Sanger sequencing and fragment analysis

Multiplex detection of respiratory pathogens using fragment analysis by capillary electrophoresis

We show that:

- Multiplex PCR coupled with fragment analysis by capillary electrophoresis (CE) can be used for concurrent detection of several respiratory RNA viruses in a single sample, with high sensitivity and specificity
- The protocol can be optimized for quantitative measurements of the queried genomic input
- This approach can be widely applied to other areas of research where simultaneous detection of multiple genomic sequences is desired

Introduction

PCR is a workhorse of modern molecular biology, but it can be tedious to investigate multiple analytes in a single sample. Methods that can analyze multiple sequences from a single sample would simplify workflows and conserve precious samples. A large number of targets can be analyzed using fragment analysis by separating amplicons based on both size and fluorescent label.

Since the onset of the SARS-CoV-2 crisis in 2020, there has been an increased interest in analytical tools that enable the simultaneous detection of various respiratory pathogens in a given sample. Like other coronaviruses, SARS-CoV-2 infection initiates in the respiratory tract and then may spread to different organs in the body. While a large percentage of infected subjects remain asymptomatic, in those who do show symptoms, the infection typically begins with common flu-like indications such as coughing and general malaise. Characterizing the etiological agent of infection therefore becomes critical, since knowledge of the agent guides disease surveillance and prevention. CE offers a versatile solution to detect multiple species in a single reaction through a fragment analysis–based approach.

Fragment analysis is a genetic analysis method that can separate fluorescently labeled genomic products from multiplex PCR based on their size, down to a 1 bp difference. This enables potential multiplexing of a large number of targets from a single sample. In addition, it may be possible to PCR-amplify different samples with different sets of fluorescent labels, allowing up to 4 samples to be analyzed in a single capillary. Although a sample multiplexing approach would need to be fully optimized, it could be very powerful for detecting and distinguishing pathogens efficiently across many different samples.

Here we demonstrate the utility of fragment analysis for multiplex PCR reactions. As an example, we designed a panel that could detect as many as 10 viral respiratory pathogens, including SARS-CoV-2, influenza, and respiratory syncytial virus, in a single sample. We show that with both DNA and RNA as input, most pathogens can be detected at a lower limit of detection (LOD) of 100 copies/µL, and that an external, nonspecific RNA can be used to control for reverse transcription and PCR amplification. Given its advantages in both multiplexing capabilities and sensitivity, the fragment analysis methodology can be adapted and utilized in any scenario where the detection of several genomic sequences from a single sample is desired.

Materials and methods

Workflow for detection of multiple respiratory pathogens using fragment analysis

The workflow for fragment analysis is straightforward and begins with sample collection for RNA isolation (Figure 1). RNA extracted from pathogen-containing samples is reverse-transcribed to cDNA. A pool of several fluorescently labeled primer sets targeting specific respiratory pathogens is used to generate fluorescently labeled amplicons of varying lengths by multiplex PCR. The PCR products are then subjected to fragment analysis by CE, based on their size and different fluorescent colors. Once the sample is collected, data can be obtained in as little as 3 hours from start to finish.

A

B

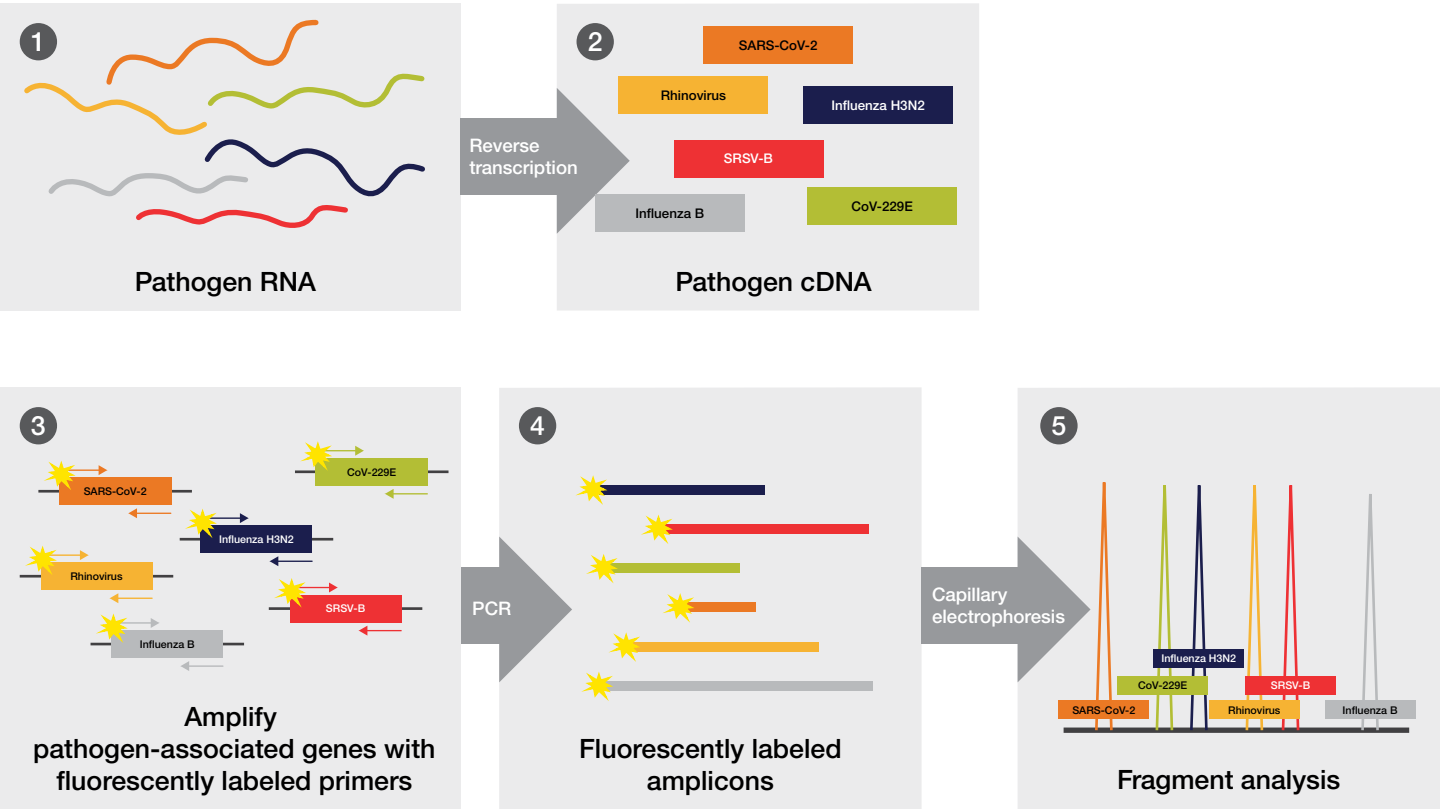

Figure 1. Simultaneous detection of multiple pathogens using multiplex PCR and fragment analysis. (A) General workflow. **(B)** Schematic of experimental procedure.

Primer design

To illustrate how multiplex PCR could be used for pathogen detection, we designed primers against common respiratory viruses. To help ensure optimal performance of the assay, candidate primer sequences were chosen to maximize detection of target organisms/strains and minimize detection of off-target sequences. This was done by performing a BLAST™ search of candidate primer sequences against sequences deposited in databases. 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, are shown in Table 1. 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. All forward primers were synthesized with a 6-FAM™ label at the 5′ end. The targets were distinguished by the different sizes of the amplicons.

Table 1. *In silico* analysis of potential cross-reactivity of the forward and reverse primers designed against the SARS-CoV-2 S gene with other pathogens.

Organism	Strain/isolate	GenBank accession no.	Assay	Percent homology, forward primer	Percent homology, reverse primer
Bat coronavirus	16B0133	KY938558.1	Nprot_r	95	0
Bat SARS coronavirus	HKU3-1	DQ022305.2	Nprot_r	95	0
Bat SARS coronavirus	HKU3-12	GQ153547.1	Nprot_r	95	0
Bat SARS coronavirus	HKU3-2	DQ84199.1	Nprot_r	95	0
Bat SARS coronavirus	HKU3-3	DQ084200.1	Nprot_r	95	0
Bat SARS coronavirus	NA	KF294457.1	Orf1ab_r	0	96
Neisseria meningitidis	R977	OAEK01000203.1	Orf1ab_r	0	96
Neisseria meningitidis	R970	OAEQ01000164.1	Orf1ab_r	0	96
Neisseria meningitidis	R985	OAEH01000146.1	Orf1ab_r	0	96
Neisseria meningitidis	DE10444	CP012392.1	Orf1ab_r	0	96
Neisseria meningitidis	FDAARGOS_212	CP020423.2	Orf1ab_r	0	96
Neisseria meningitidis	M22425	CP031329.1	Orf1ab_r	0	96
Neisseria meningitidis	M21955	CP031330.1	Orf1ab_r	0	96
Leptospira interrogans	RCA	CP022538.1	Sprot_r	0	92
Leptospira interrogans	ATCC 43642	FTNA01000026.1	Sprot_r	0	92
Leptospira interrogans	401	JMDJ01000175.1	Sprot_r	0	92
Leptospira interrogans	H78Shuang4	JQOL01000072.1	Sprot_r	0	92
Leptospira interrogans	56662	JQOM01000004.1	Sprot_r	0	92
Leptospira interrogans	56666	JQON01000188.1	Sprot_r	0	92
Leptospira interrogans	56673	JQOO01000240.1	Sprot_r	0	92
Chlamydia psittaci	GIMC 2005:CpsCP1	CP024451.1	Orf1ab_r	89	63
Chlamydia psittaci	GIMC 2003:Cps25SM	CP024453.1	Orf1ab_r	89	63
Chlamydia psittaci	GIMC 2004:CpsAP23	CP024455.1	Orf1ab_r	89	63

Results

Detection of multiple pathogen-derived DNA sequences by fragment analysis

A pool of synthetic DNA fragments synthesized by Invitrogen™ GeneArt™ Gene Synthesis Services, encoding specific genes of the respiratory pathogens rhinovirus, respiratory syncytial virus B (RSV-B), coronavirus 229E, coronavirus HKU1, coronavirus NL63, SARS-CoV-2, influenza A H1, influenza A H3, and influenza B, was used as input for multiplex PCR at 0.8 µM per individual primer. The multiplex primer pool contained probes for three different SARS-CoV-2 genes (S, N, and *orf1ab*). Although primers were designed for MERS, SARS-CoV-1, and RSV-A, DNA fragments were not synthesized for these targets. To confirm the primers were functional, each primer was tested individually against its cognate target (data not shown).

Next, the fragment pool was serially diluted from 10^6 copies per pathogen to 10^4, 10^3, and 10^2 copies per pathogen to determine the LOD. PCR was performed using Applied Biosystems™ TaqMan™ Multiplex Master Mix on the Applied Biosystems™ Veriti™ 96-Well Fast Thermal Cycler. Then 1 µL of the PCR product was used for CE-based fragment analysis on the Applied Biosystems™ SeqStudio™ Flex Genetic Analyzer using Applied Biosystems™ POP-7™ Polymer on 50 cm capillaries.

All of the pathogen sequences were detected at 100 copies per reaction. However, at the lower target concentrations, differential PCR efficiency could be detected for some pathogen sequences. For example, influenza B showed reduced peak height relative to the other targets at 10^3 and fewer copies (Figure 2). This phenomenon could be due to sequence-specific differences in primer kinetics and efficiencies, which would be particularly detectable at low target concentrations. This should be anticipated when designing a multiplex PCR panel.

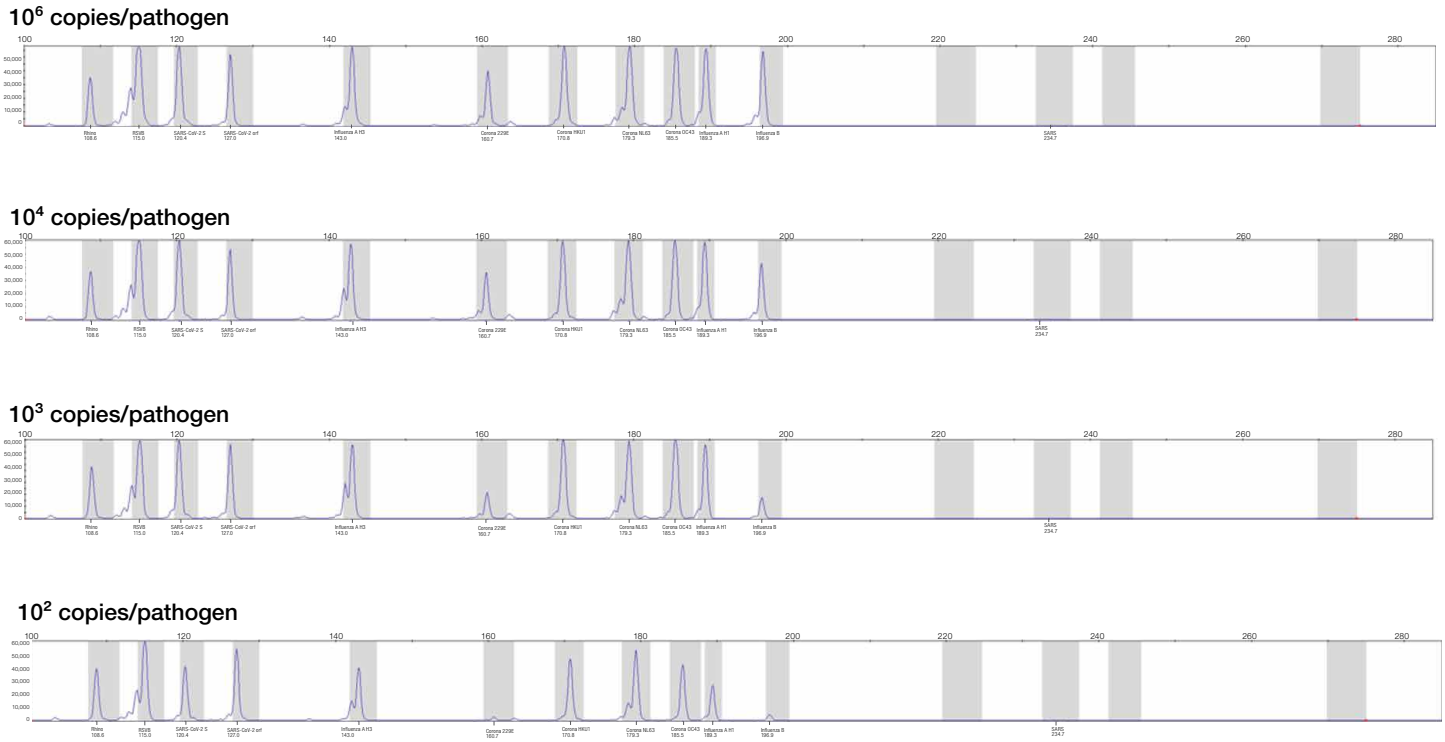

Figure 2. Simultaneous detection of pathogen-derived DNA sequences by fragment analysis. The pool of DNA sequences was diluted to various concentrations ranging from 10^6 copies per pathogen to 10^2 copies per pathogen and used as input for the assay.

Detection of multiple pathogen-derived cDNA sequences by fragment analysis

Purified RNA samples from ATCC for RSV-B, OC43, rhinovirus, influenza B, influenza A H3, coronavirus NL63, and coronavirus 229E, as well as inactivated SARS-CoV-2 from ZeptoMetrix Corporation, were combined at 10^5 copies each to create a genomic RNA mix. This mix was used in a single reverse transcription reaction with Invitrogen™ SuperScript™ IV VILO™ Master Mix.

The DNA was PCR amplified using the multiplex primer pool against the pathogens' RNA sequences, as described above. As was done with the DNA fragment samples, each individual primer pair was tested against its cognate target to verify functionality (data not shown). Then 1 µL of the PCR product was used for fragment analysis on the SeqStudio Flex Genetic Analyzer. Robust amplification of most of the input pathogen target sequences was observed, as seen by the fragment peaks with no off-target amplification (Figure 3).

Unexpectedly, amplification of coronavirus NL63 was inhibited in the multiplex setting, in contrast to its amplification in the singleplex reaction. It is possible that the primer efficiency for coronavirus NL63 was impacted in a multiplexed input scenario, or that other templates were preferentially amplified over coronavirus NL63, a phenomenon known as PCR selection [1]. This finding highlights the variability that can be observed for the same template/primer pair in single vs. multiplex settings, and illustrates that optimization of a panel may be necessary.

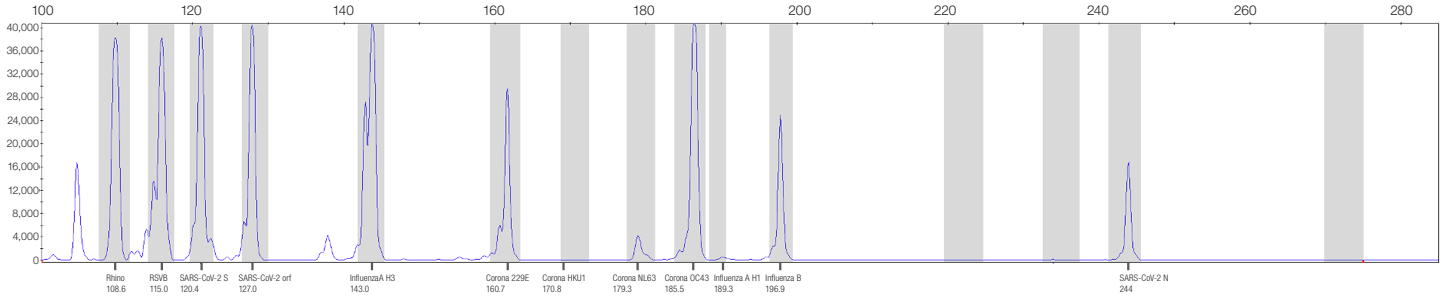

Figure 3. Detection of several respiratory pathogens by fragment analysis, using genomic RNA as input. Pooled genomic RNA at 10^5 copies per pathogen was reverse-transcribed into cDNA, followed by multiplex PCR and CE-based fragment analysis.

Determination of LOD for individual pathogens

To determine the LOD of individual pathogen sequences, purified RNA from RSV-B, rhinovirus, and SARS-CoV-2 was diluted to 10^5, 10^3, and 10^2 copies/µL. Each sample was spiked with 3,000 copies of VetMAX Xeno RNA. The dilutions were reverse-transcribed and PCR amplified using the multiplex primer pool at 0.8 µM per primer, as described above.

Fragment analysis revealed that rhinovirus and RSV-B were detectable at concentrations as low as 100 copies/µL (Figure 4). The three SARS-CoV-2 genes (S, N, and *orf1ab*) were amplified; while at 10^5 copies/µL all three genes were detectable, the amplification between the genes became variable at lower concentrations, suggesting variable primer efficiencies (Figure 4B).

The VetMAX Xeno RNA was robustly amplified in all samples, and no cross-reactivity was observed for any of the detected pathogen sequences. Note that both signal saturation and primer efficiency can impact the fluorescent signal obtained for a target.

For quantitative fluorescent PCR (qfPCR), see panel design and protocol optimization guidelines in the appendix.

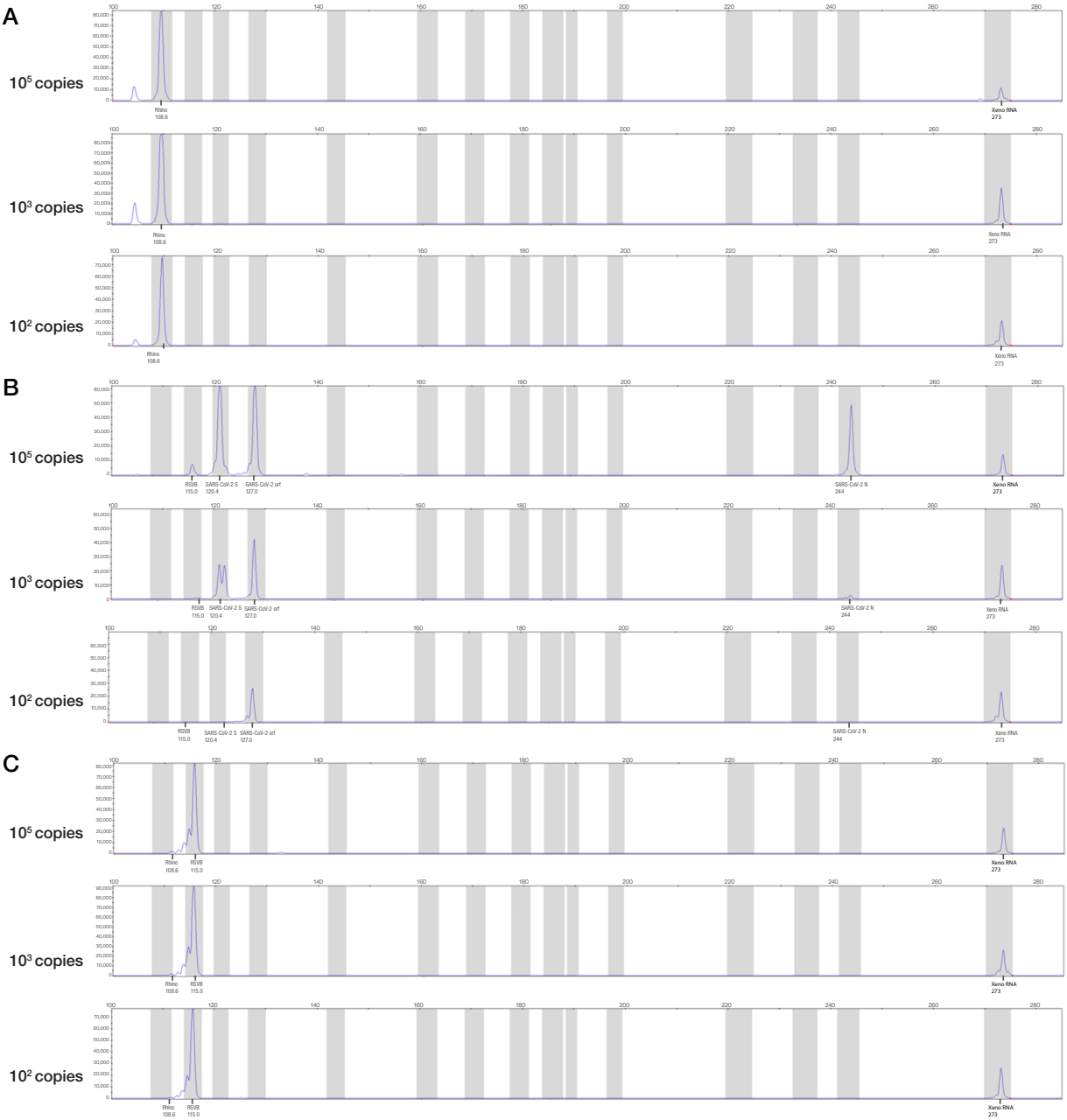

Figure 4. LOD for the assay for various respiratory pathogens. Each pathogen RNA sample was diluted to 10^5, 10^3, and 10^2 copies/µL, and a VetMAX Xeno RNA was spiked in at 3,000 copies for cDNA synthesis followed by PCR and capillary electrophoresis. **(A)** LOD for rhinovirus; **(B)** LOD for SARS-CoV-2 (S, N, and *orf1ab*); **(C)** LOD for RSV-B.

Data reproducibility across various genetic analyzers

Data reproducibility is of paramount importance in scientific research. Different instruments used to perform an assay under the same analytical conditions and settings should ideally give very similar results for a given sample. This provides confidence in the data generated and gives the researcher great flexibility to run their experiments on any genetic analyzer that is available and accessible.

To compare and contrast the fragment analysis data generated across different genetic analyzers, we ran the same sample plate on the Applied Biosystems™ SeqStudio™ 24 Flex Genetic Analyzer, Applied Biosystems™ 3500xL Genetic Analyzer, and SeqStudio Genetic Analyzer. Pooled RNA from ATCC and a Xeno RNA control were reverse transcribed and amplified using a multiplex primer pool as described above. Fragment analysis by capillary electrophoresis was performed using 1 µL of the PCR product. Similar detection profiles were generated on all three instruments, showing high concordance in pathogen identification (Figure 5).

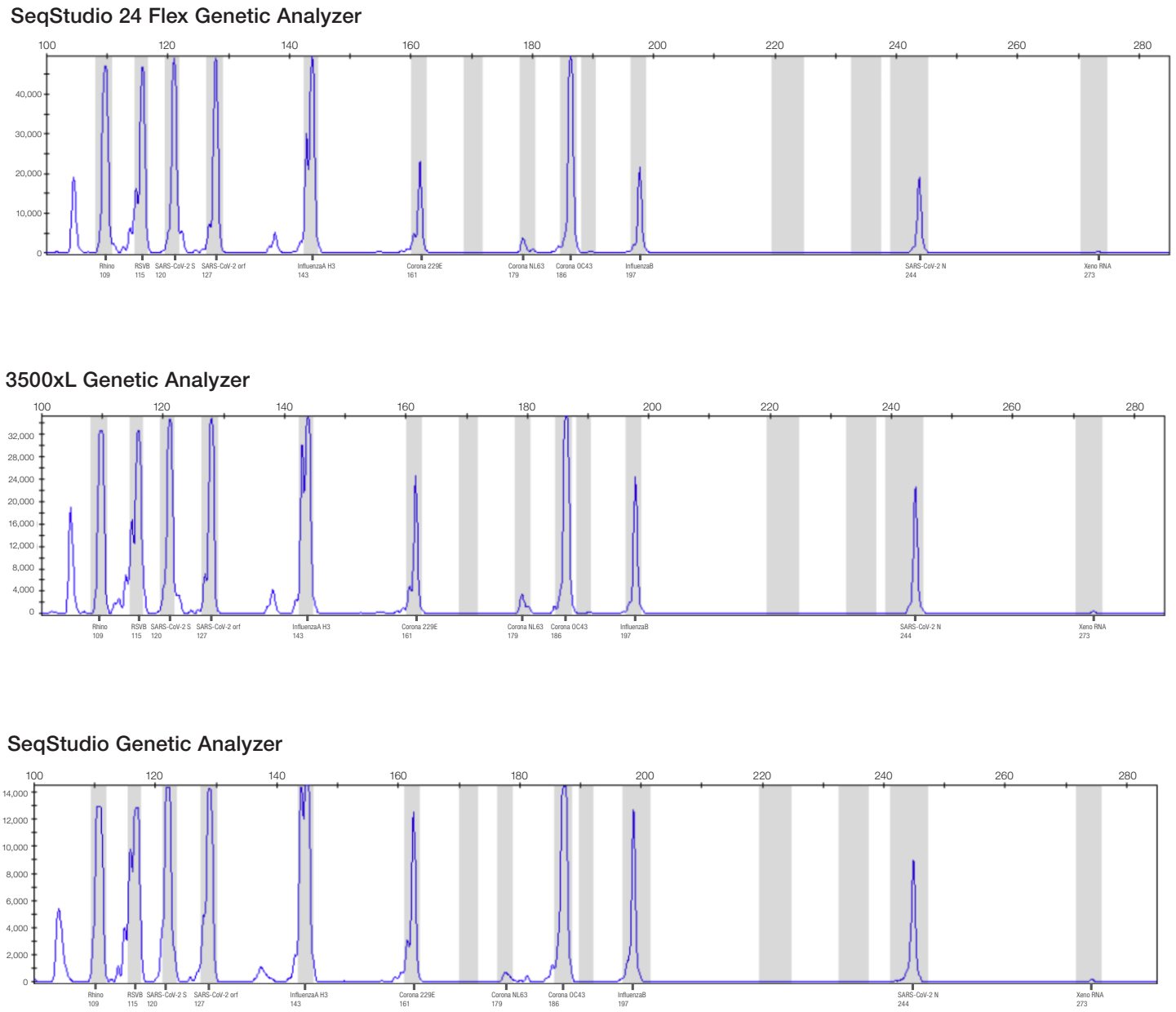

Figure 5. Detection profiles of pooled genomic RNA from respiratory pathogens on the SeqStudio 24 Flex, 3500xL, and SeqStudio genetic analyzers.

Conclusions

Genetic analysis technologies that can be scaled to detect multiple pathogens across multiple samples are necessary for timely surveillance and identification of infectious etiological agents. Here we have shown that multiplex PCR followed by fragment analysis is a powerful tool to detect several RNA viruses in a single sample with high sensitivity and specificity. Furthermore, the approach can be scaled up to accommodate a large number of targets for a given sample and can be broadly applied to other research areas where simultaneous detection of several targets, at the DNA or RNA level, is desirable.

For instance, fragment analysis can be used to determine genetic aberrations such as translocation for multiple rearrangements in cancer [2]; various mutations can be screened across a particular gene, by designing primers against the suspect region, to determine the etiology of the disease [3]; and genetic polymorphisms can be identified across several variants simultaneously to predict pharmacological responses to drugs [4]. These are just a few examples that illustrate the versatility and applications of the approach.

Appendix: Panel design and protocol optimization guidelines

- **Panel design:** Primers in this demonstration panel were chosen for optimal discrimination of the target sequences in a given amplicon size. The chosen size range was 100–300 nucleotides; larger amplicons are possible but might lead to more variable PCR results under these conditions. For best results, we recommend separating the different peaks by at least 5 nucleotides. For this demonstration, we synthesized primers with a single fluorophore; however, primer sets can be distributed across the four different fluorescent channels (for 6-FAM, VIC™, NED™, and PET™ labels) to increase the number of targets detectable and the target resolution, if needed.
- **Including an internal control:** An internal control should be included. This can be an exogenous target that is spiked into the PCR reaction before amplification. For panels detecting DNA targets, simply choose a target that will produce a unique size fragment in the same size range as the targets that will be queried. For panels detecting RNA targets, an exogenous RNA, such as the VetMAX Xeno RNA used here, can be used as a control for reverse transcription and PCR efficiency to normalize the results against. For these controls, it is best to design an amplicon that will produce a fragment slightly larger than the queried targets—this ensures the cDNA synthesis and PCR will cover the desired lengths.
- **Primer cross-reactivity determination:** Primer pairs should be tested individually as well as against other primer pairs to ensure that there is no cross-reactivity during PCR against the target and nontarget sequences. This can be achieved by testing the primer pairs individually and in combinations of multiples with other primer pairs present in the multiplex set.
- **LOQ and LOD:** Determine the limits of quantitation (LOQ) and detection (LOD) for an assay by titrating the amount of input RNA in both directions, to where the signal starts to plateau or becomes undetectable, respectively. This establishes the range for semiquantitative estimation of target input.
- **Primer concentration optimization for quantitative input measurements/qfPCR:** If the intended application is target quantitation, a standard curve using a known DNA amount is required to determine the peak intensity to its corresponding input. Primer concentrations should be adjusted such that the peak intensities for the same input of any target sequence are similar for all primer/target pairs. If primer concentration adjustment does not yield the desired results, design new primers against the target and test again.

References

1. Elnifro EM et al. (2000) Multiplex PCR: optimization and application in diagnostic virology. *Clin Microbiol Rev* 13(4): 559-570.
2. Algeciras-Schimnich A et al. (2010) Evaluation of the PAX8/PPARG translocation in follicular thyroid cancer with a 4-color reverse-transcription PCR assay and automated high-resolution fragment analysis. *Clin Chem* 56(3): 391-398.
3. Furtado LV et al. (2013) A multiplexed fragment analysis-based assay for detection of JAK2 exon 12 mutations. *J Mol Diagn* 15(5): 592-599.
4. Bouvet R et al. (2020) PharmFrag: An easy and fast multiplex pharmacogenetics assay to simultaneously analyze 9 genetic polymorphisms involved in response variability of anticancer drugs. *Int J Mol Sci* 21(24): 9650.

Ordering information

Product	Quantity	Cat. No.
VetMAX Xeno Internal Positive Control RNA	500 reactions	A29761
SuperScript IV VILO Master Mix	500 reactions	11756500
TaqMan Multiplex Master Mix	5 mL	4461882
Veriti 96-Well Fast Thermal Cycler	1 instrument	4375305
SeqStudio 8 Flex Genetic Analyzer	1 system	A53627
SeqStudio 24 Flex Genetic Analyzer	1 system	A53630
SeqStudio Genetic Analyzer	1 system	A35644
3500 Genetic Analyzer	1 system	4440462
3500xL Genetic Analyzer	1 system	4440463

Learn more at **thermofisher.com/ce**