DNASCRIPT

APPLICATION NOTE

SARS-CoV-2 Amplicon Sequencing

SARS-CoV-2 sequencing data have been used in the development of diagnostic assays, vaccines, and therapeutics and to support biosurveillance efforts. In this Application Note we demonstrate the use of primers produced with DNA Script's novel Enzymatic DNA Synthesis (EDS) technology for whole-genome amplicon sequencing and phylogenetic analysis of SARS-CoV-2 isolates using the ARTIC network protocol. EDS primers performed comparably to primers sourced from commercial suppliers. EDS enables broad-based, rapid access to synthetic DNA, thereby accelerating the pace of iteration and innovation in the life sciences.

INTRODUCTION

The COVID-19 pandemic precipitated one of the most concentrated scientific efforts ever focused on a single pathogen. Worldwide, academic, government, and commercial institutions pivoted from their usual endeavors to study the epidemiology, genomics, biochemistry, and evolution of the SARS-CoV-2 virus and develop diagnostic tests, treatments, and vaccines. One unforeseen consequence of these efforts was a global bottleneck in the supply of synthetic DNA from suppliers relying on highly centralized phosphoramidite-based production and third-party logistics.

DNA Script has developed a novel enzymatic method to synthesize oligos (Figure 1), which uses a highly engineered terminal nucleotidyl transferase (TdT) enzyme, reversibly-terminated nucleotides, and a solid support.[1] This technology enables decentralized, same-day, on-demand DNA oligo synthesis using a benchtop printer in a standard laboratory environment, without the need for specialized technical skills.

To demonstrate functionally equivalent performance in genomics and life science applications, we synthesized EDS oligos for loop-mediated isothermal amplification (LAMP, Figure 2) and NGS assays for the detection and characterization of the SARS-CoV-2 virus. In this study, we report on the performance of EDS primers in the ARTIC network's nCoV-19 amplicon sequencing protocol.[2]

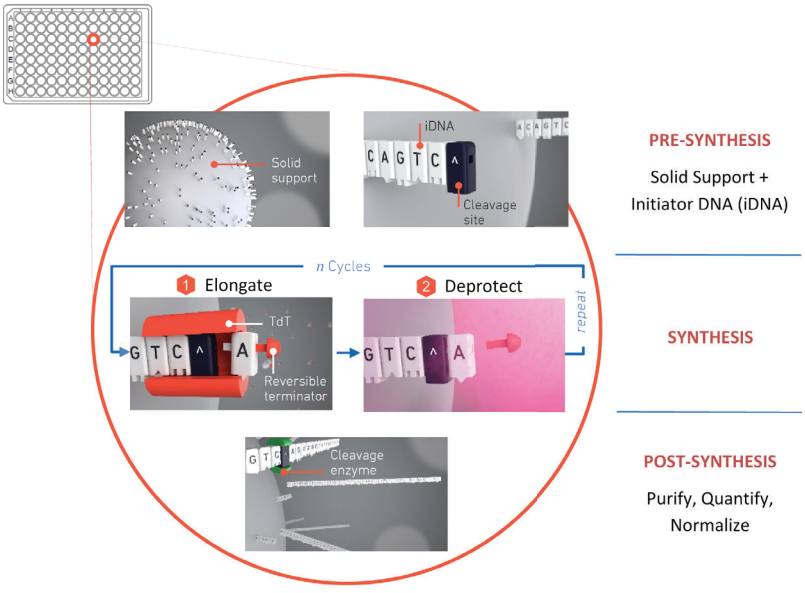

FIGURE 1. OVERVIEW OF THE EDS PROCESS. DNA oligos are synthesized in the "natural" 5'- to 3'-direction in a cyclic, two-step process. During the first step, **elongation**, a single nucleotide is added to the iDNA in each well of the synthesis plate. In the second step, **deprotection**, reversible terminators are removed, leaving each strand ready for the next base addition. Steps 1 and 2 of the process are repeated until the longest oligo on the plate is completed. Desired sequences are enzymatically cleaved from the iDNA, desalted, quantified and purified to yield molecular biology-ready oligos.

The Jackson Laboratory

Laboratorio de MEDICINA MOLECULAR

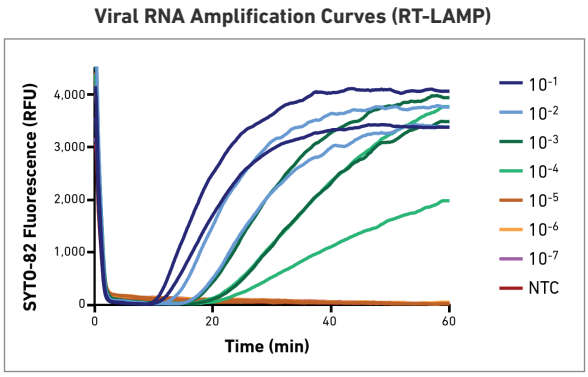

FIGURE 2. REAL-TIME DETECTION OF THE SARS-COV-2 VIRUS WITH EDS PRIMERS.

Amplification curves were generated from a 10-fold dilution series of viral RNA, using a real-time loop-mediated isothermal amplification (RT-LAMP) assay· designed for low-cost, point-of-care SARS-CoV-2 testing.[3]

Data courtesy of Patrick Tabeling, École supérieure de physique et de chimie industrielles de la Ville de Paris.

MATERIALS AND METHODS

Synthesis of the ARTIC V3 Primer Panel

The ARTIC V3 primer panel comprises 218 primers, split into two pools to cover the 30 kb SARS-CoV-2 genome in overlapping amplicons.[2] In V3, the original panel (V1, 98 amplicons) is supplemented with 11 "alternate" primer pairs to improve coverage of challenging regions.[4] The V3 panel was synthesized and normalized using DNA Script's SYNTAX System (EDS primers). Since EDS generates primers with a 5'-phosphate, primers were dephosphorylated prior to formulating the two pools. The same panel, synthesized using conventional phosphoramidite chemistry, was obtained from two commercial suppliers (CS1 and CS2).

Synthetic RNA Controls

Two synthetic SARS-CoV-2 RNA controls (corresponding to the Wuhan and Australian strains MN908947.3 and MT007544.1, respectively) were obtained from a commercial supplier. Each control consists of six non-overlapping ~5 kb fragments, covering 99.9% of the SARS-CoV-2 genome.[5]

Clinical Isolates

The five clinical samples analyzed in this study (Table 1) were collected as part of an epidemiological research study conducted in Mexico.[6] All patients tested positive using a RT-qPCR assay. Four of the five individuals had traveled to other countries before testing positive.

Amplicon Library Preparation, Sequencing, and Data Analysis

Amplicons were generated and libraries constructed as described in the ARTIC protocol for Illumina sequencing (V.2).[7] For synthetic RNA controls, an equivalent of 10,000 genome copies were used in the reverse transcription (RT) reaction. Multiplexed, paired-end sequencing (2 x 150 bp) was performed using an Illumina HiSeq instrument.

TABLE 1. DETAILS OF CLINICAL ISOLATES.

Sample name	RT-qPCR Cq value	RIN	Country visited
Mx 18.5	18.48	2.5	Italy
Mx 24.3	24.27	N/A	None
Mx 25.6	25.62	2.4	Dominican Republic
Mx 28.6	28.57	3.9	USA
Mx 30.9	30.86	N/A	France

Sample collection, RNA extraction, and the RT-qPCR assay are described elsewhere.[6] RNA Integrity Number (RIN) was determined with a Bioanalyzer 2100 instrument and RNA 6000 Pico Kit (Agilent Technologies). N/A indicates that a RIN score could not be determined.

For the analysis of clinical isolates, 5 µL of RNA extracted from clinical swabs were used in each RT reaction. Amplicons were generated using 25 cycles of PCR. Libraries were prepared with the KAPA HyperPrep Kit (Roche) following the manufacturer's recommended protocol with six cycles of library amplification. Multiplexed, paired-end sequencing (2 x 300 bp) was performed on an Illumina MiSeq instrument, generating 1 – 2 million reads per sample.

Data analysis was performed with the Geneious Prime software using BBtools[8] plugins. Consensus sequences were compiled using the following parameters: minimum coverage = 100, minimum variant frequency = 0.8, and minimum P-value = 10^{-6}. Phylogenetic analysis was performed with the Nextstrain command-line interface,[9] and results were visualized on the Auspice platform.[10]

RESULTS AND DISCUSSION

Amplification of Synthetic RNA Controls

As an initial proof of concept, amplicons were generated from synthetic SARS-CoV-2 RNA controls. Sequencing data generated with the EDS primers and two sets of commercial primers (CS1 and CS2) were compared with respect to coverage uniformity (Figure 3), and the detection of known variants between the Wuhan and Australian strains (Table 2). Results suggested no significant difference in the performance of primers produced by EDS vs. conventional chemistry.

Analysis of Clinical Isolates

In the second part of the study, five clinical samples from a Mexican epidemiological cohort were processed with the ARTIC V3 primer panel, synthesized by EDS and a commercial supplier (CS1). As shown in Figure 4, similar coverage uniformity was obtained with the EDS and CS1 panels for the four samples with the highest viral load (RT-qPCR Cq values <30). For the Mx 30.9 sample, coverage was much more variable within and between the EDS and CS1 panels. This was attributed to the low viral titer and sample integrity rather than primer quality.

Consensus sequences for the Mx 18.5, Mx 24.3, Mx 25.6, and Mx 28.6 isolates were used for variant calling (Figure 5 and Table 3) and phylogenetic analysis (Figure 6). Identical variant calls were obtained with the EDS and CS1 panels for the four isolates with RT-qPCR Cq values <30. Variant calls for Mx 30.9 differed in two positions and were deemed to be unreliable due to the low viral titer and experimental variation.

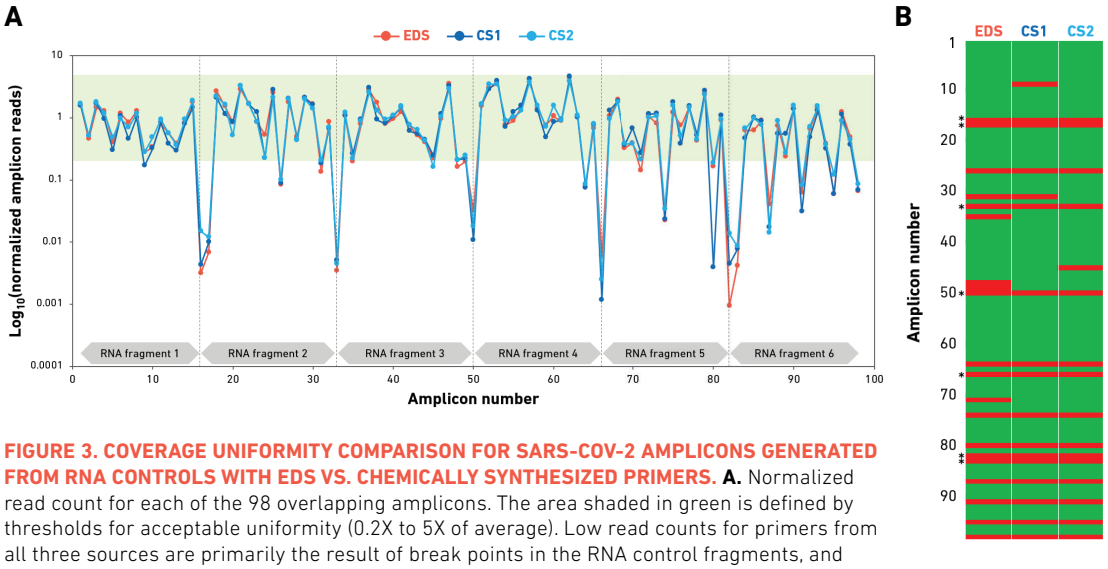

FIGURE 3. COVERAGE UNIFORMITY COMPARISON FOR SARS-COV-2 AMPLICONS GENERATED FROM RNA CONTROLS WITH EDS VS. CHEMICALLY SYNTHESIZED PRIMERS. A. Normalized read count for each of the 98 overlapping amplicons. The area shaded in green is defined by thresholds for acceptable uniformity (0.2X to 5X of average). Low read counts for primers from all three sources are primarily the result of break points in the RNA control fragments, and impact coverage of amplicons 16, 17, 33, 50, 66 and 82 – 83. **B.** Qualitative representation of coverage data. Amplicons covered within the 0.2X – 5X range are indicated in green, under-represented amplicons (<0.2X) in red, and over-represented amplicons (>5X) in cyan (none present). Coverage for amplicons marked with an * is impacted by the break points in the synthetic RNA.

TABLE 2. VARIANT CALLING SUMMARY FOR SYNTHETIC RNA CONTROLS.

Variant (coordinates)	Frequency (%)			Coverage (X)		
	EDS	CS1	CS2	EDS	CS1	CS2
T > C (19,065)	99.8	99.9	99.9	41,262	38,029	38,642
T > G (22,303)	99.8	99.6	99.8	31,335	32,992	39,369
G > T (26,144)	99.6	99.6	99.8	35,260	28,994	27,975
ΔCGATCGAGTG (29,750 – 29,759)	~99.4	98.2	~99.5	~2,536	2,200	~3,100

Consensus sequences generated from the Wuhan and Australian synthetic RNA controls were aligned against the Wuhan reference sequence (NBCI genome NC_045512). The Wuhan control (MN908947.3) sequence was identical to the reference. The four variants listed above (described in Genbank ID MT007544.1) were detected in the Australian control. Values preceded by ~ are the average of a range.

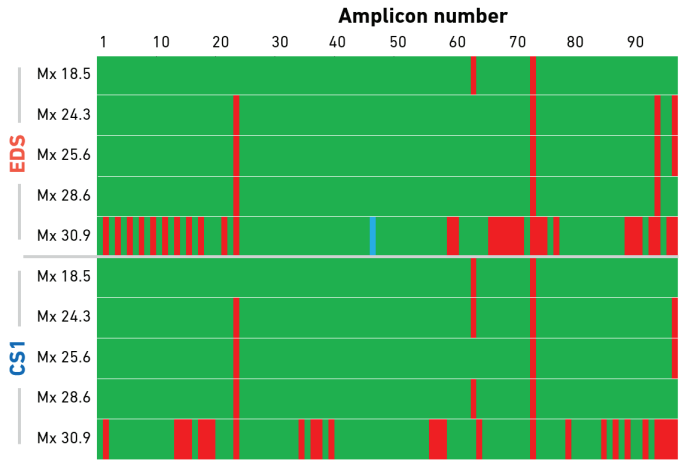

FIGURE 4. COVERAGE UNIFORMITY COMPARISON FOR CLINICAL SAMPLES.

Amplicons covered within a 0.2- to 5-fold range of average coverage are indicated in green, under-represented amplicons (<0.2X average) in red, and over-represented amplicons (>5X average) in cyan. EDS primers slightly outperformed CS1 primers for the four samples with RT-qPCR Cq values <30. Results were more variable for the Mx 30.9 sample, presumably due to poor RNA quality.

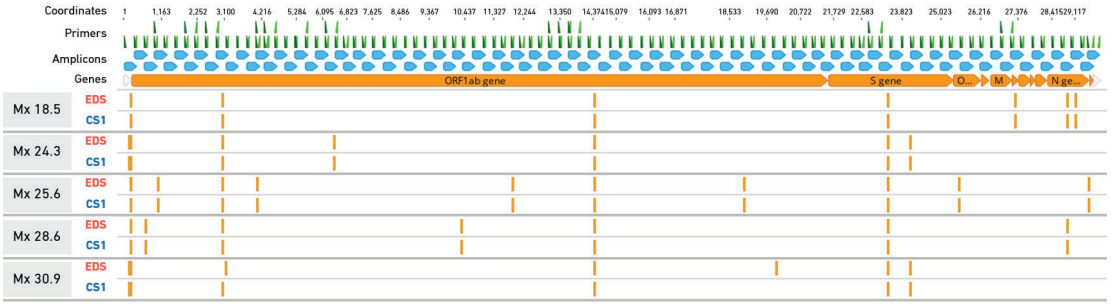

FIGURE 5. VARIANT CALLING COMPARISON FOR THE FIVE CLINICAL ISOLATES.

The consensus sequence for each sample, generated with either EDS or CS1 primers, was aligned against the Wuhan reference sequence (NCBI genome NC_045512), and is shown in order of decreasing viral load. Variant calls were identical for the first four samples. Detailed frequency and coverage statistics are given in Table 3. As shown in Figure 4, amplicon drop-out was observed with primers from both sources for the sample with the lowest viral titer (Mx 30.9), resulting in lower-confidence variant calls.

TABLE 3. VARIANT CALLING SUMMARY FOR CLINICAL ISOLATES.

Variant (coordinates)	Frequency (%)		Coverage (X)	
	EDS	CS1	EDS	CS1
Mx 18.5				
C > T (241)	99.8	99.9	17,433	22,546
C > T (3,037)	99.8	99.7	12,062	15,723
C > T (14,408)	99.9	99.9	25,106	25,097
A > G (23,403)	99.9	99.9	29,475	28,000
T > C (27,299)	99.8	99.7	31,419	37,177
GGG > AAC (28,881 – 28,883)	99.7	99.8	~4,093	~7,811
T > C (29,148)	99.8	99.9	23,965	29,638
MX25.6				
C > T (241)	100	100	23,488	27,608
C > T (1,059)	100	100	18,362	17,346
C > T (3,037)	99.9	99.9	9,734	12,265
C > T (4,113)	100	100	44,397	40,199
C > T (11,916)	99.6	100	47,610	50,237
C > T (14,408)	100	100	23,257	20,649
C > T (18,998)	99.9	100	20,597	20,790
A > G (23,403)	100	100	36,423	31,990
G > T (25,563)	100	99.9	25,918	27,139
G > A (29,540)	100	99.9	9,426	6,861

Variant (coordinates)	Frequency (%)		Coverage (X)	
	EDS	CS1	EDS	CS1
Mx 24.3				
A > G (187)	100	100	20,535	24,274
C > T (241)	99.8	99.9	20,515	24,259
C > T (3,037)	99.8	99.5	11,172	13,488
G > T (6,446)	99.8	99.9	32,190	29,065
C > T (14,408)	99.8	99.9	27,779	25,631
A > G (23,403)	99.9	99.9	37,610	33,923
T > C (24,076)	99.9	99.9	26,748	23,054
Mx 28.6				
C > T (241)	100	100	16,661	25,925
ΔAAGTCATTT (686 – 694)	~99.8	~99.7	~27,661	~27,047
C > T (3,037)	99.9	99.9	9,489	8,700
G > A (10,360)	99.9	100	10,200	16,976
C > T (14,408)	100	100	21,107	20,566
A > G (23,403)	100	95.3	19,509	29,369
GGG > AAC (28,881 – 28,883)	99.9	99.9	2,251	~6,071

Values preceded by ~ are the average of a range.

Mx 18.5, Mx 24.3, Mx 25.6, and Mx 28.6 consensus sequences were included in a phylogenetic analysis of 192 SARS-CoV-2 isolates collected from different geographical locations between December 2019 and June 2020. Of these, 146 sequences were retrieved from the Nextstrain ncov database,[11] whereas 46 additional sequences for Mexican isolates were obtained from the NCBI database.[12] As shown in Figure 6, the four sequences from this study clustered together in clade 20A, the predominant clade based on the genomes from the Nextstrain database. Mx 18.5 and Mx 28.6 clustered more closely together (clade 20B), while Mx 25.6 was the most divergent (clade 20C). The four Mx isolates did not cluster with sequences/clades from countries visited by the patients prior to testing positive in Mexico. Further analysis of these strains is ongoing.

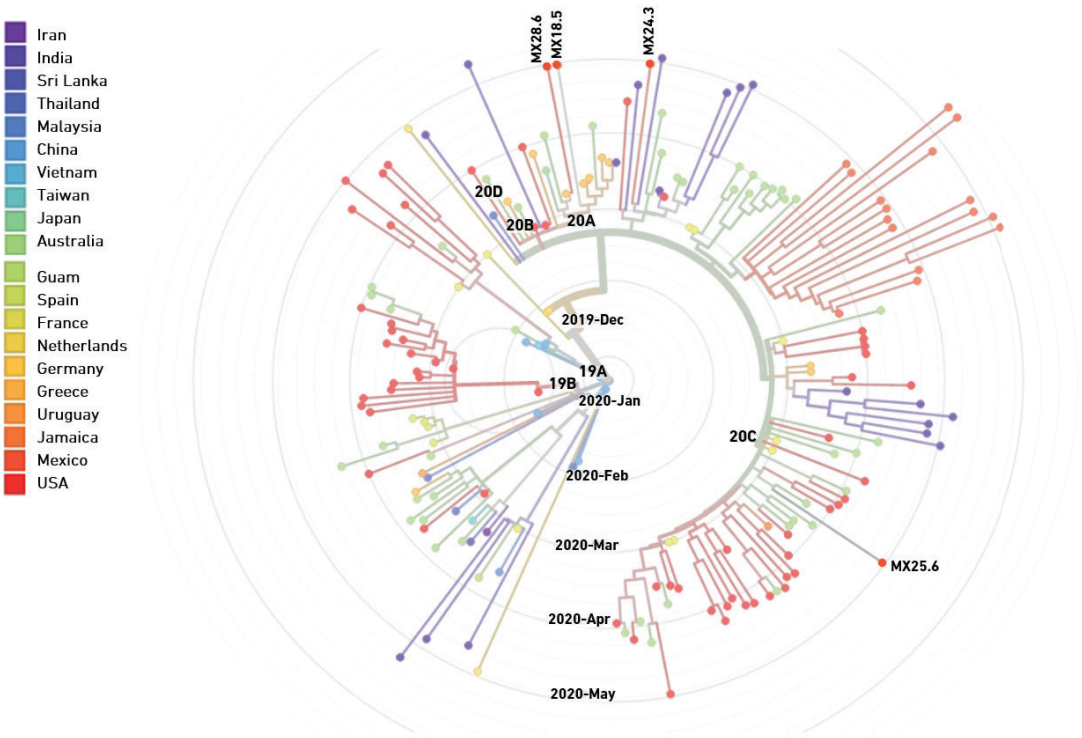

FIGURE 6. PHYLOGENETIC ANALYSIS OF THE FOUR MEXICAN ISOLATES WITH RT-QPCR CQ VALUES <30.

Between December 2019 and June 2020, 192 SARS-CoV-2 sequences were collected from different geographical locations. Three of the four isolates with RT-qPCR values <30 clustered closely together. None of the isolates clustered with sequences/clades from countries to which patients had traveled prior to testing positive.

CONCLUSION

In this study, we have demonstrated the use of DNA Script's novel EDS technology to generate primer panels for amplicon sequencing applications.

The ARTIC nCoV-19 V3 primer panel produced by EDS performed comparably to primers synthesized by phosphoramidite chemistry, when used both with synthetic RNA templates and clinical samples. Minor differences in the coverage of individual amplicons with primers from different sources had no impact on variant calling when sample quality was sufficient. Low-confidence variant calls were obtained for a clinical isolate with an RT-qPCR Cq value >30 with both EDS and commercial primers. Amplicon sequencing with EDS primers yielded high-quality consensus sequences for use in phylogenetic analysis.

With DNA Script's SYNTAX System, same-day, in-house production of DNA oligos is possible in any molecular biology laboratory. This capability enables rapid iteration, higher productivity, and predictable schedules to facilitate the development of molecular assays (whether for research or diagnostics), tools, and reagents for functional studies[13] and ultimately therapies and vaccines. In addition, it offers full control over sequence information and independence from third-party vendors, thereby lowering the risk of delays or contamination.

This study is one of several demonstrating that oligos produced by EDS support a wide range of molecular biology and genomics applications, with comparable functional performance to oligos sourced from commercial suppliers.[14]

REFERENCES

1. DNA Script. Enzymatic DNA Synthesis Technical Note (TN900100). https://www.dnascript.com/wp-content/uploads/2021/04/DNAScript_EDS_3Pillars_TN900100.pdf.
2. https://artic.network/ncov-2019.
3. Garneret P. et al. PLoS ONE 2021; 16(1):e0243712. doi: 10.1371/journal.pone.0243712.
4. https://artic.network/resources/ncov/ncov-amplicon-v3.pdf
5. https://www.twistbioscience.com/products/ngs/synthetic-viral-controls?tab=sars-cov-2-controls.
6. Martinez-Fierro ML, et al. Am. J. Infect. Control 2021; 49:15-20; doi: 10.1016/j.ajic.2020.10.002.
7. https://www.protocols.io/view/covid-19-artic-v3-illumina-library-construction-an-be3wjgpe?version_warning=no
8. https://jgi.doe.gov/data-and-tools/bbtools/.
9. https://github.com/nextstrain
10. https://auspice.us
11. https://github.com/nextstrain/ncov
12. https://www.ncbi.nlm.nih.gov/genome/
13. DNA Script. SARS CoV-2 Spike Protein Mutagenesis Application Note (ML1135). https://www.dnascript.com/wp-content/uploads/2021/10/DNAScript_SARS-CoV2-Spike-Protein-Mutagenesis_ML1135.pdf
14. https://www.dnascript.com/applications/

ACKNOWLEDGMENTS

This study was performed in collaboration with Prof. Chia-Lin Wei, Dr. ChewYee Ngan, and Rahul Maurya from Genome Technologies at The Jackson Laboratory for Genomic Medicine. Clinical samples and expertise were provided by Margarita L. Martinez-Fierro and Prof. Idalia Garza-Veloz from the Molecular Medicine Laboratory, Academic Unit of Human Medicine and Health Sciences, Autonomous University of Zacatecas, and Alba SH Centeno-Ramirez from the Public Health Laboratory, Zacatecas State Health Services.

Financé par GOUVERNEMENT Liberté Égalité Fraternité — France Relance — Financé par l'Union européenne NextGenerationEU

contact@DNAScript.com
www.DNAScript.com

US Headquarters
DNA Script, Inc.
650 Gateway Blvd.
South San Francisco, CA 94080
+1 (650) 457-0844

Europe Headquarters
DNA Script, SAS
67 avenue de Fontainebleau
94270, Le Kremlin-Bicêtre, France
+33 1 56 20 56 00

ML1134. Version 1.1.
Dec 2022.