-19 variants. *Science (New York, N.Y.)*, *371*(6535), 1215.

[2] Callaway E. (2020). The coronavirus is mutating - does it matter? *Nature*, *585*(7824), 174–177.

[3] Ascoli C. A. (2021). Could mutations of SARS-CoV-2 suppress diagnostic detection? *Nature biotechnology*, *39*(3), 274–275.

[4] Ziegler K, Steininger P, Ziegler R, Steinmann J, Korn K, Ensser A. SARS-CoV-2 samples may escape detection because of a single point mutation in the N gene. Euro Surveill. 2020 Oct;25(39):2001650.

[5] Ravi, N., Cortade, D. L., Ng, E., & Wang, S. X. (2020). Diagnostics for SARS-CoV-2 detection: A comprehensive review of the FDA-EUA COVID-19 testing landscape. *Biosensors & bioelectronics*, *165*, 112454.

[6] Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M. L., Mulders, D. G., Haagmans, B. L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J. L., Ellis, J., Zambon, M., Peiris, M., … Drosten, C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, *25*(3), 2000045.

[7] Li, P. E., Myers Y Gutiérrez, A., Davenport, K., Flynn, M., Hu, B., Lo, C. C., Player Jackson, E., Shakya, M., Xu, Y., Gans, J., & Chain, P. (2020). A Public Website for the Automated Assessment and Validation of SARS-CoV-2 Diagnostic PCR Assays. *Bioinformatics (Oxford, England)*, btaa710. Advance online publication.

[8] Wang, K., Li, H., Xu, Y., Shao, Q., Yi, J., Wang, R., Cai, W., Hang, X., Zhang, C., Cai, H., & Qu, W. (2019). MFEprimer-3.0: quality control for PCR primers. *Nucleic acids research*, *47*(W1), W610–W613.

[9] Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro surveillance: bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, *22*(13), 30494.

[10] Lieberman, J. A., Pepper, G., Naccache, S. N., Huang, M. L., Jerome, K. R., & Greninger, A. L. (2020). Comparison of Commercially Available and Laboratory-Developed Assays for *In Vitro* Detection of SARS-CoV-2 in Clinical Laboratories. *Journal of clinical microbiology*, *58*(8), e00821-20.

[11] Qu, W., Zhou, Y., Zhang, Y., Lu, Y., Wang, X., Zhao, D., Yang, Y., & Zhang, C. (2012). MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. *Nucleic acids research*, *40*(Web Server issue), W205–W208.

[12] von Ahsen, N., Wittwer, C. T., & Schütz, E. (2011). Monovalent and divalent salt correction algorithms for Tm prediction--recommendations for Primer3 usage. *Briefings in bioinformatics*, *12*(5), 514–517.

[13] Lampa, S., Dahlö, M., Alvarsson, J., & Spjuth, O. (2019). SciPipe: A workflow library for agile development of complex and dynamic bioinformatics pipelines. *GigaScience*, *8*(5), giz044.

[14] Mukhyala, K., & Masselot, A. (2014). Visualization of protein sequence features using JavaScript and SVG with pViz.js. *Bioinformatics (Oxford, England)*, *30*(23), 3408–3409.

[15] Rangan, R., Zheludev, I. N., Hagey, R. J., Pham, E. A., Wayment-Steele, H. K., Glenn, J. S., & Das, R. (2020). RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *RNA (New York, N.Y.)*, *26*(8), 937–959.

[16] Forster, P., Forster, L., Renfrew, C., & Forster, M. (2020). Phylogenetic network analysis of SARS-CoV-2 genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(17), 9241–9243.

[17] Peñarrubia, L., Ruiz, M., Porco, R., Rao, S. N., Juanola-Falgarona, M., Manissero, D., López-Fontanals, M., & Pareja, J. (2020). Multiple assays in a real-time RT-PCR SARS-CoV-2 panel can mitigate the risk of loss of sensitivity by new genomic variants during the COVID-19 outbreak. *International journal of infectious diseases: IJID : official publication of the International Society for Infectious Diseases*, *97*, 225–229.

[18] Webb Hooper, M., Nápoles, A. M., & Pérez-Stable, E. J. (2020). COVID-19 and Racial/Ethnic Disparities. *JAMA*, *323*(24), 2466–2467.

[19] Hu, B., Guo, H., Zhou, P., & Shi, Z. L. (2021). Characteristics of SARS-CoV-2 and COVID-19. *Nature reviews. Microbiology*, *19*(3), 141–154.

[20] Dinnes, J., Deeks, J. J., Adriano, A., Berhane, S., Davenport, C., Dittrich, S., Emperador, D., Takwoingi, Y., Cunningham, J., Beese, S., Dretzke, J., Ferrante di Ruffano, L., Harris, I. M., Price, M. J., Taylor-Phillips, S., Hooft, L., Leeflang, M. M., Spijker, R., Van den Bruel, A., & Cochrane COVID-19 Diagnostic Test Accuracy Group (2020). Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *The Cochrane database of systematic reviews*, *8*, CD013705.