

Raman Molecular Fingerprints of SARS-CoV-2 British Variant and the Concept of *Raman Barcode*

Giuseppe Pezzotti,* Francesco Boschetto, Eriko Ohgitani, Yuki Fujita, Masaharu Shin-Ya, Tetsuya Adachi, Toshiro Yamamoto, Narisato Kanamura, Elia Marin, Wenliang Zhu, Ichiro Nishimura, and Osam Mazda*

The multiple mutations of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus have created variants with structural differences in both their spike and nucleocapsid proteins. While the functional relevance of these mutations is under continuous scrutiny, current findings have documented their detrimental impact in terms of affinity with host receptors, antibody resistance, and diagnostic sensitivity. Raman spectra collected on two British variant sub-types found in Japan (QK002 and QHN001) are compared with that of the original Japanese isolate (JPN/TY/WK-521), and found bold vibrational differences. These included: i) fractions of sulfur-containing amino acid rotamers, ii) hydrophobic interactions of tyrosine phenol ring, iii) apparent fractions of RNA purines and pyrimidines, and iv) protein secondary structures. Building upon molecular scale results and their statistical validations, the authors propose to represent virus variants with a barcode specially tailored on Raman spectrum. Raman spectroscopy enables fast identification of virus variants, while the Raman barcode facilitates electronic recordkeeping and translates molecular characteristics into information rapidly accessible by users.

1. Introduction

SARS-CoV-2 continues to mutate in order to evade responses of the human immune system, and the mutations should be carefully monitored as they can reduce the effectiveness of vaccines and conceal themselves in diagnostic tests. Among the variants isolated so far, the British B.1.1.7 lineage shows a substantial transmission advantage over other lineages (i.e., 50–100% higher reproduction number).^[1–3] This transmission advantage, and the rapid increase in other strains with similar characteristics,^[4,5] escalates the challenges in controlling the COVID-19 pandemic. An official survey conducted recently by the Japanese National Institute of Infectious Diseases (JNIID) in Osaka and Hyogo Prefectures found that >80% of the cases reported by the beginning of May 2021 were from SARS-CoV-2 variants.^[6] The city of Kobe, in Hyogo

G. Pezzotti, F. Boschetto, Y. Fujita, E. Marin, W. Zhu
Ceramic Physics Laboratory
Kyoto Institute of Technology
Sakyo-ku, Matsugasaki, Kyoto 606-8585, Japan
E-mail: pezzotti@kit.ac.jp

G. Pezzotti, F. Boschetto, E. Ohgitani, M. Shin-Ya, O. Mazda
Department of Immunology
Graduate School of Medical Science
Kyoto Prefectural University of Medicine
Kamigyo-ku, 465 Kajii-cho, Kyoto 602-8566, Japan
E-mail: mazda@koto.kpu-m.ac.jp


G. Pezzotti
Department of Orthopedic Surgery
Tokyo Medical University
6-7-1 Nishi-Shinjuku, Shinjuku-ku, Tokyo 160-0023, Japan

G. Pezzotti
The Center for Advanced Medical Engineering and Informatics
Osaka University
2-2 Yamadaoka, Suita, Osaka 565-0854, Japan

G. Pezzotti
Institute of Biomaterials and Bioengineering
Tokyo Medical and Dental University
2-3-10 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-0062, Japan

G. Pezzotti, T. Adachi, T. Yamamoto, N. Kanamura, E. Marin
Department of Dental Medicine
Graduate School of Medical Science
Kyoto Prefectural University of Medicine
Kamigyo-ku, Kyoto 602-8566, Japan

I. Nishimura
Division of Advanced Prosthodontics
The Jane and Jerry Weintraub Center for Reconstructive Biotechnology
UCLA School of Dentistry
Los Angeles, CA 90095, USA

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/advs.202103287>

© 2021 The Authors. Advanced Science published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.202103287

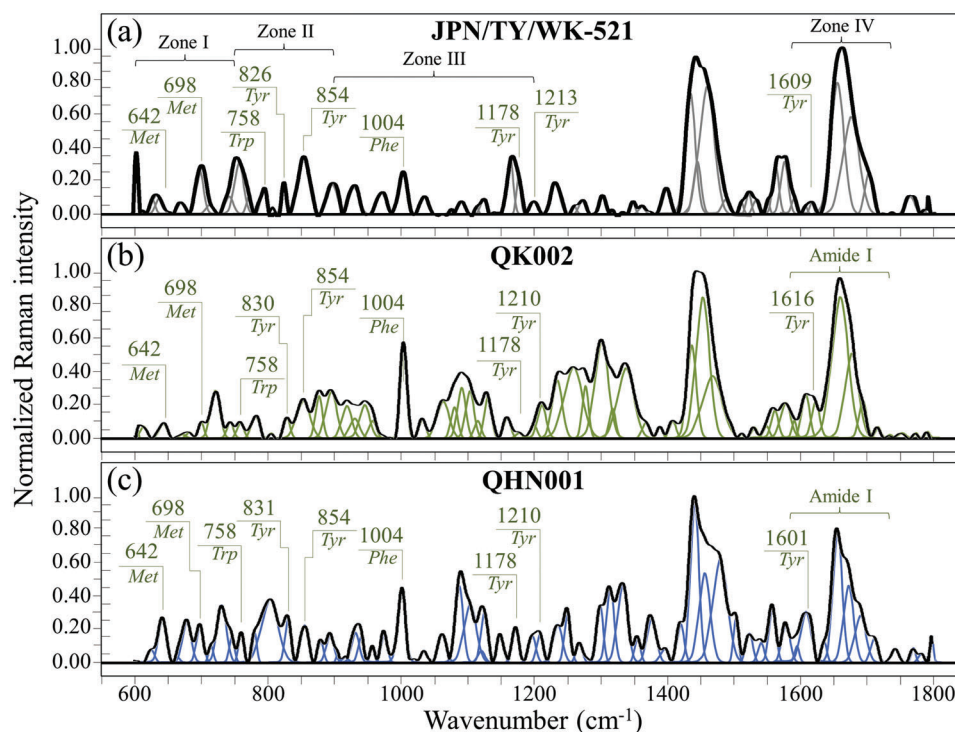


Figure 1. Raman spectra in the frequency interval 600–1800 cm^{-1} of a) the original Japanese isolate JPN/TY/WK-521, b) variant QK002, and c) variant QHN001. Spectra are normalized with respect to the Amide I signal and deconvoluted into Gaussian-Lorentzian band components according to Equation (m-1) in Methods. Four Zones are emphasized in a) and labels show frequencies at maximum of selected bands (Met, Tyr, and Phe are abbreviations for methionine, tyrosine, and phenylalanine, respectively).

Prefecture, recorded $\approx 70\%$ of the reported cases of the British variant. The survey clearly indicates an increasing predominance of cases from SARS-CoV-2 British variants in the Kansai region. Another study recently published by the JNIID examined 803 infections caused by the SARS-CoV-2 British (B.1.1.7 lineage) variant found in Japan and estimated this variant as 1.32 times more infectious than the original Japanese isolate.^[7] However, the British variant may have been excluded from early data and so the estimated infectivity data could be lower than the real ones.

The point mutations so far developed in SARS-CoV-2 have been analyzed for protein coding genes,^[8] and genomic sequences of SARS-CoV-2 reference genomes deposited to a sequence database, which is continuously updated for nucleotides isolates collected from different hosts.^[9–11] High-specificity detection of SARS-CoV-2 virions is usually achieved through real-time reverse transcription-quantitative polymerase chain reactions or serological enzyme-linked immunosorbent assays. However, both methods are time-consuming, fail to distinguish between the presence of virus and its inactivated state, and are incapable of detecting variants. The standard method adopted so far for detecting SARS-CoV-2 variants is next-generation sequencing,^[12] which is also time-consuming, expensive, and thus of limited utility in detecting and monitoring large-scale variants. The above circumstances call for new testing platforms capable of detecting SARS-CoV-2 variants in a rapid and cost-effective way.

In this study, we applied Raman spectroscopy to identify the SARS-CoV-2 British variants found in Japan as distinct from

the original Japanese isolate. Upon preliminary statistical validations based on Pearson's correlation coefficient and using refined instrumentation and spectroscopic procedures, Raman spectra could be recorded in a time frame in the order of the tens of seconds. Clear differences were recorded between the spectra of variant sub-types and that of the original isolate, thus proving the high sensitivity of the Raman approach in locating virus variants and related sub-types. The Raman spectrum directly links to the molecular structure of the virions and provides a chemical fingerprint for different variants. This could be the missing key in on-site virology diagnostics. A spectroscopic analysis based on a machine-learning algorithm specifically crafted for the Raman spectrum is proposed here together with a *Raman barcode* approach to classifying different isolates. Our findings advance the state of the art of Raman spectroscopy in virology by presenting for the first time high-resolution Raman spectra of the SARS-CoV-2 original isolate and the British variants, while proving the possibility of prompt on-site discrimination among different viral variants through vibrational assessments.

2. Results

Figure 1 shows Raman spectra collected on the original Japanese isolate JPN/TY/WK-521 (a), and two British variant sub-types, QK002 (b), and QHN001 (c). An explanatory draft of the virus sample and sample/probe interaction is given in Figure M-1 of Methods (Supporting Information). Four frequency intervals are shown and labeled as Zones I–IV at 600–750, 750–900, 900–

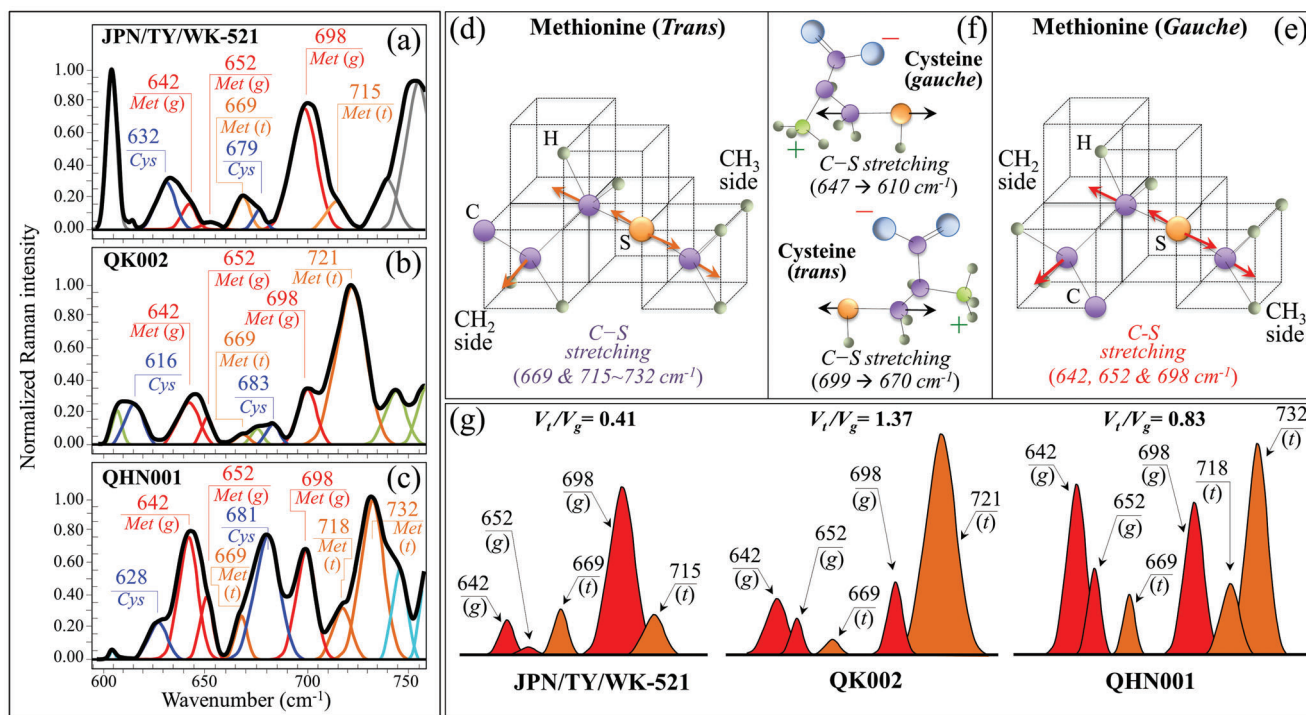


Figure 2. Low-frequency Zone I (600–750 cm^{-1}) of the Raman spectra of a) JPN/TY/WK-521, b) QK002, and c) QHN001 viral strains; spectra are deconvoluted into a sequence of Gaussian-Lorentzian sub-bands (frequencies for selected bands shown in inset). The abbreviations Met and Cys refer to methionine and cysteine, respectively, while (t) and (g) locate trans and gauche rotamers, respectively. Structures and C–S stretching vibrational modes/frequencies of d) trans and e) gauche methionine rotamers; and, f), gauche and trans rotamers of cysteine and C–S stretching vibrational modes/frequencies. In (g), components are shown that represent signals from different rotameric configurations of methionine structure with related trans-to-gauche ratios, V_t/V_g ratios (given in inset together with labels of band frequencies and types of rotamer).

1200, and 1600–1750 cm^{-1} , respectively. The spectra were normalized with respect to their Amide II signal at 1460 cm^{-1} , and deconvoluted into band components according to the algorithm given by Equation (m-1) in Methods (Supporting Information). Frequencies at maximum and proposed vibrational origins are given in the Supporting Information (see Figure S1 and Tables S1a–d, Supporting Information). The spectra appear very different to each other, proving that Raman spectroscopy captured fundamental differences in the molecular structure of variants.

2.1. Rotamers of S-Containing Amino Acids: Spectral Zone I

Figure 2a–c shows the low-frequency Zone I for the Raman spectra of JPN/TY/WK-521, QK002, and QHN001 viral strains. This zone is dominated by vibrational signals relating to the C–S bond,^[13] which is only found in methionine and cysteine incorporated into viral proteins. The structures and main vibrational modes of the methionine and cysteine rotamers are shown in Figure 2d–g (see labels). Reported substitutions in the spike-protein amino acid sequence of the British variants do not mention methionine or cysteine but only spell asparagine-to-tyrosine and aspartate-to-glycine substitutions.^[14,15] However, high-resolution spectra in Figure 2 show significant differences in the fractions of S-containing amino acid rotamers between the two British variant sub-types and the original Japanese isolate, and between the variant sub-types themselves. The Ra-

man spectrum is very sensitive to molecular symmetry, which appears as pronounced spectral differences. This characteristic, coupled with the high sensitivity of the Raman spectrum to C–S bonds, reveals the distinctive structural details of the S-containing molecules in SARS-CoV-2 viral strains. Figure 2d,e show *trans* and *gauche* methionine rotamers and their related C–S stretching modes,^[16,17] C–S stretching bands from methionine residues are found at 642, 652, 669, 698, and 715–732 cm^{-1} (see Figure 2). Bands at 642 and 652 cm^{-1} represent C–S bond stretching vibrations on the CH_2 side of molecules in *gauche* rotameric configuration, while the signal at 669 cm^{-1} arises from the same vibrational mode on the CH_2 side for molecules in *trans* configuration. The signal at ≈ 642 cm^{-1} incorporates contributions from tyrosine (ring deformation mode).^[18] Conversely, signals at 698, and 715–732 cm^{-1} relate to C–S stretching on the CH_3 carboxyl side in *gauche* and *trans* configurations (see Table S1a, Supporting Information).^[16,17]

Methionine is one of the most hydrophobic amino acids present in nature, therefore, its residues are generally located at buried sites in the hydrophobic core of globular proteins, or in links with lipid bilayers in membrane-spanning protein domains. Accordingly, the frequencies of the C–S bands are not expected to show any marked dependence on environmental pH. However, the molecular symmetry characteristics of methionine residues of the spike protein play a key-role in a number of functions during the virus lifetime^[19,20] Importantly, different SARS-CoV-2 isolates possess different fractions of methionine

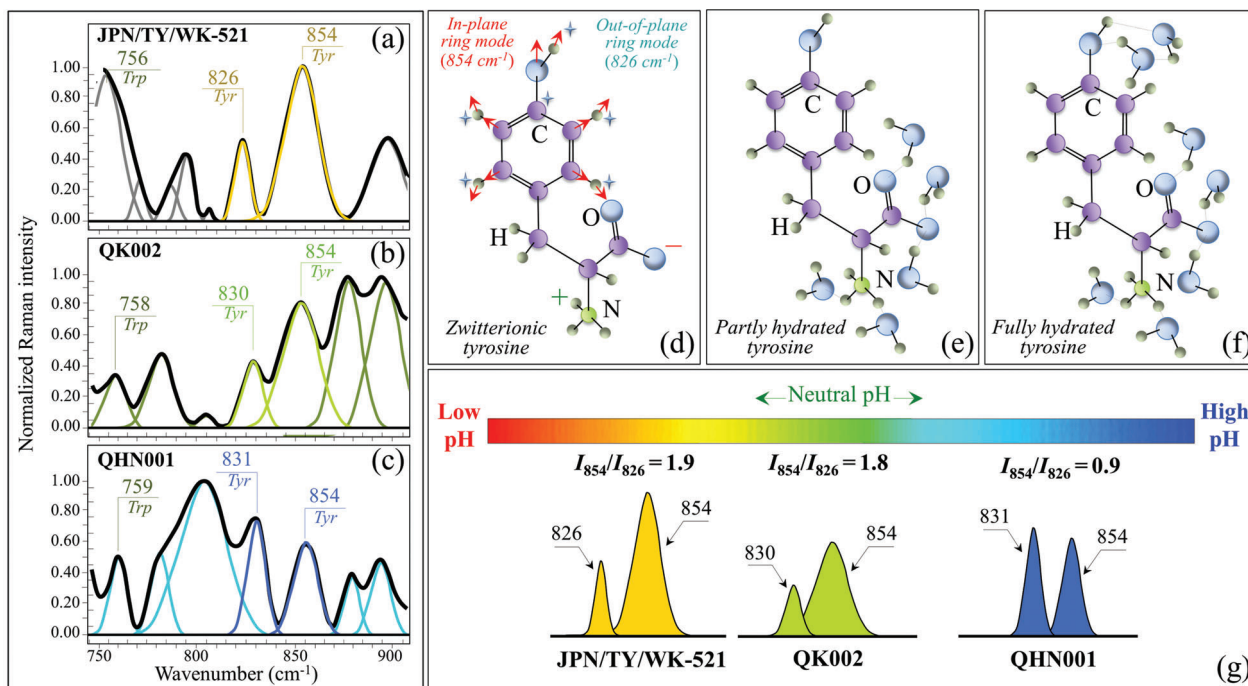


Figure 3. Spectral Zone II (750–900 cm⁻¹) of the Raman spectra of a) JPN/TY/WK-521, b) QK002, and c) QHN001 viral strains; spectra are deconvoluted into a sequence of Gaussian-Lorentzian sub-bands (frequencies for selected bands shown in inset). The abbreviations Trp and Tyr refer to tryptophan and tyrosine, respectively. Structure of tyrosine in d) zwitterionic, e) non-hydrated, and f) fully hydrated condition are shown together with in-plane and out-of-plane vibrational modes of the phenol ring according to Ref. [27]. In (g), components are shown from which the Raman ratio, I_{854}/I_{826} , together with the expected values of environmental pH at the surface of different SARS-CoV-2 viral strains (see labels in inset).

rotamers, a consequence of the different amino acid sequences to which methionine links. Different *trans-to-gauche* ratios, V_t/V_g , could be computed for different viral strains (0.41, 1.37, and 0.83, for JPN/TY/WK-521, QK002, and QHN001 isolates; see labels in Figure 2g). These different values reflect the different chirality of S-containing amino acids, peculiar to individual strains and are efficient in variant recognition.

The zwitterionic structure of monoclinic cysteine rotamers and the expected frequencies for their C-S stretching mode in molecules with different chirality characteristics are shown in Figure 2f.[21–24] Unlike methionine, the thiol side chain in cysteine participates as a nucleophile in environmental reactions and is susceptible to oxidation. C-S stretching bands are expected at ≈ 630 , and ≈ 670 cm⁻¹. [21–24] We indeed observed C-S bands at ≈ 670 cm⁻¹ (distinct from those of methionine) in all investigated strains. Signals at 632 and 628 cm⁻¹ were observed in the JPN/TY/WK-521 isolate and QHN001 variants, respectively (see labels in Figure 2), while only a band of cysteine located at 616 cm⁻¹ (COO⁻ rocking)[24] (missing in JPN/TY/WK-521 and QHN001) could be observed in the QK002 variant. This signal senses different pH environments at the virion surface, as discussed later.

2.2. Tyrosine Phenol Ring as a Sensor of Interface pH: Spectral Zone II

The important feature found in Zone II (Figure 3) relates to the intensity ratio of two bands of tyrosine (Fermi doublet) located

at 854 and 826 cm⁻¹ (spectra from JPN/TY/WK-521, QK002, and QHN001 isolates in (a), (b), and (c)). This ratio, referred to as I_{854}/I_{826} , is diagnostic of the H-bonding environment around tyrosine units, the lower the ratio the more hydrophobic the environment in which the tyrosine residue is embedded.[25–27] According to Hernandez et al.,[27] the components of the tyrosine doublet originate from two independent vibrational modes of the phenol ring, namely, in-plane ring breathing (854 cm⁻¹) and out-of-plane C-H bending (826 cm⁻¹) (see Table S1b, Supporting Information). The I_{854}/I_{826} ratio is a sensor of the hydrophobic/hydrophilic balance in environmental interactions at the virion surface. A low I_{854}/I_{826} ratio represents a hydrophobic tyrosine configuration, tyrosine being most hydrophobic in an alkaline environment, and vice versa for an acidic environment. The different structures of tyrosine are shown in Figure 3d (zwitterionic), (e) (non-hydrated), and (f) (fully hydrated) together with in-plane and out-of-plane vibrational modes of the phenol ring.[27] Figure 3g shows the relationship between the Raman ratio, I_{854}/I_{826} , and environmental pH, as detected for different SARS-CoV-2 viral strains. The I_{854}/I_{826} intensity ratio experienced a maximum value in the spectrum of the JPN/TY/WK-521 isolate (1.9), while the ratios found for the QK002, and QHN001 variant sub-types were 1.8, and 0.9, respectively. The above values suggest that the British variants have shifted from the acidic environment of the original isolate toward neutral-to-alkaline surface protonation conditions. Viruses possess a pH-dependent surface charge in polar media (e.g., water). This electrostatic charge is key in their mobility and governs their sorption processes. According to a previous study,[28] we assigned the character of a

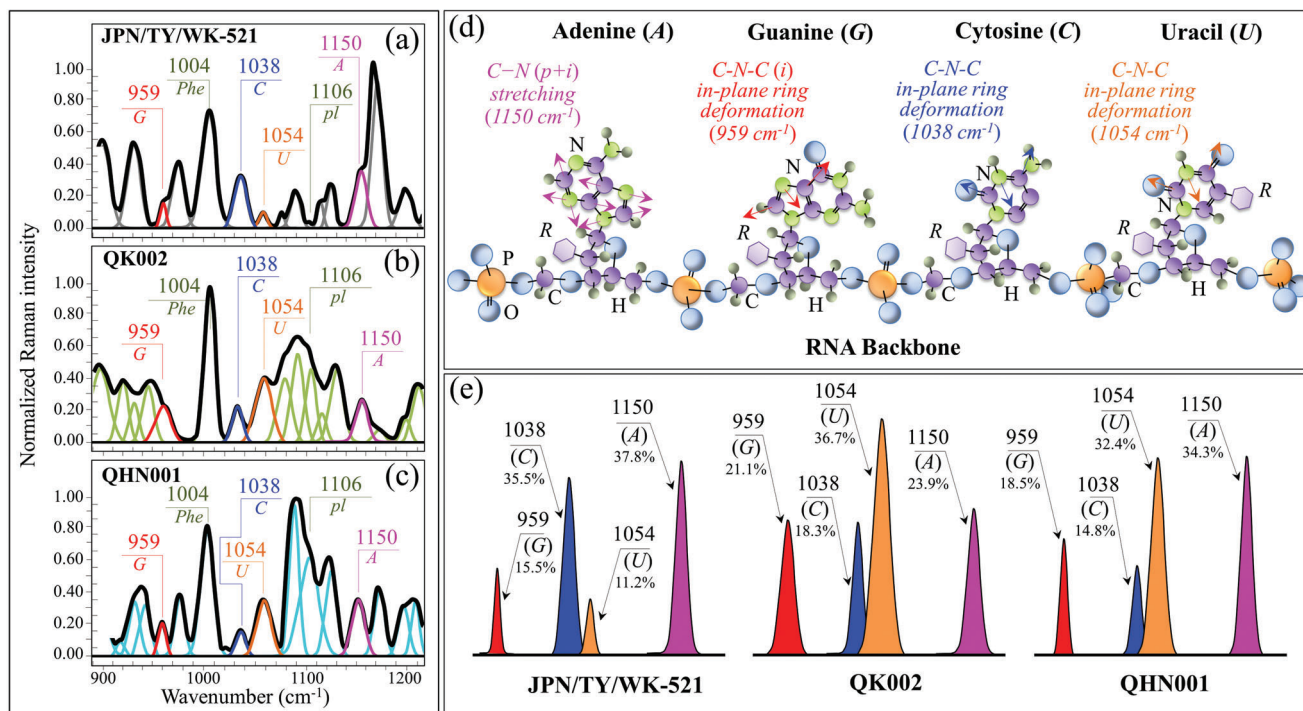


Figure 4. Spectral Zone III (900–1200 cm^{-1}) of the Raman spectra of a) JPN/TY/WK-521, b) QK002, and c) QHN001 viral strains; spectra are deconvoluted into a sequence of Gaussian-Lorentzian sub-bands (frequencies for selected bands shown in inset). The abbreviations G, C, U, pl, and A refer to guanine, cytosine, uracil, phosphodiester linkages, and adenosine, respectively. d) Schematic draft of purines and pyrimidines with phosphodiester linkages together with the ring vibrational modes selected to locate different RNA bases; in (e), signals are shown, which are used to estimate the fractions of different purine and pyrimidines (shown in inset together with the frequencies of the selected signals) found in different strains. The abbreviations *p* and *i* (in brackets in (d)) refer to pyridine and imidazole rings, respectively.

strong hydrogen-bond donor to the tyrosine phenoxyl proton in the QHN001 isolate, while the tyrosine phenoxyl oxygen in both JPN/TY/WK-521 and QK002 virions behave as a strong hydrogen-bond acceptor. In other words, the tyrosine doublet ratio can be taken as a marker of the virion/environment equilibrium in aqueous chemistry. Again, this characteristic clearly differed for different SARS-CoV-2 viral strains.

2.3. “Raman Fractions” of RNA Bases: Spectral Zone III

Figure 4a–c show Zone III Raman spectra for the JPN/TY/WK-521 original isolate, and the QK002 and QHN001 sub-types (labeling and vibrational origins in Figure S1 and Table S1c, Supporting Information). Besides containing the main Raman signal of phenylalanine at 1004 cm^{-1} (symmetric ring breathing), this spectral zone incorporates ring-related signals from individual RNA purines and pyrimidines. Both cytosine (C) and uracil (U) pyrimidines presented isolated bands from C-N-C in-plane deformation of their heterocyclic aromatic ring centered at 1038 and 1054 cm^{-1} .^[29–31] The same type of vibration appeared in the guanine (G) purine at 959 cm^{-1} , while a cumulative signal from C–N stretching modes in both imidazole and pyridine rings of the adenine (A) purine was found at 1150 cm^{-1} .^[31–33] Note that the above signals could clearly be detected in the spectra of all strains, although purine signals appeared as shoulder bands in

the spectrum of the QK002 variant. Figure 4d shows schematic drafts of purines and pyrimidines in the RNA sequence together with their selected vibrational fingerprints.

Fingerprint signals were selected to minimize overlap with signals from different molecules, according to the machine-learning algorithm described in the Supporting Information. The relative intensity ratio of selected fingerprints bands from purines and pyrimidines in Zone III showed quite different values for different strains (see spectra in Figure 4). In Figure 4e, the intensity differences of fingerprint signals are schematically drawn for different strains, together with their fractional intensity values (computed by assuming as 100% the sum of selected signals from all purine and pyrimidine bases). Remarkably, the sum of the fingerprint Raman intensities of different bases was the same for the two different variant sub-types (i.e., within an experimental scatter of $\pm 1\%$), but it was much larger than that found in the original Japanese isolate. This suggests that “reshuffling” of genome components occurred differently in the two variants, and that both variants underwent significant alteration of the total amount of RNA genome as compared to the original isolate. In viruses, the band at 1106 cm^{-1} represents phosphodiester linkages (pl) in the RNA backbone (see spectra in Figure 4 and Table S1c, Supporting Information).^[34] The intensity of this band is proportional to the number of nucleotides units of ordered structures and, consequently, to the number of nucleotides involved in secondary interactions, and accordingly, both variants presented a much

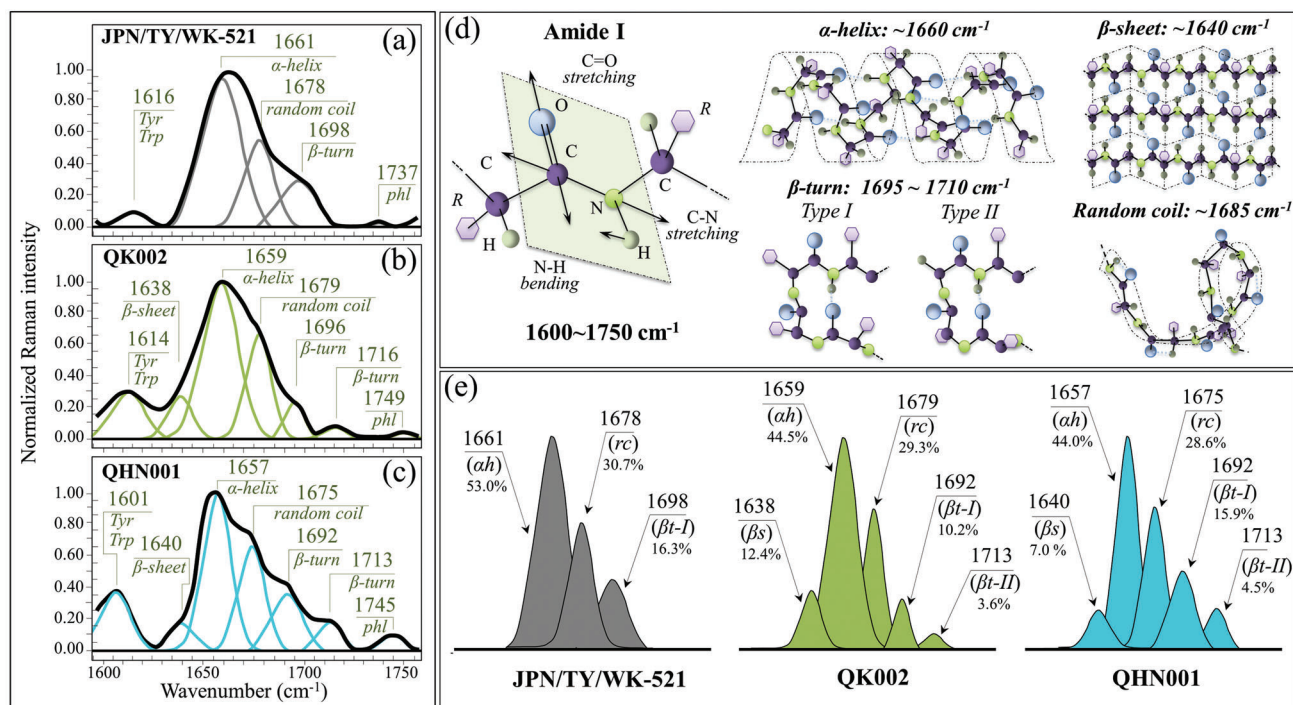


Figure 5. Spectral Zone IV (Amide I, 1600–1750 cm⁻¹) of the Raman spectra of a) JPN/TY/WK-521, b) QK002, and c) QHN001 viral strains; spectra are deconvoluted into a sequence of Gaussian-Lorentzian sub-bands (frequencies for selected bands shown in inset). The abbreviations Trp and Tyr refer to tryptophan and tyrosine, respectively. d) Schematic drafts of the Amide I vibrational mode, the different secondary structure of proteins and related frequencies; in e), signals are shown, which are used to estimate the fractions of different protein secondary structures (shown in inset together with the frequencies of the selected signals) found in different strains. The abbreviations βs, αh, rc, and βt-I, βt-II, and phl represent β-sheet, α-helix, random coil, two types of β-turn rotamers, and phospholipids, respectively.

higher probability of secondary interaction as compared to the original isolate. Labels in the inset of Figure 4e give the computed fractions of individual bases. It should be emphasized that the fractions computed from the intensities of the Raman bands do not represent the actual fractions computed by genome analyses (see comparison between fractions computed by Raman and genome analyses in Table S2, Supporting Information). Unlike conventional genome analyses, base pairing influences the computed “Raman fractions”. However, this characteristic appears to enhance the differences between different isolates, and it is useful to distinguish variants. The segmented composition of influenza viral genomes promotes reassortment; however, not all permutations of segment reassortment occur at the same frequency, with certain reassortment events being observed at higher rates.^[35] Since viral nucleoproteins associate at preferential RNA sites,^[36] variations in genome composition are linked to changes in the secondary structure of proteins, as discussed below.

2.4. Protein Secondary Structure: Spectral Zone IV

In Figure 5a–c, deconvoluted Raman spectra in the Amide I frequency region (1600–1750 cm⁻¹) are shown for the original JPN/TY/WK-521 Japanese isolate, and the two British variant subtypes QK002, and QHN001 (see labels). Deconvoluted Amide

I band components, centered at 1638–1640, 1657–1661, 1675–1679, 1692–1698, and 1713–1716 cm⁻¹, were assigned to β-sheet (βs), α-helix (αh), random coil (rc), and β-turn rotamers (βt-I and βt-II) (see Figure S1 and Table S1d, Supporting Information).^[36] Figure 5d shows schematic drafts of the secondary structures of proteins and their expected ranges of vibrational frequencies. In Figure 5e, Amide I signals are compared for spectra recorded different isolates (see labels). Despite a common feature in the predominance of α-helix component in all strains, the Amide I spectra of the British variants showed clearly different morphologies. The differences reflected substantial dissimilarity in the protein secondary structure of the different strains. The JPN/TY/WK-521 original isolate lacked signals from β-sheet and β-turn Type II rotamer, which instead appeared in both British variant subtypes. The fractional amounts of different protein structures computed from relative intensities of the Amide I spectrum are given in the inset of Figure 5e. According to Eker et al.,^[37] a hydrophobic environment leads to a propensity to form β-sheet. Our results corroborate this, since the two British variant subtypes, which newly included the β-sheet structure, also experienced a lowered I_{854}/I_{826} tyrosine ratio (i.e., a marker of increased environmental hydrophobicity) as compared with the original Japanese isolate. The conformational preference of viral proteins has key implications in peptide immunogenicity and can be swiftly captured by Raman spectroscopic assessments in the Amide I zone.

3. Discussion

3.1. Importance of the Raman Technology in the COVID-19 Pandemic

Regional surveys by the Japanese Government showed a surge in the British coronavirus variants with N501Y mutation in the Kansai area in the first half of 2021, and indicate their possible increased transmissibility as compared to the original isolate. In-surgence of new variants is already occurring and will likely continue in the near future in countries with a low percent of vaccinated population. Our Raman analysis discovered bold differences in signals from RNA bases in both QK002 and QHN001 British variant sub-types. The Raman fractional data did not fully correspond to the genomic data of the variants. However, they represented a new set of “Raman genomic” data distinctive of the variants. Once implemented into a vibrational library, the “Raman genome” could provide insightful and promptly obtainable information as compared to conventional genomic analyses. RNA sequencing has a wide-ranging impact from diagnosis and pathogenesis to vaccine design and viral ecology.^[38] This study shows for the first time the possibility to use, on-site and in real time, the unique sensitivity of Raman spectroscopy in locating molecular rotamers, protonation conditions of the viral surface, genomic differences, and protein secondary structures for discriminating between different variants of the SARS-CoV-2 virus. This study is in line with our previous studies of variants of Influenza virus.^[39] The Raman characteristics yield molecular-scale information uniquely discriminating among different variants. Facilitating real-time data access and exchange is an essential step in tracking variants and in estimating their spread and evolutionary mutation rate. Raman technology could provide an invaluable contribution to the presently available practices of traceability, tracking, and record-keeping of viral spread.

Important aspects of the present work are the spectroscopic identification of different variants and the observation of spectroscopic differences between SARS-CoV-2 viral sub-types. With the COVID-19 pandemic rapidly expanding, the fast evolution of variants into viral sub-types may be difficult to monitor and standardize. From a purely virological point of view, the discovery of multiplication of sub-types within the same variant is an unequivocal proof that the so-called “founder effect” plays a role at least as important as that of selection pressure. In other words, the competitive advantage of a given variant with respect to viral replication, transmission, or escape from immunity overlaps locally inherited genomic mutations. As a result, a variant sub-type might survive regardless of its actual viral fitness. As widely recognized, it is the interplay of natural selection and chance events that defines virus evolution in local communities and leads to the generation of a number of variants and their sub-types. In this context, we emphasize that it is both a remarkable and fortunate circumstance that Raman spectroscopy could distinguish among virus sub-types by virtue of its high sensitivity to molecular symmetry, base pairing, and protein secondary structures. This property of the Raman method could, in principle, allow geographical maps of viral outbreaks in downstream analyses (to support or even replace real-time tracing), and plotting spatiotemporal visualization of viral dynamics. We anticipate here the possibility to detect additional sub-types for the British variant as well as for

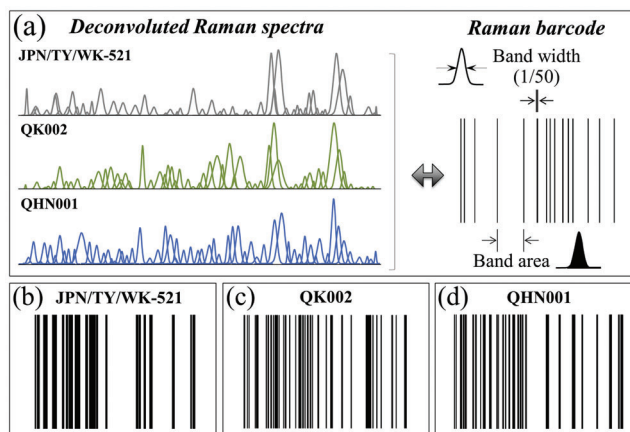


Figure 6. a) Sequences of Raman Gaussian-Lorentzian bands are shown as deconvoluted from average Raman spectra recorded on the original Japanese isolate and two British variants (see labels); and, algorithm to convert a band sequence into a barcode. In (b–d), barcodes are shown that are constructed for the sub-band sequences in (a).

South African, Brazilian, and Indian variants, according to our preliminary unpublished studies.

From a purely structural point of view, the fractional balance between the two S-containing amino acids, methionine, and cysteine, in the protein structure of virion sub-types should conceivably be the result of genomic interplay with the cellular environment in which the specific viral sub-type was developed. Note also that the catalytic and enzymatic activity of cysteine residues bound by disulfide bridges is key in both protein folding and stability; thus, its abundance directly links to viral fitness. The above two arguments reasonably justify the existence of a variety of virus sub-types with different methionine/cysteine fractions as modulated by both genomic and founder-effect-driven kinetics.

3.2. Benefits of Electronic Bar Coding as a Measure of Pandemic Control

In view of the present findings, we propose to use a barcode with the Raman spectrum of different variants and their sub-types in order to enable efficient electronic recordkeeping and increase users’ accessibility to the emergence and transmission characteristics of variants through apps and user-friendly software. Barcodes could be matched to the Raman spectrum in a number of different ways, and provide the Raman spectroscopic method with the flexibility and the swiftness necessary to inform users about viral mutations periodically developed under different local conditions. Examples of Raman spectrum/barcode matching are given in **Figure 6**. In **Figure 6a**, sequences of Raman Gaussian-Lorentzian bands are shown as deconvoluted from average Raman spectra recorded on the original Japanese isolate and two British variant sub-types. An algorithm is also shown, which converts the band sequence into a barcode by assigning to each band a line with thickness equal to 1/50 of the sub-band width and a distance from the successive line proportional to the band area. The “Raman barcodes” unambiguously locate the virus variants/sub-types (**Figure 6b**) and can be decrypted into easily readable information through appropriate apps.

The use of electronic bar coding has already been set in practice in a number of hospitals and health systems in order to reduce medication errors. The increasing popularity of this technology arises from the need to add an additional level of verification into medication administration processes.^[40]

Also in the context of COVID-19 pandemic, a “genetic barcode” has been proposed, which uses information from a global database of genetic data extracted from coronavirus tests.^[41] Similar to the “genetic barcode”, the “Raman barcode” is capable to locate the presence of a multiplicity of slightly different versions of the virus among infected people as sub-types within their general classification into individual variants. Such details could become of fundamental importance in future management of the pandemic because small genetic differences could be key in future drugs and vaccine developments. Moreover, the Raman barcode technology could be useful in tracking whether new cases of infection are the result of local evolution/transmission or come from different parts of the world. The “Raman barcode” shares with the “genetic barcode” the capacity of thoroughly locating genetically significant virus sub-types of individual variants, but might surpass it in swiftness and flexibility regarding on-site diagnostics. In summary, the authors believe that the introduction of the “Raman barcode” approach into capillary networks would be an important contribution to the real-time management of information concerning virus evolution during pandemics. In the immediate future, once automatically linked to clouds, it could report on-site how the virus is evolving under the selective pressure of the polyclonal immune response in vaccinated people.

3.3. Statistical Assessments and Limitations of the Present Study

In order to validate the clear differences in average spectra detected among different variants/sub-types (see Figure 1), which we used here in discussing differences at the molecular scale, we also attempted a statistical evaluation of the spectral data collected on different sub-types of the British variant as compared to the original Japanese isolate (see statistical assessments in the Supporting Information). An approach based on the Pearson's correlation coefficient^[42] (PC), r , was adopted, according to the following considerations:

- i) The PC value has widely been used for pixel-to-pixel image correlations and to compare similarities between Raman spectra considered as strings of intensity values detected for different CCD pixels.^[43]
- ii) Among the available statistical methods, the one based on PC appears to be most suitable to assess real on-site situations, in which a single spectrum collected on-site from a patient, needs to be matched with a multitude of average spectra from a library.^[44]

The PC algorithm was designed to select the best matching spectrum based on a similarity criterion (see mathematical formulation in Equation (S1), Supporting Information), with $r = 1$ stemming for perfectly identical spectra and $r = 0$ indicating a complete mismatch. The results of the PC-based statistical analysis performed in this study are summarized in Table S3, Sup-

porting Information, in terms of the r values and related standard deviations. A cross comparison of r values among variants and sub-types systematically confirmed an almost identical pattern ($r > 0.920$) when comparing spectra collected on samples of the same variant/sub-type, while low PC scores ($r < 0.420$) were retrieved for any mixed matching among different variants. All statistical matches of spectra collected on the same variant presented a quite small standard deviation (i.e., in the order of 10^{-2}), while matching spectra from different variants involved larger standard deviations (in the order of 10^{-1}) in addition to low r values (see Table S3, Supporting Information). An interesting result was that cross-matches between different sub-types of the British variant gave high r values (i.e., comparable with those recorded for spectra collected on the same sub-type), but low standard deviations (i.e., comparable with those recorded for different variants). In order to support the statistical analysis discussed above, comparisons between average Raman spectra and spectra collected at individual locations on the same variant/sub-type were made, as shown in the Supporting Information (from Figures S2–S11, Supporting Information). Related spectroscopic analyses are given for fractions of sulfur-containing amino acid rotamers, tyrosine phenol ring, apparent fractions of RNA purines and pyrimidines, and protein secondary structures. Individual spectra collected at different locations generally showed agreement with the corresponding average one in agreement with Pearson's statistical assessments. In particular, sub-types of the British variant were closer to their respective average spectra (PC = 0.985 ± 0.027 and 0.996 ± 0.001 for QHN001 and QK002, respectively) as compared to the matching shown by the original Japanese isolate JPN/TY/WK-521 with its average spectrum (PC = 0.925 ± 0.065) (see Table S3, Supporting Information). However, some anomalous spectra were locally found with interesting spectral variations. Such anomalous spectra, although having no statistical significance, unveiled some interesting structural features. These additional aspects are discussed in the Supporting Information.

Although the present statistical assessment should only be considered as a preliminary one, it strengthens our thesis of high sensitivity for a Raman approach in SARS-CoV-2 viral speciation. While mathematical methods remain to be rigorously tested and standardized to confirm the elimination of any eventual influence of spectral background and signal-to-noise ratio, we consider the present results as quite encouraging. We have suggested refined procedures of spectral pre-treatments in the Supporting Information, which should critically be evaluated in future works. There are additional points to be assessed in future evaluations. First, spectroscopic experiments on samples from on-site patients should be performed to confirm the possibility of matching them with high reliability to average spectra from a library. Note, in this context, that the different variants/sub-types investigated in this study were all obtained from the Japanese National Institute of Infectious Diseases after a rigorous classification through genomic analyses based on a large number of biological replicates. Second, a precise spectroscopic scrutiny should be made of the *in vitro* process of viral propagation, which we applied in the present study for achieving a fixed titer concentration from the small fractions of virions received per each variant/sub-type, in order to check whether such process could have any in-

fluence on the original virus identity through additional interplay between genome-driven mutations and chance events.

Despite the above limitations, the present spectroscopic study shows for the first time the existence of significant molecular gradients in the population of SARS-CoV-2 virion structures belonging to different variants/sub-types. Similar to humans, such “somatic” differences ultimately cluster into well-defined average patterns with high statistical reliability.

4. Conclusion

In conclusion, we showed that Raman analyses could provide virologists with insightful and rapid information on SARS-CoV-2 variants and their sub-types through a clear view on virus structure at the molecular scale. The high sensitivity of Raman spectroscopy in virus speciation and the importance of molecular symmetry in SARS-CoV-2 virus classification could also be exploited in the case of other viral strains and pathogenic species. Data obtained from Raman measurements are particularly informative regarding the molecular details of conformational isomerism in viral proteins, fractional amounts of RNA purine and pyrimidine bases, protonation conditions of the viral surface, and protein secondary structures. The Raman spectrum encrypts the solvent-dependent structural propensity of viral proteins, directly linking to site-specific interactions with the key molecular subgroups that define the viral morphogenetic pathway. While the construction of a wide Raman library of SARS-CoV-2 variants/sub-types is presently ongoing, it is hoped that the experimental evidence presented here will provide impetus for developing high-sensitivity portable Raman equipment for fast on-site characterizations of virus variants.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

SARS-CoV-2 strains (JPN/TY/WK-521, QK002, and QHN001) were kindly provided by the National Institute of Infectious Diseases. Raman experiments were conducted at the Kyoto Integrated Science & Technology Bio-Analysis Center. This work was supported by AMED (Grant Number 20he1122006j0001).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

Raman, barcode, SARS-CoV-2, variants, fingerprints

Received: August 1, 2021

Revised: November 4, 2021

Published online: December 7, 2021

- [1] E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. Laydon, G. Dabrera, A. O'Toole, R. Amato, M. Ragonnet-Cronin, I. Harrison, B. Jackson, C. V. Ariani, O. Boyd, N. J. Loman, J. T. McCrone, S. Goncalves, D. Jorgensen, R. Myers, V. Hill, D. K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D. P. Kwiatkowski, The COVID-19 Genomics UK (COG-UK) consortium, S. Flaxman, O. Ratmann, et al., *Nature* **2021**, 593, 266.
- [2] K. Leung, M. H. Shum, G. M. Leung, T. T. Lam, J. T. Wu, *Eurosurveillance* **2021**, 26, 2002106.
- [3] N. G. Davies, S. Abbott, R. C. Barnard, C. I. Jarvis, A. J. Kucharski, J. D. Munday, C. A. B. Pearson, T. W. Russell, D. C. Tully, A. D. Washburne, T. Wenseleers, A. Gimma, W. Waites, K. L. M. Wong, K. van Zandvoort, J. D. Silverman, CMMID COVID-19 Working Group, COVID-19 Genomics UK (COG-UK) Consortium, K. Diaz-Ordaz, R. Keogh, R. M. Eggo, S. Funk, M. Jit, K. E. Atkins, W. J. Edmunds, *Science* **2021**, 372, 149.
- [4] H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, S. Pillay, E. J. San, N. Msomi, K. Mlisana, A. von Gottberg, S. Walaza, M. Allam, A. Ismail, T. Mohale, A. J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, F. Petruccione, A. Sigal, D. Hardie, G. Marais, N.-y. Hsiao, S. Korsman, M.-A. Davies, L. Tyers, I. Mudau, D. York, et al., *Nature* **2021**, 592, 438.
- [5] E. C. Sabino, L. F. Buss, M. P. S. Carvalho, C. A. Preete, Jr., M. A. E. Crispim, N. A. Fraiji, R. H. M. Pereira, K. V. Parag, P. da Silva Peixoto, M. U. G. Kraemer, M. K. Oikawa, T. Salomon, Z. M. Cucunuba, M. C. Castro, A. A. de Souza Santos, V. H. Nashimento, H. S. Pereira, N. M. Ferguson, O. G. Pybus, A. Kucharski, M. P. Busch, C. Dye, N. R. Faria, *Lancet* **2021**, 397, 452.
- [6] <https://www.niid.go.jp/niid/ja/diseases/ka/corona-virus/2019-ncov/10344-covid19-ab33th.html> (accessed: November 2021).
- [7] <https://www.niid.go.jp/niid/en/2019-ncov-e/10334-covid19-ab31th-en.html> (accessed: November 2021).
- [8] T. T. Nguyen, P. N. Pathirana, T. Nguyen, Q. V. H. Nguyen, A. Bhatti, D. C. Nguyen, D. T. Nguyen, N. D. N., D. Creighton, M. Abdelrazek, *Sci. Rep.* **2021**, 11, 3487.
- [9] H. Tegally, E. Wilkinson, J. Lessells, J. Giandhari, S. Pillay, N. Msomi, K. Mlisana, J. N. Bhiman, A. von Gottberg, S. Walaza, V. Fonseca, M. Allam, A. Ismail, A. J. Glass, S. Engelbrecht, G. Van Zyl, W. Preiser, C. Williamson, F. Petruccione, A. Sigal, I. Gazy, D. Hardie, N.-y. Hsiao, D. Martin, D. York, D. Goedhals, E. J. San, M. Giovanetti, J. Laurencio, L. C. Alcantara Jr., T. D. Oliveira, *Nat. Med.* **2021**, 27, 440.
- [10] T. Koyama, D. Platt, L. Parida, *Bull. V. H. O.* **2020**, 98, 495.
- [11] D. Cyranoski, *Nature* **2021**, 589, 337.
- [12] A. D. Radford, D. Chapman, L. Dixon, J. Chantrey, A. C. Darby, N. Hall, *J. Gen. Virol.* **2012**, 93, 1853.
- [13] G. Zhu, X. Zhu, Q. Fan, X. Wan, *Spectrochimica Acta Part A* **2011**, 78, 1187.
- [14] F. Tian, B. Tong, L. Sun, S. Shi, B. Zheng, Z. Wang, X. Dong, P. Zheng, *bioRxiv* **2021**, <https://www.biorxiv.org/content/10.1101/2021.10.12.464114v1>
- [15] K. Qu, X. Xiong, K. A. Ciazynska, A. P. Carter, J. A. G. Briggs, *Nature* **2020**, 588, 498.
- [16] E. Podstawka, Y. Ozaki, L. M. Proniewicz, *Appl. Spectrosc.* **2004**, 58, 581.
- [17] X. Cao, G. Fischer, *J. Phys. Chem.* **2002**, 106, 41.
- [18] M. Tsuboi, Y. Ezaki, M. Aida, M. Suzuki, A. Yimit, K. Ushizawa, T. Ueda, *Biospectrosc.* **1998**, 4, 61.
- [19] P. Svancarova, T. Betakova, *Virol. J.* **2018**, 15, 187.
- [20] Z. Ye, T. Liu, D. P. Offringa, J. McInnis, R. A. Levandowski, *J. Virol.* **1999**, 73, 7467.
- [21] G. D. Fleming, J. J. Finnerty, M. Campos-Vallette, F. Celis, A. E. Aliaga, C. Fredes, R. Koch, *J. Raman Spectrosc.* **2009**, 40, 632.
- [22] S. F. Parker, *Chem. Phys.* **2013**, 424, 75.
- [23] P. Bazylewski, R. Divigalpitaya, G. Fanchini, *RSC Adv.* **2017**, 7, 2964.

- [24] B. Hernandez, F. Pfluger, A. Adenier, S. K. Kruglik, M. Ghomi, *Phys. Chem. Chem. Phys.* **2011**, *13*, 17284.
- [25] P. G. Hildebrandt, R. A. Copeland, T. G. Spiro, J. Otlewski, M. Laskowski Jr., F. G. Prendergast, *Biochem* **1988**, *27*, 5426.
- [26] C. Krafft, W. Hinrichs, P. Orth, W. Saenger, H. Welfle, *Biophys. J.* **1998**, *74*, 63.
- [27] B. Hernandez, Y.-M. Coic, F. Pluger, S. G. Kruglik, M. Ghomi, *J. Raman Spectrosc.* **2016**, *47*, 210.
- [28] Z. Arp, D. Autrey, J. Laane, S. A. Overman, G. J. Thomas Jr., *Biochem* **2001**, *40*, 2522.
- [29] S. Sanchez Cortes, J. V. Garcia-Ramos, *J. Raman Spectrosc.* **1992**, *23*, 61.
- [30] M. Mathlouthi, A. M. Seuvre, J. L. Koenig, *Carbohydrate Res.* **1986**, *146*, 15.
- [31] F. Madzharova, Z. Heiner, M. Guehlke, J. Kneipp, *J. Phys. Chem. C* **2016**, *120*, 15415.
- [32] R. P. Lopes, M. P. M. Marques, R. Valero, J. Tomkinson, L. A. E. Batista de Carvalho, *J. Spectrosc.* **2012**, *27*, 273.
- [33] R. P. Lopes, R. Valero, J. Tomkinson, M. P. M. Marques, L. A. E. Batista de Carvalho, *New J. Chem.* **2013**, *37*, 2691.
- [34] G. J. Thomas Jr., K. A. Hartman, *Biochim. Biophys. Acta* **1973**, *312*, 37.
- [35] M. Gerber, C. Isel, V. Moules, R. Marquet, *Trends Microbiol.* **2014**, *22*, 446.
- [36] M. Voicescu, S. Ionescu, C. L. Nistor, *Spectrochimica Acta Part A: Mole. Biomole. Spectrosc.* **2017**, *170*, 1.
- [37] F. Eker, X. Cao, L. Nafie, Q. Huang, R. Schweitzer-Stenner, *J. Phys. Chem. B* **2003**, *107*, 358.
- [38] S. Di Giorgio, F. Martignano, M. G. Torcia, G. Mattiuz, S. G. Conticello, *Sci. Adv.* **2020**, *6*, eabb5813.
- [39] G. Pezzotti, *J. Raman Spectrosc.* **2021**, <https://doi.org/10.1002/jrs.6204>
- [40] E. Truitt, R. Thompson, D. Blazey-Martin, D. NiSai, D. Salem, *Hosp. Pharm.* **2016**, *51*, 474.
- [41] Z. Zhao, B. A. Sokhansanj, C. Malhotra, K. Zheng, G. L. Rosen, *PLoS Comput. Biol.* **2020**, *16*, e1008269.
- [42] K. Pearson, *Philos. Trans. R. Soc. London* **1896**, *187*, 253.
- [43] T. Mu, S. Wang, T. Li, B. Wang, X. Ma, B. Huang, L. Zhu, J. Guo, *IEEE J. Sel. Top. Quantum Electron.* **2019**, *25*, 1.
- [44] A. Z. Samuel, R. Mukojima, S. Horii, M. Ando, S. Egashira, T. Nakashima, M. Iwatsuki, H. Takeyama, *ACS Omega* **2021**, *6*, 2060.