



Ultrarapid On-Site Detection of SARS-CoV-2 Infection Using Simple ATR-FTIR Spectroscopy and an Analysis Algorithm: High Sensitivity and Specificity

Valério G. Barauna,* Maneesh N. Singh, Leonardo Leal Barbosa, Wena Dantas Marcarini, Paula Frizera Vassallo, Jose Geraldo Mill, Rodrigo Ribeiro-Rodrigues, Luciene C. G. Campos, Patrick H. Warnke, and Francis L. Martin*



Cite This: *Anal. Chem.* 2021, 93, 2950–2958



Read Online

ACCESS |



Metrics & More

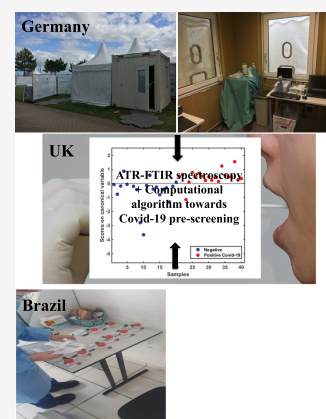


Article Recommendations



Supporting Information

ABSTRACT: There is an urgent need for ultrarapid testing regimens to detect the severe acute respiratory syndrome coronavirus 2 [SARS-CoV-2] infections in real-time within seconds to stop its spread. Current testing approaches for this RNA virus focus primarily on diagnosis by RT-qPCR, which is time-consuming, costly, often inaccurate, and impractical for general population rollout due to the need for laboratory processing. The latency until the test result arrives with the patient has led to further virus spread. Furthermore, latest antigen rapid tests still require 15–30 min processing time and are challenging to handle. Despite increased polymerase chain reaction (PCR)-test and antigen-test efforts, the pandemic continues to evolve worldwide. Herein, we developed a superfast, reagent-free, and nondestructive approach of attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy with subsequent chemometric analysis toward the prescreening of virus-infected samples. Contrived saliva samples spiked with inactivated γ -irradiated COVID-19 virus particles at levels down to 1582 copies/mL generated infrared (IR) spectra with a good signal-to-noise ratio. Predominant virus spectral peaks are tentatively associated with nucleic acid bands, including RNA. At low copy numbers, the presence of a virus particle was found to be capable of modifying the IR spectral signature of saliva, again with discriminating wavenumbers primarily associated with RNA. Discrimination was also achievable following ATR-FTIR spectral analysis of swabs immersed in saliva variously spiked with virus. Next, we nested our test system in a clinical setting wherein participants were recruited to provide demographic details, symptoms, parallel RT-qPCR testing, and the acquisition of pharyngeal swabs for ATR-FTIR spectral analysis. Initial categorization of swab samples into negative versus positive COVID-19 infection was based on symptoms and PCR results ($n = 111$ negatives and 70 positives). Following training and validation (using $n = 61$ negatives and 20 positives) of a genetic algorithm-linear discriminant analysis (GA-LDA) algorithm, a blind sensitivity of 95% and specificity of 89% was achieved. This prompt approach generates results within 2 min and is applicable in areas with increased people traffic that require sudden test results such as airports, events, or gate controls.



INTRODUCTION

In early 2020, a new strain of coronavirus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), more commonly known as causing the COVID-19 disease, gave rise to a global pandemic.¹ Starting from an epidemic outbreak in Wuhan (China), the virus quickly spread westwards toward Europe and the USA² with serious health and socioeconomic consequences worldwide.³ SARS-CoV-2 exhibits a high propensity for infectious spread throughout populations.² Every COVID-19 positive case, if not contained, can readily spread to two or more people giving a virulent R number.⁴ Some countries, such as South Korea, initially successfully fought the COVID-19 outbreak. This is based on the key aspects⁵ of: (a) prevention, via good cleaning practices and isolation of potential cases; (b) testing, to identify those infected and to precisely isolate risk cases; and (c) antiviral treatment and, in the future, a

vaccine. Testing is fundamental to identify infected people and regions of risk.⁶ This can enable intelligent isolation of areas without affecting an entire country's economy and allow allocation of resources to more strategically fight the disease, with more ventilators, medication, and medical staff assigned to regions with more diagnosed cases.

The main challenges for testing are the cost and, in particular, time required for each test result. Gold-standard diagnosis by RT-qPCR is costly with a shortage of testing facilities even in

Received: November 1, 2020

Accepted: January 12, 2021

Published: January 22, 2021



developed countries and can take >2 days to get the result because specimens have to be transported for processing to often distant laboratories.⁷ This is not suitable for mass testing.⁸ Despite globally increased polymerase chain reaction (PCR)-test efforts, the pandemic was not brought to a halt. In contrast, there is a recurrence and second wave of the disease because many infectious patients spread the disease while waiting on their PCR-test results. There are some companies developing quicker and lower-cost tests based on novel sensors.⁹ Alternative antigen- or antibody-detection approaches remain unproven, again maybe creating a statistical bias that could directly affect public health policies.¹⁰ Whilst initial reports claim a sensitivity of 100%, more recent studies exhibit levels as low as 30%;¹¹ others report 72 or 81% in different settings.¹² At this stage, experimental claims of sensitivity and specificity for antigen tests remain to be robustly proven.¹³ Thus, there is still a need to develop COVID-19 test approaches that can deliver results in real-time and on-site.

Vibrational spectroscopy, including attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy, has been widely used to discriminate and classify normal and pathological populations using different cell types, tissues, or biofluids.^{14–16} Readily accessible biofluids, such as blood plasma/serum, saliva, or urine, are considered ideal for clinical implementation due to routine methods of collection, as well as minimal sample preparation.¹⁷ Interrogation of samples with infrared (IR) spectroscopic techniques allows for the generation of a “spectral fingerprint”, which subsequently facilitates the discrimination of the different populations and identification of potential biomarkers.¹⁸ In the past few years, biofluid-based ATR-FTIR spectroscopy has been used for diagnosing, screening, or monitoring the progression/regression in a variety of diseases.¹⁹ Spectroscopic techniques are rapid, cost-effective, and non-destructive, which make them a perfect candidate for translation to clinic, even as an adjunct to more established methods.

As a readily accessible noninvasive biofluid, saliva is an ideal candidate to facilitate disease detection; indeed, oral health has long been known to be an indicator of whole organism health.²⁰ Herein, ATR-FTIR spectroscopy was used to interrogate saliva samples on pharyngeal swabs taken from individuals with or without suspected infection with COVID-19. Unlike many tests developed using laboratory-based contrived specimens, we trialed the approach in clinical settings on real-world samples. Our goal was to differentiate individuals with active infection based on a series of spectral biomarkers. We also took into consideration symptoms and other demographic features of our participants as confounding factors (see the [Supporting Information](#)). We propose a new, ultrarapid on-site method to detect COVID-19 based on pharyngeal swabs using IR light, with potential for ready implementation in general population settings. This approach is not designed to replace existing diagnostic methods such as RT-qPCR but to serve as a rapid prescreening tool to allow the ready movement of population interactions typical of an open economy.

METHODS

Ethical Approval. This study was carried out in agreement with the Helsinki declaration and authorized by the Hospitals Directive due to the emergency situation.²¹ Ethical approval for the investigation was granted by the Ethics Committee Federal University of Espirito Santo (#0993920.1.0000.5071 and #31411420.9.0000.8207). Full ethical approval was given to undertake the studies described herein. All procedures and

possible risks were explained to participants before they provided written consent.

Participant Recruitment and Swab Collection. Pharyngeal cotton swabs (FirstLab, Brazil) were from individuals who came to one of the six hospitals participating in the study and met the criteria for suspected cases according to the State Health Secretary and World Health Organization (WHO) guidelines between June and September 2020. For all participants, demographic data (age, gender, pre-existing medical conditions, symptoms, and date of symptoms' onset) were collected (see the [Supporting Information](#)). Exclusion criteria from this study were those with inconclusive RT-qPCR results after two rounds of RT-qPCR.

For the gold-standard protocol via diagnosis by RT-qPCR, a nasopharyngeal swab was collected from participants by inserting a rayon swab with a plastic shaft into the nostril parallel to the palate. The swab was inserted to a location equidistant from the nostril and the outer opening of the ear and was gently scraped for a few seconds to absorb secretions. The swab was then placed immediately into a sterile tube containing a viral transport medium. RT-qPCR was performed in the Central Laboratory from the Health Secretary of Espirito Santo (LACEN-SESA) to allow definitive diagnosis of COVID-19 infection.

For ATR-FTIR spectroscopy, a pharyngeal swab was collected from participants by inserting a cotton swab into the mouth and scrapping the tonsils, the tongue, and the inner part of the cheek. The swab was then placed immediately into a sterile tube and stored on ice until analysis.

Parallel RT-qPCR Testing. Samples were taken simultaneously as nasopharyngeal swabs for PCR testing. In the clinical setting, all PCRs were locally or nationally approved tests. All samples were analyzed at the same state-approved laboratory.

Nucleic acid extraction and real-time RT-qPCR for virus detection were performed to allow the identification of SARS-CoV-2. The extraction of total nucleic acid (DNA and RNA) from collected samples was performed using the BioGene Extraction kit (Bioclin, K204-4, Brazil), following the manufacturer's instructions. Specimens were handled under the laboratory biosafety guidance required for the novel coronavirus (2019-nCoV) designated by WHO at the Central Laboratory of the Espirito Santo state (LACEN-ES). A combination of four tests was employed to detect viral RNA. The first was using the IDT (Integrated DNA Technologies; Coralville, IA) kit, which is developed in association with the CDC and employs primers and probes for the N1, N2, and RP genes. The second was Maccura (designed by Maccura Biotechnology Co., Hi-tech Zone, Chengdu, China), which is a single-well triple target assay and identifies three genes from SARS-CoV-2 (E, N, and ORF1ab) and provides a separate positive internal control (IC). The third was the Molecular SARS-CoV-2 (E/RP genes) kit (Instituto de Tecnologia em Imunobiológicos, Bio-Manguinhos, FioCruz, RJ, Brazil), which uses primers and probes, as reported by Corman et al.²² The detection of viral RNA was carried out on an ABI 7500 real-time PCR machine (Applied Biosystems, Weiterstadt, Germany) using the published protocol and the sequence of primers and probe for E gene and RNase P. Lastly, the IBMP (Instituto de Biologia Molecular do Paraná—FioCRUZ) kit was employed, which is a single-well test and detects N and ORF1ab genes, and uses the RP gene as an internal control.

All assays were performed using manufacturers' recommendations. First, all samples were tested in a single-well assay

(IBMP or Maccura) for the qPCR run, and interpretations of all results were added to a spreadsheet, together with the values of Cts obtained. Samples with inconclusive results, either by nonamplification in the internal control or by nonamplification of another gene, were tested with the other two qPCR kits (IDT or Bio-manguinhos gene E). If the PCR result remained inconclusive, the result and Ct values were added to the spreadsheet as a negative.

ATR-FTIR Spectral Analyses of Pharyngeal Swabs. FTIR spectra data (wavenumber range 4000–650 cm^{-1}) for each swab (see Table S1) were obtained by directly placing the saliva swab on a portable Agilent Cary 630 FTIR Spectrometer equipped with an ATR ZnSe crystal (Agilent, Santa Clara, CA) and Microlab PC software run from a dedicated laptop. Each whole spectrum contains 1798 points (1.86 cm^{-1} spectral resolution). For every ATR-FTIR spectroscopic measurement, three spectra were obtained from each saliva swab. Each swab analysis was performed with 32 coadditions, interspersed with 32 background scans. After each analysis, the swab was removed from the crystal and the crystal was cleaned with MilliQ water and 70% alcohol, thus avoiding intersample contamination. The robustness of this analytical method was trialed in a 2 week demonstration in Kiel (September 2020) during which 625 pharyngeal swabs were taken and analyzed using ATR-FTIR spectroscopy. Only three swab analyses in the entire spectral data set generated outliers (data not shown).

Spiking Experiments to Determine the Limit of Detection (LoD). SARS-CoV-2 virus-particle stock solutions were a generous donation from Prof. Viviane F. Botosso (Instituto Butantan, Brazil) and Prof. Edison L. Durigon (Department of Microbiology, Universidade de São Paulo, Brazil). In brief, a SARS-CoV-2 isolate [HIAE-02: SARS-CoV-2/SP02/human/2020/BRA (GenBank accession number MT 126808.1)] was employed. The virus was propagated in African green monkey kidney Vero cells (ATCC CCL-81) maintained in a Dulbecco's modified Eagle medium (DMEM) supplemented with 5% fetal bovine serum (FBS, Gibco), 1% nonessential amino acids (NEAA), 1% sodium pyruvate (Sigma-Aldrich Co., Deisenhofen, Germany), and incubated in a humidified atmosphere at 37 °C and 5% CO_2 . The viral particle was isolated from the supernatant. To clarify and remove cellular waste or debris in the supernatant, centrifugation and diafiltration cycles were performed. Aliquots of the clear supernatant were transferred to cryogenic tubes and stored at -80 °C in a freezer prior to irradiation treatment. The full protocol for γ -irradiated inactivated virus particle remains to be published.

For spiking experiments, γ -irradiated inactivated SARS-CoV-2 virus particles (from a stock solution of 1×10^5 copies/mL deionized water) were mixed in various copy number concentrations in saliva taken from a 42-year-old male classified as a negative for infection. The following protocols were undertaken:

- (1) γ -Radiation-inactivated COVID-19 virus solution (4 μl) was applied to the ATR diamond and allowed to dry for 4–5 min. Then, serial dilutions of the virus in deionized water were analyzed in a similar fashion.
- (2) A series of serial dilutions in saliva from a negative study participant were generated and applied to the ATR diamond and allowed to dry for 4–5 min.
- (3) Saliva (15 μl) spiked with γ -radiation-inactivated virus (step (2)) was added to a cotton swab. The saliva cotton

swab was then applied straight to the ATR diamond and immediately analyzed.

Data Preprocessing and Analysis. Preprocessing²³ and data analysis were carried out using MATLAB 2014b (The Math Works, MA). Data analysis was performed using three MATLAB toolboxes: PLS Toolbox version 7.9.3 for preprocessing (Eigenvector Research, Inc.), GA-LDA for feature selection (available at <https://doi.org/10.6084/m9.figshare.3479003.v1>), and the Classification Toolbox for MATLAB used for graphical outputs of an LDA algorithm (available at <https://michem.unimib.it/download/matlab-toolboxes/classification-toolbox-for-matlab/>).²⁴ The spectra were preprocessed by truncating the fingerprint region (1800–900 cm^{-1}), followed by Savitzky–Golay smoothing (9 point window, 2nd order polynomial fitting), automatic weighted least-squares baseline correction, and vector normalization. The triplicate replicate spectra per sample were averaged before model construction. Toward exploratory data analyses, following preprocessing of raw spectra, spectral data were mean-centered and evaluated by means of principal component analysis (PCA).²⁵ PCA is an unsupervised technique that reduces the spectral data space to principal components (PCs) responsible for the majority of variance in the original data set. Each PC is orthogonal to each other, where the first PC accounts to the maximum explained variance followed by the second PC and so on. The PCs are composed of scores and loadings, where the first represents the variance in the sample direction, thus used to assess similarities/dissimilarities among the samples; the latter represents the contribution of each variable for the model decomposition, thus used to find important spectral markers. This technique looks for inherent similarities/differences and provides a scores matrix representing the overall “identity” of each sample, a loading matrix representing the spectral profile in each PC, and a residual matrix containing the unexplained data. Scores information can be used for exploratory analysis providing possible classification between data classes.

PCA was the method of choice for analyzing saliva samples spiked with an inactivated virus particle. It is simple, fast, and combines exploratory analysis, data reduction, and feature extraction into one single method. PCA scores were used to explore overall data set variance and any clustering related to the limit of detection, while the loadings on the first two PCs were used to derive specific biomarkers indicative of the infection category.

A genetic algorithm (GA) is a variable selection technique used to reduce the spectral data space into a few variables and works by simulating the data throughout an evolutionary process.^{26,27} The original space is maintained for both algorithms and no transformation is made as in PCA. Therefore, the selected variables have the same meaning of the original ones (i.e., wavenumbers), and they are responsible for the region where there are more differences between the classes being analyzed or, in other words, between the chemical changes.

For all classification models, samples were divided into training (50 designated negative and 50 designated positive for COVID-19 infection based on symptoms and RT-PCR; see Tables S2 and S3) and validation ($n = 61$ designated negative and 20 designated positive for COVID-19 infection based on symptoms and RT-PCR) sets by applying the Kennard–Stone (KS) uniform sampling selection algorithm.²³ The training samples were used in the modeling procedure, whereas the prediction set was only used in the final classification evaluation

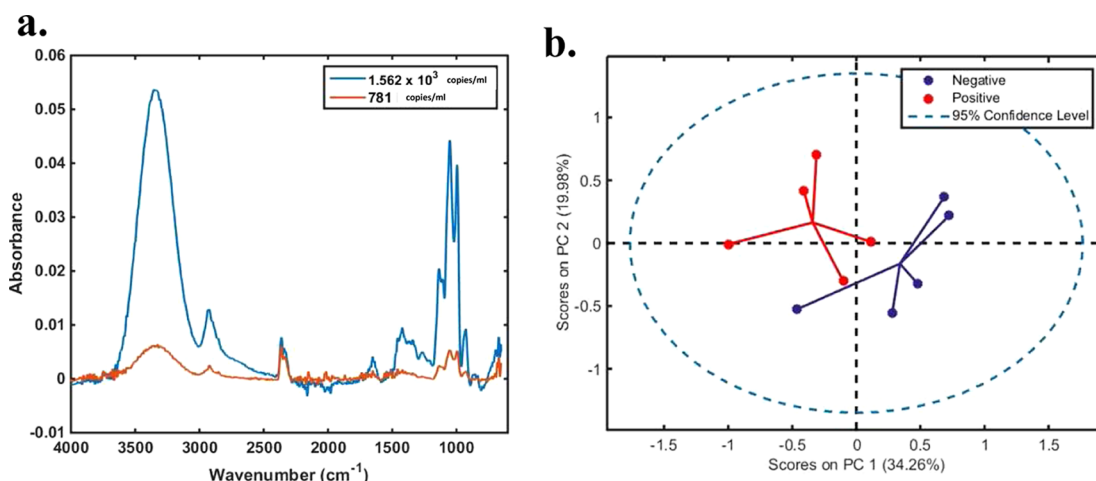


Figure 1. Preliminary SARS-CoV-2 analyses employing ATR-FTIR spectroscopy. (a) Spectra of pure virus in saliva (whole virus inactivated by γ -radiation) [e SARS.CoV2/SP02.2020.HIAE.Br (GenBank accession number MT 126808.1)]. (b) Graphical demonstration of separation in a PCA scores plot of positive and negative samples in RNA-extracted samples prepared for PCR analyzed by biospectroscopy.

using the LDA discriminant approach. The optimal number of variables for GA was determined with an average risk G of LDA misclassification. Such cost function is calculated in a subset of the training set as

$$G = \frac{1}{N_V} \sum_{n=1}^{N_V} g_n \quad (1)$$

where g_n is defined as

$$g_n = \frac{r^2(x_n, m_{I(n)})}{\min_{I(m) \neq I(n)} r^2(x_n, m_{I(m)})} \quad (2)$$

where the numerator is the squared Mahalanobis distance between the object x_n and the sample mean $m_{I(n)}$ of its true class, and the denominator is the squared Mahalanobis distance between the object x_n and the mean of the closest wrong class.^{25,28}

The GA calculations were performed during 100 generations with 200 chromosomes each. One-point crossover and mutation probabilities were set to 60 and 10%, respectively. GA is a nondeterministic algorithm, which can give different results by running the same equation/model. Therefore, the algorithm was repeated three times, starting from random initial populations, with the best solution resulting from the three realizations of GA employed.

Sensitivity (the probability that a test result will be positive when the disease is present) and specificity (the probability that a test result will be negative when the disease is not present) were given by the following equations

$$\text{sensitivity (\%)} = \frac{TP}{TP + FN} \times 100 \quad (3)$$

$$\text{specificity (\%)} = \frac{TN}{TN + FP} \times 100 \quad (4)$$

where TP is defined as true positive, FN as false negative, TN as true negative, and FP as false positive.

RESULTS AND DISCUSSION

Spiking of Saliva with Inactivated COVID-19 Virus.

Figure 1a shows a typical spectrum of inactivated γ -irradiated COVID-19 virus particles [e SARS.CoV2/SP02.2020.HIAE.Br

(GenBank accession number MT 126808.1)];²¹ at 1582 copies/mL, an ATR-FTIR spectrum with a good signal-to-noise ratio (SNR) is obtained. This was to assess the limit of detection (LoD) for biospectroscopy to ascertain the minimum concentration at which the virus could be detected by IR spectroscopy. Below this level, the SNR becomes poor and noisy. This clearly points to the ability of ATR-FTIR spectroscopy to extract a unique viral fingerprint consistent of spectral features associated with a pure virus spectrum. It is interesting to note that the predominant spectral peaks can be tentatively associated with nucleic acid bands, including RNA. Following nucleic acid (RNA/DNA) extraction of saliva samples obtained from participants either positive ($n = 5$) or negative ($n = 5$) for COVID-19, clear segregation of spectral data points is obtained using exploratory principal component analysis (PCA) (Figure 1b).

Control saliva from a human participant (male, 42 years and RT-qPCR negative) was spiked with various numbers of inactivated γ -irradiated COVID-19 virus particles (Figure 2). At low copy numbers, the virus particle is clearly capable of modifying the IR spectral signature of saliva (Figure 2a,b). Examination of control saliva in comparison with saliva-spiked inactivated virus particle at various copy number levels highlighted an ability to detect virus particle-induced spectral alterations at levels that would be considered extremely low in the pharyngeal cavity of infected humans (symptomatic or asymptomatic). Even more compellingly, when this is examined using basic multivariate analysis (i.e., PCA), the IR spectral signature of pure inactivated virus segregates away from control saliva in a scores plot (Figure 2c). When saliva is spiked with exceptionally low levels of virus (781 copies/mL; a1 cluster below), the spectral points cocluster with control saliva spectral points, suggesting no differences. However, at a level of 12 500 copies/mL (a2 cluster below), there is segregation from the control. It is critical to note that the loading plot specifically identifies RNA as being proportional to virus levels (Figure 2b). The loadings on PC1 show the bands responsible for the increase in the virus concentration (nucleic acid bands), and the loadings on PC2 show the bands responsible for discrimination between saliva and virus (amide I and amide II bands present in saliva but not virus). Other, primarily protein-associated bands

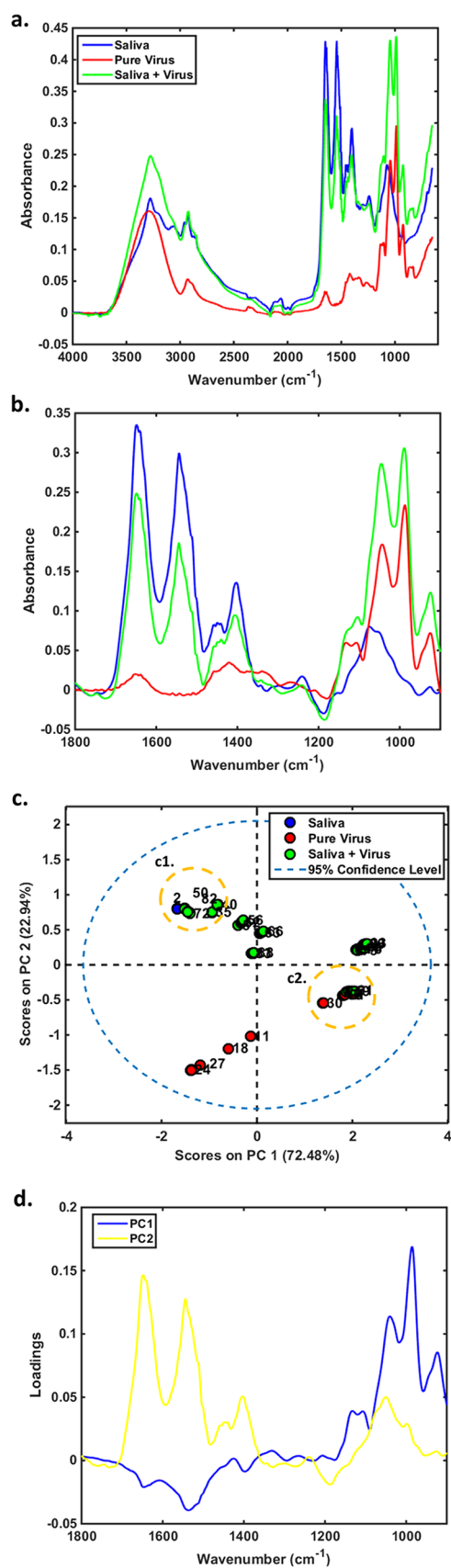


Figure 2. Spiking of control saliva with SARS-CoV-2 virus and analyses using ATR-FTIR spectroscopy. (a) Average raw spectra and (b) preprocessed spectra for saliva ($n = 2$), pure SARS-CoV-2 virus in different concentrations ($n = 28$, 1×10^5 – 98 copies/mL), and saliva + virus in different concentrations ($n = 63$, 1×10^5 – 24 copies/mL). (c)

Figure 2. continued

PCA scores and (d) PCA loadings on PC1 versus PC2 for the preprocessed data. Insets: (c1) mix between saliva and saliva + virus for low concentration (≤ 781 copies/mL) and (c2) mix between pure virus and saliva + virus for high concentration ($\geq 1.25 \times 10^4$ copies/mL). Preprocessing: Savitzky–Golay (SG) smoothing (7 point window, 2nd order polynomial fitting) and baseline correction. The loadings on PC1 show the bands responsible for the increase in the virus concentration (nucleic acid bands), and the loadings on PC2 show the bands responsible for discrimination between saliva and virus (amide I and amide II bands present in saliva but not virus).

discriminate the saliva from the virus—we believe this to be the first report of its kind using biospectroscopy.

Furthermore, in the complex milieu of a saliva sample, which will undoubtedly contain a range of complex constituents including aqueous, exfoliated cellular material, and postinfection immunoglobulins such as IgA and other individual or contaminating factors, a multivariate chemometric approach can still extract the viral-associated discriminating features. Following this, Figure 3 shows the analysis of swabs spiked with either saliva with or without spiking with γ -irradiated COVID-19 virus particles. Figure 3a,b shows spectra with a good SNR. In consequent PCA scores plots, the spectral data points for virus-spiked saliva swabs segregate away from swab or control saliva swab categories (Figure 3c). This is achieved at low copy numbers. The loadings on PC1 show the bands responsible for separation between swab + saliva and swab + saliva + virus (amide I and amide II band of proteins) and the loadings on PC2 show the bands responsible for variation of virus concentration (amide I, amide II, and nucleic acids bands) (Figure 3d). Different from saliva, the swab sample contains bands on the nucleic acids region plus amide I and amide II that may come from the saliva itself.

GA-LDA Segregation of Categories: COVID-19 Infected versus Uninfected. Categorization into the negative (designated not infected by COVID-19) and positive (designated as infected) categories was based on a series of RT-qPCR tests, primarily carried out at Central Laboratory of Espirito Santo State alongside symptoms/outcome (see Tables S2 and S3). Follow-up showed that COVID-19 participants required hospitalization. A Ct < 37 in RT-qPCR designated a PCR-positive result (Figure 4; see the Supporting Information). Metaparameters such as gender, age, or smoking habits were not incorporated into model construction; some participant details such as the presence or the absence of symptoms (see Tables S2 and S3) were used for designation of negative versus positive categories.

The chemometric technique of genetic algorithm-linear discriminant analysis (GA-LDA) was applied toward classification^{25,28,29} of negative versus positive for COVID-19 infection (Figure 5). The classification ratios achieved after GA-LDA were a sensitivity of 95% and specificity of 89% (Table 1). Figure 5a,b shows the full raw spectra across the entire mid-IR spectral range and raw spectra in the fingerprint region for all negative and COVID-19 positive swab samples ($n = 111$ negatives and 70 positives). Spectra were preprocessed [Savitzky–Golay smoothing (9 point window, 2nd order polynomial fitting), automatic weighted least-squares baseline correction, and vector normalization] in the fingerprint region (Figure 5c). Training and validation of GA-LDA were undertaken using 50 negatives and 50 positives; the GA-LDA

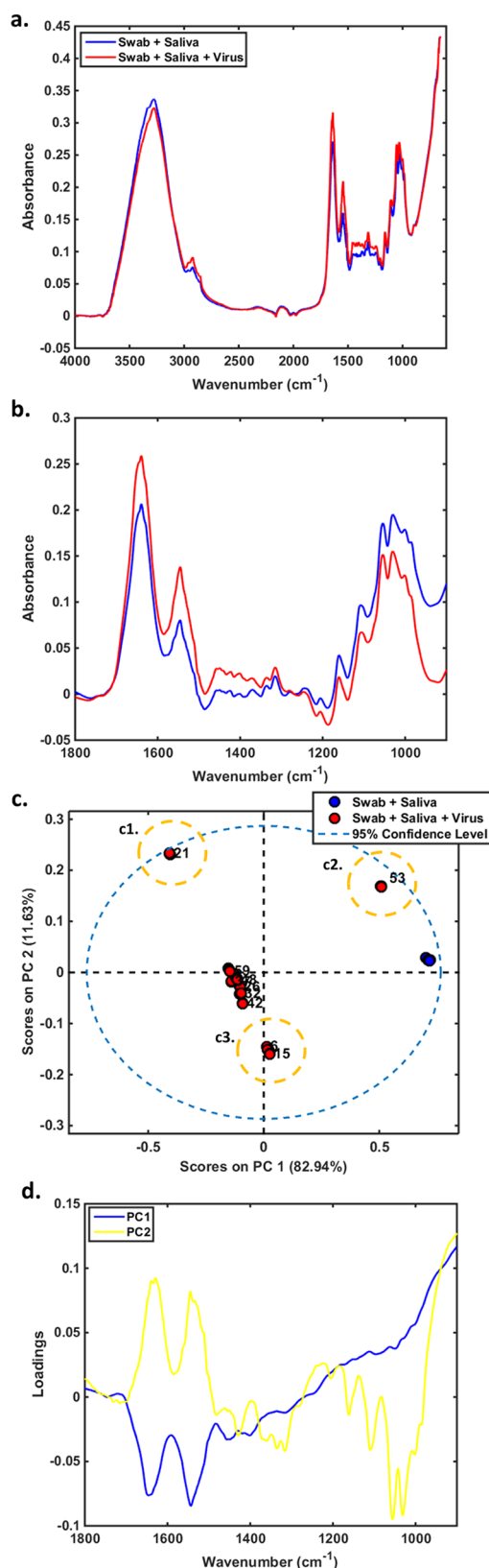


Figure 3. Spiking of swab with saliva or saliva + SARS-CoV-2 virus prior to analyses using ATR-FTIR spectroscopy. (a) Average raw spectra and (b) preprocessed spectra for swab + saliva ($n = 5$) and swab + saliva + virus ($n = 54$, 1×10^5 – 98 copies/mL). (c) PCA scores and (d) PCA loadings on PC1 versus PC2 for the preprocessed data. Insets: (c1) virus concentration around 6.25×10^3 copies/mL, (c2) virus concentration around 1.56×10^3 copies/mL, and (c3) virus

Figure 3. continued

concentration ≤ 781 copies/mL. Preprocessing: Savitzky–Golay (SG) smoothing (7 point window, 2nd order polynomial fitting) and baseline correction. The loadings on PC1 show the bands responsible for separation between swab + saliva and swab + saliva + virus (amide I and amide II band of proteins) and the loadings on PC2 show the bands responsible for variation of virus concentration (amide I, amide II, and nucleic acids bands). Different from saliva, the swab sample contains bands on the nucleic acids region plus amide I and amide II that may come from the saliva itself.

scores plot for the validation set ($n = 61$ negatives and 20 positives) is shown in Figure 5d.

Consequently, five GA-LDA-selected variables were identified, each which significantly ($P < 0.01$) discriminates negative and COVID-19 positive swab samples (Figure 6 and Table 2). Using saliva swab-based vibrational spectroscopy, we achieved results with significant clinical relevance. ATR-FTIR spectroscopy has been proven to be capable of distinguishing between patient and healthy group negative and COVID-19 positive swab samples. A plausible mechanistic basis for this is that the prominent distinguishing features extracted are primarily associated with nucleic acids, RNA in particular. Four of five molecular tentative RNA assignments are higher in the negative group compared to the positive category. A plausible hypothesis is that the 1429 cm^{-1} increase is associated with a virus, e.g., a simple RNA virus. The corresponding decreases at 1220 , 1084 , 1069 , and 1041 cm^{-1} may be associated with a response of the host organism to the virus infection.

Future work will extend toward validation for regulatory approval. We believe that existing spectrometers and current manufacturing capacity of HeNe lasers will be sufficient for the significant rollout of our technology in the event of it being adopted. Even with the vaccines (efficacy and safety remain to be proven), which are being adopted and the possible emergence of herd immunity (given that the SARS-CoV-2 virus is a mutating RNA virus, this may not occur), there will remain a need for risk reduction strategies to get worldwide economies moving again. In a demonstration in September 2020 at the Kiel (Schleswig-Holstein, Germany) Regatta, which is an annual sailing event (plus Olympic qualifier), we trialed this method (data not shown). Over 2 weeks, some 620 pharyngeal swabs from workers and athletes (from some 40 countries) were analyzed; in comparison with RT-qPCR, we had 100% specificity, which excluded the possibility of cross-reactivity. We believe this approach could translate as a viable option in the battle against SARS-CoV-2.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.0c04608>.

An excel spreadsheet containing spectral data for the negative and positive categories for SARS-CoV-2 infection alongside study participant demographic information (XLSX)

■ AUTHOR INFORMATION

Corresponding Authors

Valério G. Barauna – Department of Physiological Sciences, Federal University of Espírito Santo, 29075-910 Vitória, Brazil; Email: barauna2@gmail.com

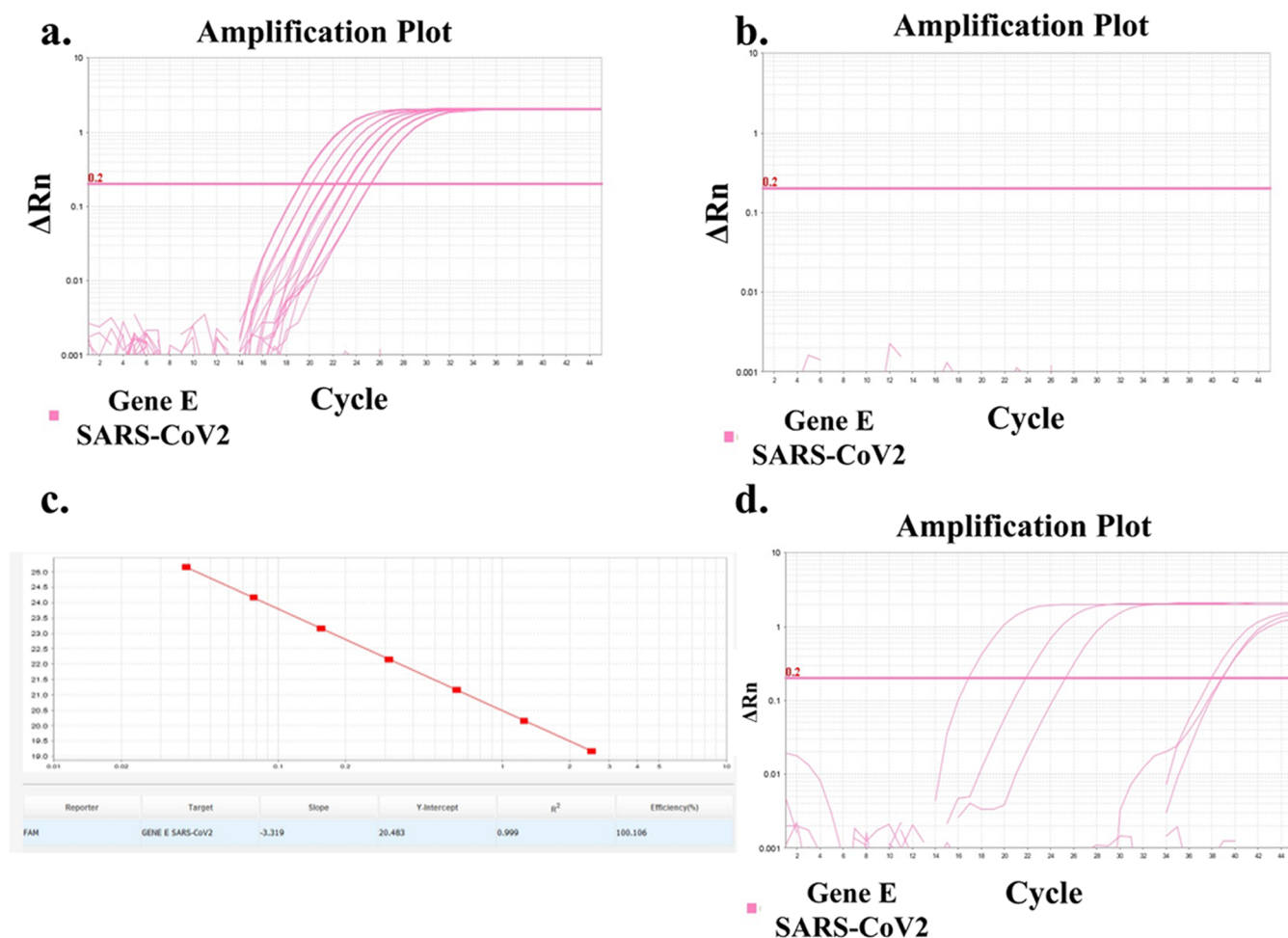


Figure 4. RT-qPCR of samples. A tiered RT-qPCR system was employed for the analyses of parallel nasopharyngeal samples taken from study participants. (a) The samples were considered positive if the E gene was amplified with Ct < 37. (b) A standard curve is below alongside a negative control. (c) The efficiency curve to determine the threshold. (d) Three positive samples are shown juxtaposed with three negative samples (no amplification) and three with a Ct > 37 (negative/inconclusive).

Francis L. Martin – Biocel UK Ltd., Hull HU10 6TS, U.K.;
 orcid.org/0000-0001-8562-4944; Email: flm13@biocel.uk

Patrick H. Warnke – Praxisklinik am Ballastkai, 24937 Flensburg, Germany; Department of OMF-Surgery, Christian-Albrechts-University of Kiel, 24118 Kiel, Germany

Authors

Maneesh N. Singh – Biocel UK Ltd., Hull HU10 6TS, U.K.
 Leonardo Leal Barbosa – Department of Physiological Sciences, Federal University of Espírito Santo, 29075-910 Vitoria, Brazil
 Wena Dantas Marcarini – Department of Physiological Sciences, Federal University of Espírito Santo, 29075-910 Vitoria, Brazil
 Paula Frizzera Vassallo – Department of Physiological Sciences, Federal University of Espírito Santo, 29075-910 Vitoria, Brazil; Clinical Hospital, Federal University of Minas Gerais, 31270-901 Belo Horizonte, Brazil
 Jose Geraldo Mill – Department of Physiological Sciences, Federal University of Espírito Santo, 29075-910 Vitoria, Brazil
 Rodrigo Ribeiro-Rodrigues – Núcleo de Doenças Infecciosas, Federal University of Espírito Santo, 29075-910 Vitoria, Brazil
 Luciene C. G. Campos – Department of Biological Science, Santa Cruz State University, 45662-900 Bahia, Brazil

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.analchem.0c04608>

Author Contributions

V.G.B., P.F.V., and J.G.M. were involved in experimental design, writing, review and editing, supervision, project administration, and funding acquisition; V.G.B., L.L.B., and W.D.M. were involved in data acquisition, analysis, and interpretation; M.N.S. and P.H.W. provided clinical expertise, data analytical input, and context throughout the course of the study; R.R.-R. undertook and coordinated the RT-qPCR analyses and validation; F.L.M. codesigned the study, was involved in funding acquisition, was principal investigator, and wrote the manuscript. All authors approved the finalized manuscript.

Notes

The authors declare the following competing financial interest(s): M.N.S., P.H.W., and F.L.M. hold positions and shareholdings in Biocel UK Ltd. and its subsidiary companies; these companies are developing a spectrochemical test for SARS-CoV-2 prescreening for commercial gain. To this end, they are also currently seeking regulatory approval through the

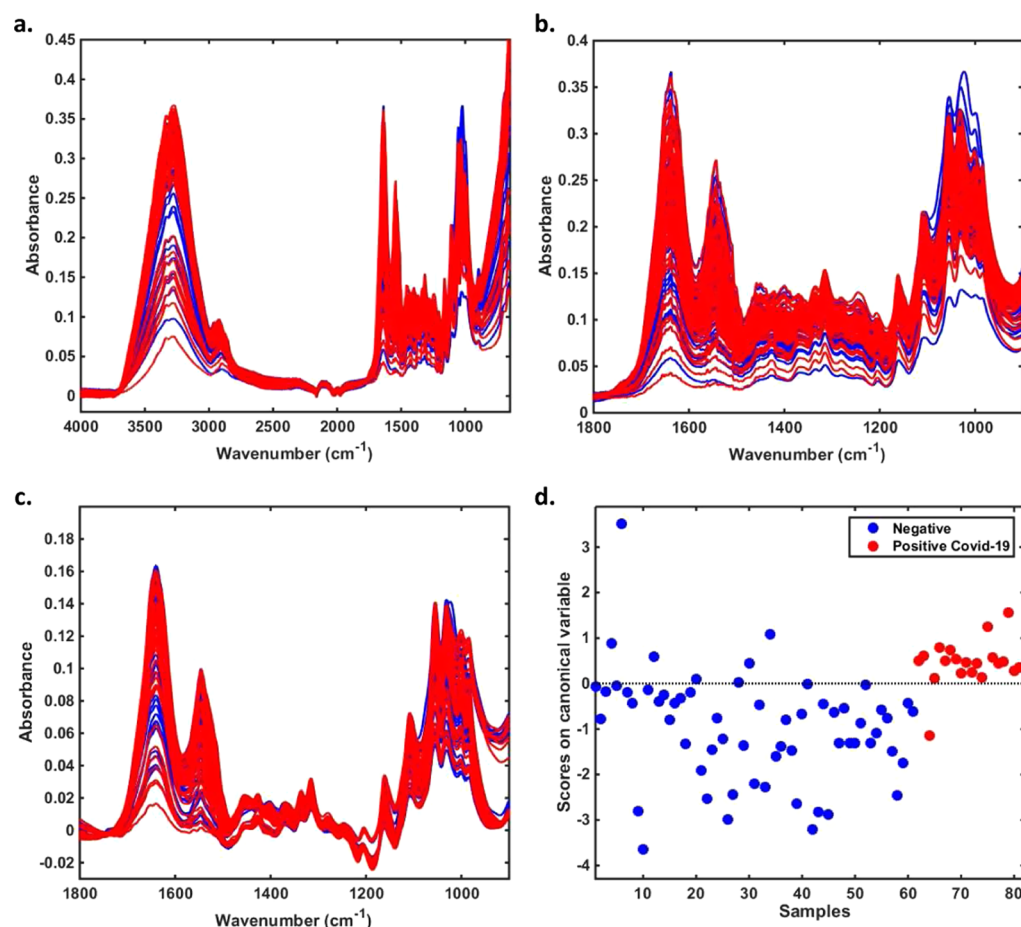


Figure 5. Analyses of pharyngeal swabs using ATR-FTIR spectroscopy in a clinical setting. (a) Full raw spectra and (b) raw spectra in the fingerprint region for all negative and COVID-19 positive swab samples ($n = 111$ negatives and 70 positives). (c) Preprocessed spectra [Savitzky–Golay smoothing (9 point window, 2nd order polynomial fitting), automatic weighted least-squares baseline correction, and vector normalization] in the fingerprint region and the (d) GA-LDA score plot for the validation set ($n = 61$ negatives and 20 positives).

Table 1. Confusion Matrix Showing the Number of Patients and Figures of Merit for the Validation Set Using GA-LDA Algorithm Preprocessing^a

	baseline correction and vector normalization ^b	
	predicted negative	predicted positive
negative	54	7
positive	1	19
parameters		
accuracy	90% (95% CI, 76–97%)	
sensitivity	95% (95% CI, 73–100%)	
specificity	89% (95% CI, 77–95%)	
F-score	92% (95% CI, 75–97%)	

^a95% CI: 95% confidence interval. ^bPreprocessing: Savitzky–Golay smoothing (9 point window, 2nd order polynomial fitting), automatic weighted least-squares baseline correction, and vector normalization.

requisite US [Food and Drug Administration (FDA)], UK [Medicines and Healthcare products Regulatory Agency (MHRA)], and European (CE marking) bodies. The other co-authors do not have a conflict of interest to declare.

ACKNOWLEDGMENTS

The authors would like to thank LabPetro (UFES, Brazil) for technical support with the ATR-FTIR spectroscopy. The authors also would like to thank all of the hospitals (Vila

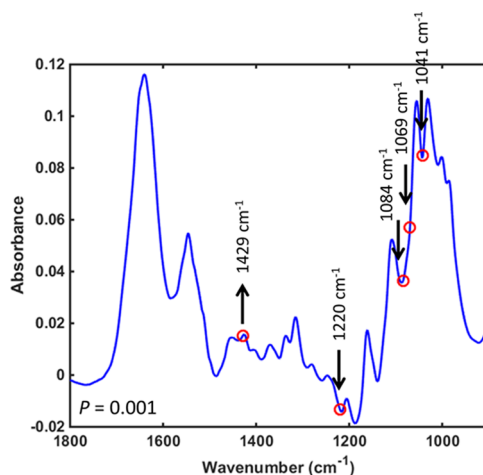


Figure 6. GA-LDA selected variables. Arrow \uparrow : higher absorbance in the COVID-19 positive class. Arrow \downarrow : higher absorbance in the COVID-19 negative class. P -value calculated using a MANOVA test with all five GA-LDA-selected variables between all negative and positive samples.

Velha Hospital, Hospital Universitário Cassiano Antonio Moraes, Pronto Atendimento da Glória, Hospital Roberto Arnizaut Silveiras), employees, nurses, doctors, and patients that accepted to participate in this study. The authors would like to

Table 2. Selected Variables by GA-LDA With Their Respective Tentative Assignments

variable (cm ⁻¹)	tentative assignment
1429	$\delta(\text{CH}_2)$ polysaccharides
1220	asymmetric PO ₂ ⁻ stretching in RNA and DNA
1084	symmetric PO ₂ ⁻ stretching in nucleic acids
1069	C–O stretching in ribose
1041	symmetric PO ₂ ⁻ stretching in nucleic acids

thank the Instituto Capixaba de Ensino, Pesquisa e Inovação em Saúde (ICEPi), and the Secretaria Estadual de Saúde do Espírito Santo (SESA) for clinical data of patients and RT-PCR results. This study was supported by FAPES (#151/2020), CNPq (#401870/2020-0), and UESC (#073.11012.2020.0007594-29). The authors also thank Professor Edison Luiz Durigon and Dr. Danielle BL de Oliveira from the Department of Microbiology at the Institute of Biomedical Science (ICB) at the University of São Paulo, USP (Brazil) for providing the inactivated SARs-CoV-2 virus strain (SARs-CoV2/SP02.2020.HIAE.Br GenBank accession number MT 126808.1).

REFERENCES

- (1) Li, G.; Fan, Y.; Lai, Y.; Han, T.; Li, Z.; Zhou, P.; Pan, P.; Wang, W.; Hu, D.; Liu, X.; Zhang, Q.; Wu, J. *J. Med. Virol.* **2020**, *92*, 424–432.
- (2) Khadke, S.; Ahmed, N.; Ahmed, N.; Ratts, R.; Raju, S.; Gallogly, M.; de Lima, M.; Sohail, M. *Virol. J.* **2020**, *17*, No. 154.
- (3) Maffettone, P.; Oldani, C. *Global Policy* **2020**, *11*, 501–507.
- (4) Bryant, P.; Elofsson, A. *PeerJ* **2020**, *8*, No. e9879.
- (5) Means, A. R.; Wagner, A. D.; Kern, E.; Newman, L. P.; Weiner, B. *J. Front. Public Health* **2020**, *8*, No. 462.
- (6) Vandenberg, O.; Martiny, D.; Rochas, O.; van Belkum, A.; Kozlakidis, Z. *Nat. Rev. Microbiol.* **2020**, DOI: 10.1038/s41579-020-00461-z.
- (7) Steel, J. J.; Sitko, J. C.; Adkins, M. G.; Hasstedt, S. C.; Rohrer, J. W.; Almand, E. A. *Biotechniques* **2020**, *69*, 245–248.
- (8) Pokhrel, P.; Hu, C.; Mao, H. *ACS Sens.* **2020**, *5*, 2283–2296.
- (9) Zhou, L.; Chandrasekaran, A. R.; Punnoose, J. A.; Bonenfant, G.; Charles, S.; Levchenko, O.; Badu, P.; Cavaliere, C.; Pager, C. T.; Halvorsen, K. *Sci. Adv.* **2020**, *6*, No. eabc6246.
- (10) Morel, C.; Lindahl, O.; Özenci, V. *J. Global Health* **2020**, *10*, No. 020330.
- (11) Scohy, A.; Anantharajah, A.; Bodéus, M.; Kabamba-Mukadi, B.; Verroken, A.; Rodriguez-Villalobos, H. *J. Clin. Virol.* **2020**, *129*, No. 104455.
- (12) Gremmels, H.; Winkel, B. M. F.; Schuurman, R.; Rosingh, A.; Rigter, N. A. M.; Rodriguez, O.; Ubijaa, J.; Wensing, A. M. J.; Bonten, M. J. M.; Hofstra, L. M.; et al. *EClinicalMedicine* **2020**, No. 100677.
- (13) Fitzpatrick, M. C.; Pandey, A.; Wells, C. R.; Sah, P.; Galvani, A. P. *Lancet* **2020**, *397*, 24–25.
- (14) Movasaghi, Z.; Rehman, S.; ur Rehman, I. *Appl. Spectrosc. Rev.* **2008**, *43*, 134–179.
- (15) Baker, M. J.; Trevisan, J.; Bassan, P.; Bhargava, R.; Butler, H. J.; Dorling, K. M.; Fielden, P. R.; Fogarty, S. W.; Fullwood, N. J.; Heys, K. A.; Hughes, C.; Lasch, P.; Martin-Hirsch, P. L.; Obinaju, B.; Sockalingum, G. D.; Sulé-Suso, J.; Strong, R. J.; Walsh, M. J.; Wood, B. R.; Gardner, P.; Martin, F. L. *Nat. Protoc.* **2014**, *9*, 1771–1791.
- (16) Mitchell, A. L.; Gajjar, K. B.; Theophilou, G.; Martin, F. L.; Martin-Hirsch, P. L. *J. Biophotonics* **2014**, *7*, 153–165.
- (17) Freitas, D. L. D.; Câmara, I. M.; Silva, P. P.; Wanderley, N. R. S.; Alves, M. B. C.; Morais, C. L. M.; Martin, F. L.; Lajus, T. B. P.; Lima, K. M. G. *Sci. Rep.* **2020**, *10*, No. 12818.
- (18) Martin, F. L.; Kelly, J. G.; Llabjani, V.; Martin-Hirsch, P. L.; Patel, I. I.; Trevisan, J.; Fullwood, N. J.; Walsh, M. J. *Nat. Protoc.* **2010**, *5*, 1748–1760.

(19) Maitra, I.; Morais, C. L. M.; Lima, K. M. G.; Ashton, K. M.; Date, R. S.; Martin, F. L. *Analyst* **2019**, *144*, 7447–7456.

(20) Paraskevaidi, M.; Allsop, D.; Karim, S.; Martin, F. L.; Crean, S. J. *Clin. Med.* **2020**, *9*, No. 1673.

(21) Araujo, D. B.; Machado, R. R. G.; Amgarten, D. E.; Malta, F. M.; de Araujo, G. G.; Monteiro, C. O.; Candido, E. D.; Soares, C. P.; de Menezes, F. G.; Pires, A. C. C.; Santana, R. A. F.; Viana, A. O.; Dorlass, E.; Thomazelli, L.; Ferreira, L. C. S.; Botosso, V. F.; Carvalho, C. R. G.; Oliveira, D. B. L.; Pinho, J. R. R.; Durigon, E. L. *Mem. Inst. Oswaldo Cruz.* **2020**, *115*, No. e200342.

(22) Corman, V. M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D. K.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M. L.; Mulders, D. G.; Haagmans, B. L.; van der Veer, B.; van den Brink, S.; Wijsman, L.; Goderski, G.; Romette, J. L.; Ellis, J.; Zambon, M.; Peiris, M.; Goossens, H.; Reusken, C.; Koopmans, M. P.; Drosten, C. *Eurosurveillance* **2020**, *25*, No. 2000045.

(23) Kennard, R. W.; Stone, L. A. *Technometrics* **1969**, *11*, 137–148.

(24) Ballabio, D.; Consonni, V. *Anal. Methods* **2013**, *5*, 3790–3798.

(25) Lima, K. M. G.; Gajjar, K.; Valasoulis, G.; Nasioutziki, M.; Kyrgiou, M.; Karakitsos, P.; Paraskevaidis, E.; Martin-Hirsch, P. L.; Martin, F. L. *Anal. Methods* **2014**, *6*, 9643–9652.

(26) Bro, R.; Smilde, A. K. *Anal. Methods* **2014**, *6*, 2812–2831.

(27) Paraskevaidi, M.; Morais, C. L. M.; Lima, K. M. G.; Snowden, J. S.; Saxon, J. A.; Richardson, A. M. T.; Jones, M.; Mann, D. M. A.; Allsop, D.; Martin-Hirsch, P. L.; Martin, F. L. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E7929–E7938.

(28) Lima, K. M.; Gajjar, K. B.; Martin-Hirsch, P. L.; Martin, F. L. *Biotechnol. Prog.* **2015**, *31*, 832–839.

(29) Morais, C. L. M.; Lima, K. M. G.; Singh, M.; Martin, F. L. *Nat. Protoc.* **2020**, *15*, 2143–2162.