Multiplexed target fragment analysis for detection of viral pathogens, including SARS-CoV-2

In this application note, we show:

- Strategies for designing fragment analysis–based target multiplexing solutions

- How fragment analysis by capillary electrophoresis (CE) can provide a sensitive method for detecting pathogens

- A simple method for detecting SARS-CoV-2 viral sequences
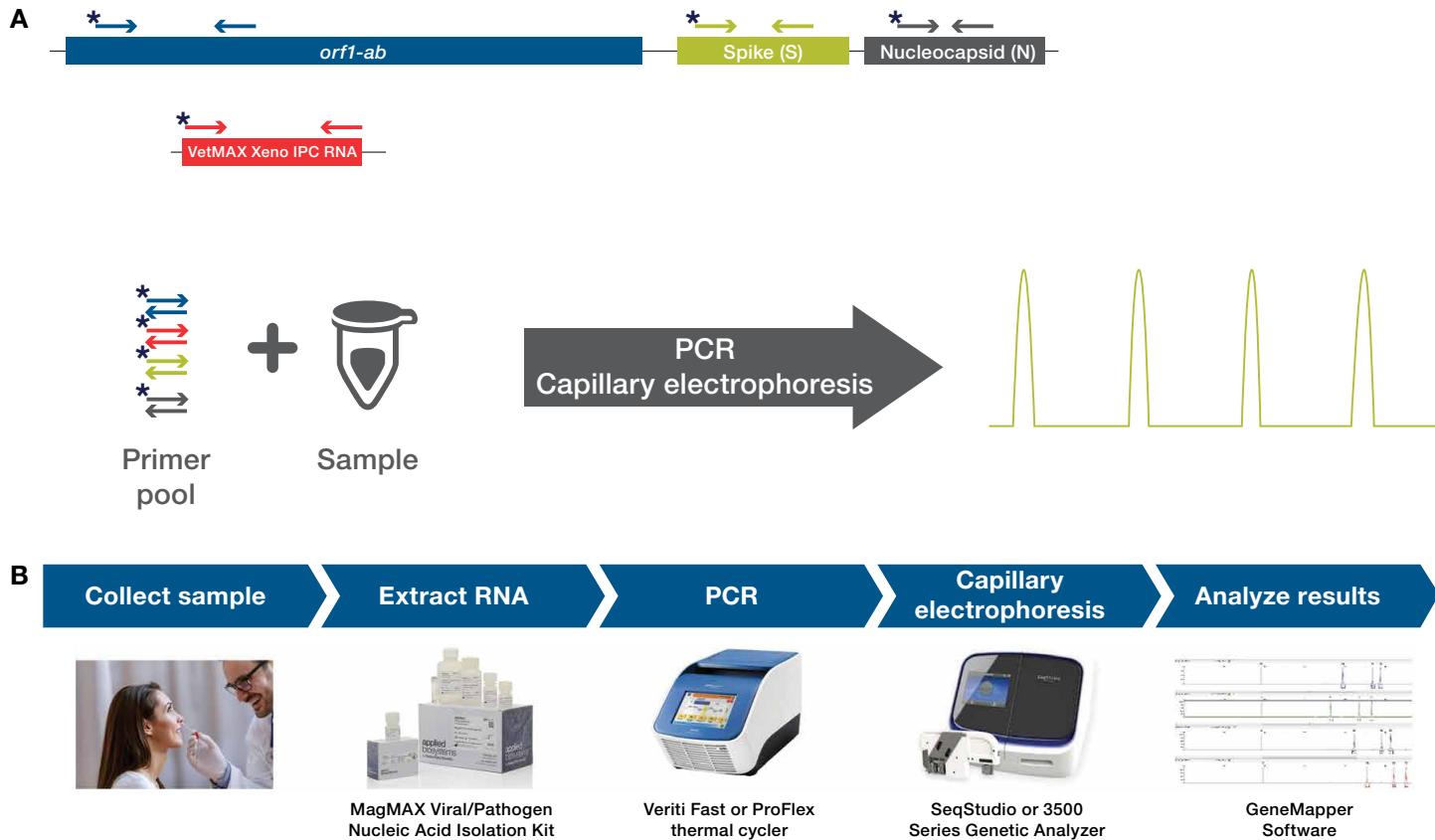
Introduction

Infectious diseases can present significant global problems. The recent SARS-CoV-2 crisis has disrupted economic and personal activities, and continues to pose a serious risk to certain populations, including the elderly and those with underlying conditions. It is therefore critical to rapidly study individuals who might be infected so that appropriate control and containment procedures can be determined.

A wide variety of genetic analysis techniques can be used to detect pathogenic organisms, including viruses. Next-generation sequencing (NGS) is ideal for discovering new organisms and new variants of known strains. However, NGS technologies require sophisticated sample preparation and data analysis protocols. As a result, they are more suited for discovery and epidemiological studies. On the other hand, quantitative real-time PCR (qPCR) methods utilize simplified workflows that yield clear results when identity of the pathogens needs to be confirmed.

Multiplexed qPCR solutions have been developed that screen for small numbers of relevant pathogens (for example, see Ramanan et al. [1]). Nevertheless, the relatively small capacity of multiplexed qPCR reactions can limit throughput when large numbers of target sequences or pathogens need to be detected. A rapid, simple, and sensitive method is needed in which multiple targets can be processed and analyzed at the same time.

Fragment analysis by CE provides an attractive method for rapid analysis of multiple targets (Figure 1A). For target multiplexing, PCR amplicons can be designed such that the length and 5′-terminal fluorophore of the amplicon are unique to a given target. A single PCR reaction can be set up with a pool of fluorescently labeled primer sets, and the resulting differently sized PCR fragments can be separated and detected by CE. High resolution of target fragment lengths is achieved in CE by running a size standard in the same injection, ensuring confidence in results.

Here we describe a simple workflow in which multiple amplicons from a pathogenic virus can be analyzed using fragment analysis (Figure 1B). To illustrate this approach, we analyze targets derived from SARS-CoV-2. We show how positive-control synthetic DNA sequences can be used to define and resolve the fragment sizes of the target amplicons. We illustrate the sensitivity of the method using known quantities of the SARS-CoV-2 RNA genome. For details on the protocol, see "**Detection of RNA from SARS-CoV-2 using fragment analysis**" [2].

Figure 1. Workflow for a fragment analysis–based pathogen detection solution. (A) Primers are designed to target regions of pathogens and a spike-in control (e.g., Applied Biosystems™ VetMAX™ Xeno™ Internal Positive Control (IPC) RNA). As an example, we show the primer pairs for SARS-CoV-2. For each primer pair, the forward primer is labeled with a fluorophore (*). The primers are pooled and mixed with a sample, which is then amplified and analyzed by capillary electrophoresis. The targets are distinguished by the different sizes of the amplicons. **(B)** Thermo Fisher Scientific has all the tools needed to perform the analysis. Once the sample is collected, data can be obtained in as little as 2 hours.

Description of method

The analysis of multiple amplicons using CE-based fragment analysis is a widely used strategy in many different applications, from human identification to research on infectious and inherited diseases to oncology research. Briefly, for this application, after identifying suitable target sequences, amplicons should be designed to be 150–500 nucleotides long. Each amplicon should have a unique length, differing from other amplicons by at least 5 nucleotides. The forward primer should be labeled at the 5´ end with the fluorescent dye, and the reverse primer should be unlabeled. Reverse transcription polymerase chain reactions (RT-PCR) are then set up using a system that allows for cDNA synthesis followed by endpoint PCR in a single tube (e.g., Applied Biosystems™ TaqMan® Fast Virus 1-Step Master Mix). The nucleic acid samples are subjected to PCR and the resulting fragments separated by CE. The appearance or absence of pathogen-specific fragments of the expected sizes determines the presence or absence of the pathogen in the sample (Table 1). The height of the peaks can be a semi-quantitative indicator of the abundance of the targets in a sample. The principles and applications of fluorescent DNA fragment analysis using high-resolution CE are explained in our **Fragment Analysis User Guide** [3].

Table 1. Suggestion for interpreting peaks arising from analysis of SARS-CoV-2 sequences by fragment analysis. Similar guidelines have been used to define detection in a qPCR-based SARS-CoV-2 test. Parameters for other detection assays may be defined according to the specific testing needs and characteristics of the pathogens of interest.

Xeno RNA control peak	SARS-CoV-2 RNA peaks	Interpretation of results
Present	None	**Negative**—no SARS-CoV-2 RNA detected
Present	3 peaks	**Positive**–SARS-CoV-2 RNA detected
Present	2 peaks	**Positive**–SARS-CoV-2 RNA detected
Present	1 peak	**Indeterminate**—retest the same purified sample
No peak	No peak	**Invalid**—no SARS-CoV-2 RNA detected; retest the same purified sample

Here we illustrate this method using SARS-CoV-2 as an example. For redundancy, we designed primers that target three different regions of the viral genome: the nucleocapsid protein (N) gene, the spike protein (S) gene, and *orf1-ab*. Examples of the primers that were used in this study to amplify SARS-CoV-2 sequences are shown in Figure 2. For development of the method, we synthesized the target amplicons (see ref. 2). In addition, we obtained purified viral RNA from commercial sources (BEI). As a control for cDNA synthesis and detection, we used the VetMAX Xeno IPC RNA. RT-PCR was performed using samples with viral nucleic acid sequences mixed with the pool of primers (final concentration 80 pM each), TaqMan Fast Virus 1-Step Master Mix, 0.5 µL VetMAX Xeno IPC RNA, and 1 ng human genomic DNA (optional to help prevent stochastic loss of small amounts of target) in a 10 µL reaction. Amplicons were generated on either an Applied Biosystems™ Veriti™ Fast or ProFlex™ thermal cycler and analyzed on either a SeqStudio™ or 3500 Genetic Analyzer. Results were analyzed using Applied Biosystems™ GeneMapper™ Software. For more details on the method, see ref. 2.

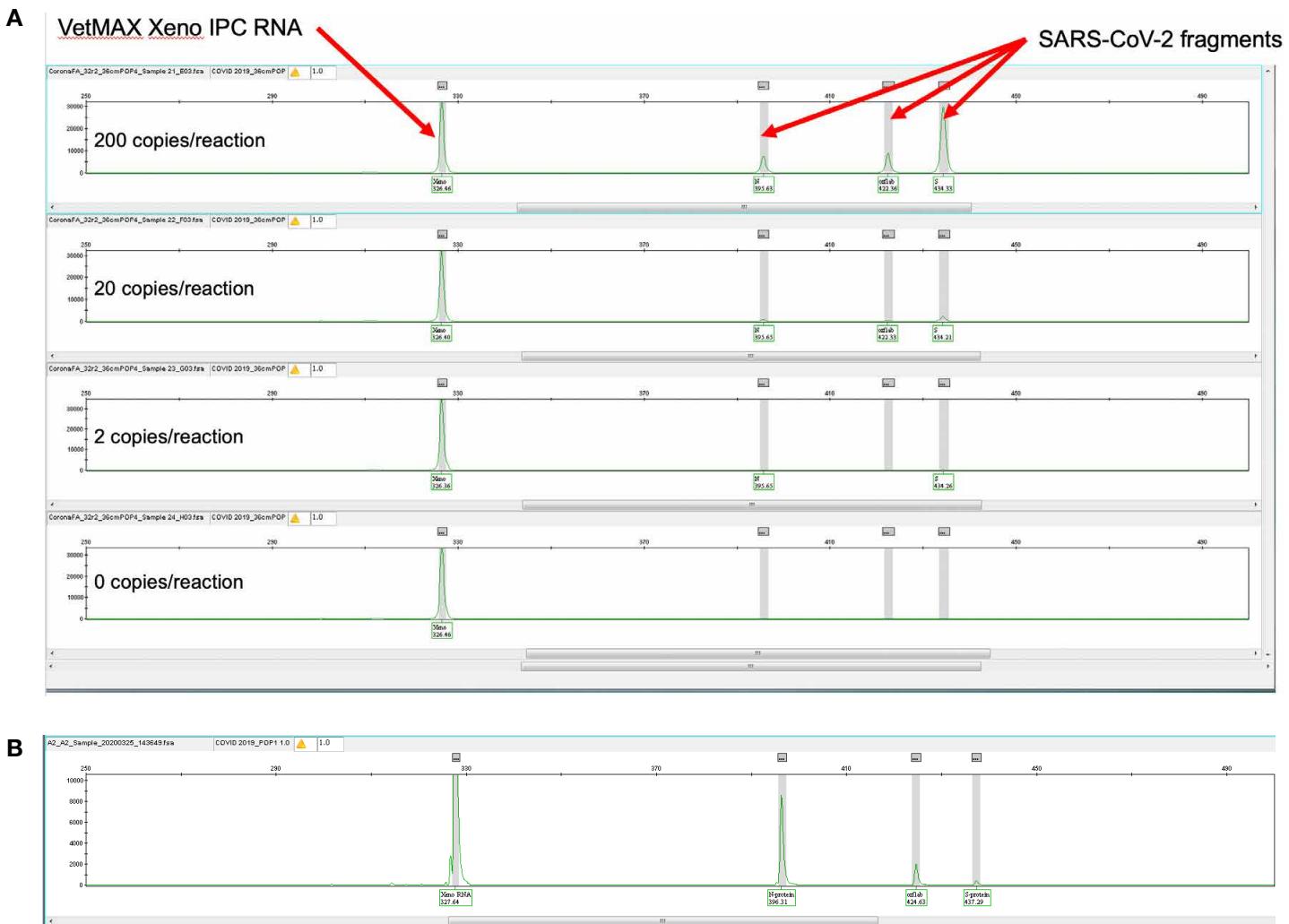| Primer name	Sequence
Polyprotein *orf1-ab* forward	5´-TGCCTGGAATATTGGTGAA-3´
Polyprotein *orf1-ab* reverse	5´-ACAATTTCACAAGCACAGGTTGAG-3´
S gene forward	5´-TCGAAGACCCAGTCCCTACTTATT-3´
S gene reverse	5´-CTGAAGAAGAATCACCAGGAGTCAA-3´
N gene forward	5´-GGACCAGGAACTAATCAGACAAGGA-3´
N gene reverse	5´-TTAGGCCTGAGTTGAGTCAGC-3´
VetMAX Xeno RNA control forward	5´-GCTGACTCCAGTGGTGAAAC-3´
VetMAX Xeno RNA control reverse	5´-ACCCTTGCTAGTAGGTGTAGATTCTC-3´

Figure 2. Primer sequences used to detect SARS-CoV-2 in this study. Forward primers were labeled with the VIC™ fluorophore, and the reverse primers were unlabeled. Oligos can be ordered desalted and dried.

Results

Demonstration of method

We illustrate how to develop and characterize this method by utilizing noninfectious synthetic DNA fragments and purified RNA genomes from SARS-CoV-2. Initially, we analyzed different dilutions of a pool of DNA fragments and a synthetic RNA control. In the presence of the viral RNA sequences, three distinct SARS-CoV-2 amplicons were generated (Figure 3). As expected, the level of fluorescence decreases with decreasing amounts of template input, with one fragment (*orf1-ab*) dropping below detectable levels at 2 copies per reaction. Two controls verified the successful outcome of the assay: the no-template control (NTC) did not produce any peaks in the range of SARS-CoV-2 fragment sizes, and the VetMAX Xeno RNA control was reverse-transcribed

and detected in all samples. Other results (not shown) have demonstrated that detecting the VetMAX Xeno RNA control depends on reverse transcriptase activity in the 1-step RT master mix. Thus, this combination of primers can amplify SARS-CoV-2 fragments in a multiplex reaction. Since multiple pathogen-specific fragments may be detectable in an assay, it is necessary to define what constitutes a pathogen-positive sample versus a pathogen-negative sample early in the design process. For the SARS-CoV-2 test, we utilized the guidelines set forth in Table 1. Similar rules can be defined for any other pathogen test, depending on the number of sequences chosen, the pathogen being analyzed, and the needs of the researchers.

Figure 3. Electropherogram traces of analysis of SARS-CoV-2 target concentrations. (A) Synthetic DNA targets corresponding to the nucleocapsid protein (N), spike protein (S), and *orf1-ab* genes, along with a spiked-in VetMAX Xeno RNA control, were amplified using fluorescent primers and separated by capillary electrophoresis. Note the clear size separation of the chosen sequences and the presence of the amplicon from the VetMAX Xeno RNA control in the same capillary. The numbers in the boxes indicate the measured size of the peak in nucleotides. **(B)** Viral RNA analyzed at 50 copies/reaction. Note that the amplicon of the N gene transcript is the most abundant, followed by those of *orf1-ab* and the S gene transcript. The y-axis was magnified to focus on the SARS-CoV-2 peaks.

In silico inclusivity/exclusivity

To help ensure optimal performance of an assay, candidate primer sequences should be chosen to maximize detection of target organisms/strains and minimize detection of off-target sequences. Typically, this is done by performing a BLAST search of candidate primer sequences against sequences deposited in databases. To illustrate an example, we performed a BLAST search of the NCBI Genbank viral database and complete bacterial and fungal genome database using the forward and reverse primers for SARS-CoV-2. Some of the sequence homologies are shown in Table 2. For simplicity, we show only a portion of all the results returned but illustrate the highest homologies that were detected. Two of the reverse primers showed

high homology with other pathogens, while the forward primer for the N gene had high homology with other coronavirus sequences. Importantly, because there was no homology with the corresponding reverse primer, no off-target amplification product would be expected. In other cases, there was lower homology for both primers of a pair. However, the PCR annealing conditions are such that any homologies less than 80% should not bind, so no off-target amplification product would be expected. And even if the organism was present and anomalous priming did occur, it is highly unlikely that the 425 nt amplicon would be produced by the *orf1-ab* primers.

Table 2. *In silico* analysis of potential cross-reactivity with other pathogens. Results of a BLAST search of the NCBI Genbank viral database and complete genomes of selected bacteria and fungi using the primers designed for SARS-CoV-2. Only the highest homologies are shown. Note that although some primers have >80% homology, the partner primer is either not homologous or has low enough homology to be predicted to not bind in the PCR reaction. FP = forward primer, RP = reverse primer.

Organism	Strain	Genbank accession No.	Assay	% Homology test FP	% Homology test RP
Bat coronavirus	16B0133	KY938558.1	Nprot_r	95	0
Bat SARS coronavirus HKU3-1	HKU3-1	DQ022305.2	Nprot_r	95	0
Bat SARS coronavirus HKU3-12	HKU3-12	GQ153547.1	Nprot_r	95	0
Bat SARS coronavirus HKU3-2	NA	DQ084199.1	Nprot_r	95	0
Bat SARS coronavirus HKU3-3	NA	DQ084200.1	Nprot_r	95	0
Bat SARS-like coronavirus	NA	KF294457.1	Orf1ab_r	0	96
Neisseria animalis	strain = *Neisseria meningitidis*	OQEK01000203.1	Orf1ab_r	0	96
Neisseria canis	strain = *Neisseria meningitidis*	OAEQ01000164.1	Orf1ab_r	0	96
Neisseria macacae	strain = *Neisseria meningitidis*	OAEH01000146.1	Orf1ab_r	0	96
Neisseria meningitidis	strain = DE10444	CP012392.1	Orf1ab_r	0	96
Neisseria meningitidis	strain = FDAARGOS_212	CP020423.2	Orf1ab_r	0	96
Neisseria meningitidis	strain = M22425	CP031329.1	Orf1ab_r	0	96
Neisseria meningitidis	strain = M21955	CP031330.1	Orf1ab_r	0	96
Leptospira interrogans	strain = RCA	CP022538.1	Sprot_r	0	92
Leptospira interrogans	strain = ATCC 43642	FTNA01000026.1	Sprot_r	0	92
Leptospira interrogans	strain = 401	JMDJ01000175.1	Sprot_r	0	92
Leptospira interrogans	strain = H78Shuang4	JQOL01000072.1	Sprot_r	0	92
Leptospira interrogans	strain = 56662	JQOM01000004.1	Sprot_r	0	92
Leptospira interrogans	strain = 56666	JQON01000188.1	Sprot_r	0	92
Leptospira interrogans	strain = 56673	JQOO01000240.1	Sprot_r	0	92
Chlamydia psittaci	strain = GIMC 2005:CpsCP1	CP024451.1	Orf1ab_r	89	63
Chlamydia psittaci	strain = GIMC 2003:Cps25SM	CP024453.1	Orf1ab_r	89	63
Chlamydia psittaci	strain = GIMC 2004:CpsAP23	CP024455.1	Orf1ab_r	89	63

Another aspect of a successful pathogen detection test is to ensure maximal strain coverage. Databases dedicated to the pathogen of interest are likely to exist and can be checked to make sure the primers detect the desired family of strains. To illustrate this, we queried the database of SARS-CoV-2 sequences deposited at GISAID using the fragment analysis test primer pairs. We found that the designed primers cover 100% of the SARS-CoV-2 S gene and *orf1-ab* sequences, and 99.68% of the N gene sequences, that were deposited at the time of the query (n = 5,377). Performing similar checks on other pathogens is therefore recommended.

Cross-reactivity

Although examining sequence similarity *in silico* is an important quality check and provides information on potential cross-reactivity, it is equally important to verify that the assay produces no cross-reactivity *in vitro* using samples that are genetically similar. To test the specificity of the SARS-CoV-2 primers, we obtained commercial preparations of respiratory pathogens (ZeptoMetrix). These are either pools of inactivated pathogens of humans, including related coronaviruses, or single-organism isolates of inactivated severe acute respiratory syndrome (SARS) or Middle East respiratory syndrome (MERS) coronaviruses. Purified viral RNA from aliquots of the pools or single isolates were prepared using the Applied Biosystems™

MagMAX™ Viral/Pathogen Nucleic Acid Isolation Kit and analyzed with the SARS-CoV-2 fragment analysis primers. We analyzed 5 ng of purified RNA from these samples and found no peaks specific to SARS-CoV-2 in any of the samples, illustrating the specificity of the assay (Figure 4). The identity of related organisms that should be checked will vary depending on the assay and needs of the investigator. Thus, although the *in silico* analysis is an important step, *in vitro* verification that the assay lacks cross-reactivity against either pathologically or genetically related organisms provides additional confidence that positive results are due to the presence of the intended target.

A

Organism	Result
Influenza A H1N1	No positive call
Influenza A H1 (New Caledonia)	No positive call
Influenza A H3 (Brisbane)	No positive call
Influenza B (Florida)	No positive call
Parainfluenza type 1	No positive call
Parainfluenza type 2	No positive call
Parainfluenza type 3	No positive call
Parainfluenza type 4A	No positive call
Parainfluenza type 4b	No positive call
Respiratory syncytial virus A	No positive call
Respiratory syncytial virus B	No positive call
Mycoplasma pneumoniae M129	No positive call
Chlamydia pneumoniae	No positive call

Organism	Result
Coronavirus HKU-1	No positive call
Coronavirus NL63	No positive call
Coronavirus OC43	No positive call
Coronavirus 229E	No positive call
Metapneumovirus (Peru 2003)	No positive call
Rhinovirus (1A)	No positive call
Adenovirus type 3	No positive call
Legionella pneumophila (Philadelphia)	No positive call
Human bocavirus	No positive call
Negative pool	No positive call
SARS	No positive call
MERS	No positive call

B

Figure 4. Cross-reactivity with other pathogens. (A) List of organisms in ZeptoMetrix™ NATtrol™ respiratory pathogen panel pools and individual NATtrol preparations (SARS and MERS) analyzed using the SARS-CoV-2 sequences described above. No positive calls were made on organisms in any of the pools or individual preparations. **(B)** Representative electropherogram traces of analysis of RNA purified from MERS (top), SARS-CoV-2 (middle), and no-template control (bottom) samples.

Limit of detection

Establishing the limit of detection (LOD) is an important way to evaluate performance of an assay. Although this can be defined in many ways, a commonly used metric is to define the concentration for which a true positive detection can be made in at least 95% of samples. A straightforward way to determine the LOD is to analyze serial dilutions of the sample to find an approximate LOD, then verify that concentration with multiple replicates of concentrations above and below that approximation. For example, initial experiments using SARS-CoV-2 viral RNA genomes suggested we could detect around 5–10 copies per reaction. To verify that result, we set up technical replicates of 24 reactions, using 0, 2, 5, and 10 copies/reaction (Table 3). These results demonstrated that we could detect 10 viral RNA genome copies in ~95% of the reactions. Although it was possible to detect 2 and 5 copies in some of the samples, other samples in the same concentration range were indeterminate or negative. This dropped the percentage that scored positive below the 95% threshold. Thus, even though the assay may be capable of detecting 2–5 copies/reaction, 10 copies/reaction is achievable 95% of the time and is therefore an accurate measurement of the LOD for this assay. Of course, other pathogens and primer sets may produce different results; therefore it is necessary to determine LOD for the specific assay system that will be used.
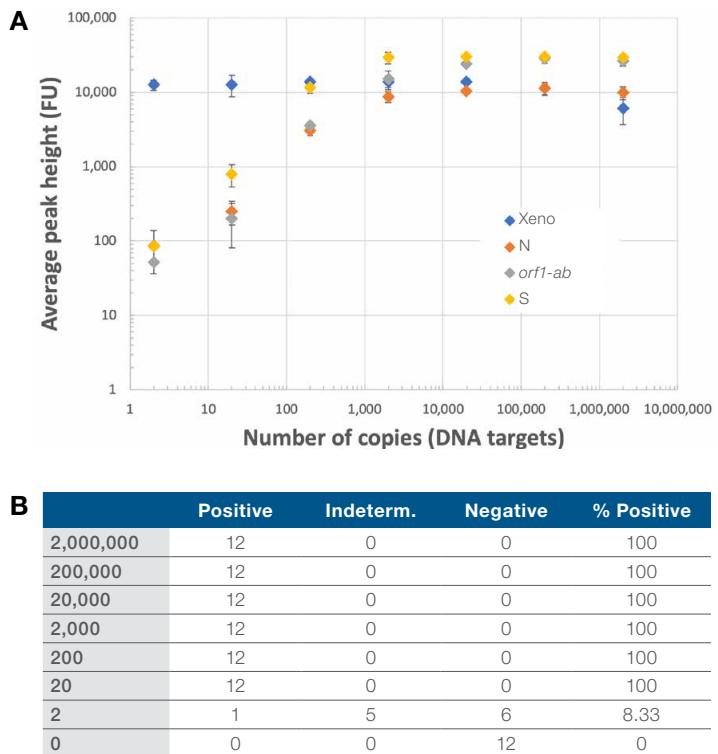
Table 3. Limit of detection for the SARS-CoV-2 assay. Limit of detection was determined by analyzing the indicated number of viral RNA genomes in 23 of 24 replicate reactions for each input amount. Percentages of negative, indeterminate, and positive, as well as false-negative (FN%) and false-positive (FP%), are shown. Note that 10 copies/reaction can be detected 95% of the time.

	Negative	Indeterm.	Positive	FN%	FP%
0 copies	78.3	21.7	0.0	NA	0
2 copies	8.7	56.5	34.8	20.0	NA
5 copies	8.7	43.5	47.8	15.4	NA
10 copies	4.3	0.0	95.7	4.3	NA

Dynamic range

Another consideration in designing a pathogen detection test is the target concentration range at which a positive signal can be detected. This could be determined in one of two ways, depending on the needs of the assay. One way is to determine the concentrations at which a positive signal can be quantifiably detected. However, in pathogen detection research assays, it is not always necessary to determine exact quantities of the pathogen; it may be sufficient to know whether the

pathogen is present or absent in a sample. To illustrate how to determine the dynamic range of a fragment analysis–based method, we made a dilution series of 12 replicate concentrations of SARS-CoV-2, from 2×10^6 copies/reaction to 2 copies/reaction. To get an approximation of the abundance of targets, we measured the peak heights of the target amplicons at each concentration (Figure 5A). Note that for the concentration range between 2 and 2×10^3 copies/reaction, there is a strong correlation between the input copy number and peak height (R^2 averaged 0.96). For input amounts higher than 2×10^3 copies/reaction, however, the fluorescence detector became saturated, and signal level measured was not correlated with the number of copies present. Nevertheless, it was clear that viral target sequences were present in those samples and they could therefore be scored positive based on the criteria described above (Figure 5B).
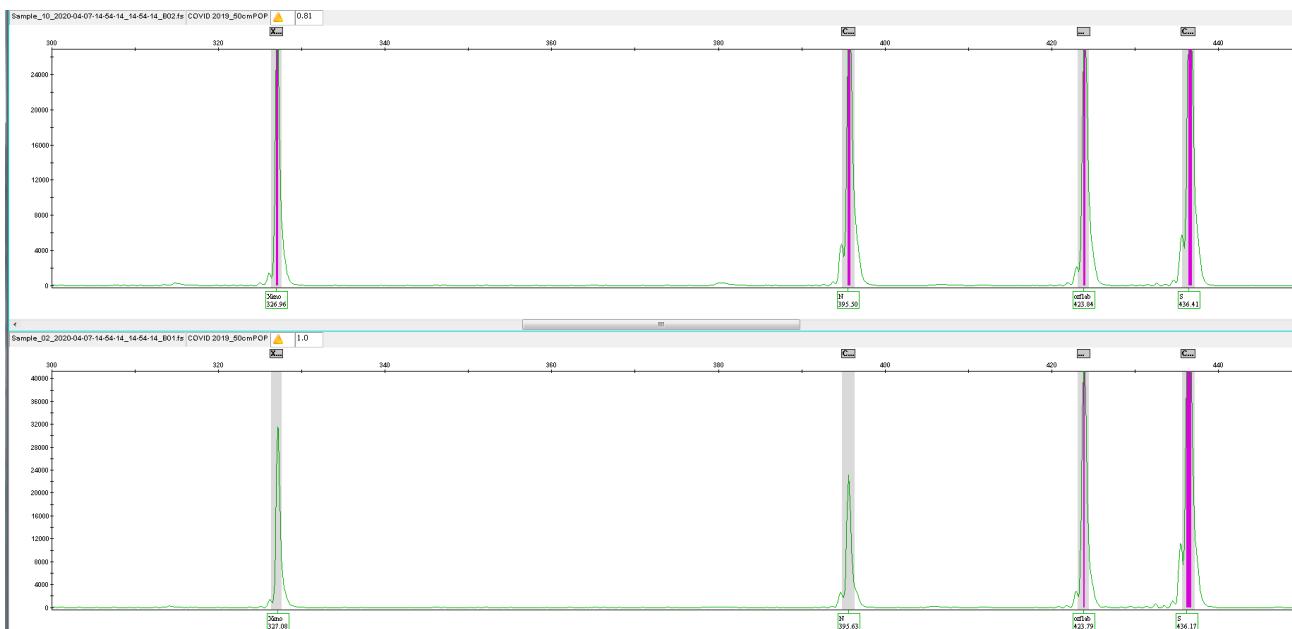
B

	Positive	Indeterm.	Negative	% Positive
2,000,000	12	0	0	100
200,000	12	0	0	100
20,000	12	0	0	100
2,000	12	0	0	100
200	12	0	0	100
20	12	0	0	100
2	1	5	6	8.33
0	0	0	12	0

Figure 5. Dynamic range of assay that uses SARS-CoV-2 primers. **(A)** Plot of peak height versus copy number. Above 2,000 copies/reaction, the detector measures positive signal but is not quantitative. At 2,000 copies and below, there is a strong correlation between peak height and copy number. Regression (red line on graph) can define a relationship of $y = ax^b$, where y = peak height, x = input copy number, and a and b are amplicon-specific constants. Overall, the correlation average R^2 is 0.96 for all SARS-CoV-2 peaks. Note that the blue markers are the VetMAX Xeno RNA control, spiked in at a constant concentration across all input ranges. **(B)** Using the criteria described in Table 1, positive samples can be detected over at least 6 orders of magnitude of input.

If a quantitative determination for highly concentrated samples is required, there are several options. The easiest method is to dilute the samples in a buffer such as TE and repeat the assay. If the peak height of the diluted sample falls within the previously determined quantitative range, then the approximate copy number in the undiluted sample can be determined. Another method is to attenuate peak heights by limiting the amount of primers. For example, one of the highest-expressed targets during SARS-CoV-2 infection is the gene encoding the nucleocapsid protein (N) [4]. This can be seen in the peak heights relative to *orf1-ab* and the spike protein (S) gene in SARS-CoV-2 viral RNA preps (Figure 3). To attenuate the N gene signal, we made a primer mix where the concentrations of the N gene and VetMAX Xeno RNA fluorescent primers were reduced, for a concentration ratio of 0.3:0.3:1:1 for

VetMAX Xeno RNA:N gene:*orf1-ab*:S gene. The reverse primer concentrations were not altered in either mix. In a sample containing 2 x 10⁴ copies of RNA targets, the unattenuated primer set produced detector saturation of all peaks (Figure 6). However, the same concentration of targets analyzed by the attenuated primer set had VetMAX Xeno RNA control and N gene peaks reduced to below saturation, while leaving the *orf1-ab* and S gene peaks unaffected. Thus, primer attenuation may be an effective strategy for analyzing highly concentrated samples. However, doing so may affect the LOD, since low levels of target may not get amplified using attenuated primers. Determining whether or not to use attenuated primers will therefore depend on the relative concentrations of targets and will require balancing the need for sensitivity vs. high-concentration quantification.

Figure 6. Primer attenuation can reduce very high signals. 20,000 copies of SARS-CoV-2 RNA targets were analyzed using either (top) unattenuated primers (1:1:1:1 ratio of fluorescent forward primers for VetMAX Xeno RNA control, N gene, *orf1-ab*, and S gene) or (bottom) attenuated primers (0.3:0.3:1:1). Note that when unattenuated primers were used, all the peaks were saturated (pink lines). However, attenuating the primers for the VetMAX Xeno RNA and N gene reduced their amplicons' overall signal to below saturation, while not affecting the other two peaks. This strategy may be used to reduce signals from samples that have very high concentrations of target sequences.

applied biosystems

Summary

In this application note, we have outlined a general strategy for designing a multiplexed target fragment analysis solution for detecting pathogenic sequences. We suggest designing amplicons for target sequences that are between 150 and 500 nucleotides long and including a process control target that can be spiked into reactions at the beginning of the protocol and analyzed along with the specific targets. We suggest performing *in silico* analyses to characterize the potential cross-reactivity with other pathogens and strains that will be covered by the assay. We illustrate how the LOD and dynamic range may be analyzed and adjusted, if needed. Finally, we show how all these suggestions can be implemented by defining a protocol that may be used to detect SARS-CoV-2 sequences in research samples. In general, a fragment analysis–based, multiplexed target approach using CE gives investigators another tool for analyzing the presence of genomic sequences of pathogenic organisms using a method that is rapid, simple, and sensitive.

References

1. Ramanan P et al. (2017) Syndromic Panel-based Testing in Clinical Microbiology. *Clin Microbiol Rev* 15:31(1):e00024-17. doi:10.1128/CMR.00024-17.

2. Detection of RNA from SARS-CoV-2 using fragment analysis. **https://assets. thermofisher.com/TFS-Assets/GSD/Scientific%C2%A0Guides/detection-sarscov2-fragment-analysis-protocol.pdf**

3. DNA fragment analysis by capillary electrophoresis. **https://www.thermofisher.com/us/en/home/life-science/sequencing/fragment-analysis.html**

4. Kim D et al. (2020) The architecture of SARS-CoV-2 transcriptome. *Cell* 181:914-921. **https://doi.org/10.1016/j.cell.2020.04.011**.

Ordering information

Product	Quantity	Cat. No.
MagMAX Viral/Pathogen Nucleic Acid Isolation Kit	1 kit	A42352
TaqMan Fast Virus 1-Step Master Mix	1 mL	4444432
MicroAmp Clear Adhesive Film	100 films	4306311
5′-labeled/unlabeled primer pairs	10 nmol	450056
VetMAX Xeno Internal Positive Control RNA	100 reactions	A29763
Veriti 96-Well Fast Thermal Cycler	1 instrument	4375305
Hi-Di Formamide	25 mL	4311320
DS-33 Matrix Standard Kit (Dye Set G5)	8 runs	4345833
GeneScan 600 LIZ Dye Size Standard v2.0	800 reactions	4408399
SeqStudio Genetic Analyzer	1 system	A35644
3500 Genetic Analyzer for Fragment Analysis	1 system	A30468
GeneMapper Software 6, full installation	1 license	4475073

Find out more at **thermofisher.com/sangercovid19**

Thermo Fisher
SCIENTIFIC