DNASCRIPT

Enzymatic DNA Synthesis (EDS) enables decentralized and same-day access to DNA for critical study, developments and diagnosis to fight SARS-CoV-2

Benoit Derrien[1], Chew Yee Ngan[2], Rahul Maurya[2], Sophie Romero[1], Nadège Tardieu[1], Henri Lachaize[1], Maëllys Kevin[1], Maryke Appel[1], Margarita L Martinez-Fierro[3], Idalia Garza-Veloz[3], Alba SH Centeno-Ramirez[4], Xavier Godron[1] and Chia-Lin Wei[2]

[1]DNA Script, [2]Genome Technologies at The Jackson Laboratory for Genomic Medicine, [3]Molecular Medicine Laboratory, Academic Unit of Human Medicine and Health Sciences, Autonomous University of Zacatecas, [4]Public Health Laboratory, Zacatecas State Health Services

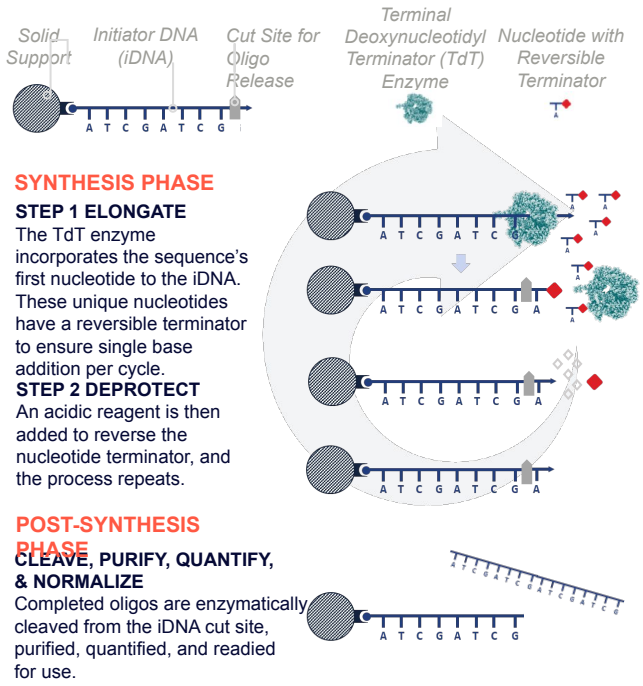

Figure 1. Overview of EDS process.
The cyclic 3-step synthesis process is performed on a solid support. Unlike conventional phosphoramidite chemistry, oligos are synthesized in the 5'→ 3' direction and are delivered by default with a 5'-phosphate. Reversible terminators have a simple blocking group that does not leave scars on the DNA after deprotection.

INTRODUCTION

The 2019 - 2021 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 this effort was a global bottleneck in the supply of synthetic DNA, which currently relies on highly centralized phosphoramidite-based production and third-party logistics.

We have previously reported on a novel enzymatic DNA synthesis (EDS) method (Figure 1), which utilizes a highly engineered TdT enzyme, reversibly-terminated nucleotides and a solid support. This technology enables decentralized, same-day, on-demand DNA oligo production with a benchtop "printer" – in a standard laboratory environment, requiring no specialized technical skills. To demonstrate that the technology is mature enough to support genomics and life science applications, we have synthesized DNA oligos for LAMP (Figure 2), NGS and FISH 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 hCoV-19 amplicon sequencing protocol (https://artic.network/ncov-2019).

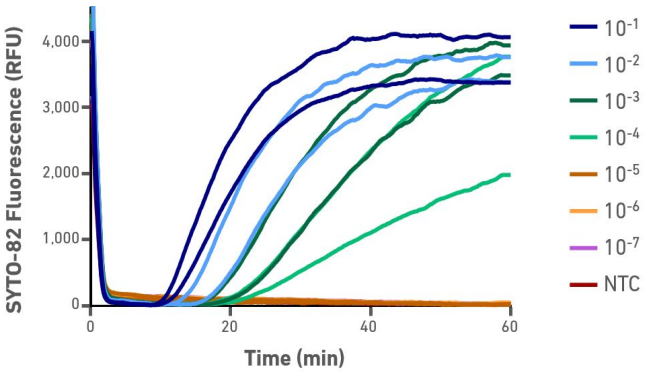

Figure 2. Real-time detection of SARS-CoV-2, from a dilution series (10^{-1} to 10^{-7}) of viral RNA. Data was generated in an RT-LAMP assay performed with primers produced by EDS.

Data courtesy of Patrick Tabeling, ESPCI PSL.

DNASCRIPT

The Jackson Laboratory

Laboratorio de MEDICINA MOLECULAR

Amplification of Synthetic RNA Controls

- The ARTIC v3 primer set (218 primers, designed to cover the 30-kb SARS-CoV-2 genome in 98 overlapping amplicons) was synthesized on DNA Script's SYNTAX™ EDS System (**EDS** primers), or ordered from commercial suppliers (**CS1** and **CS2** primers).
- 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 provider. Each control consists of six non-overlapping ~5 kb fragments covering 99.9% of the SARS-CoV-2 genome.
- Amplicons were generated with the EDS, CS1 and CS2 primers according to the ARTIC v3 protocol (10,000 genome copies per reaction). Libraries were prepared and sequenced according to the standard protocol.
- Data analysis was performed with Geneious Prime software using BBtools plugins (https://jgi.doe.gov/data-and-tools/bbtools/).
- Data generated with primers from different sources were compared with respect to coverage uniformity (Figure 3), and the detection of known variants (Table 1).

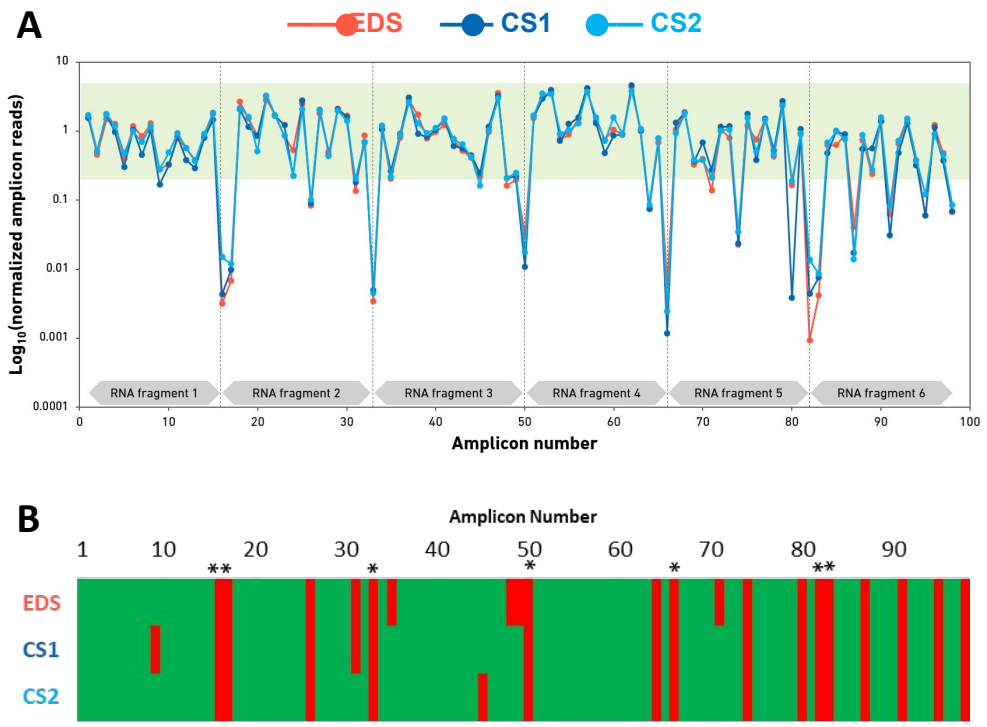

Figure 3. Coverage uniformity comparison for SARS-CoV-2 amplicons generated with EDS vs. chemically synthesized primers. **A.** Normalized read count for each of the 98 amplicons. The area shaded in green is defined by thresholds for acceptable uniformity (0.2X to 5X of average). Significant dips in uniformity with primers from all three sources correspond to break points in the RNA control fragments, and impact coverage of amplicons 16, 17, 33, 50, 66 and 82 – 83. In **B.** coverage data is represented more qualitatively. 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**. Coverage for amplicons marked with an * is impacted by the break points in the synthetic RNA.

Table 1. Variant calling summary.

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

Amplicon 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 sequence. The four variants listed above (described in Genbank ID MT007544.1) were detected in the Australian control. Values marked with an ~ are the average of a range.

Results suggest no significant difference in the performance of primers produced by EDS vs. conventional phosphoramidite chemistry.

DNASCRIPT

Amplification of Clinical Isolates for Epidemiological Research

The Jackson Laboratory

Laboratorio de MEDICINA MOLECULAR

- Five clinical RNA isolates were processed with the ARTIC v3 panel, using primers were produced either by **EDS** or chemical synthesis (**CS1**). Samples (**Table 2**) originated from a cohort of Mexican patients who tested positive using the CDC RT-qPCR assay. Four of the five individuals travelled before testing positive.
- Libraries were constructed following the recommended ARTIC protocol. Sequencing (2 x 300 bp) was performed on an Illumina® MiSeq® instrument (1– 2 million reads per library). Data were analysed as described above.

Table 2. Details of clinical isolates.

Sample Name	RT-qPCR Cq value	RIN score	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 in Martinez-Fierro ML, et al. Am. J. Infect. Control 2021; 49:15-20; doi: 10.1016/j.ajic.2020.10.002.

- Identical variant calls were obtained with the EDS and CS1 panels for the four isolates with qPCR Cq values <30 (**Figure 5 and Table 3**). Variant calls for Mx 30.9 differed slightly and were deemed to be unreliable due to the low viral titer and experimental variation.

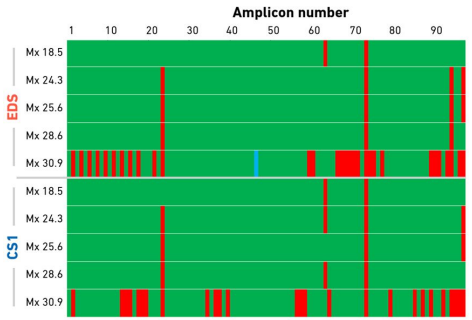

Figure 4. Coverage uniformity comparison for clinical samples. Amplicons covered within the 0.2X – 5X range are indicated in **green**, under-represented amplicons (<0.2X) in **red** and over-represented amplicons (>5X) in **cyan**.

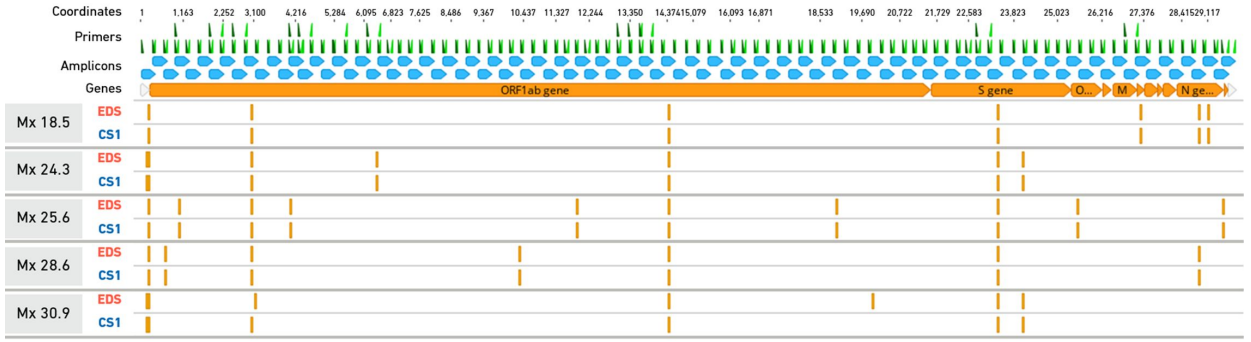

Figure 5. Variant calling analysis for the five clinical isolates. The consensus sequence for each sample, generated with either EDS or CS1 primers, was aligned against the Wuhan reference sequence (GenBank ID 908947.3); and is shown in order of decreasing viral load. Variant calls were identical for 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 titre (Mx 30.9), resulting in lower-confidence variant calls.

Table 3. Variant calling summary for clinical samples.

Variant (coordinates)	Frequency (%)		Coverage (X)	
	EDS	CS1	EDS	CS1
Mx18.5				
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

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

Variant (coordinates)	Frequency (%)		Coverage (X)	
	EDS	CS1	EDS	CS1
Mx 25.6				
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 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 marked with an ~ are the average of a range.

DNASCRIPT

Amplification of Clinical Isolates for Epidemiological Research

The Jackson Laboratory

Laboratorio de MEDICINA MOLECULAR

- For the four samples with the highest viral load, coverage uniformity (**Figure 4, previous slide**) was highly comparable for the EDS and CS1 panels. 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 quality, rather than primer quality.
- Consensus sequences for the Mx 18.5, Mx 24.3, Mx 25.6 and Mx 28.6 isolates were included in a phylogenetic analysis of SARS-CoV-2 strains collected between December 2019 and June 2020 (Figure 6). No obvious correlation with isolates from countries to which patients had travelled was observed. Further analysis of these strains are ongoing.

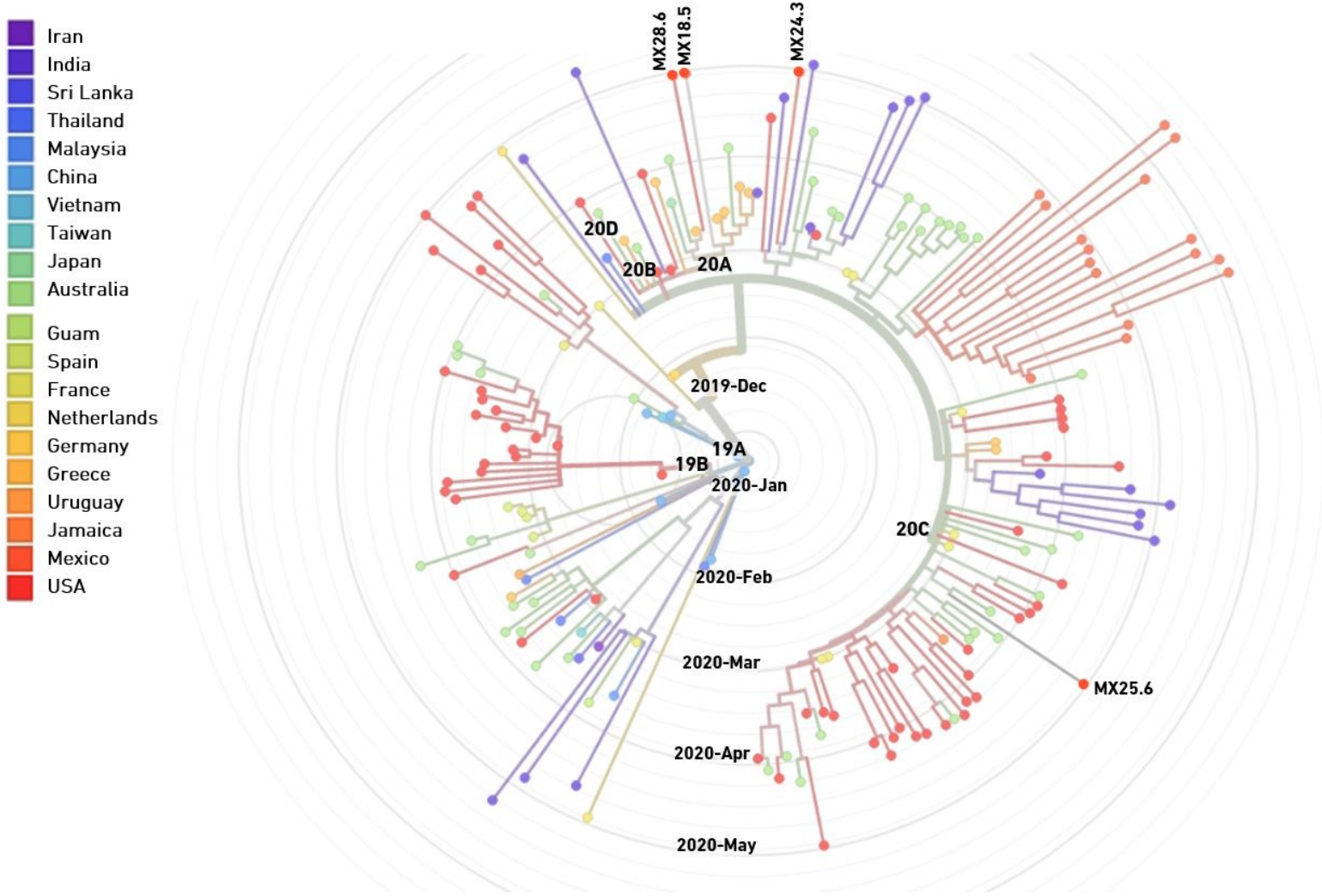

Figure 6. Phylogenetic analysis of the four Mexican isolates with RT-qPCR Cq values <30, in the context of 192 SARS-CoV-2 sequences collected from different geographical locations between December 2019 and June 2020. Of these, 146 sequences were retrieved from the ncov Nextstrain database (https://github.com/nextstrain/ncov), whereas 46 sequences for other Mexican isolates (n = 46) were obtained from the NBCI database (https://www.ncbi.nlm.nih.gov/genome/). The four sequences from this study clustered together in clade 20A, which is 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). Analysis was performed with Nextstrain command-line interface (https://github.com/nextstrain) and results were visualized with Auspice (https://auspice.us).

DNASCRIPT

The Jackson Laboratory

Laboratorio de MEDICINA MOLECULAR

CONCLUSIONS

- Primers for the ARTIC hCoV-19 amplicon sequencing protocol, synthesized using novel enzymatic DNA synthesis (EDS) technology, performed comparably to primers obtained from commercial suppliers, with both RNA controls and clinical samples.
- Minor differences in the coverage of individual amplicons when using EDS primers vs. those synthesized by conventional phosphoramidite chemistry
  has no impact on variant calling when sample quality is sufficient. Low-confidence variant calls were obtained for a clinical isolate with an RT-qPCR Cq value >30 with primers from both sources.
- EDS technology enables decentralized oligo synthesis to support rapid iteration when developing molecular assays. In addition, it offers full control over sequence information and independence from third-party vendors and logistics.
- This study demonstrates that the technology support genomics and molecular biology applications.

APPLY FOR OUR GRANT PROGRAM

The SYNTAX System, the world's first benchtop enzymatic DNA printer, is commercially available for order.
contact@DNAScript.com

JOIN OUR COMMUNITY