White Paper

Agilent
Trusted Answers

Quality Control Steps for Viral Detection and Sequencing

Abstract

With the onset of the SARS-CoV-2 outbreak in 2019, fast and high-throughput identification and detection of the virus quickly became a necessity to aid in the understanding of where and how fast the virus was spreading. Viral sequencing is necessary not only for detection, but to trace, monitor and identify new variants. Next-generation sequencing (NGS) applications are available with a number of established protocols, and quality control steps are essential to determine the integrity, size and concentration of the starting nucleic acids and libraries, which helps ensure successful and reliable sequencing results. In this white paper, citations referenced from a variety of publications demonstrate the uses of Agilent automated electrophoresis systems for efficient and robust sample quality control (QC) when integrated in low- and high-throughput viral sequencing NGS workflows.

Introduction

Many sequencing platforms have been developed that enable detection of known and novel viruses in complex biological samples. Reliability and quality of the sequencing data can depend greatly on the integrity of the starting material and library preparation process. Viral genomes can vary widely in size and content, consisting of single-stranded or double-stranded RNA or DNA. Often, viral nucleic acids are extracted with the host genome and the viral component is present at comparatively lower levels. Optimizing the protocols for extracting viral nucleic acids and amplifying viral genomes separately from the host nucleic acids is necessary since viral and host genomes are so diverse.

In general, quality control (QC) of the input DNA and RNA from any source can help determine which samples are suitable for sequencing. This includes checking integrity, size, and concentration of the starting nucleic acid. The many different extraction methods and kits available introduce a high variability in size, distribution, and concentration of the extracted nucleic acids. Thus, the integrity of the nucleic acid starting material can vary greatly. For example, samples derived from formalin-fixed paraffin embedded (FFPE) tissue and older materials are often degraded due to chemical fixation, time, increased temperature, or UV radiation. Furthermore, RNA and DNA are prone to degradation by enzyme digestion, and improper handling. In addition, pre-analytical procedures including collecting, processing, storing, and shipping of biospecimens can affect the integrity of nucleic acids. If a sample is significantly degraded with only small fragments existing, it will result in poor sequencing data, loss of coding areas of interest, and gaps in the full-length RNA and gDNA. Processing and environmental impact can never be entirely controlled, and many factors can affect the sample, thus making it a necessity to monitor the quality of the starting material.

Agilent offers a suite of automated electrophoresis instruments, including the Bioanalyzer, Fragment Analyzer, and TapeStation systems which are ideal for nucleic acid quality control. Results are automatically presented in a digital gel image, electropherogram, and table that includes size, concentration, and molarity. Each instrument requires only 1-2 µL of sample for analysis, allowing for conservation of the original nucleic acid sample for viral sequencing analysis. The high sensitivity of the instruments allows detection of very low concentrated samples. In addition, both low- and high-throughput facilities are covered with the array of Agilent automated electrophoresis instruments.

Overview of QC steps in library preparation workflow

There are several common QC checkpoints that can apply to all library preparation workflows before sequencing. Figure 1 outlines different approaches for NGS sample preparation and potential QC steps during library preparation from part of the discussion in the Advanced Virus Detection Technologies Interest group (AVDTIG)[1]. Orange dots represent QC steps in the library preparation workflow. Knowing the starting quality of the nucleic acid helps provide guidance as to which samples are suited for a particular research study or workflow. QC also aids in directing changes needed to optimize a workflow, such as input concentration, fragmentation conditions, the amount of library used in enrichment, and the number of PCR cycles to be used in amplification steps.[2,3] The Agilent automated electrophoresis instruments provide vital QC information for the first three QC steps: size, integrity and quantity of the starting material, throughout sample preparation (cDNA synthesis), and the final library sample.

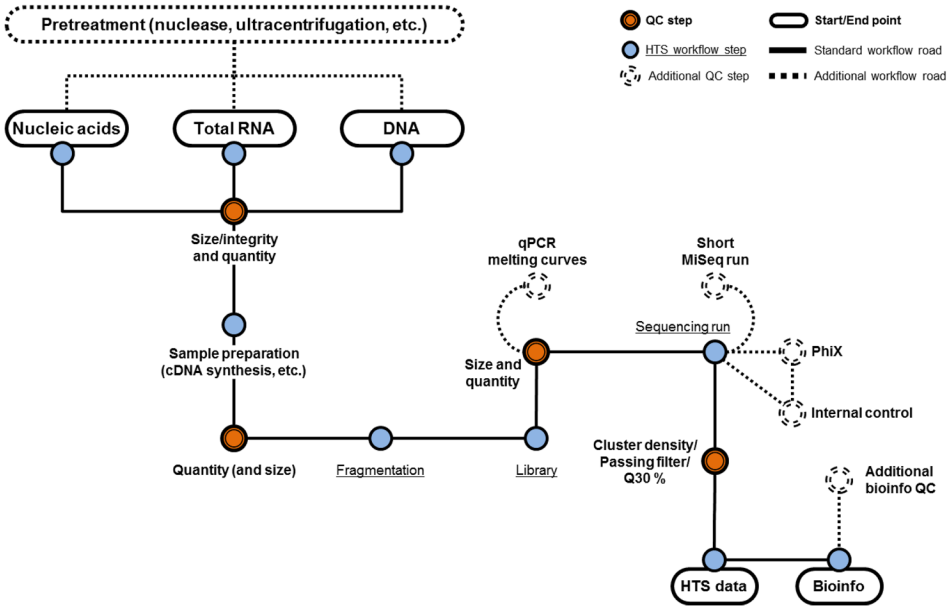

Figure 1. Different approaches for NGS or high-throughput sequencing (HTS) sample preparation and potential quality control (QC) opportunities. Experimental steps and potential QC steps are represented by blue and orange circles respectively. This figure has been reproduced from Ng et al.[1] This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Viral RNA extraction – size, integrity, quantity

RNA is easily degraded through handling and environmental factors, including heat and RNases. Thus, it is important to assess RNA quality after extraction. Over the years, the Bioanalyzer RNA integrity number (RIN) has become an expected method for determining RNA integrity. Indeed, Ng et al., states that "*Ideally, total RNA should be checked for integrity and purity using a microfluidics electrophoresis assay (e.g., Bioanalyzer) … and evaluating with the RNA integrity number (RIN).*"[1] All of the Agilent automated electrophoresis instruments have a comparable quality metric for the assessment of RNA integrity.[4, 5] The Bioanalyzer RIN, TapeStation RNA integrity number equivalent (RIN^e), and the Femto Pulse and Fragment Analyzer RNA quality number (RQN)[6], assist in assessing the presence or absence of RNA degradation. All three of the RNA quality scores, the RIN, RIN^e, and RQN, have scoring from 10 to 1, where 10 indicates the highest possible RNA quality.

Reliable large-scale testing of SARS-CoV-2 became necessary with the 2019-2020 worldwide outbreak. Aynaud et al.[7], established a COVID-19 screening using Systemic Parallel Analysis of RNA coupled to Sequencing (SPAR-Seq) for patients with low- and high-viral load. Extracted total RNA quality was assessed by the Bioanalyzer. Archival samples that failed sequencing were thought to be due to "*lost RNA integrity upon repeated freeze-thaw cycles"* and *"low viral RNA levels*".[7] Determining the total RNA quality and concentration prior to sequencing provided a potential explanation for the failed sequencing results. In addition, knowledge of the total RNA quality can alert researchers to the possibility of poor sequencing results prior to testing.

Indeed, Li et al.[8], also identified the need for an efficient workflow for SARS-CoV-2 whole-genome sequencing. They developed a workflow utilizing a one-step RT-PCR amplification with 39 primer sets on a microfluidic platform. The amplicons were subsequently sequenced to obtain a full genome sequence for SARS-CoV-2. The RT-PCR products were analyzed with the 4200 TapeStation system and High Sensitivity D5000 kit to determine the quality and quantity of the amplicons.

Primer sets were tested for genome RT-PCR amplification with a serial dilution of input RNA from SARS-CoV-2. The TapeStation system confirmed the expected results of a single amplicon with an approximate size of 1 kb (Figure 2). This result was seen for all input RNA concentrations, with the band intensities correlating with the copy numbers of SARS-CoV-2 in the serial dilutions. The TapeStation results helped determine the range of input RNA required for successful results.

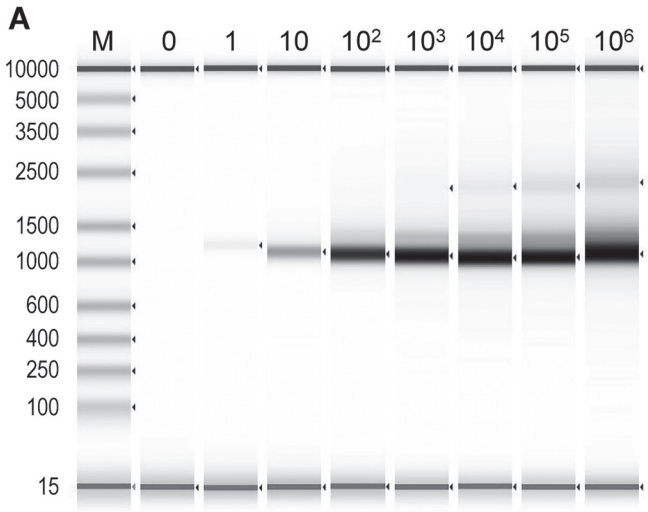

Figure 2. Agilent TapeStation system analysis of SARS-CoV-2 whole-genome RT-PCR amplification. (A) Digital gel image of RT-PCR products from left to right, 0, 1, 10, 10^2, 10^3, 10^4, 10^5, and 10^6, copies of SARS-CoV-2 RNA from isolate R4717. This figure was excerpted from Li et al.[8] Permission has been obtained from the publisher.

Ochoa et al.[9] screened 15-year-old archived Davidson's-fixed paraffin-embedded (DFPE) shrimp tissues for the presence of Taura syndrome virus (TSV) and conducted phylogenetic analyses by RT-qPCR. Because the samples were old, very precious, and stored by DFPE, initial QC analysis of the total RNA was essential after extraction to ensure the best possible library preparation and sequencing results. RNA samples "*were analyzed using an automated electrophoresis TapeStation system to investigate the degree of RNA degradation*" (Figure 3). The samples showed a low RIN^e number, which was to be expected due to the DFPE storage conditions. The RIN^e number, along with concentration, aided in determining which samples were tested and the amount of total RNA sample needed for successful amplification.

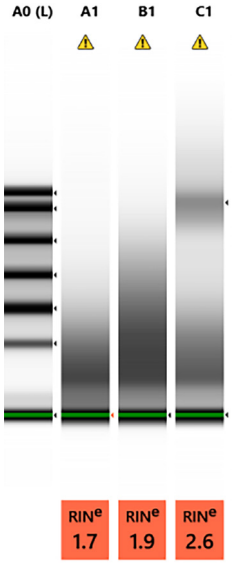

Figure 3. Gel image from the Agilent TapeStation system of three RNA samples extracted from DFPE. All three samples presented very low RNA Integrity Numbers (RIN^e) that ranged from 1.7-2.6, indicating a high level of degradation and fragmentation. This figure has been reproduced from Ochoa et al[9]. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Quality control of viral DNA or cDNA

Often library preparation workflows require fragmentation of DNA or cDNA. It is important to do a quality check and ensure the fragmented DNA is in the correct sizing range before proceeding. As seen in Figure 1, the second necessary QC step is verifying quantity and size of the DNA or cDNA sample. The author states that "*Checking that the fragments are of the expected size is of paramount importance for the success of the sequencing run. This can be achieved by visualizing the size profile of the fragmented nucleic acids using a microfluidic electrophoresis system (e.g., Bioanalyzer) and purifying the sample if necessary. Classical electrophoresis is still used because it allows simultaneous visualization and extraction of the region of interest but leads to a risk of cross contaminations between samples.*"[1] After size selection, confirmation of the fragment size is necessary. Sample quality control with the Bioanalyzer, Fragment Analyzer, and TapeStation system requires only 1-2 µL, conserving the original sample for sequencing. The TapeStation, and Fragment Analyzer systems utilize separate channels for analysis of each sample, eliminating any possible cross-contamination during analysis. In addition, the high resolution of these instruments allows for visualization and sizing to confirm the expected size range of the sample.

When performing RNA sequencing, sample QC of not only the starting RNA, but also of the subsequent cDNA is important (Figure 1). Often, viral sequencing is performed to determine conservation of genome regions and to classify genus as viruses are found in new host species. The mouse kidney parvovirus (MKPV), a member of the provisional Chapparvovirus genus, was investigated to determine global distribution and describe a closely related full-length Chapparvovirus from a primate kidney.[10] Several transcripts and primers were compared. Before sequencing, the "*sizes and yields of RACE (rapid amplification of cDNA ends) products were determined using a Fragment Analyzer equipped with a 55 cm electrophoresis capillary and reagents capable of resolving dsDNA fragments between 35 and 1,500 bp*"[10] (Figure 4). The Fragment Analyzer displayed the number of dominate products from each reaction and allowed for the reaction methods to be optimized.

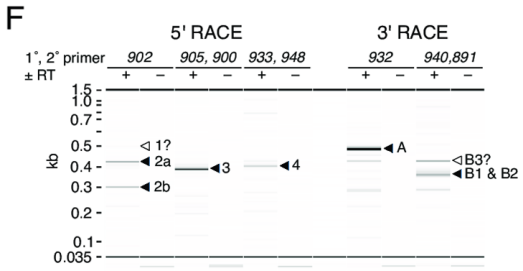

Figure 4. The Agilent Fragment Analyzer system identified major 5′ and 3′ RACE products, indicated by black arrows. This figure was excerpted from Lee et al.[10] © 2020 Lee et al. This is an open access article distributed under the terms of the Creative Commons Attribution License.

Fragmentation and library synthesis

Library size and concentration can directly affect the reliability and amount of sequencing data produced. Thus, it is optimal to QC libraries before sequencing to determine if the library is the appropriate size for the sequencing method. Fragments shorter than the read length will result in the possibility of sequencing through the insert and past the adapter.[11] In addition, some sequencing platforms have a bias towards shorter reads, with a greater affinity for clustering shorter fragments, so it is important to QC before sequencing. The fragment size distribution and quantity of the sequencing library can be checked by the Bioanalyzer, Fragment Analyzer, or TapeStation system, as was also recommended by Ng et al.[1]

Epidemics of H3N8 and H3N2 influenza A viruses in dogs can also transfect other animals and humans, necessitating the importance of efficient laboratory testing. Mitchell et al.[12] compared universal amplification primers on positive influenza A virus specimens from dogs, horses, and a cat, followed by sequencing to identify the subtype of the influenza A virus strains. Quality of the DNA fragmentation and different library preparation methods A and B, each with different primers and amplification settings were assessed by the Fragment Analyzer and the TapeStation system.

The Fragment Analyzer revealed that libraries from method A resulted in a peak at 144-146 bp, suggesting the presence of "*empty adapters*" (i.e. adapters with no insert, also known as adapter dimers), while libraries from method B did not have this peak, suggesting that these samples did not contain empty adapters. The Fragment Analyzer was also used to confirm the fragment distribution of the library (Figure 5). Ideal libraries have an even fragment distribution and a high percentage of the relative sample concentration as seen with method B (Figure 5B), while the presence of multiple peaks in addition to the adapter peaks in method A was of concern (Figure 5A). Indeed, method B libraries resulted in higher quality sequencing results than method A, "*suggesting that the difference in library preparation protocol led to more consistent fragment sizes*"[12] and higher-quality sequencing results. The presence of empty adapters in method A libraries "*likely led to the lower read quality and mapping depth observed for this method*".[12] Knowledge of the library size and distribution provided by the Fragment Analyzer and TapeStation systems provided information as to why some of the sequencing results were of poor quality.

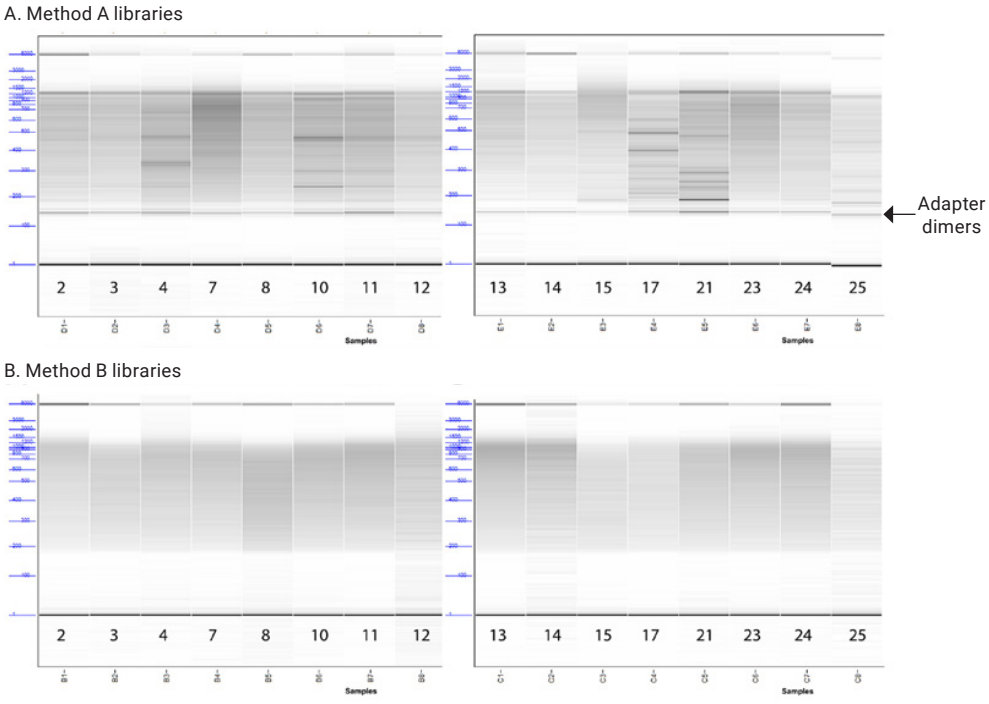

Figure 5. NGS libraries were analyzed on the Agilent Fragment Analyzer system prior to sequencing. (A) Method A (B) Method B. Method B has evenly distributed libraries with no adapters or fragment peaks, as seen in method A. This figure has been adapted from Mitchell et al.[12], p 195, © 2020 by The Author(s); reprinted by permission of SAGE Publications.

The worldwide spread of COVID-19 has created the need for a reliable system for testing thousands of patients in the larger urban areas. Anyaud et al.[7], described a COVID-19 screening system using Systemic Parallel Analysis of RNA coupled to Sequencing (C19-SPAR-Seq) for tens of thousands of patient samples in a single instrument run. Quality control metrics were set for low-viral and high-viral loads. Library quality was assessed with the Agilent 5200 Fragment Analyzer for all sample types. "*Due to NSA (non-specific amplification) products in the Fragment Analyzer profile in the test cohort and pilot cohort, we performed size selection purification (220-350 bp)*"[7] (Figure 6). The Fragment Analyzer provided essential quality control analysis by detecting NSA products that would not have been detected otherwise. This finding provided insight for optimizing the method development, leading to the addition of a necessary size-selection step in the library preparation workflow.

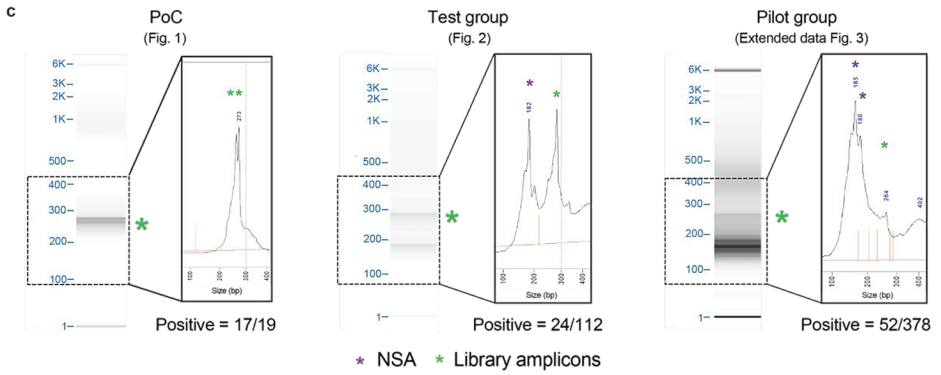

Figure 6. The Agilent Fragment Analyzer system detected non-specific amplification (NSA, purple star) in the test and pilot group. This finding lead to optimization of the library method with addition of a size-selection step in the library preparation workflow. This figure was excerpted from Aynaud et al.[7] This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Viral variant detection

The most common genetic variations to date in SARS CoV-2 are single or point mutations. Indels have also been identified, albeit less frequently, which provide information about the robustness of the virus' functionality. Scientists in Uruguay analyzed an outbreak of 14 patients and detected a 12-nucleotide deletion in the ORF7a accessory gene.[13] Quality control and length of the sequencing libraries were assessed on the 5200 Fragment Analyzer. In addition, the deletion was identified with the Fragment Analyzer and the High Sensitivity NGS kit by detecting a size difference between the variant and wild-type amplicons (Figure 7).

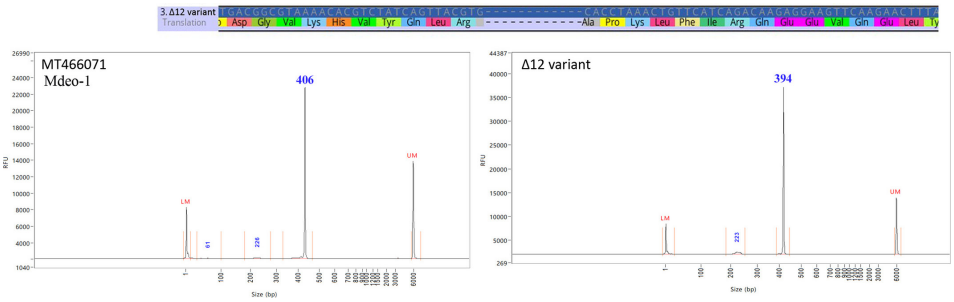

Figure 7. SARS-CoV-2 wildtype and variant amplicon analysis with the Agilent Fragment Analyzer system. Electropherograms display Mdeo-1 reference sequence (MT466071) and Δ12 variant, left and right, respectively. This figure was excerpted from Panzera et al.[13] This article is being made freely available through PubMed Central as part of the COVID-19 public health emergency response. It can be used for unrestricted research re-use and analysis in any form or by any means with acknowledgement of the original source, for the duration of the public health emergency.

Viral detection

Sequencing of viruses is important for determining strains, tracking genetic evolution, and identifying new emergent variants. However, detection of viruses can be accomplished by faster, less-expensive techniques. A platform for low-cost and rapid detection of viral RNA with DNA nano switches was developed by Zhou et al.[14] DNA nanoswitches are comprised of selected DNA sequences on an oligonucleotide backbone that undergo a conformational change from linear to looped, upon binding a target sequence. Gel electrophoresis and a common nucleic acid stain is used to detect the presence of viral RNA through a migration shift caused by the looped nanoswitch. The Zika virus was used as a model system to demonstrate nonenzymatic detection of viral RNA with selective detection between related viruses and viral strains. They demonstrated adaptability by developing DNA nanoswitches to detect SARS-CoV-2 RNA. Viral RNA can be long and have secondary structures that interfere with nanoswitch detection. Because of this, the Zika and SARS-CoV-2 RNA were chemically fragmented into RNA pieces shorter than 200 nt. The fragmented RNA was subjected to QC with the Fragment Analyzer system to ensure that RNA fragments were shorter than 200 nt before the samples were subjected to detection with the nanoswitches (Figure 8).

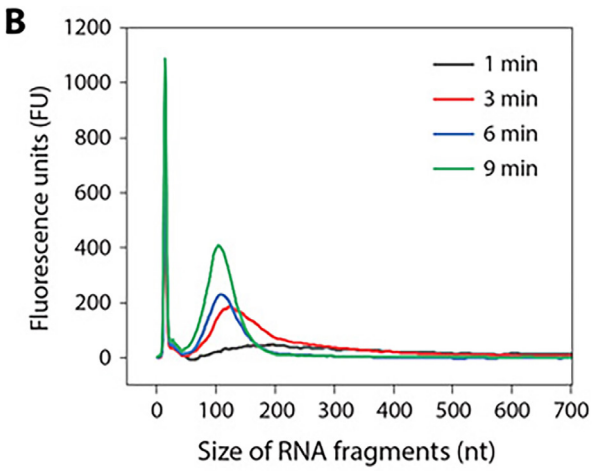

Figure 8. Detection of viral RNA using DNA nanoswitches. Zika RNA was fragmented at 94 °C for 1, 3, 6, and 9 minutes. The Agilent Fragment Analyzer system was used to confirm RNA fragments smaller than 200 nt. This figure was excerpted from Zhou et al.[14] This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license.

A reliable diagnostic assay is crucial for early detection of new COVID-19 cases. Etievant et al. compared sensitivity and specificity of different RT-PCR assays developed by referral laboratories and published by the World Health Organization.[15] Overall, the RT-PCR assays performed well for SARS-CoV-2 detection. False positives from the E Charité and N2 US CDC assays were further explored with the Bioanalyzer system and the DNA 1000 kit. The false positive samples from E Charité displayed a peak at 121 bp, the expected size of a positive sample, *"that might be derived from a contamination (amplicon size at 121 bp), but could be associated with an aspecific amplification (amplicon size at 84 bp)"*[15] (Figure 9). Likewise, the N2 US CDC false positive sample displayed a peak at 73 bp, the expected size for a positive amplification. The Bioanalyzer system helped to determine an explanation as to why the two assays were reporting false positive tests.

Conclusion

Viral sequencing and detection have always been vital for monitoring new variants and identifying outbreaks. Lately, viral sequencing has become a necessity and an important everyday occurrence with the outbreak of COVID-19. Quality control steps of the viral nucleic acid, cDNA synthesis, and library preparation remain a necessity in providing reliable sequencing data. These same sample QC steps, in addition to others are also utilized in new high-throughput detection methods. The Agilent automated electrophoresis instruments, Bioanalyzer, Fragment Analyzer, and TapeStation systems provide sensitive and accurate quality metrics, including integrity, size and concentration for viral sequencing samples.

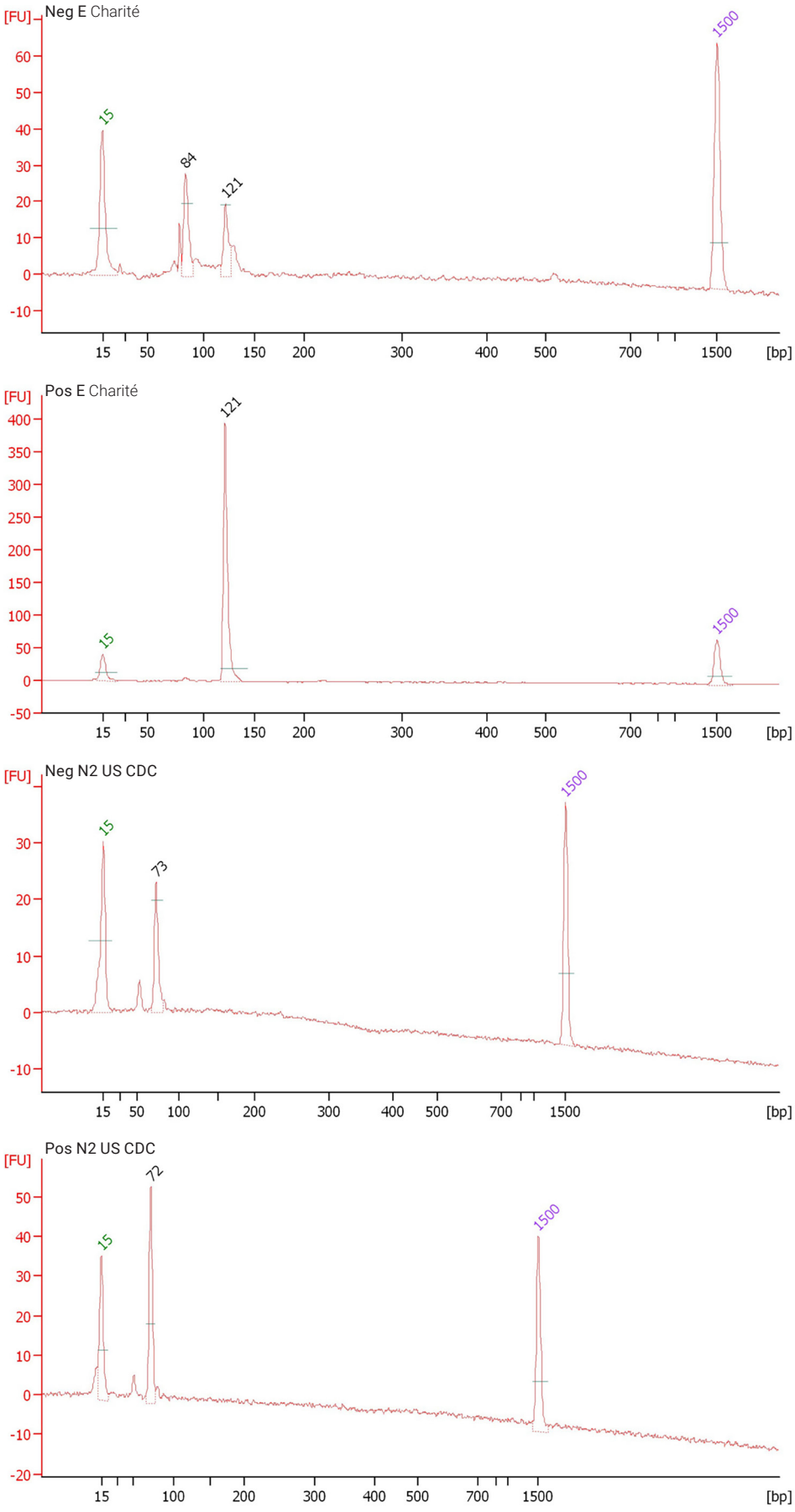

Figure 9. Electropherograms of amplicon sizes obtained using Agilent Bioanalyzer and the DNA 1000 kit for one positive sample (pos) and one negative sample (neg) for E Charité (Germany) and N2 US CDC (United States). This figure has been reproduced from Etievant et al.[15] This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license.

References

1. Ng, S.; Braxton, C.; Eloit, M.; Feng, S.; Fragnoud, R.; Mallet, L.; Mee, E.; Sathiamoorthy, S.; Vandeputte, O.; Khan, A. Current Perspectives on High-Throughput Sequencing (HTS) for Adventitious Virus Detection: Upstream Sample Processing and Library Preparation. *Viruses* **2018**, 10, 566.
2. The Importance of Quality Control in NGS Library Preparation Workflows: A Review of Applications using the Agilent Automated Electrophoresis Portfolio. *Agilent Technologies white paper*, publication number 5994-2281EN, **2020**.
3. Quality Control of NGS Libraries with Daisy Chains. *Agilent Technologies application note*, publication number 5994-2233EN, **2020**.
4. Performance Characteristics of the RNA and the High Sensitivity RNA ScreenTape Assays for the 4150 TapeStation System. *Agilent Technologies technical overview*, publication number 5994-1038EN, **2019**.
5. Comparison of RIN and RQN for the Agilent 2100 Bioanalyzer and the 5200 Fragment Analyzer Systems. *Agilent Technologies application note*, publication number 5994-1860EN, **2020**.
6. Analysis of Total RNA on the Agilent Femto Pulse System with the Agilent Ultra Sensitivity RNA Kit. *Agilent Technologies technical overview*, publication number 5994-3669EN, **2021**.
7. Aynaud, M.; Hernandez, J.; Barutcu, S.; Braunschweig, U.; Chan, K.; Pearson, J.; Trcka, D.; Prosser, S.; Kim, J.; Barrios-Rodiles, M.; Jen, M.; Song, S.; Shen, J.; Bruce, C.; Hazlett, B.; Poutanen, S.; Attisano, L.; Bremner, R.; Blencowe, B.; Mazzulli, T.; Han, H.; Pelletier, L.; Wrana, J. A Multiplexed, Next Generation Sequencing Platform for High-Throughput Detection of SARS-CoV-2. *Nature Communications* **2021**, 12, 1405.
8. Li, T.; Chung, H.K.; Pireku, P.K.; Beitzel, B.F.; Sanborn, M.A.; Tang, C.Y.; Hammer, R.D.; Ritter, D.; Wan, X.; Berry, I.M.; Hang, J. Rapid High-Throughput Whole-Genome Sequencing of SARS-CoV-2 by Using One-Step Reverse Transcription-PCR Amplification with an Integrated Microfluidic System and Next-Generation Sequencing. *J. Clin. Microbiol*. **2021**, 59: e02784-20.
9. Ochoa, L.; Cruz-Flores, R.; Dhar, A.K. Detection and Phylogenetic Analysis of Taura Syndrome Virus from Archived Davidson's-Fixed Paraffin-Embedded Shrimp Tissue. *Viruses* **2020**, 12, 1030.
10. Lee, Q.; Padula, M.P.; Pinello, N.; Williams, S.H.; O'Rourke, M.B.; Fumagalli, M.J.; Orkin, J.D.; Song, R.; Shaban, B.; Brenner, O.; Pimanda, J.E.; Weninger, W.; Marciel de Souza, W.; Melin, A.D.; Wong, J.J.L.; Crim, M.J.; Monette, S.; Roediger, B.; Jolly, C.J. Murine and Related Chapparvoviruses are Nephron-trophic and Produce Novel Accessory Proteins in Infected Kidneys. *PLOS Pathogens* **2020**, 16(1): e1008262.
11. Preventative Control of Sequencing Through the Insert with the Agilent 5200 Fragment Analyzer System. *Agilent Technologies application note,* publication number 5994-3985EN, **2021**.
12. Mitchell, P.K.; Cronk, B.D.; Voorhees, I.E.H.; Rothenheber, D.; Anderson, R.R.; Chan, T.H.; Wasik, B.R.; Dubovi, E.J.; Parrish, C.R.; Goodman, L.B. Method Comparison of Targeted Influenza A Virus Typing and Whole-Genome Sequencing from Respiratory Specimens of Companion Animals. *J. Vet. Diag. Invest.* **2020**, 1-11.
13. Panzera, Y.; Ramos, N.; Frabasile, S.; Calleros, L.; Marandino, A.; Tomás, G.; Techera, C.; Brecco, S.; Fuques, E.; Goñi, N.; Rama, V.; Coppola, L.; Chiparelli, H.; Sorhouet, C.; Mogdasy, C.; Aribiza, J.; Delfraro, A.; Pérez, R. A Deletion in SARS-CoV-2 ORF7 Identified in COVID-19 Outbreak in Uruguay. *Transbound Emerg. Dis*. **2021**, 00, 1-8.
14. Zhou, L.; Chandrasekaran, A.R.; Punnoose, J.A.; Bonenfant, G.; Charles, S.; Levchenko, O.; Badu, P.; Cavaliere, C.; Pager, C.T.; Halvorsen, K. Programmable Low-Cost DNA-Based Platform for Viral RNA Detection. *Sci. Adv.* **2020**, 6: eabc6246.
15. Etievant, S.; Bal, A.; Escuret, V.; Brengel-Pesce, K.; Bouscambert, M.; Cheynet, V.; Generenaz, L.; Oriol, G.; Destras, G.; Billaud, G.; Josset, L.; Frobert, E.; Morfin, F.; Gaymard, A. Performance Assessment of SARS-CoV-2 PCR Assays Developed by WHO Referral Laboratories. *J. Clin. Med*. **2020**, 9, 1871.

www.agilent.com/en/solutions/genomics-applications-solutions/sample-qc/ngs

PR7000-8198

Published in the USA, September 20, 2021
5994-4064EN

Agilent
Trusted Answers